



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE TECNOLOGIA (CTC)  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Luis Felipe de Deus

**Estimativa dos níveis de glicose através do processamento de sinais de  
Fotopletismografia (PPG) e modelos de inteligência artificial**

Florianópolis  
2023

Luis Felipe de Deus

**Estimativa dos níveis de glicose através do processamento de sinais de Fotopletismografia (PPG) e modelos de inteligência artificial**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de mestre em Engenharia Elétrica.

Orientador: Prof. Cesar Ramos Rodrigues, Dr.

Coorientador: Prof. Jefferson Luiz Brum Marques, Dr.

Florianópolis

2023

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

de Deus, Luis Felipe

Estimativa dos níveis de glicose através do  
processamento de sinais de Fotopletismografia (PPG) e  
modelos de inteligência artificial / Luis Felipe de Deus ;  
orientador, Cesar Ramos Rodrigues, coorientador, Jefferson  
Luiz Brum Marques, 2023.

79 p.

Dissertação (mestrado) - Universidade Federal de Santa  
Catarina, Centro Tecnológico, Programa de Pós-Graduação em  
Engenharia Elétrica, Florianópolis, 2023.

Inclui referências.

1. Engenharia Elétrica. 2. Glicose. 3.  
Fotopletismografia. 4. Inteligência Artificial. I.  
Rodrigues, Cesar Ramos. II. Marques, Jefferson Luiz Brum.  
III. Universidade Federal de Santa Catarina. Programa de  
Pós-Graduação em Engenharia Elétrica. IV. Título.

Luis Felipe de Deus

**Estimativa dos níveis de glicose através do processamento de sinais de Fotopletismografia (PPG) e modelos de inteligência artificial**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Cesar Ramos Rodrigues, Dr.  
Universidade Federal de Santa Catarina (UFSC)

Profa. Daniela Ota Hisayasu Suzuki, Dra.  
Universidade Federal de Santa Catarina (UFSC)

Prof. Cesar Augusto Prior, Dr.  
Universidade Federal de Santa Maria (UFSM)

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Engenharia Elétrica.

---

Prof. Telles Brunelli Lazzarin, Dr.  
Coordenação do Programa de  
Pós-Graduação

---

Prof. Cesar Ramos Rodrigues, Dr.  
Orientador

Florianópolis, 2023.

Este trabalho é dedicado aos meus pais, Daniel e  
Roselei, o meu maior exemplo.

## **AGRADECIMENTOS**

A realização deste trabalho representa a conclusão de outra etapa, a qual demandou muito esforço, dedicação e força para não desistir, bem como o auxílio e compreensão de diversas pessoas. De maneira especial agradeço:

- Primeiramente, agradeço a Deus pela oportunidade que me deste de aprender, pela força e determinação durante esta jornada.
- Aos meus pais Daniel e Roselei, por tudo aquilo que sou hoje, pelas oportunidades, amor, carinho e compreensão, pelo suporte e pela chance de alcançar meus sonhos.
- Aos meus familiares, em especial a memória dos meus amados avós, Lorivo e Laureci, que sempre me inspiraram com seu amor, sabedoria e apoio incondicional.
- À Universidade Federal de Santa Catarina, que abriu suas portas e proporcionou infraestrutura e corpo docente capacitado para a minha formação acadêmica.
- Ao Hospital Universitário Polydoro de São Thiago da Universidade Federal de Santa Catarina (HU-UFSC/EBSERH). Em especial às enfermeiras Adnairdes Cabral de Sena e Susana Terezinha de Quadros pela assistência e ajuda, principalmente durante a etapa de coleta de dados.
- Ao meu orientador, Prof. Cesar Ramos Rodrigues, primeiro pelo convite de realizar o mestrado nesta universidade, pela dedicação e orientação com este trabalho.

*“Everything is theoretically impossible,  
until it is done.”  
(Robert A. Heinlein)*

## RESUMO

A glicose desempenha um papel essencial como fonte de energia para o organismo, sendo obtida principalmente através dos alimentos consumidos. Para que as células possam utilizar a glicose, é essencial que os níveis de açúcar no sangue sejam regulados de forma adequada. A insulina, um hormônio produzido pelas células  $\beta$  do pâncreas, desempenha um papel crucial no controle da glicose sanguínea. No entanto, a diabetes mellitus é uma condição em que ocorre uma disfunção na produção ou no uso da insulina pelo corpo. Isso resulta em desequilíbrios nos níveis de glicose, podendo levar a complicações graves, como doenças renais, ataques cardíacos e acidente vascular cerebral. Segundo a Organização Mundial de Saúde, a diabetes é a nona maior causa de mortes no mundo, sendo responsável por 1,5 milhão de óbitos diretamente relacionados à doença. Portanto, é de extrema importância monitorar regularmente os níveis de glicose no sangue, especialmente para pessoas com diagnóstico de diabetes, como forma de detectar precocemente possíveis complicações de saúde. Atualmente, os métodos utilizados para monitoramento contínuo da glicose são invasivos, caros e desconfortáveis. Este estudo propõe um método não invasivo para estimativa da glicose utilizando o sinal de fotopletismografia em conjunto com algoritmos de Inteligência Artificial. A pesquisa foi conduzida em duas bases de dados distintas: uma coletada durante o estudo (UFSC-BGL) e outra obtida através do conjunto MIMIC III (MIMIC-BGL). Foram realizados experimentos para determinar a melhor abordagem na estimativa da glicose com base nos sinais de fotopletismografia. Os sinais foram segmentados em janelas de 10 segundos, sem sobreposição, e um total de 51 características nos domínios da frequência, não linear e morfológico foram extraídas. Os dados foram divididos em conjuntos de treinamento (67%) e teste (33%), utilizados para treinar modelos de aprendizado de máquina e posteriormente avaliados com o conjunto de teste. Os resultados demonstraram que o modelo CatBoost alcançou um erro médio absoluto percentual de 11,91% e 73% das estimativas foram classificadas na zona de maior acurácia do gráfico de Clarke ao utilizar o conjunto de dados UFSC-BGL. Por outro lado, ao utilizar o conjunto de dados MIMIC-BGL, o modelo Support Vector Machine demonstrou o melhor desempenho, com um erro médio absoluto percentual de 20,64% e 65% das estimativas localizadas na zona de maior acurácia do gráfico de Clarke. Este estudo apresenta duas principais contribuições. A primeira é a disponibilização de dois conjuntos de dados, um deles obtido durante o curso desta pesquisa. A segunda é a criação de um componente de avaliação de qualidade de sinal robusto, capaz de identificar e remover segmentos que não satisfaçam os critérios de qualidade preestabelecidos. Além disso, como resultado desta pesquisa, foi desenvolvido um modelo computacional que estima com erro médio absoluto percentual de 11,91% os níveis de glicose a partir do sinal de fotopletismografia.

**Palavras-chave:** Glicose. Fotopletismografia. Inteligência Artificial.



## ABSTRACT

Glucose plays an essential role as an energy source for the body, primarily obtained through consumed food. To enable cells to utilize glucose, proper regulation of blood sugar levels is crucial. Insulin, a hormone produced by pancreatic beta cells, plays a critical role in controlling blood glucose. However, diabetes mellitus is a condition characterized by dysfunction in insulin production or utilization by the body, resulting in imbalances in glucose levels and potentially leading to serious complications such as kidney disease, heart attacks, and strokes. According to the World Health Organization, diabetes is the ninth leading cause of death worldwide, responsible for 1.5 million deaths directly related to the disease. Therefore, regular monitoring of blood glucose levels is of utmost importance, especially for individuals diagnosed with diabetes, as a means of early detection of potential health complications. Currently, the methods used for continuous glucose monitoring are invasive, expensive, and uncomfortable. This study proposes a non-invasive method for glucose estimation using Photoplethysmography signals in conjunction with Artificial Intelligence algorithms. The research was conducted using two distinct databases: one collected during the study (UFSC-BGL) and another obtained from the MIMIC III dataset (MIMIC-BGL). Experiments were performed to determine the best approach for glucose estimation based on the Photoplethysmography signals. The signals were segmented into non-overlapping 10-second windows, and a total of 51 features in the frequency, nonlinear, and morphological domains were extracted. The data were divided into training (67%) and testing (33%) sets, used to train machine learning models and subsequently evaluated using the test set. The results demonstrated that the CatBoost model achieved a mean absolute percentage error of 11.91% and 73% of the estimates were classified within the most accurate zone of the Clarke error grid when using the UFSC-BGL dataset. On the other hand, when using the MIMIC-BGL dataset, the Support Vector Machine model exhibited the best performance, with a mean absolute percentage error of 20.64% and 65% of the estimates falling within the most accurate zone of the Clarke error grid. This study presents two main contributions. The first, is the provision of two datasets, one of which was obtained during the course of this research. The second, is the creation of a robust signal quality assessment component capable of identifying and removing segments that do not meet pre-established quality criteria. Additionally, as a result of this research, a computational model has been developed that estimates glucose levels with mean absolute percentage error of 11.91% from the photoplethysmography signal.

**Keywords:** Blood Glucose. Photoplethysmography. Artificial Intelligence.

## LISTA DE FIGURAS

Figura 1 – Esquemático para aquisição do sinal de PPG por meio de transmissão e reflexão à esquerda; representação da variação da luz através dos tecidos do corpo humano à direita. . . . .	19
Figura 2 – Exemplo de resposta em frequência para os filtros: a) Butterworth b) Chebyshev I c) Elíptico. . . . .	20
Figura 3 – Exemplo de gráfico de erro de Clarke. . . . .	24
Figura 4 – Fluxograma de execução para o conjunto de dados MIMIC-BGL. . .	29
Figura 5 – Equipamento para extração do sinal de PPG, oxímetro CONTEC CMS50E e <i>software</i> de aquisição. . . . .	32
Figura 6 – Equipamento para aferição da glicose, glicosímetro digital GTech. .	32
Figura 7 – Gráficos obtidos a partir do conjunto UFSC-BGL. . . . .	33
Figura 8 – Fluxograma de execução deste trabalho. . . . .	34
Figura 9 – Fluxograma detalhado do pré-processamento. . . . .	35
Figura 10 – Fluxograma do particionamento dos dados em treino e teste. . . .	38
Figura 11 – Segmento participante 006 @60Hz, glicose de 114 mg/dL. a) sinal <i>raw</i> , b) sinal filtrado com Savitzky-Golay + Butterworth, c) sinal normalizado + inversão de fase, d) sinal com remoção de linha base . .	42
Figura 12 – Extração de características em domínios distintos. Participante 006 @60Hz, glicose de 114 mg/dL. . . . .	43
Figura 13 – Extração de características do pulso do sinal de PPG, VPG e APG, destacados em azul, vermelho e verde. Participante 006 pulso reamostrado a @200 Hz, com glicose de 114 mg/dL. . . . .	44
Figura 14 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de treino e teste para o conjunto UFSC-BGL. . . . .	44
Figura 15 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de treino e teste para o conjunto MIMIC-BGL. . . . .	45
Figura 16 – Histograma de importância de cada característica para o modelo Random Forest utilizando UFSC-BGL e MIMIC-BGL. . . . .	47
Figura 17 – Gráfico de dispersão dos valores preditos de glicose agrupados para cada participante, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL. Linha 1:1 e Linha de melhor ajuste realçadas respectivamente em vermelho tracejado e verde sólido. . . . .	48

Figura 18 – Gráfico de Clarke dos valores preditos de glicose agrupados para cada participante, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL.	49
Figura 19 – Gráfico de segmento dos valores reais e preditos pelo modelo para três participantes diferentes. ID: 12122, 18846 e 16590 . . . . .	51
Figura 20 – Gráfico de dispersão dos valores preditos de glicose para cada segmento de 10 segundos, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL. Linha 1:1 e Linha de melhor ajuste realçadas respectivamente em vermelho tracejado e verde sólido. . . . .	65
Figura 21 – Gráfico de Clarke dos valores preditos de glicose para cada segmento de 10 segundos, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL.	66
Figura 22 – Gráfico de segmento dos valores reais e preditos pelo modelo para três participantes diferentes. ID: 12795, 17692 e 10814 . . . . .	67

## LISTA DE TABELAS

Tabela 1 – Disposição do conjunto de dados MIMIC-BGL em termos de participantes e amostras para cada um dos subgrupos. . . . .	30
Tabela 2 – Exemplo de tabela estruturada para os dados do conjunto MIMIC-BGL. . . . .	30
Tabela 3 – Estatísticas dos participantes presentes no conjunto UFSC-BGL. . . . .	33
Tabela 4 – Dados estruturados de características extraídas de pulsos do sinal de PPG. . . . .	42
Tabela 5 – Resultados dos modelos utilizando o conjunto UFSC-BGL com 26 participantes no conjunto de teste. . . . .	46
Tabela 6 – Resultados dos modelos utilizando o conjunto MIMIC-BGL com 576 amostras. . . . .	46
Tabela 7 – Resultados dos modelos utilizando o conjunto UFSC+MIMIC com 602 amostras. . . . .	46
Tabela 8 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke para o conjunto UFSC-BGL. . . . .	50
Tabela 9 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke para o conjunto MIMIC-BGL. . . . .	50
Tabela 10 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke para o conjunto UFSC+MIMIC. . . . .	50
Tabela 11 – Comparação entre resultados deste trabalho e outros autores. . . . .	55
Tabela 12 – Resultados dos modelos utilizando o conjunto UFSC-BGL para predições de segmentos de 10 segundos no conjunto de teste. . . . .	64
Tabela 13 – Resultados dos modelos utilizando o conjunto MIMIC-BGL para predições de segmentos de 10 segundos no conjunto de teste. . . . .	64
Tabela 14 – Resultados dos modelos utilizando o conjunto UFSC+MIMIC para predições de segmentos de 10 segundos no conjunto de teste. . . . .	64
Tabela 15 – Resultado da distribuição das predições de cada segmento no modelo de erro de Clarke em cada um dos conjuntos. . . . .	64
Tabela 16 – Características extraídas do segmento de 10 segundos do sinal de PPG e posteriormente utilizadas nos modelos de aprendizado de máquina. . . . .	68
Tabela 17 – Características extraídas dos pulsos do sinal de PPG e posteriormente utilizadas nos modelos de aprendizado de máquina. . . . .	69

## LISTA DE ABREVIATURAS E SIGLAS

AI	Inteligência Artificial ( <i>Artificial Intelligence</i> )
BGL	Nível de Glicose no Sangue ( <i>Blood Glucose Level</i> )
CEG	Clarke Error Grid
LED	Diodo Emissor de Luz ( <i>Light Emitting Diode</i> )
MAE	Erro Médio Absoluto ( <i>Mean Absolute Error</i> )
MAPE	Erro Médio Absoluto Percentual ( <i>Mean Absolute Percentage Error</i> )
ME	Erro Médio ( <i>Mean Error</i> )
MFCC	Coeficientes Cepstrais de Frequência de Mel ( <i>Mel Frequency Cepstral Coefficients</i> )
mg/dL	Miligramas por Decilitro
MIMIC III	<i>Multiparameter Intelligent Monitoring in Intensive Care III</i>
OMS	Organização Mundial de Saúde ( <i>World Health Organization</i> )
PPG	Fotopletismografia ( <i>Photoplethysmography</i> )
r <sup>2</sup>	Coeficiente de Determinação
RF	<i>Random Forest</i>
RMSE	Raiz do Erro Médio Quadrático ( <i>Root Mean Square Error</i> )
SSQI	Índice de Qualidade do Sinal baseado em Assimetria ( <i>Skewness Signal Quality Index</i> )
SVM	<i>Support Vector Machine</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
1.1	OBJETIVOS	16
1.1.1	<b>Objetivos Específicos</b>	<b>17</b>
1.1.2	<b>Organização do texto</b>	<b>17</b>
<b>2</b>	<b>REVISÃO TEÓRICA</b>	<b>18</b>
2.1	FOTOPLETISMOGRAFIA	18
2.2	PROCESSAMENTO DE SINAIS	18
<b>2.2.1</b>	<b>Filtro Savitzky-Golay</b>	<b>19</b>
<b>2.2.2</b>	<b>Filtro Butterworth</b>	<b>20</b>
<b>2.2.3</b>	<b>Normalização Min Max</b>	<b>20</b>
2.3	INTELIGÊNCIA ARTIFICIAL	21
2.4	MÉTRICAS DE AVALIAÇÃO	22
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>25</b>
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	<b>27</b>
4.1	<i>CONJUNTOS DE DADOS</i>	27
<b>4.1.1</b>	<b>MIMIC-BGL</b>	<b>27</b>
<b>4.1.2</b>	<b>UFSC-BGL</b>	<b>30</b>
4.1.2.1	Protocolo	31
4.2	MÉTODO	33
<b>4.2.1</b>	<b>Janelamento</b>	<b>33</b>
<b>4.2.2</b>	<b>Pré-processamento</b>	<b>34</b>
<b>4.2.3</b>	<b>Avaliação da qualidade do sinal</b>	<b>35</b>
<b>4.2.4</b>	<b>Extração de características</b>	<b>36</b>
<b>4.2.5</b>	<b>Particionamento de dados</b>	<b>37</b>
<b>4.2.6</b>	<b>Aprendizado de Máquina</b>	<b>38</b>
<b>4.2.7</b>	<b>Outras abordagens</b>	<b>39</b>
4.2.7.1	<i>Mel Frequency Cepstral Coefficients</i>	39
4.2.7.2	<i>Spectrogram</i>	40
4.2.7.3	<i>Scalogram</i>	40
<b>5</b>	<b>RESULTADOS</b>	<b>41</b>
<b>6</b>	<b>DISCUSSÃO</b>	<b>52</b>
<b>7</b>	<b>CONCLUSÃO</b>	<b>57</b>
	<b>REFERÊNCIAS</b>	<b>59</b>
	<b>APÊNDICE A – RESULTADOS AVALIANDO SEGMENTOS DE 10 SEGUNDOS</b>	<b>64</b>
	<b>APÊNDICE B – CARACTERÍSTICAS EXTRAÍDAS DO SINAL DE FOTOPLETISMOGRAFIA</b>	<b>68</b>

<b>APÊNDICE C – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) . . . . .</b>	<b>70</b>
<b>APÊNDICE D – QUESTIONÁRIOS APLICADOS DURANTE A COLETA DE DADOS . . . . .</b>	<b>75</b>

## 1 INTRODUÇÃO

A Diabetes Mellitus é uma doença crônica que afeta milhões de pessoas em todo o mundo e, atualmente, não possui cura. Essa condição ocorre quando o pâncreas não é capaz de produzir insulina ou quando o corpo não consegue utilizá-la adequadamente (DEFRONZO; BONADONNA; FERRANNINI, s.d.). A insulina, um hormônio peptídico secretado pelas células  $\beta$  do pâncreas, desempenha um papel fundamental na regulação dos níveis de glicose no sangue (BGL). Sua principal função é facilitar a captação de glicose pelas células, que será utilizada posteriormente como fonte de energia (WILCOX, 2005). A glicose é uma molécula de açúcar utilizada pelo organismo para produzir energia, sendo encontrada em diversos alimentos, como frutas, vegetais, grãos e produtos lácteos. Uma vez que as moléculas de glicose entram na corrente sanguínea, são transportadas para as células do corpo, onde são convertidas em energia. Além disso, a glicose também pode ser armazenada no fígado e nos músculos na forma de glicogênio. Essas reservas de glicose são liberadas quando o corpo precisa de energia adicional, como durante exercícios intensos ou quando o corpo fica períodos prolongados sem alimento (SOCIEDADE BRASILEIRA DE DIABETES, 2023; (WHO), 2016; MARATHE; GAO; CLOSE, 2017).

A diabetes é caracterizada por uma regulação inadequada dos níveis de glicose, o que pode resultar em diversas complicações, como doenças cardíacas ou renais (SOCIEDADE BRASILEIRA DE DIABETES, 2023; (WHO), 2016; ISLAM *et al.*, 2021). De acordo com a Organização Mundial de Saúde (OMS), o número de pessoas portadoras de diabetes cresceu de 108 milhões em 1980 para 422 milhões em 2014. Além disso, é uma das principais causas de cegueira, insuficiência renal, ataques cardíacos, acidente vascular cerebral e amputação de membros inferiores. Atualmente, é a 9ª maior entre as causas de mortes ao redor do mundo, com 1,5 milhões de casos diretamente causados pela diabetes (WORLD HEALTH ORGANIZATION, 2020).

O método de referência para aferição da glicose (*Gold standard*) é a análise laboratorial de sangue. Neste método, uma amostra de sangue é coletada do paciente e processada em um laboratório. Existem dois tipos de testes laboratoriais, o da glicose no sangue em jejum (FPG) e o de hemoglobina glicada (HbA1c), este último sendo o mais preciso (SACKS, 2011). No processo, a amostra de sangue é analisada para determinar a concentração exata de glicose presente. Embora seja um método que envolve custos elevados, tanto financeiro como de tempo, em comparação com outros dispositivos de monitoramento de glicose, o teste de glicose em laboratório é amplamente utilizado devido à sua precisão e confiabilidade. Tais características, tornam este método extremamente relevante em contextos clínicos e de pesquisa, permitindo diagnóstico e tratamento adequados, especialmente para portadores de diabetes.

Embora os métodos de análise laboratorial sejam considerados os mais pre-



cisos, existem situações em que o tempo é crucial, e a precisão se torna um fator secundário. Nesses contextos, o método tradicional que utiliza tiras reagentes torna-se relevante. Esse método emprega um dispositivo de menor custo, se comparado à análise laboratorial, conhecido como glicosímetro para aferir os níveis de glicose. No entanto, é importante ter em mente que esse procedimento requer cuidados adicionais, pois muitas vezes é feito pelo próprio usuário. Tarefa esta que envolve a coleta de uma amostra de sangue do paciente (VASHIST *et al.*, 2011). Além do glicosímetro, é necessário o uso de tiras reagentes descartáveis para a coleta da amostra de sangue, bem como um dispositivo de punção (lanceta) e material médico básico, como álcool, luvas e curativo, para prevenir infecções.

O método tradicional possui muitas vantagens, como a simplicidade e o tempo necessário para uma aferição. No entanto, ainda é um método invasivo onde é necessário a perfuração, geralmente na ponta de um dedo da mão, para extração da amostra de sangue capilar. Este procedimento se torna quase insuportável para pessoas que necessitam aferir seus níveis de glicose três ou mais vezes ao dia, todos os dias. Uma alternativa para este problema são monitores de uso contínuo, como por exemplo o desenvolvido pela empresa Abbott®. O FreeStyle Libre é um dispositivo composto por um pequeno sensor que é colocado na parte posterior do braço do indivíduo. Esse sensor possui uma fina agulha que é inserida na pele, permitindo a medição dos níveis de glicose presentes no fluido intersticial, o líquido que envolve as células da pele.

Os três métodos apresentados possuem suas vantagens e desvantagens, porém todos compartilham da mesma premissa, dispositivos invasivos. Existe uma crescente demanda em dispositivos vestíveis focados para o monitoramento de condições de saúde de forma não invasiva (NASIRI, 2019). Estes dispositivos não têm a finalidade de efetuar diagnósticos, e sim de prover informações mínimas para o acompanhamento de parâmetros relacionados à saúde (biomarcadores).

Uma das principais ferramentas para aquisição de biomarcadores, através de tecnologias vestíveis, é a técnica de Fotopletismografia (PPG, do inglês). Através da fotopletismografia é possível extrair informações relacionadas a variação do fluxo sanguíneo, e por consequência, da atividade cardíaca. O PPG é uma técnica de baixo custo, não invasiva, baseada na iluminação de uma região do corpo humano que contém vasos sanguíneos, com uma mínima camada de tecido e gordura, como por exemplo um dos dedos da mão (LIANG *et al.*, 2018).

## 1.1 OBJETIVOS

O objetivo deste projeto é a implementação de um modelo computacional para estimativa dos níveis de glicose em termos de miligramas por decilitro (mg/dL) de forma não invasiva e contínua através do sinal de Fotopletismografia.

### 1.1.1 Objetivos Específicos

Dentre os objetivos específicos deste trabalho, é possível citar:

1. Aquisição de um conjunto de dados que engloba participantes com glicose alterada e normal, onde duas variáveis são mensuradas, o sinal de fotopleletismografia e o valor de referência de glicose em mg/dL.
2. Analisar técnicas de processamento de sinais voltadas para a filtragem e remoção de ruído do sinal de fotopleletismografia.
3. Analisar características do sinal de fotopleletismografia que possam sofrer alterações relacionadas aos níveis de glicose.
4. Pesquisar e implementar abordagens relacionadas a estimativa da glicose através do sinal de fotopleletismografia.

### 1.1.2 Organização do texto

Este trabalho propõe a construção de um modelo computacional para estimativa não invasiva dos níveis de glicose através do sinal de fotopleletismografia. A composição deste trabalho se apresenta em 7 capítulos. O capítulo 2 apresenta uma breve revisão teórica, com o objetivo de esclarecer ao leitor os conceitos explorados neste trabalho.

No capítulo 3, é apresentada uma revisão do estado atual da literatura relacionada ao objetivo deste trabalho. Em seguida, o capítulo 4 descreve a metodologia adotada, incluindo detalhes dos blocos que compõem este projeto e as abordagens desenvolvidas neste estudo. Além disso, são fornecidas informações sobre as bases de dados utilizadas e a organização dos dados.

O capítulo 5, por sua vez, apresenta os resultados obtidos através de gráficos e tabelas que buscam descrever os experimentos realizados. O que segue com uma discussão no capítulo 6, onde o leitor é apresentado a uma análise que busca realçar e elucidar os resultados descritos.

O capítulo 7 conclui o trabalho revisando brevemente os objetivos propostos em contraste aos resultados obtidos, além de promover um paralelo com possíveis aplicações desta pesquisa.

## 2 REVISÃO TEÓRICA

Este capítulo tem por objetivo elucidar os principais conceitos tratados neste trabalho. Primeiramente, uma breve contextualização sobre a fonte primária de dados deste trabalho, o sinal de fotopletismografia. Além das diferentes técnicas de processamento de sinais utilizadas para remover artefatos deste sinal. Por fim, este trabalho busca situar o leitor no âmbito dos algoritmos de inteligência artificial e as métricas utilizadas para avaliar sua performance neste trabalho.

### 2.1 FOTOPLETISMOGRAFIA

Fotopletismografia é um método não invasivo usado para extrair sinais fisiológicos do sistema cardiovascular humano (ELGENDI *et al.*, 2018). Baseado na iluminação de uma área do corpo, normalmente a ponta do dedo, ou o pulso, onde é refletida a variação do fluxo sanguíneo entre as fases sistólica e diastólica do ciclo cardíaco do paciente (UFP, 2017).

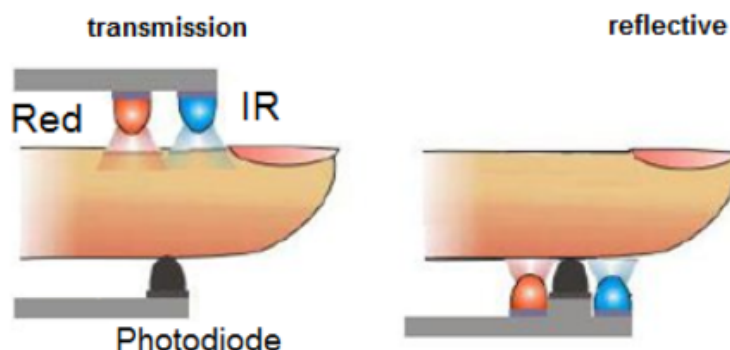
O PPG produz um sinal óptico relacionado a pulsações volumétricas arteriais, é composto por uma componente pulsátil (AC), atribuída às mudanças cardíacas sincronizadas no volume sanguíneo a cada batimento cardíaco. Esta componente é sobreposta a uma linha de base de variação lenta (DC), com vários componentes de baixa frequência atribuídos à respiração, atividade do sistema nervoso simpático e termorregulação (ALLEN, 2007).

Esta técnica é considerada como tendo um grande potencial em aplicações clínicas devido à sua simplicidade, por ser um método não invasivo e por seu baixo custo monetário (KAVSAOGLU; POLAT; BOZKURT, 2016). A construção desse sistema pode ser de duas formas: Através de um Diodo Emissor de Luz (LED, do inglês), como demonstra a Fig.1, onde um transmissor é colocado em uma das extremidades e um fotodiodo receptor na outra, ou através do princípio da reflexão, onde ambos os componentes estão no mesmo lado.

### 2.2 PROCESSAMENTO DE SINAIS

Os sinais de PPG são obtidos por meio de um processo que envolve a incidência de luz, seja de forma direta ou indireta no corpo humano. No entanto, diversas variáveis, como taxa de gordura corporal, coloração da pele e até mesmo a respiração, podem introduzir ruídos significativos no sinal de PPG, tornando inviável a utilização do sinal em sua forma bruta (*raw*). Portanto, é necessário realizar um estágio de pré-processamento, no qual técnicas de processamento de sinais são aplicadas. Após este estágio, o sinal estará preparado para a extração de características relevantes (*features*).

Figura 1 – Esquemático para aquisição do sinal de PPG por meio de transmissão e reflexão à esquerda; representação da variação da luz através dos tecidos do corpo humano à direita.



Fonte: Adaptado de (TAMURA *et al.*, 2014).

Um estudo proposto por (ELGENDI *et al.*, 2018), buscou avaliar o impacto e performance de filtros digitais aplicados ao sinal de PPG. A avaliação contou com nove (9) tipos de filtros com dez (10) diferentes ordens para cada um, resultando em 90 configurações diferentes. Os autores concluíram então, que o filtro Chebyshev II seguido pelo filtro Butterworth, ambos de 4<sup>a</sup> ordem, obtiveram a melhor performance, baseado no indicador SSQI (*Skewness Signal Quality Index*) proposto no estudo. A Eq. (1) demonstra o método de avaliação.

$$SSQI = \frac{1}{n} \times \sum_{i=1}^n \frac{[x_i - \mu_x]^3}{\sigma} \quad (1)$$

Onde “N” é o numero de pontos do sinal de PPG,  $\mu$  e  $\sigma$  são respectivamente a média e o desvio padrão de  $x_j$ .

Entretanto, nos testes realizados neste trabalho, constatou-se que a técnica que obteve o melhor desempenho na redução de ruídos dos sinais de PPG foi uma combinação dos filtros Savitzky-Golay (SAVITZKY; GOLAY, 1964) e Butterworth (SELESNICK; BURRUS, 1998). A configuração utilizada no filtro Savitzky-Golay foi de 19 *frames* e 4<sup>a</sup> ordem, enquanto no filtro Butterworth foi de 4<sup>a</sup> ordem com frequência de corte entre 0,5 Hz e 15 Hz.

### 2.2.1 Filtro Savitzky-Golay

Proposto por Abraham Savitzky e Marcel Golay, o filtro Savitzky-Golay é um tipo de filtro digital que pode ser aplicado a um conjunto pontos com o objetivo de suavizar os dados, ou seja, aumentar a precisão sem distorcer a tendência (SAVITZKY; GOLAY, 1964).

Como descrito por Schafer (SCHAFER, 2011), o filtro é baseado na aproximação polinomial dos mínimos quadrados. A base do estudo mostra que ajustar um polinômio a um conjunto de amostras de entrada e depois avaliar o polinômio resultante em um único ponto dentro do intervalo de aproximação é equivalente a uma convolução discreta com uma resposta de impulso fixa.

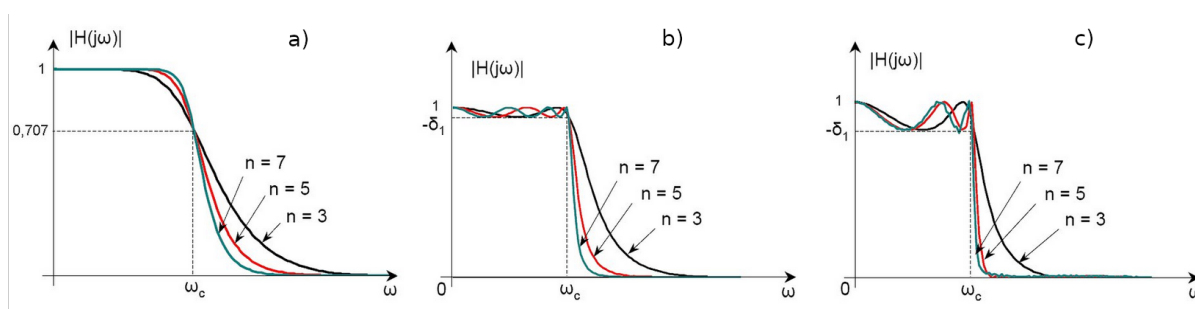
### 2.2.2 Filtro Butterworth

O filtro do tipo Butterworth é amplamente utilizado na área de eletrônica e processamento digital de sinais devido suas características de frequência e fase. Proposto inicialmente pelo engenheiro britânico Stephen Butterworth (SELESNICK; BURRUS, 1998), foi desenvolvido com o objetivo de ter uma resposta em frequência o mais plana o quanto for matematicamente possível na banda passante.

A resposta em frequência do filtro passa-baixas Butterworth é uma função da sua ordem “n” e da frequência de corte  $\omega_c$ , descrito na equação (2) (MIHAI, B.; MIHAI, P., 2015). A Fig.2 por sua vez, demonstra uma comparação da resposta em frequência em três diferentes ordens para os filtros digitais Butterworth, Chebyshev I e Elíptico, semelhantes e amplamente usados nesta categoria.

$$|H(j\omega)| = \frac{1}{\sqrt{1 + (\frac{\omega}{\omega_c})^{2n}}} \quad (2)$$

Figura 2 – Exemplo de resposta em frequência para os filtros: a) Butterworth b) Chebyshev I c) Elíptico.



Fonte: Adaptado de (MIHAI, B.; MIHAI, P., 2015).

### 2.2.3 Normalização Min Max

A última etapa de pré-processamento empregada, é a normalização do sinal, esta técnica faz-se necessária, para evitar diferentes magnitudes em amplitude.

O sinal de PPG foi normalizado entre 0 e 1 através da técnica de normalização “Min Max” que como o nome sugere, utiliza o valor mínimo e máximo de amplitude do sinal como parâmetros, como descreve a Eq.(3)

$$PPGn[i] = \frac{PPGf[i] - \min(PPGf)}{\max(PPGf) - \min(PPGf)} \quad (3)$$

- **PPGn**: PPG normalizado
- **PPGf**: PPG filtrado

### 2.3 INTELIGÊNCIA ARTIFICIAL

Os algoritmos de Inteligência Artificial (AI, do inglês) têm desempenhado um papel crucial na área biomédica, proporcionando avanços significativos e promissores. Esses algoritmos encontram aplicação em uma diversidade de áreas e cenários, como diagnóstico médico, descoberta de medicamentos, monitoramento de saúde, análise de imagens médicas e previsão de doenças. Em diagnóstico médico, por exemplo, os algoritmos de AI são capazes de analisar grandes quantidades de dados clínicos e históricos do paciente para auxiliar médicos na tomada de decisões precisas e rápidas (JIANG *et al.*, 2017; MAGRABI *et al.*, 2019).

Além disso, algoritmos de AI são utilizados na análise de imagens médicas, permitindo a detecção precoce de doenças e o auxílio na interpretação de exames radiológicos (GORE, 2020; HOSNY *et al.*, 2018; DUONG *et al.*, 2019). Essas tecnologias têm o potencial de melhorar a eficiência dos sistemas de saúde, personalizar tratamentos e melhorar os resultados para os pacientes. Com a contínua evolução da AI, espera-se que esses algoritmos desempenhem um papel cada vez mais importante no avanço da medicina e na melhoria da qualidade de vida das pessoas.

Enquanto a inteligência artificial pode ser definida, de modo amplo, como a ciência capaz de mimetizar as habilidades humanas, o aprendizado de máquina (*Machine Learning*) é uma vertente específica da AI que treina máquinas para aprender com dados. Na atualidade, não há recurso maior em abundância do que dados estruturados e não estruturados. Na segunda metade do século 20, o aprendizado de máquina evoluiu como uma sub-área da AI com algoritmos de auto aprendizagem que derivam conhecimento a partir de dados para fazer previsões futuras (RASCHKA; MIRJALILI, 2019).

Em vez de programadores construírem rotinas manualmente, o aprendizado de máquina oferece uma alternativa mais eficiente para adquirir conhecimento a partir de dados, para gradualmente melhorar a performance das previsões. Neste trabalho, a linguagem de programação Python na sua versão 3.8 foi utilizada para o desenvolvimento, bem como o pacote Scikit-Learn versão 1.2.2 (PEDREGOSA *et al.*, 2011) para os modelos de inteligência artificial.

## 2.4 MÉTRICAS DE AVALIAÇÃO

Independentemente do modelo empregado para a estimativa da glicose, é essencial utilizar métricas apropriadas para avaliar o desempenho do mesmo. Essas métricas proporcionam uma base objetiva para medir a eficácia e a qualidade dos resultados obtidos. Ao utilizar métricas de avaliação, é possível comparar diferentes modelos, identificar suas limitações e realizar melhorias. Além disso, as métricas auxiliam na tomada de decisões embasadas, permitindo a seleção do modelo mais adequado para atingir os objetivos desejados.

Dessa forma, a utilização de métricas de avaliação é um componente essencial em qualquer análise de desempenho de modelo, contribuindo para aprimorar a eficiência e a confiabilidade dos resultados obtidos. Neste trabalho, as seguintes métricas foram utilizadas:

- **Mean Error (ME):** O erro médio descreve o valor médio entre a diferença do valor real com o valor estimado Eq.(4).

$$ME = \frac{1}{n} \times \sum_{i=1}^n y_i - \hat{y}_i \quad (4)$$

- **Mean Absolute Error (MAE):** O erro médio absoluto descreve o valor médio entre a diferença absoluta do valor real com o valor estimado Eq.(5).

$$MAE = \frac{1}{n} \times \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

- **Mean Absolute Percentage Error (MAPE):** O erro absoluto percentual descreve o erro em termos de uma porcentagem em relação ao valor de referência. Eq.(6).

$$MAPE = \frac{1}{n} \times \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (6)$$

- **Root Mean Squared Error (RMSE):** A raiz do erro quadrático médio, é uma medida quadrática que busca realçar erros extremos que seriam suavizados pela média em uma amostra grande, Eq.(7).

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

- **Coefficient of determination (r2):** O coeficiente de determinação avalia a proporção da variação da variável dependente que pode ser explicada pela

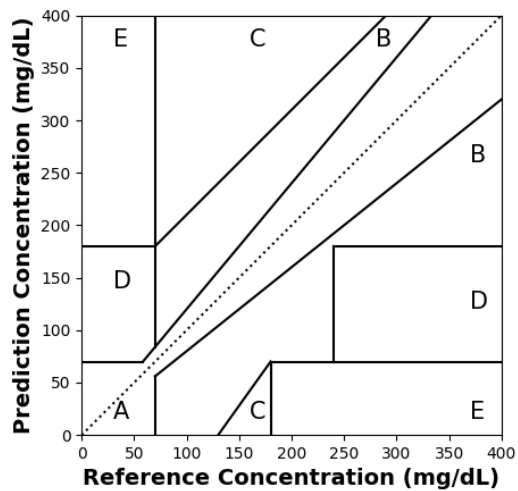
variável independente. O coeficiente varia de 0 a 1, onde 0 indica que o modelo não é capaz de explicar a variabilidade dos dados e 1 indica que o modelo se ajusta perfeitamente aos dados.

- **Clarke Error Grid (CEG):** O método de erro de Clarke (CLARKE, William L *et al.*, 1987)(CLARKE, William L., 2005)(SENGUPTA *et al.*, 2021) é uma métrica utilizada para avaliar a precisão de sistemas de monitoramento contínuo de glicose. Neste método, a partir de um gráfico de dispersão, zonas de precisão são delimitadas com o objetivo de avaliar as estimativas categoricamente, como demonstra a Fig.3:

- Zona A: É a faixa de maior precisão, inclui valores com uma discrepância pequena, dentro de  $\pm 20\%$  em relação aos valores de referência. Indica um ajuste adequado, uma estimativa nesta zona não modifica uma ação clínica.
- Zona B: Nesta zona, os valores de erro estão em uma faixa de  $\pm 20\%$  a  $\pm 40\%$  em relação aos valores de referência. Essa zona ainda é aceitável, mas indica uma precisão um pouco inferior. Representa uma alteração de ação clínica, como modificação de medicação, mas não tem efeito na alteração do diagnóstico.
- Zona C: Nesta zona, os valores de erro estão em uma faixa de  $\pm 40\%$  a  $\pm 60\%$  em relação aos valores de referência. Essa zona indica uma precisão moderada. Representa uma alteração de ação clínica como prescrição de medicamentos, também pode alterar o diagnóstico do paciente.
- Zona D: Nesta zona, os valores de erro estão em uma faixa de  $\pm 60\%$  a  $\pm 80\%$  em relação aos valores de referência. Essa zona indica uma precisão baixa. Representa uma alteração de ação clínica na prescrição de medicamentos, e pode ter risco clínico significativo.
- Zona E: Nesta zona, a discrepância é a maior de todas. Ela engloba valores que estão fora da faixa de  $\pm 80\%$  em relação aos valores de referência. Essa zona indica uma falta de precisão significativa. Representa uma alteração de ação clínica, e pode ter consequências perigosas.



Figura 3 – Exemplo de gráfico de erro de Clarke.



Fonte: Adaptado de (CLARKE, William L *et al.*, 1987).

### 3 REVISÃO BIBLIOGRÁFICA

Neste capítulo, será apresentada uma revisão da literatura relacionada ao tema deste trabalho. Serão discutidos diferentes estudos em que diversos autores utilizaram técnicas e abordagens variadas para estimar os níveis de glicose. O objetivo é fornecer uma visão geral das pesquisas anteriores e destacar as contribuições e descobertas relevantes nessa área.

É do entendimento dos autores que, até o momento, o número de estudos relacionados à estimativa não invasiva dos níveis de glicose é limitado. Além disso, a maioria dos estudos publicados utilizou dados não disponíveis publicamente. Essa escassez de pesquisas e a falta de disponibilidade de dados abertos representam um desafio para avançar nessa área e desenvolver abordagens mais robustas e generalizáveis. As abordagens utilizadas para estimativa dos níveis de glicose variam desde a utilização exclusiva do sinal de fotopletismografia até a combinação desse sinal com outras informações biomédicas.

Yen *et al.* (YEN *et al.*, 2022), buscou estimar os valores de glicose, em termos de mg/dL, através de características extraídas do sinal de PPG utilizando dois comprimentos de ondas: vermelho (660 nm) e infra-vermelho (900 nm), juntamente com o sinal de impedância bioelétrica. Ao total, 30 atributos foram extraídos de 40 participantes, sendo 12 do sinal de PPG, 11 do sinal de bioimpedância e 7 de características do participante, como idade, altura e peso. Entretanto, os valores de hemoglobina (g/L) também foram usados no treinamento do modelo, o que pode ter um alto impacto no resultado final. Contudo, o modelo de AI utilizado foi uma rede neural com topologia *back propagation* que obteve erro médio absoluto de 5,09 mg/dL e fator de correlação  $r^2$  de 0,99.

Em um estudo anterior conduzido por Habbu *et al.* (HABBU *et al.*, 2019), foi utilizada uma abordagem baseada em uma rede neural totalmente conectada para estimar os níveis de glicose. Sinais de PPG de 3 minutos de 611 pacientes foram coletados durante o estudo, dos quais foram obtidas janelas de 1 minuto com 50% de sobreposição. A partir dos segmentos de PPG, coeficientes cepstrais foram extraídos usando o espectro de frequências. Um banco de filtros de mel (MFB) com 15 filtros foi utilizado para obter as energias cepstrais. A Transformada Discreta do Cosseno (DCT) foi aplicada ao logaritmo dos coeficientes MFB, resultando nos coeficientes cepstrais. Os 8 primeiros coeficientes foram selecionados, juntamente com a média, desvio padrão e variância, formando um vetor de 24 características que foram utilizadas como entrada para a rede neural. O fator de correlação e o gráfico de Clarke foram utilizados como métrica. Resultados demonstraram  $r^2$  de 0,90 e 85,20% das predições concentradas na zona A do gráfico de Clarke.

De forma semelhante, Silverio *et al.* (ALONSO-SILVERIO *et al.*, 2021) tam-

bém se propuseram a estimar os níveis de glicose utilizando os coeficientes cepstrais obtidos a partir do sinal de PPG, seguindo o método mencionado por Habbu *et al.* (HABBU *et al.*, 2019). Neste estudo, os autores realizaram uma comparação de métodos, tanto na extração de características quanto nos modelos de aprendizado de máquina. Foram utilizados três métodos diferentes para a extração de características: os coeficientes cepstrais (MFCC), combinados com a redução de dimensionalidade por Análise de Componentes Principais (PCA) e também com Análise de Componentes Independentes (ICA). Na etapa de treinamento, 10 modelos de aprendizado de máquina foram testados. Os dados utilizados neste estudo foram coletados de 122 participantes, sendo 72 deles utilizados para o treinamento e 50 para a validação. O treinamento dos modelos foi realizado utilizando a técnica de *grid search* com validação cruzada de 5 iterações nos 72 participantes de treinamento. Os resultados foram obtidos a partir dos 50 participantes restantes, onde o melhor modelo encontrado foi o algoritmo AdaBoost utilizando a técnica de MFCC, com um MAE reportado de 11,62 mg/dL e 92% das amostras localizadas na zona A do gráfico de Clark.

Por outro lado, Zhang *et al.* (ZHANG *et al.*, 2019) e, mais recentemente, Islam *et al.* (ISLAM *et al.*, 2021), investigaram a relação entre os níveis de glicose e o sinal de rPPG. O sinal de rPPG é semelhante ao sinal de PPG, com a diferença na forma de aquisição. Enquanto o sinal de PPG é extraído por meio de um sensor óptico e uma fonte de luz, como um LED, o sinal de rPPG é obtido através da câmera de um celular, onde o dedo é posicionado sobre o *flash* e a câmera, e um vídeo é capturado. Através da análise das variações de cor nos quadros (*frames*) do vídeo, é possível extrair o sinal de rPPG. Zhang *et al.* coletaram sinais de 14 voluntários, enquanto Islam *et al.* coletaram sinais de 22 participantes. Na etapa de extração de características, ambos os estudos utilizaram características morfológicas do sinal, sendo que Islam *et al.* também utilizaram as características da primeira e segunda derivadas do sinal. Em relação aos resultados, Zhang *et al.* alcançaram uma acurácia de 86,2% utilizando o modelo KNN para classificar os participantes em dois grupos: normal e com diabetes. Por outro lado, Islam *et al.* utilizaram o método de Regressão Parcial dos Mínimos Quadrados (PLS) e obtiveram um Erro Médio de 17,02 mg/dL.

Outra abordagem é a de Guzman *et al.* (GUZMAN; CAZARES; MARTINEZ-TORTEYA, 2021), que utilizaram o sinal de PPG para extrair características relacionadas a variabilidade cardíaca (HRV). Entretanto, dados de apenas 8 participantes foram coletados. Foram explorados os domínios do tempo, frequência e não-linear, juntamente com informações demográficas como idade, gênero, altura, peso. Além disso, medidas fisiológicas foram utilizadas, como pressão arterial e saturação de oxigênio. Um total de 96 atributos foram extraídos, onde o conjunto de dados foi dividido em 80% para treino e 20% para teste. Resultados demonstraram que o modelo com a melhor performance foi o SVM com erro médio absoluto de 16,45 mg/dL.

## 4 MATERIAIS E MÉTODOS

Este capítulo descreve a metodologia empregada no desenvolvimento deste projeto. Primeiramente, são apresentadas as bases de dados utilizadas, seguidas por uma visão geral do método proposto. Cada um dos principais blocos desse método será detalhadamente abordado, destacando seu objetivo principal, que é a estimativa dos níveis de glicose.

### 4.1 CONJUNTOS DE DADOS

A disponibilidade de um volume adequado de dados para o treinamento é, sem dúvida, um dos componentes essenciais para qualquer modelo de inteligência artificial baseado em aprendizado supervisionado. Para este trabalho, foram selecionadas duas bases de dados, a primeira sendo publicamente disponível extraída do extenso *Multiparameter Intelligent Monitoring in Intensive Care III* (MIMIC III), denominado por este trabalho como: MIMIC-BGL (JOHNSON *et al.*, 2016) (ARÉVALO *et al.*, 2021). Enquanto a segunda foi coletada durante esta pesquisa através da Universidade Federal de Santa Catarina (UFSC) e do Hospital Universitário Polydoro de São Thiago da Universidade Federal de Santa Catarina (HU-UFSC/EBSERH), denominado UFSC-BGL.

#### 4.1.1 MIMIC-BGL

O MIMIC-III é um conjunto de dados de saúde disponibilizado pelo projeto PhysioNet (JOHNSON *et al.*, 2016). Ele contém informações clínicas detalhadas de pacientes admitidos nas unidades de terapia intensiva (UTIs) do Beth Israel Deaconess Medical Center (BIDMC) entre 2001 e 2012.

MIMIC-III contém uma vasta quantidade de dados, com cerca de 40.000 pacientes adultos e mais de 70.000 registros. Esses registros incluem uma variedade de informações, como identificação de admissão, dados demográficos, diagnósticos, medições de sinais vitais, resultados de exames laboratoriais, prescrições de medicamentos, notas de enfermagem, entre outros. Entre os sinais vitais, tipicamente pode se encontrar: Eletrocardiograma (ECG), Pressão Arterial (ABP) e Fotopletismografia, juntamente com registros numéricos, como: Oxigenação Sanguínea (SpO<sub>2</sub>), Pressão Sistólica e Diastólica (SBP, DBP) e frequência cardíaca (HR).

Do que se refere a diversidade de dados, o conjunto abrange uma ampla gama de informações clínicas, como resultados de exames de imagem, dados de monitoramento invasivo (por exemplo, pressão arterial invasiva), dados de fluxo de informações (como transfusões e eventos de ventilação mecânica), escores de gravidade e índices de prognóstico, entre outros. Para proteger a privacidade dos pacientes, todas as informações pessoais identificáveis foram removidas ou anonimizadas do conjunto de

dados. No entanto, ainda é necessário seguir rigorosos protocolos de uso e solicitar acesso autorizado ao MIMIC-III.

O MIMIC-III é fornecido em formato de banco de dados SQL (*Structured query language*), facilitando a consulta e extração de informações específicas. Além disso, é possível fazer o download de cada arquivo específico separadamente através de uma API (*Application Programming Interface*). Todavia, para o download completo dos dados, é necessário 2.4 TB (Terabytes) de armazenamento. Para facilitar a identificação, os pacientes são separados em subgrupos através do primeiro número do identificador único. Os subgrupos vão de 1 a 9, são encontrados no grupo 1 pacientes com o identificador iniciando com o dígito 1, por exemplo 11861.

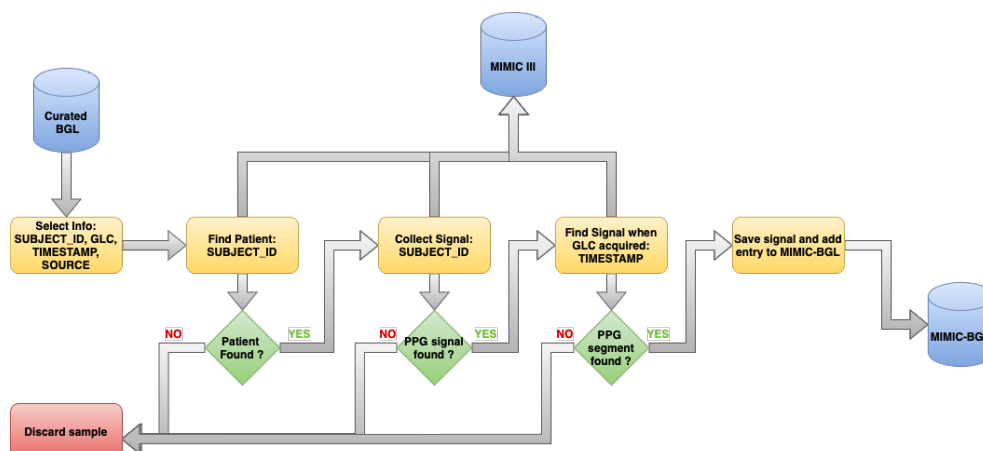
Dentre este imenso conjunto, também são descritos subconjuntos que reúnem informações específicas para um determinado problema, a partir das informações do MIMIC-III. Descrito em 2021 por (ARÉVALO *et al.*, 2021), aqui denominado como Curated-BGL, este subconjunto contém cerca de 500.000 aferições de glicemia relativo a aproximadamente 9.600 pacientes.

No entanto, esse conjunto de dados não foi originalmente concebido para a estimativa não invasiva de glicose por meio de sinais biomédicos, que é o objetivo deste trabalho. O objetivo inicial foi descrever o tratamento de insulina em pacientes em tratamento intensivo. Dentre as informações presentes, é possível citar: O identificador único do paciente; identificador único da sua entrada na UTI; valor da aferição de glicose; data e hora da aferição, codificados para evitar qualquer relação com a identidade do paciente; tipo de exame, como fita reagente ou teste de laboratório; administração ou não de insulina, entre outros.

Devido à ausência de sinais biomédicos no conjunto de dados existente (Curated-BGL), este trabalho propôs criar um novo subconjunto de dados. Esta tarefa foi executada combinando as informações de glicemia provenientes de (ARÉVALO *et al.*, 2021) com os sinais de PPG disponíveis no MIMIC III (JOHNSON *et al.*, 2016). O desenvolvimento deste novo conjunto de dados, agora descrito como MIMIC-BGL, é possível devido ao campo "SUBJECT\_ID", o identificador único do paciente que está presente em ambos conjuntos. Os passos para a obtenção do novo conjunto são descritos abaixo, bem como o fluxograma da Fig.4 exemplifica graficamente.

- Para cada entrada no conjunto de dados Curated-BGL é selecionado as informações de: identificador do paciente ("SUBJECT\_ID"), data e hora da aferição da glicose (TIMESTAMP), a referência de glicose (GLC) e o método de aferição de glicose (SOURCE).
- Utilizando o "SUBJECT\_ID", é selecionado o paciente no conjunto de dados MIMIC III.
- Uma vez encontrado, é verificado se este paciente possui registro de sinal de PPG, caso contrário a leitura é descartada.

Figura 4 – Fluxograma de execução para o conjunto de dados MIMIC-BGL.



Fonte: Autor.

- No conjunto de dados MIMIC III os sinais biomédicos são contínuos, com dias de duração, também é registrada a frequência de amostragem, bem como a data e hora do início do monitoramento, novamente, ambas codificadas.
- Utilizando o TIMESTAMP da aferição de glicose, do conjunto de dados Curated-BGL, é verificado se neste determinado momento houve registro de PPG, caso contrário a leitura é descartada.
- Em caso positivo, são extraídos 10 minutos de sinal de PPG antes e depois do momento da leitura de glicose, totalizando aproximadamente 20 minutos de sinal para cada leitura.
- Caso não exista registro de sinal 10 minutos antes ou depois, o total disponível é extraído, contanto que seja maior do que 2 minutos no total.
- É importante ressaltar que um paciente tem múltiplas aferições, logo seu identificador é mantido através de todas as aferições.

Contudo, o conjunto de dados resultante, MIMIC-BGL, conta com 98.111 registros de 2.954 pacientes. A Tab.1 demonstra o número de amostras (SAMPLES), participantes (SUBJECTS) para cada um dos subgrupos do conjunto, bem como a porcentagem da distribuição de amostras com glicose maior ou igual a 200 mg/dL (GLC  $\geq$  200) e menor que 200 mg/dL (GLC  $<$  200).

Os respectivos sinais de PPG foram salvos em arquivos binários totalizando aproximadamente 120 GB de armazenamento, e suas referências foram salvas em arquivos estruturados como descrito na Tab.2.

O campo descrito como SIGNAME engloba o nome do arquivo binário que contém o sinal de PPG, a nomenclatura do arquivo leva o identificador único do paciente,

Tabela 1 – Disposição do conjunto de dados MIMIC-BGL em termos de participantes e amostras para cada um dos subgrupos.

SUB_DIR	SAMPLES	SUBJECTS	GLC < 200	GLC >= 200
p01	2.536	68	81%	19%
p02	3.798	106	80%	20%
p03	1.146	39	83%	17%
p04	15.125	456	84%	16%
p05	16.433	481	87%	13%
p06	14.742	475	87%	13%
p07	15.421	471	88%	12%
p08	14.180	435	88%	12%
p09	14.730	423	86%	14%

Fonte: Autor.

Tabela 2 – Exemplo de tabela estruturada para os dados do conjunto MIMIC-BGL.

SUBJECT_ID	FS	GLC	TIMESTAMP	SIGNAME	SOURCE
11861	125	149	2168-12-06 01:44:00	FILENAME	FINGERSTICK
14836	125	118	2139-10-09 09:16:00	FILENAME	BLOOD
13072	125	158	2194-01-22 22:22:00	FILENAME	BLOOD
10425	125	215	2199-08-30 17:00:00	FILENAME	FINGERSTICK

Fonte: Autor.

o valor de glicose, data e hora da aferição, no seguinte formato:

SUBJECT\_ID\_GLC\_YYYY\_MM\_DD\_hh\_mm\_ss.npy.

A Tabela 1 revela a abundância de amostras disponíveis. Essa vasta quantidade de dados é fundamental para o treinamento de modelos de AI. No entanto, também apresenta um desafio em termos de tempo de processamento e treinamento. Contudo, este trabalho optou por restringir a análise aos participantes do grupo "p01", utilizando apenas 2 minutos de sinal por amostra, em vez dos 20 minutos disponíveis.

#### 4.1.2 UFSC-BGL

O conjunto de dados UFSC-BGL foi desenvolvido e coletado especificamente para este projeto. O objetivo principal foi obter um conjunto de dados abrangendo participantes com níveis normais e alterados de glicose dentro do contexto hospitalar. Além disso, no momento da concepção deste projeto, não havia nenhum conjunto de dados publicamente disponível com referência na literatura. Portanto, a criação do UFSC-BGL preencheu essa lacuna e contribuiu para a disponibilidade de dados relevantes nessa área de pesquisa.

O projeto foi aprovado em 10 de outubro de 2022 pelo Comitê de Ética em Pesquisa com Seres Humanos (CEPSH-UFSC) disponível para consulta através do identificador CAAE 60111422.5.0000.0121. O estudo abrange 80 participantes, de ambos os sexos, adultos entre 18 e 75 anos de idade. O tamanho da amostra populacional

foi estimado através do *software* estatístico GPower (ERDFELDER; FAUL; BUCHNER, 1996). Foram coletados dados de 40 participantes com glicose considerada normal e 40 participantes com glicose considerada alterada.

O critério para classificação dos grupos foi desenvolvido baseado nas diretrizes brasileiras de diabetes para testes aleatórios (SOCIEDADE BRASILEIRA DE DIABETES, 2023). Participantes com valores de glicose maiores ou iguais a 120 mg/dL em jejum ou maiores ou iguais a 140 mg/dL sem jejum foram considerados no grupo diabético. Participantes com valores de glicose menores que 120 mg/dL em jejum ou menores que 140 mg/dL sem jejum serão considerados no grupo controle (saudáveis). Para a concepção deste novo conjunto, foi necessário estabelecer um protocolo para a coleta de dados.

#### 4.1.2.1 Protocolo

O protocolo desenvolvido estabelece que, inicialmente, os participantes são abordados por um membro da equipe de projeto, devidamente identificado, onde é apresentado o Termo de Consentimento Livre e Esclarecido (TCLE), descrito no Apêndice C.

Após a concordância, o participante é conduzido até as dependências do laboratório de Endocrinologia e Metabologia do HU-UFSC/EBSERH, onde é realizado a coleta de dados.

O participante permanece sentado em uma cadeira confortável onde deverá responder dois questionários. O primeiro, trata de informações demográficas, como: idade, altura, peso, sexo. Enquanto o outro é referente à condição de saúde, tal qual medicações e tratamentos em vigor, ambos descritos no Apêndice D.

Após responder os questionários, o sinal de fotopletismografia é coletado por cinco (5) minutos através de um oxímetro de pulso do tipo grampo, da empresa CONTEC, modelo CMS50E, conforme Fig.5a. O sensor é colocado no dedo indicador da mão direita do participante, caso não for possível, o dedo médio da mão direita é utilizado.

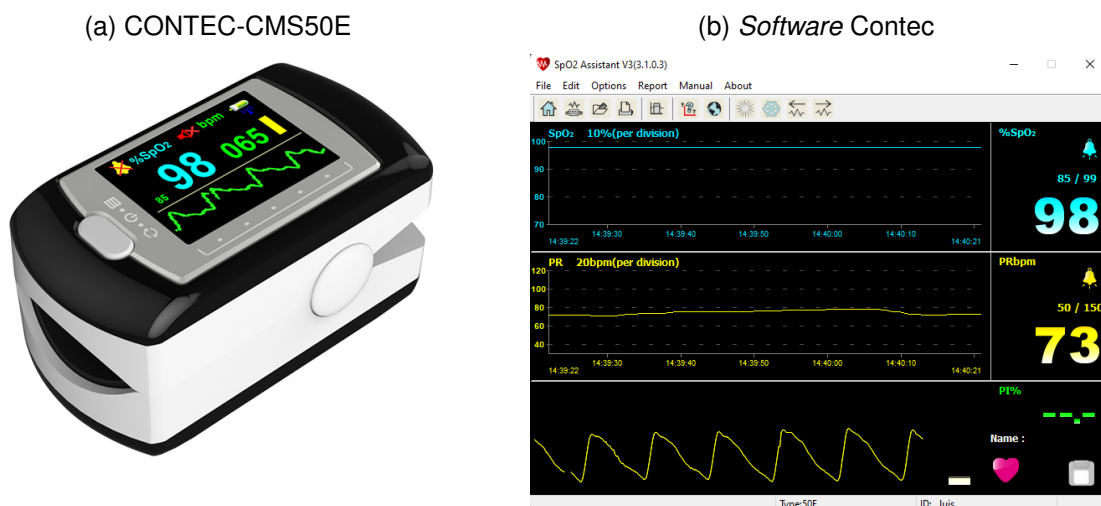
O oxímetro é plugado a um Tablet Acer por meio de conexão USB. Através do *software* proprietário da CONTEC Fig.5b, o sinal de PPG é coletado a uma taxa de amostragem de 60 Hz e salvo em um arquivo CSV (*Comma-separated values*).

Coletado o sinal de PPG, os participantes têm sua glicose aferida através de um glicosímetro digital por meio do método de tiras reagentes, conforme Fig.6. Uma amostra de sangue é extraída do participante através de uma lanceta introduzida no dedo indicador da mão esquerda, caso não seja possível, o dedo médio da mão esquerda é utilizado. A amostra de sangue é coletada por uma enfermeira credenciada com a tira reagente e introduzida no glicosímetro para obtenção do valor de glicemia em mg/dL.



Por fim, o participante recebe a devida assistência médica com esterilização e curativo no local da punção.

Figura 5 – Equipamento para extração do sinal de PPG, oxímetro CONTEC CMS50E e *software* de aquisição.



Fonte: Contec Health (CONTEC HEALTH, 2022).

Figura 6 – Equipamento para aferição da glicose, glicosímetro digital GTECH.



Fonte: GTECH (GTECH, 2022).

Em suma, dados de 80 participantes foram coletados, sendo 40 do grupo N01 (controle), considerados saudáveis e 40 do grupo D01 de teste (com glicose alterada). Um panorama geral das estatísticas demográficas dos dados pode ser encontrada na Tab.3, onde é demonstrado para cada grupo e no geral, o valor médio de glicose, idade e distribuição de gênero. A Fig.7a demonstra um histograma da distribuição dos níveis de glicose que compõem este conjunto, enquanto a Fig.7b demonstra um gráfico de

dispersão onde o eixo das abscissas se refere a idade dos participantes enquanto o eixo das ordenadas denota os valores de glicose.

Tabela 3 – Estatísticas dos participantes presentes no conjunto UFSC-BGL.

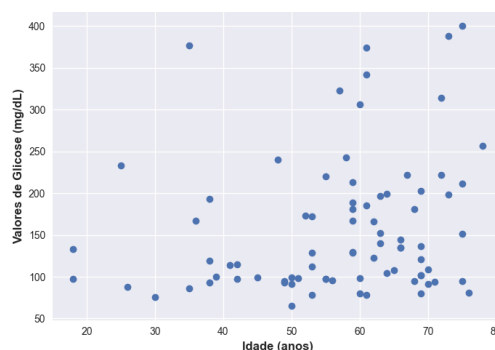
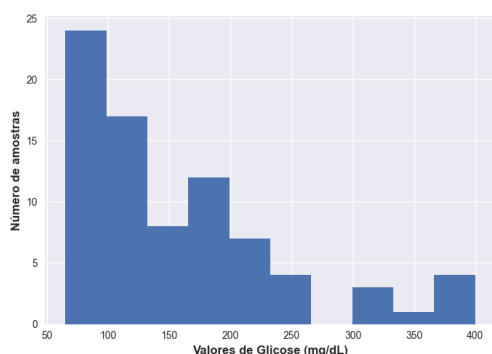
Grupo	Amostras	Glicose Média (mg/dL)	Idade Média	Gênero
<b>N01</b>	40	98,60	54,23	F 28 (70%) M 12 (30%)
<b>D01</b>	40	217,32	59,40	F 31 (77%) M 9 (23%)
<b>Geral</b>	80	157,96	56,81	F 59 (74%) M 21 (26%)

Fonte: Autor.

Figura 7 – Gráficos obtidos a partir do conjunto UFSC-BGL.

(a) Distribuição de valores de glicose.

(b) Correlação de valores de glicose com idade.



Fonte: Contec Health (CONTEC HEALTH, 2022).

## 4.2 MÉTODO

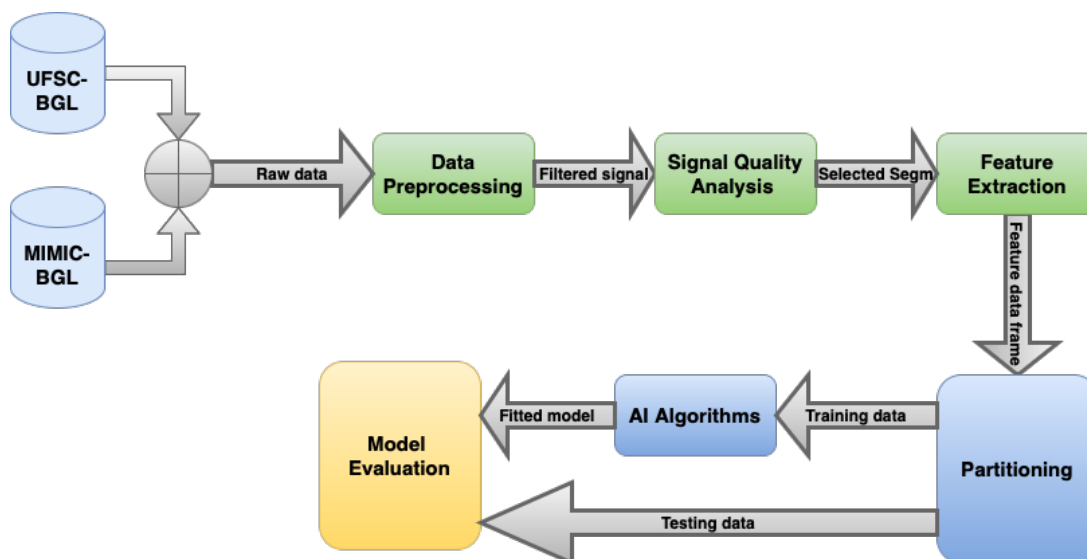
O método proposto pode ser resumido em seis etapas principais: Pré-processamento, Análise de qualidade do sinal, Extração de características, Particionamento de dados, Treinamento dos algoritmos de AI e Avaliação do método. A Fig.8 demonstra o fluxo-grama de execução deste trabalho.

### 4.2.1 Janelamento

Amplamente utilizada, a técnica de janelamento consiste em dividir um sinal de longa duração em segmentos menores. Neste trabalho, optou-se por janelas de 10 segundos devido a ser um tempo relativamente baixo para a aquisição de uma predição.

Esses segmentos podem ser processados individualmente, permitindo evitar a exclusão de um sinal inteiro devido a um trecho que contenha artefatos ou interferências. Além disso, é comum aplicar uma função matemática a cada segmento para

Figura 8 – Fluxograma de execução deste trabalho.



Fonte: Autor.

modificar o sinal. No entanto, neste trabalho, não foi utilizada nenhuma função sobre os segmentos. Entretanto, o primeiro e o último pulsos do sinal de PPG de cada segmento foram excluídos. O objetivo é eliminar qualquer inconsistência relacionada a divisão inicial dos segmentos, bem como introduzir um preenchimento de janela (*padding*).

#### 4.2.2 Pré-processamento

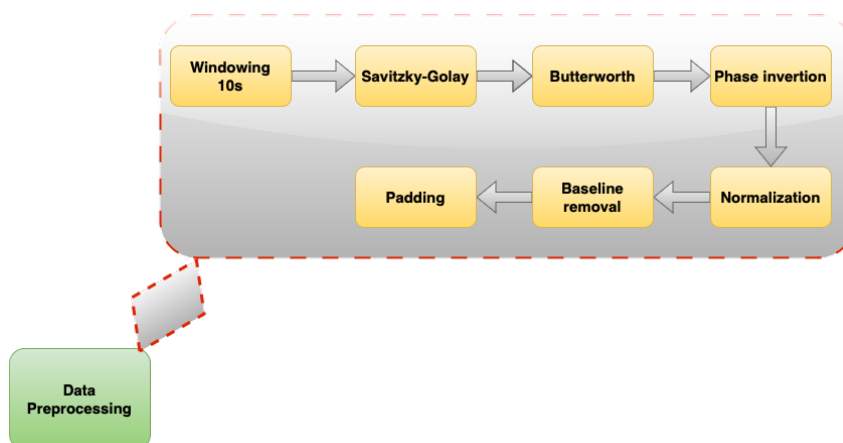
O sinal presente nas bases de dados pode ser frequentemente afetado por ruídos de alta ou baixa frequência. Além disso, também são comuns anomalias decorrentes de perda de contato do sensor ou movimentação. Esse sinal é comumente referido como “sinal *raw*” ou “sinal bruto”. Com o objetivo de remover parte desse ruído se faz necessário uma etapa de pré-processamento.

Após o sinal ser particionado, cada segmento é processado individualmente, como descrito no capítulo 2, com os filtros Savitzky-Golay e Butterworth seguidos de uma normalização em amplitude. O filtro Savitzky-Golay (SAVITZKY; GOLAY, 1964) tem o objetivo de suavizar o sinal (*smoothing*) através de padrões na série temporal. O filtro passa-banda Butterworth (SELESNICK; BURRUS, 1998), por sua vez, é aplicado para remover frequências indesejadas, provenientes da respiração ou de artefatos do dispositivo da coleta de dados. Entretanto, devido às características de inversão de fase do filtro Butterworth, o sinal teve sua fase corrigida pela inversão do segmento.

Uma vez que o segmento é filtrado, uma normalização em amplitude é aplicada através da abordagem MinMax. Por fim, o sinal teve sua linha base corrigida através do método de interpolação *cubic spline*. A Fig.9 demonstra um fluxograma dos blocos

que compõem a etapa de pré-processamento.

Figura 9 – Fluxograma detalhado do pré-processamento.



Fonte: Autor.

### 4.2.3 Avaliação da qualidade do sinal

Mesmo com o uso de técnicas avançadas de processamento de sinais, é inevitável a presença de ruídos e artefatos. Portanto, é essencial incorporar um componente capaz de avaliar e quantificar a qualidade de cada segmento do sinal. Assim sendo, identificar e remover segmentos que não atendam aos critérios de qualidade desejados.

Neste projeto, foi desenvolvido um componente robusto de avaliação da qualidade do sinal. Este, é capaz de detectar e remover segmentos que não atendam aos critérios mínimos de qualidade, evitando que segmentos espúrios sejam enviados para os algoritmos de inteligência artificial. Para isso, é utilizado um método de extração de pulsos, no qual os pulsos são separados com base no valor mínimo entre dois picos sistólicos, identificando assim o início de cada pulso. Em seguida, cada pulso é avaliado individualmente pelo método do índice de qualidade do sinal, descrito por (LI; HUANG, 2020). Esse método atribui um índice de qualidade entre 0 e 1, sendo que 1 indica uma qualidade perfeita do sinal. Além disso, um segundo filtro foi implementado visando aumentar a robustez. Neste método, características morfológicas são extraídas de cada pulso, as quais são comparadas a valores limite pré-definidos.

As características morfológicas utilizadas para filtrar segmentos com artefatos foram o pico sistólico, *Skewness* e a Área do Ponto de Inflexão (IPA). Os pulsos de PPG, em sua construção típica, apresentam seu ponto de maior amplitude no segundo quadrante, onde geralmente ocorre o pico sistólico. Logo, é incomum encontrar sinais em que o pico sistólico esteja deslocado a ponto de estar localizado no final do pulso.

Da mesma forma, o valor de *skewness*, é uma medida estatística que descreve a assimetria da distribuição de um conjunto de dados, neste caso os pontos que compõem o pulso de PPG. Ela indica a tendência da distribuição em se inclinar para um lado ou para o outro em relação à média. A IPA, por sua vez, representa a relação entre a área anterior e posterior ao pico diastólico. Em pulsos sem artefatos, espera-se que o valor dessa razão seja maior do que 1.

Contudo, para que um segmento seja considerado válido e enviado para a extração de características, 9 de 10 pulsos devem atingir os critérios. Os critérios são de SSQI de pelo menos 0,9 pelo método de Li, (LI; HUANG, 2020). Além disso, o pico sistólico deve ocorrer antes de 750 ms do tempo total do pulso, o valor de *skewness* deve estar entre 0 e 2, e o valor da IPA deve ser maior do que 1,5, os dois últimos com valores adimensionais. Os limites de valores e as características selecionadas foram determinados por meio de experimentos específicos, visando identificar os atributos com melhor capacidade de eliminar segmentos com ruídos.

#### 4.2.4 Extração de características

A extração de características de sinais biomédicos, neste caso do sinal de PPG, desempenha um papel fundamental na análise e interpretação dos dados. A utilização do sinal através da sua forma de onda nem sempre revela de forma clara e coesa correlações existentes entre o sinal biomédico e condições clínicas. Características como amplitudes, deslocamentos e larguras dos pulsos do sinal de PPG são frequentemente utilizadas para quantificar e descrever as propriedades fisiológicas subjacentes. Através da detecção e medição dessas características, é possível obter informações valiosas sobre a função cardiovascular e outras condições relacionadas.

Essas medidas morfológicas podem ser utilizadas para identificar padrões de onda específicos, avaliar a integridade do sistema circulatório ou detectar possíveis alterações e anomalias. Portanto, a extração de características morfológicas do sinal de PPG desempenha um papel crucial na compreensão e análise, contribuindo para aplicações clínicas e de monitoramento da saúde.

Neste estudo, cada segmento de PPG com duração de 10 segundos, que passou pelo filtro de qualidade do sinal, foi usado para a extração de características. Um segmento de 10 segundos de sinal normalmente contém de 10 a 15 pulsos, no entanto, nem todos os pulsos em cada segmento apresentam qualidade ideal. Para evitar a extração de características de pulsos ruidosos, foi adotada a técnica de média de todos os pulsos dentro do segmento. Essa abordagem permite obter uma representação mais confiável e robusta das características morfológicas do sinal de PPG, ao minimizar a influência de pulsos de baixa qualidade.

Uma vez que o pulso resultante é obtido, as características são extraídas, dentre elas é possível citar as seguintes:

1. SYS\_PEAK: Posição do pico sistólico.
2. DIA\_PEAK: Posição do pico diastólico.
3. RISING\_TIME: Tempo de subida entre o início do pulso e o pico sistólico.
4. NOTCH: Posição do *notch* do pulso de PPG.
5. W75: Largura do pulso de PPG em 75% de amplitude.
6. W50: Largura do pulso de PPG em 50% de amplitude.
7. MAX\_POWER\_1st: Potência da primeira componente harmônica do espectrograma.
8. MAX\_POWER\_2nd: Potência da segunda componente harmônica do espectrograma.
9. MAX\_FREQ\_1st: Frequência do pico da primeira componente harmônica do espectrograma.
10. MAX\_FREQ\_2nd: Frequência do pico da segunda componente harmônica do espectrograma.

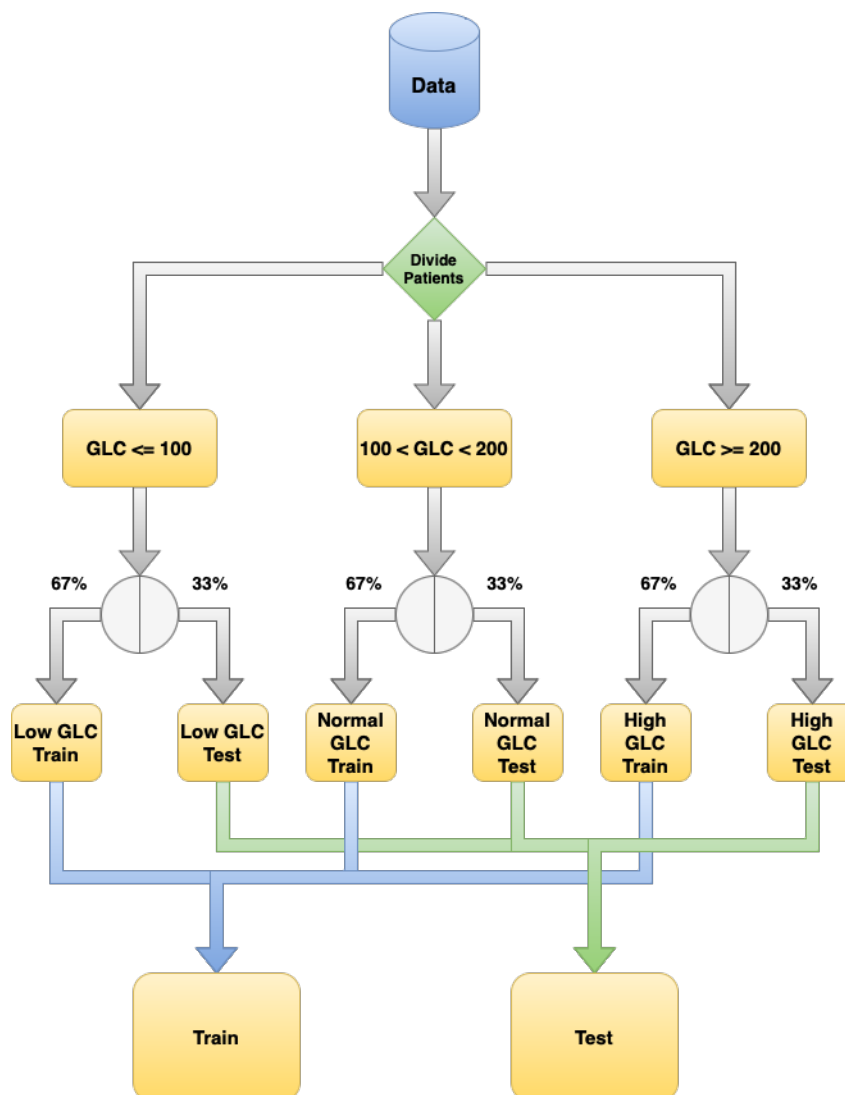
#### 4.2.5 Particionamento de dados

O particionamento dos dados em conjuntos de treinamento e teste é uma prática comum em algoritmos de aprendizado de máquina. Esse processo envolve dividir o conjunto de dados disponível em duas partes distintas: um conjunto de treinamento, que é utilizado para treinar o modelo, e um conjunto de teste, que é utilizado para avaliar o desempenho do modelo em dados não vistos anteriormente. Essa divisão é essencial para avaliar a capacidade do modelo de generalizar o conhecimento aprendido durante o treinamento e verificar a precisão em dados desconhecidos. O particionamento adequado dos dados em treinamento e teste é fundamental para garantir a validade e confiabilidade dos resultados obtidos pelo algoritmo de aprendizado de máquina.

No presente estudo, o particionamento dos dados foi realizado com base nos pacientes, utilizando seus identificadores únicos. Inicialmente, os pacientes foram divididos em três subgrupos: glicose baixa, normal e alta. Para cada um desses subgrupos, 67% dos pacientes foram selecionados para compor o conjunto de treinamento, enquanto os 33% restantes foram destinados ao conjunto de teste.

Logo, o conjunto final de treinamento foi formado pela concatenação dos pacientes de cada subgrupo de treinamento. Este método garante, primeiro que um mesmo paciente esteja presente em apenas um conjunto. Além disso, assegura uma distribuição equitativa de diferentes faixas de valores de glicose nos conjuntos de treinamento e teste. A Fig.10 demonstra o processo de particionamento dos dados.

Figura 10 – Fluxograma do particionamento dos dados em treino e teste.



Fonte: Autor.

#### 4.2.6 Aprendizado de Máquina

Neste trabalho, modelos de aprendizado de máquina foram usados para a estimativa da glicose. Os dados foram divididos, e o conjunto de treino foi utilizado no treinamento dos modelos. Os seguintes modelos apresentaram os melhores resultados: Random Forest, Support Vector Machine e CatBoost.

O Random Forest (RF) (LIAW; WIENER, 2002) é um algoritmo de aprendizado de máquina que se baseia na construção de um conjunto de árvores de decisão individuais para realizar tarefas de classificação ou regressão. Cada árvore é criada a partir de amostras aleatórias do conjunto de dados de treinamento, usando um processo conhecido como *bootstrap* (LOUPPE, 2014). Em cada árvore, as decisões são tomadas com base em um subconjunto aleatório de características. Ao combinar

as previsões de todas as árvores, o RF produz uma previsão final mais robusta e geralmente mais precisa. Essa abordagem de conjunto reduz a variância do modelo e ajuda a evitar o *overfitting*, tornando o RF um algoritmo poderoso amplamente utilizado em diversos domínios.

O Support Vector Machine (SVM) (CORTES; VAPNIK, 1995) é um algoritmo que busca encontrar um hiperplano ótimo para separar as diferentes classes de dados. O hiperplano é uma superfície de decisão que divide o espaço de atributos em duas regiões, uma para cada classe. Os pontos de dados mais próximos do hiperplano são chamados de vetores de suporte. O SVM procura maximizar a margem entre os vetores de suporte e o hiperplano, onde a margem é definida como a distância perpendicular entre eles. Essa maximização visa criar um classificador robusto e generalizável, capaz de lidar melhor com dados de teste não vistos anteriormente.

O CatBoost (DOROGUSH; ERSHOV; GULIN, 2018) é um modelo de *gradient boosting*, que combina vários modelos de aprendizado de máquina mais fracos para formar um modelo mais forte. No caso do CatBoost, os modelos fracos são árvores de decisão. Ele possui regularização embutida para evitar o *overfitting*, utilizando uma combinação de regularização L1, L2 e regularização por desvio. Essas técnicas ajudam a controlar a complexidade do modelo e a evitar ajustes excessivos aos dados de treinamento. Além disso, o CatBoost possui uma característica interessante que é a detecção automática de melhorias significativas durante o treinamento. Isso é feito por meio de um método estatístico que identifica quando adicionar uma nova árvore de decisão à sequência existente. Essa característica pode melhorar a eficiência do treinamento e reduzir o tempo necessário para alcançar um bom desempenho.

#### 4.2.7 Outras abordagens

Durante a fase de desenvolvimento deste trabalho, outras abordagens foram testadas com o objetivo de encontrar o método de melhor performance para a estimativa da glicose.

##### 4.2.7.1 Mel Frequency Cepstral Coefficients

Nesta abordagem, inspirada no método descrito por (HABBU *et al.*, 2019) e (ALONSO-SILVERIO *et al.*, 2021), foi adotada a extração de características do domínio da frequência utilizando os coeficientes cepstrais de Mel. Esses coeficientes foram calculados a partir de segmentos de 10 segundos do sinal de PPG de cada participante. Os segmentos foram processados e selecionados utilizando os mesmos métodos descritos neste estudo. Inicialmente, 32 coeficientes foram extraídos e utilizados em modelos de aprendizado de máquina, como Random Forest e Support Vector Machine.

Em um segundo experimento, apenas os 8 primeiros coeficientes foram utilizados. Os resultados obtidos mostram que os modelos foram capazes de generalizar



para um grupo relativo de participantes, no entanto, o erro médio absoluto geral não foi satisfatório, chegando a 40 mg/dL para o UFSC-BGL.

#### 4.2.7.2 Spectrogram

Ao analisar gráficos no domínio da frequência para um grupo selecionado de participantes, observou-se visualmente que os indivíduos com níveis alterados de glicose apresentavam um pico da segunda harmônica acima de 2 Hz. Entretanto, aqueles com níveis normais não exibiam esse comportamento. Com o objetivo de validar essa hipótese, o sinal de PPG de cada participante do conjunto de dados UFSC-BGL foi dividido em janelas de 10 segundos e utilizado para gerar um espectrograma. O espectrograma é obtido por meio da Transformada Rápida de Fourier (FFT) do sinal e proporciona uma representação visual de como as frequências variam ao longo do tempo.

As imagens resultantes dos espectrogramas foram utilizadas para treinar uma rede neural artificial convolucional. Foram realizados experimentos com diferentes níveis de detalhamento da imagem e variações na arquitetura da rede, geralmente com 2 a 5 blocos convolucionais. Esse método apresentou resultados promissores, porém enfrentou desafios como *overfitting* e tempos de treinamento prolongados, que poderiam exceder 24 horas.

#### 4.2.7.3 Scalogram

Similar ao método mencionado anteriormente, neste caso, o segmento de 10 segundos do sinal de PPG foi utilizado para gerar imagens a partir da Transformada Contínua de Wavelet (CWT). A CWT é uma técnica de processamento de sinais que permite analisar o conteúdo de frequência de um sinal em diferentes escalas de tempo. Ao contrário da FFT, que analisa o sinal em uma escala fixa, a CWT permite uma análise em várias escalas, adaptando-se às características de frequência que variam ao longo do tempo (AGUIAR; SOARES, 2011).

Neste estudo, a wavelet-mãe Morlet complexa foi utilizada, onde o objetivo foi aumentar o nível de informação contida nas imagens em comparação com o espectrograma. As imagens geradas foram utilizadas para treinar uma rede neural artificial convolucional. Os experimentos realizados seguiram uma abordagem semelhante a do método do espectrograma. No entanto, os resultados obtidos foram inferiores e os problemas mencionados anteriormente ainda persistiram, o que levou à interrupção desse método no estudo.

## 5 RESULTADOS

Este capítulo tem o propósito de demonstrar os resultados encontrados neste trabalho. Inicialmente, uma análise comportamental do perfil dos participantes do UFSC-BGL foi realizada baseada nos questionários aplicados.

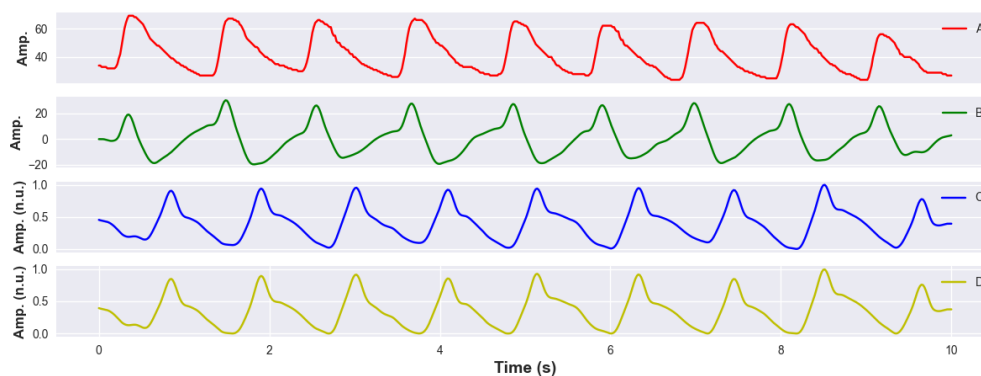
Durante o estudo, por meio de questionários, foram obtidos os seguintes resultados: 85% (68/80) dos participantes relataram fazer uso de algum tipo de medicação diariamente, sendo que 27,5% (22/80) fazem uso de insulina diariamente ou periodicamente. Além disso, 52,5% (42/80) dos participantes já receberam diagnóstico de diabetes ou pré-diabetes. Entre os 42 participantes com diagnóstico de diabetes, 35 apresentaram níveis considerados alterados de glicose durante o momento da pesquisa e 7 níveis normais. Por fim, 5 participantes apresentaram níveis alterados de glicose, mas não estavam cientes de sua possível condição de diabetes.

Como descrito anteriormente, o glicosímetro não é o padrão ouro para aferição da glicose, porém é o método mais utilizado. Em relação aos glicosímetros, 45% (36/80) dos participantes possuíam o equipamento em casa. Destes, 31 pertenciam ao grupo com níveis alterados de glicose, enquanto 9 participantes do mesmo grupo não possuíam o dispositivo. Dos participantes com glicosímetro, 20 afirmaram checar a glicose diariamente, 12 realizam essa verificação semanalmente e 4 participantes não realizam o monitoramento da glicose. Dos 20 participantes que checam diariamente, 18 afirmaram verificar a glicose duas ou mais vezes por dia, sendo que um participante checa a glicose 5 vezes por dia. Além disso, foi perguntado aos participantes com que frequência eles experimentam sintomas de hiperglicemia ou hipoglicemia. Logo, 14 participantes responderam que têm crises uma vez por semana, enquanto 3 participantes relataram ter crises quase diariamente.

Uma das principais preocupações que antecederam este estudo foi a qualidade do sinal coletado. Uma vez que possíveis interferências poderiam ocorrer, devido a problemas como taxa de gordura corporal ou a circulação sanguínea na região monitorada. Além disso, a qualidade do sinal obtido por meio de técnicas ópticas e o possível impacto da presença de colorações, como esmalte nas unhas, têm sido amplamente discutido. Durante a coleta de dados, foi observado que cinco participantes apresentavam esmalte nas unhas, com cores variando entre laranja, rosa, vermelho, cinza e preto. No entanto, constatou-se que a presença de qualquer uma dessas cores não afetou significativamente a qualidade do sinal obtido.

Um exemplo de sinal obtido, bem como de todas as etapas de pré-processamento, pode ser encontrada na Fig.11.

Figura 11 – Segmento participante 006 @60Hz, glicose de 114 mg/dL. a) sinal *raw*, b) sinal filtrado com Savitzky-Golay + Butterworth, c) sinal normalizado + inversão de fase, d) sinal com remoção de linha base



Fonte: Autor.

Utilizando o segmento do sinal de PPG após a etapa de pré-processamento, foram extraídas características do domínio da frequência. Através do gráfico de Densidade Espectral do sinal, foram coletados atributos como a frequência da primeira e segunda harmônicas, bem como a potência máxima nestes determinados momentos. A Fig. 12a ilustra o processo, realçando os pontos da primeira e segunda harmônicas.

Contudo, os pulsos do segmento do sinal foram separados e, em seguida, foi realizada uma média para obter um único pulso que representasse o segmento. A Fig. 12b ilustra o processo descrito. A partir desse pulso médio, foram extraídas as características morfológicas relevantes para a análise.

O pulso médio foi utilizado para extração de outras características, que posteriormente serão concatenadas e utilizadas nos modelos de AI. Um exemplo de processamento de atributos morfológicos é demonstrado na Fig.13.

As características descritas no Apêndice B foram extraídas de cada pulso. O resultado de cada pulso foi armazenado em um objeto do tipo *data frame*, como representa a Tab.4. Onde, além das características é descrito o Identificador do Participante (PID), Identificador da Amostra (SID) e o valor de referência de glicose (GLC)

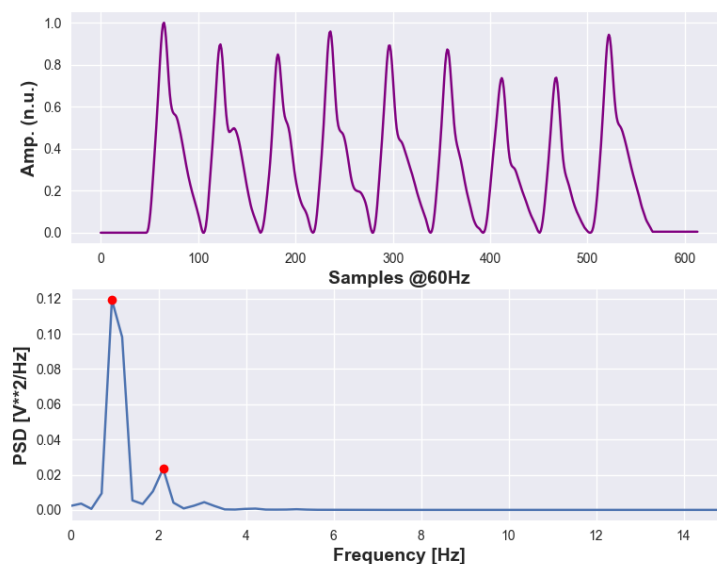
Tabela 4 – Dados estruturados de características extraídas de pulsos do sinal de PPG.

PID	SID	SYS_PEAK	DIA_PEAK	RISING_TIME	W75	W50	...	GLC
001	001_1	0,780	0,219	0,433	0,281	0,561	...	120
001	001_2	0,643	0,290	0,490	0,223	0,642	...	120
002	002_1	0,834	0,112	0,411	0,259	0,490	...	242
003	003_1	0,801	0,231	0,389	0,290	0,503	...	170

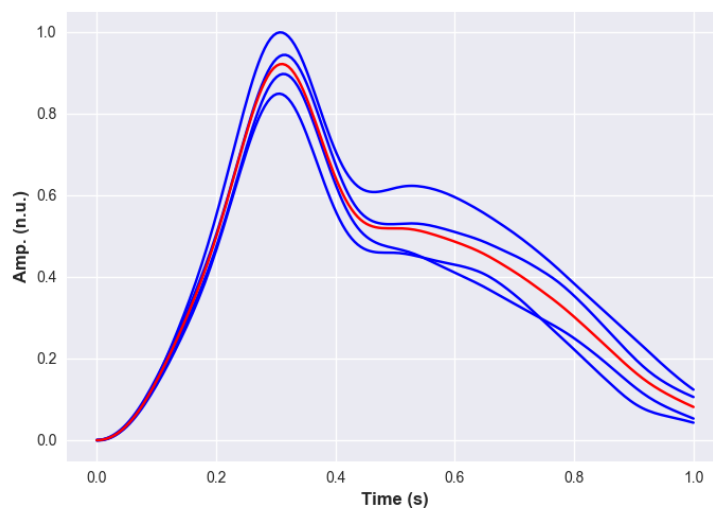
Fonte: Autor.

Figura 12 – Extração de características em domínios distintos. Participante 006 @60Hz, glicose de 114 mg/dL.

(a) Domínio da frequência. Respetivamente, segmento de sinal processado e gráfico de Densidade Espectral de Potência.



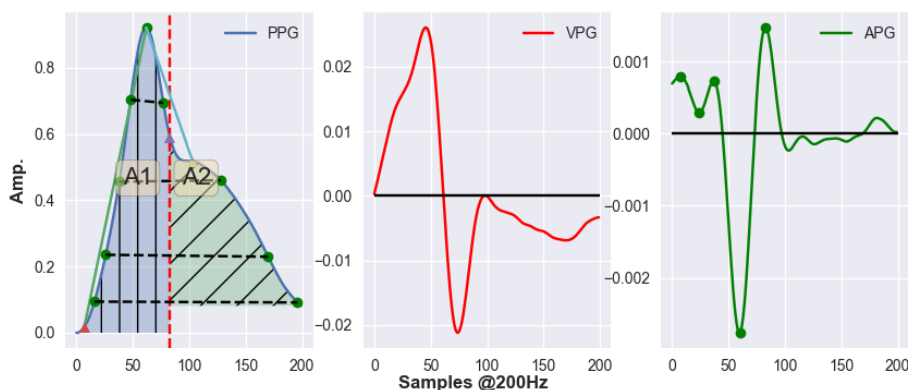
(b) Domínio do tempo: Pulsos sobrepostos (em azul) de um segmento de 10 segundos, com o pulso médio realçado em vermelho.



Fonte: Autor.

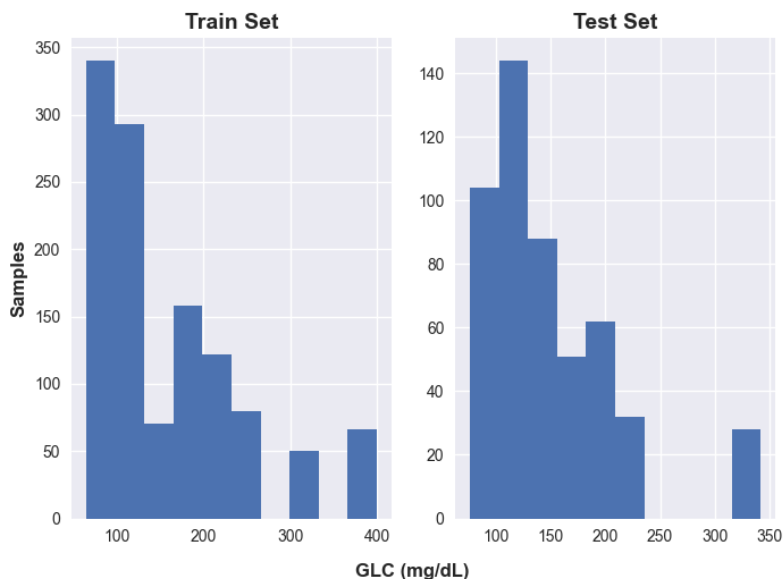
Uma vez extraídas, as características foram separadas em dois grupos: Treino e Teste. A divisão dos grupos foi de 67% para treino e 33% para teste, como descrito na Fig.10. O objetivo é garantir que um único indivíduo esteja em apenas um dos grupos. Inicialmente, as figuras Fig.14 e Fig.15 fornecem uma visualização da distribuição dos conjuntos de dados utilizados neste estudo.

Figura 13 – Extração de características do pulso do sinal de PPG, VPG e APG, destacados em azul, vermelho e verde. Participante 006 pulso re-amostrado a @200 Hz, com glicose de 114 mg/dL.



Fonte: Autor.

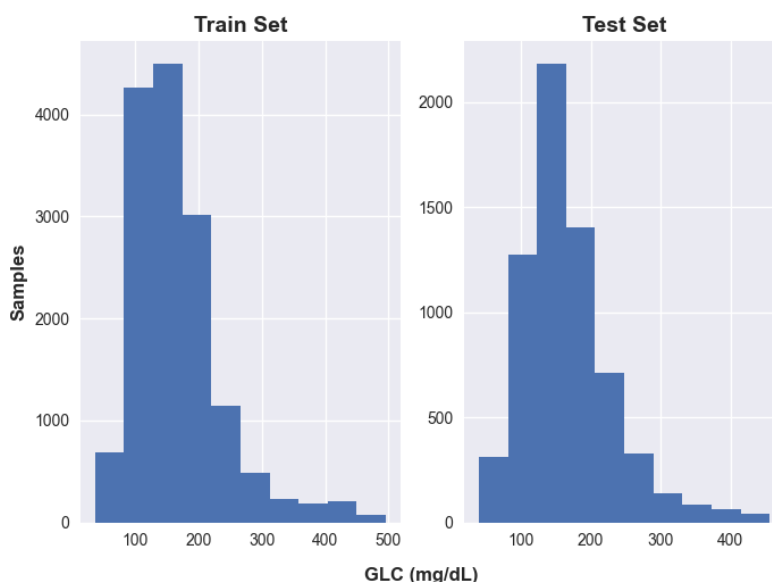
Figura 14 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de treino e teste para o conjunto UFSC-BGL.



Fonte: Autor.

Os histogramas fornecem uma descrição dos conjuntos de dados com base em cada amostra de 10 segundos do sinal de PPG. Como cada sinal disponível no conjunto UFSC-BGL possui uma duração de 5 minutos, pode haver até 30 segmentos de 10 segundos em um único sinal. Da mesma forma, no conjunto MIMIC-BGL, que possui sinais com duração de 2 minutos, pode haver até 12 segmentos.

Figura 15 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de treino e teste para o conjunto MIMIC-BGL.



Fonte: Autor.

Este trabalho adotou dois métodos para a avaliação dos experimentos. Primeiro, os resultados foram avaliados comparando-se as estimativas de cada segmento com seus respectivos valores reais. Além disso, todas as estimativas de um único participante foram agregadas em um único valor, que foi então comparado ao valor de referência. A média aritmética foi utilizada para agregar os valores estimados de cada participante.

Nesta seção, são apresentados os resultados agregados para cada participante, os resultados gerais são descritos no Apêndice A. O conjunto UFSC-BGL compreende uma única aferição de glicose para cada participante. Todavia, MIMIC-BGL possui várias aferições, com valores distintos, para um único participante, inviabilizando a agregação. Logo, as estimativas devem ser agregadas apenas para a amostra de sinal da respectiva aferição, ou seja, para todos os segmentos provenientes da amostra de sinal de 2 minutos.

Foram selecionados 10 modelos de regressão para a fase inicial de experimentos. Dentre estes, os três modelos com a melhor performance foram utilizados para a extração de resultados. As tabelas Tab.5 e 6 demonstram os resultados para cada um dos modelos, incluindo métricas de desempenho como: erro médio, erro médio absoluto, erro médio absoluto percentual e o coeficiente de determinação, respectivamente para UFSC-BGL e MIMIC-BGL. Essas métricas permitem avaliar a capacidade do modelo em fazer estimativas precisas dos níveis de glicose.

Tabela 5 – Resultados dos modelos utilizando o conjunto UFSC-BGL com 26 participantes no conjunto de teste.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
SVM	4,38	20,31	39,31	0,51	13,79
CatBoost	2,04	17,96	33,85	0,64	11,91
RF	1,23	20,00	36,98	0,57	13,66

Fonte: Autor.

Tabela 6 – Resultados dos modelos utilizando o conjunto MIMIC-BGL com 576 amostras.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
SVM	5,94	31,02	52,53	0,46	20,64
CatBoost	1,40	32,54	52,22	0,47	23,07
RF	0,92	38,60	57,89	0,34	27,65

Fonte: Autor.

Tabela 7 – Resultados dos modelos utilizando o conjunto UFSC+MIMIC com 602 amostras.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
SVM	4,41	31,11	51,72	0,39	20,75
CatBoost	-2,72	31,65	48,83	0,42	23,38
RF	-2,91	37,04	53,92	0,32	27,82

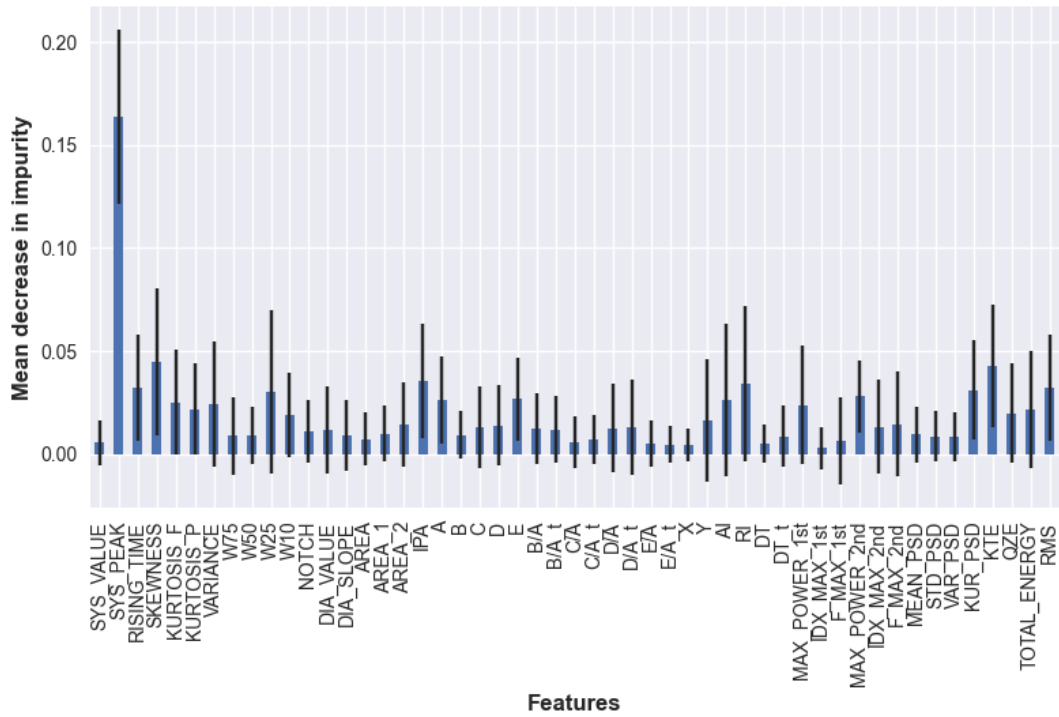
Fonte: Autor.

Através do modelo Random Forest, a importância de cada atributo pode ser estimada. A Fig.16 demonstra a importância de cada característica na estimativa através do conjunto UFSC-BGL (Fig.16a) e MIMIC-BGL (Fig.16b).

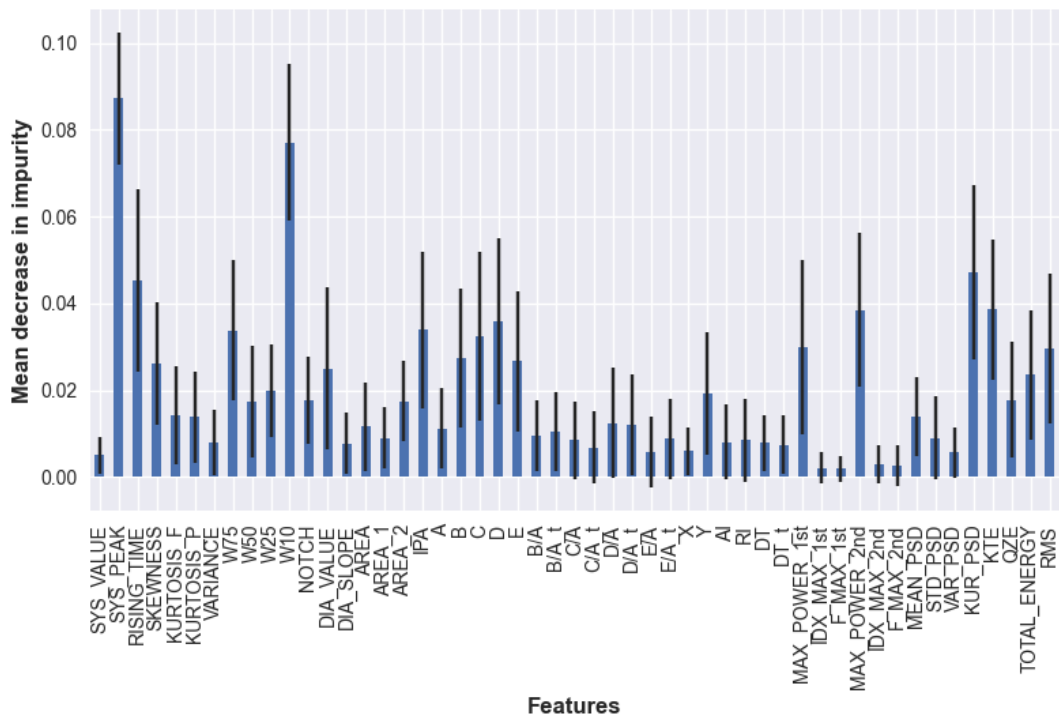
Além disso, os resultados são demonstrados por meio de um gráfico de dispersão das predições do modelo em relação aos valores reais. Essa representação gráfica permite visualizar o quão bem o modelo está se aproximando dos valores verdadeiros, identificando possíveis desvios ou tendências nas previsões. A Fig.17a demonstra as estimativas para o conjunto UFSC-BGL através do modelo CatBoost, enquanto a Fig.17b demonstra as estimativas para o conjunto MIMIC-BGL através do modelo SVM. Por fim, Fig.17c demonstra as estimativas para a concatenação de ambos (UFSC+MIMIC) através do modelo SVM.

Figura 16 – Histograma de importância de cada característica para o modelo Random Forest utilizando UFSC-BGL e MIMIC-BGL.

(a) UFSC-BGL



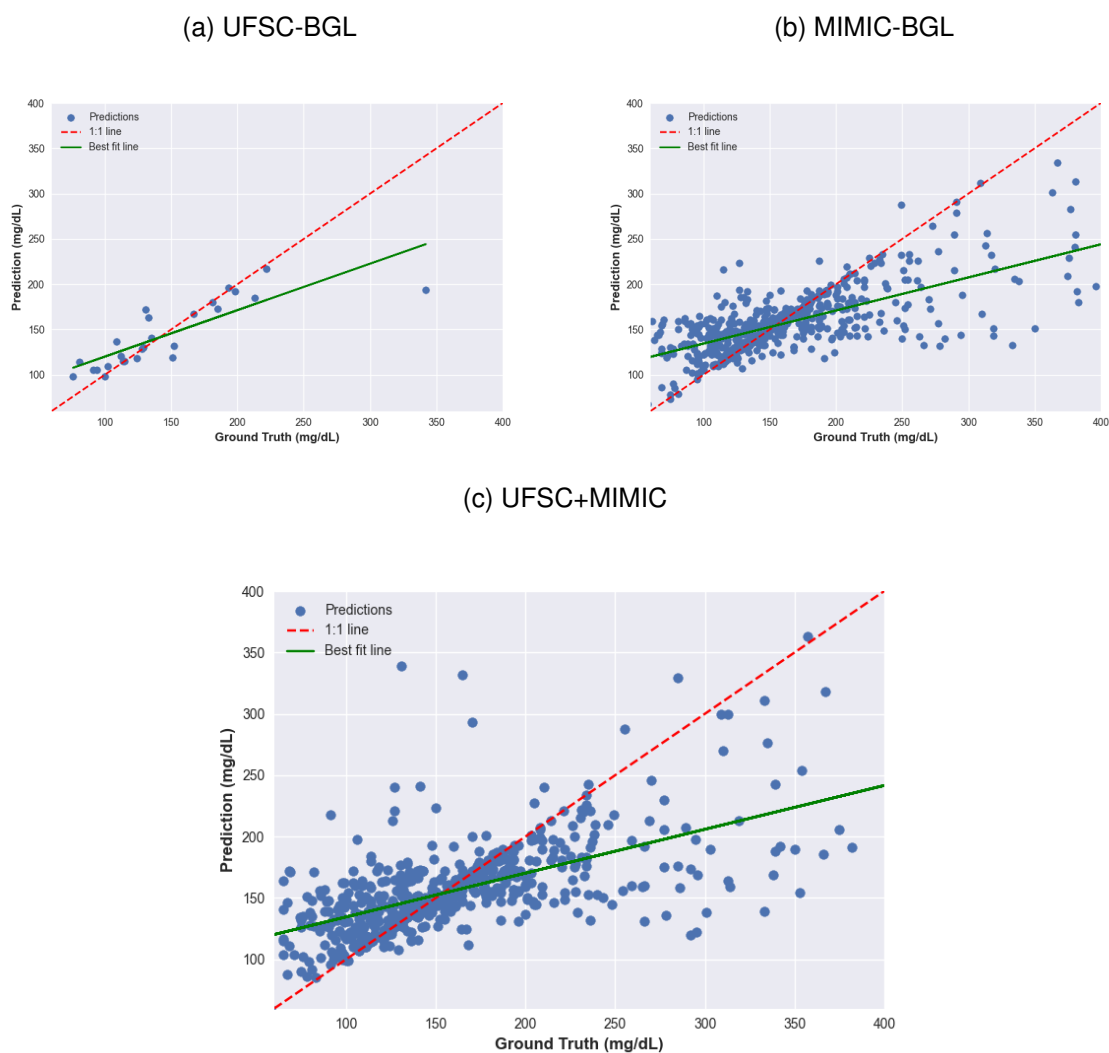
(b) MIMIC-BGL



Fonte: Autor.



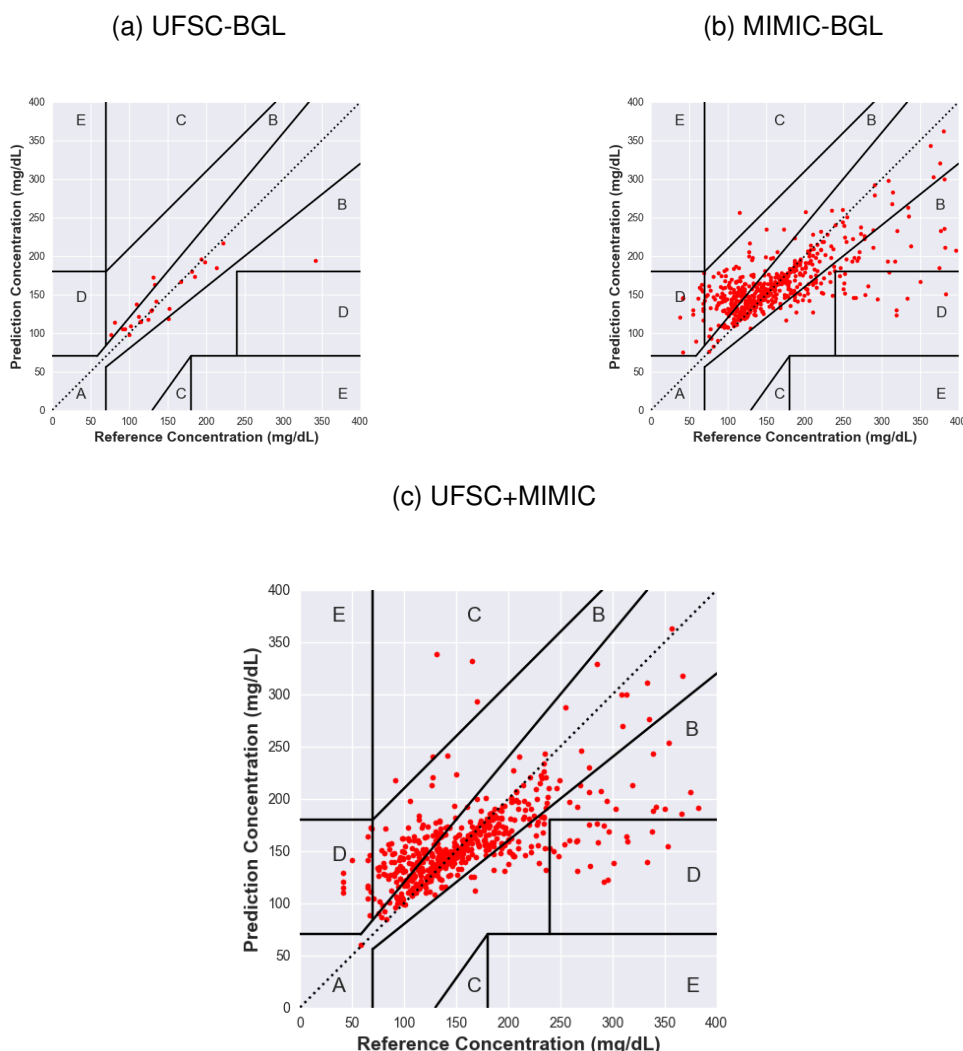
Figura 17 – Gráfico de dispersão dos valores preditos de glicose agrupados para cada participante, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL. Linha 1:1 e Linha de melhor ajuste realçadas respectivamente em vermelho tracejado e verde sólido.



Fonte: Autor.

O gráfico de erro de Clarke é uma ferramenta amplamente utilizada para avaliar o desempenho de modelos de estimativa de glicose. Esse gráfico fornece uma visão mais abrangente da precisão do modelo, mostrando a proporção de tempo em que as predições se encontram em diferentes faixas de erro. Similarmente aos gráficos previamente descritos, Fig.8 descreve o gráfico de Clarke para o UFSC-BGL através do modelo CatBoost, Fig.9 para o MIMIC-BGL através do modelo SVM e por fim, Fig.10 para o UFSC+MIMIC através do modelo SVM.

Figura 18 – Gráfico de Clarke dos valores preditos de glicose agrupados para cada participante, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL.



Fonte: Autor.

Por meio do gráfico de Clarke, é possível determinar o número de estimativas que se encontram em cada zona de precisão do gráfico. As tabelas, Tab.8, 9 e 10 demonstram os resultados da porcentagem de estimativas em cada zona de precisão, respectivamente para: UFSC-BGL, MIMIC-BGL e UFSC+MIMIC.

Conforme mencionado anteriormente, cada participante do conjunto MIMIC-BGL possui múltiplas medições de glicose. Isso permite a construção de um gráfico de segmentos que ilustra a trajetória das medições de glicose reais e estimadas para cada participante específico. A Fig.19 apresenta esse gráfico para três participantes distintos. Além disso, um exemplo adicional para outros três participantes pode ser encontrado no Apêndice A.

Tabela 8 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke para o conjunto UFSC-BGL.

<b>Modelo</b>	<b>Zona A (%)</b>	<b>Zone a (%)</b>	<b>Zona C (%)</b>	<b>Zona D (%)</b>	<b>Zona E (%)</b>
<b>RF</b>	73	27	0	0	0
<b>CatBoost</b>	73	27	0	0	0
<b>SVM</b>	77	19	0	4	0

Fonte: Autor.

Tabela 9 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke para o conjunto MIMIC-BGL.

<b>Modelo</b>	<b>Zona A (%)</b>	<b>Zona B (%)</b>	<b>Zona C (%)</b>	<b>Zona D (%)</b>	<b>Zona E (%)</b>
<b>RF</b>	55	37	0	8	0
<b>CatBoost</b>	63	30	0	7	0
<b>SVM</b>	65	27	0	7	0

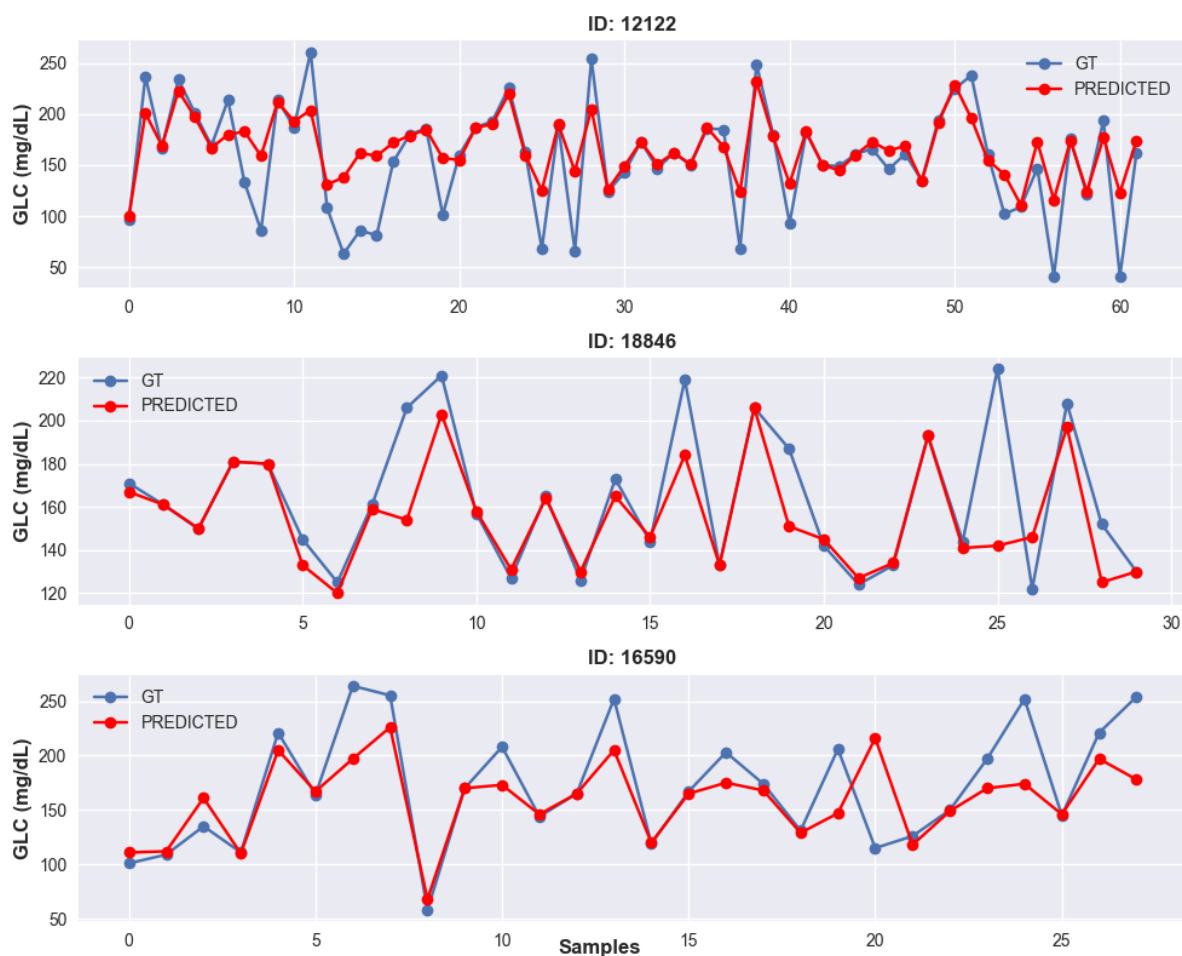
Fonte: Autor.

Tabela 10 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke para o conjunto UFSC+MIMIC.

<b>Modelo</b>	<b>Zona A (%)</b>	<b>Zona B (%)</b>	<b>Zona C (%)</b>	<b>Zona D (%)</b>	<b>Zona E (%)</b>
<b>RF</b>	55	36	0	8	0
<b>CatBoost</b>	62	33	1	5	0
<b>SVM</b>	66	27	1	6	0

Fonte: Autor.

Figura 19 – Gráfico de segmento dos valores reais e preditos pelo modelo para três participantes diferentes. ID: 12122, 18846 e 16590



Fonte: Autor.

## 6 DISCUSSÃO

Este capítulo tem como objetivo discutir os resultados obtidos no presente estudo, conforme descrito no capítulo anterior. Foram utilizados dois conjuntos de dados, inicialmente de forma separada e posteriormente combinados.

O conjunto de dados UFSC-BGL, proposto neste estudo, conta com a participação de 80 indivíduos, dos quais 40 apresentaram níveis de glicose alterados e os outros 40 apresentaram níveis considerados normais. Ao analisar os resultados, destaca-se que 22 participantes fazem uso de insulina para o controle glicêmico, sendo geralmente a opção de tratamento adotada em último caso. Outro aspecto preocupante é de que 5 participantes apresentaram níveis alterados de glicose, porém não tinham consciência de sua possível condição de diabetes. Esta falta de ciência do problema pode acarretar em riscos graves, visto que a diabetes de ser devidamente controlada.

Além disso, menos da metade dos participantes realizava o monitoramento glicêmico em casa, e apenas 20 indivíduos faziam o acompanhamento diário. Especialmente para pessoas diabéticas, o controle glicêmico é essencial como medida preventiva para diversas complicações, particularmente para evitar episódios de hiperglicemia ou hipoglicemia. No entanto, 14 participantes relataram ter crises relacionadas a esses episódios pelo menos uma vez por semana. Com base nesses dados, este estudo conclui que o uso do glicosímetro, embora seja crucial, ainda não é amplamente adotado pela população. Dentre os fatores que levam a este cenário é possível citar o custo do dispositivo e o desconforto físico associado à coleta diária de sangue.

Através da coleta de dados, foi possível investigar brevemente o efeito de colorações na área de coleta do sinal de PPG, tendo em vista sua construção óptica. Não foi observado nenhuma correlação entre a qualidade do sinal e a presença de esmalte nas unhas do participante. Todos os participantes que apresentavam este adereço tiveram seu sinal extraído sem necessidade de exclusão.

Após a conclusão das etapas de processamento e filtragem, procedeu-se à extração das características do sinal. O processo de extração de características é composto de aspectos morfológicos e nos domínios da frequência e não linear. Para extração morfológica, a soma dos pulsos de um segmento e a utilização do pulso médio como representante é utilizada. Como denota a Fig.12b, existem diferenças em cada pulso de PPG, mesmo que mínimas, estas diferenças podem tornar os modelos de AI suscetíveis a pequenas variações. Logo, a utilização da média dos pulsos busca trazer estabilidade ao modelo. Além disso, características nos domínios da frequência e não linear são utilizadas, estas extraídas do segmento de 10 segundos. O objetivo é adicionar informações de outros domínios para o modelo, além do domínio do tempo já explorado.

Os histogramas apresentados nas figuras Fig.14 e Fig.15 revelam que ambos os

conjuntos de dados, UFSC-BGL e MIMIC-BGL, possuem uma ampla faixa de valores de glicose, variando de menos de 100 mg/dL até mais de 300 mg/dL. Observa-se também que as distribuições dos conjuntos de treino e teste são semelhantes. No entanto, é importante ressaltar que ambos os conjuntos são desbalanceados. É mais frequente a ocorrência de valores normais de glicose em comparação com os valores alterados. Embora seja possível balancear os conjuntos reduzindo as amostras normais ou gerando amostras artificiais, nenhuma dessas abordagens foi adotada neste trabalho. Isso se deve ao fato de que a distribuição desigual dos dados não foi considerada um problema, uma vez que as análises realizadas neste estudo destacam os resultados para cada grupo demográfico. Diferentemente de uma abordagem de classificação, onde a distribuição desbalanceada poderia mascarar os resultados obtidos.

Com base nas características extraídas, foram realizados testes iniciais em 10 modelos de aprendizado de máquina. Dentre eles, foram selecionados três modelos para a extração de resultados e análises: Support Vector Machine, CatBoost e Random Forest. Utilizando o conjunto UFSC-BGL, os modelos apresentaram resultados semelhantes, com erro médio absoluto variando entre 18 e 20 mg/dL e erro percentual absoluto médio entre 12% e 14%. É importante ressaltar que esses valores de erro são ampliados devido à ampla faixa de valores de glicose presentes, onde a variação observada é de 60 mg/dL a 350 mg/dL. Além disso, não existem amostras suficientes em determinadas faixas de valores. Portanto, mesmo que exista uma diferença de 100 mg/dL entre um valor de referência de 350 mg/dL e um valor estimado de 250 mg/dL, a informação contida nesses valores pode ser considerada semelhante.

Entre os resultados obtidos através do UFSC-BGL, o modelo CatBoost apresentou o melhor desempenho, com MAE de 17,96 mg/dL, MAPE de 11,91% e coeficiente de determinação de 0,64. O mesmo procedimento foi realizado utilizando o conjunto MIMIC-BGL, onde o modelo SVM apresentou o melhor desempenho. Os resultados reportados foram de MAE de 31,02 mg/dL, MAPE de 20,64% e coeficiente  $r^2$  de 0,46. É possível notar que o valor do erro aumentou em comparação aos resultados obtidos com o UFSC-BGL, principalmente devido ao maior número de amostras disponíveis (576). Além disso, MIMIC-BGL é composto por indivíduos acamados em unidades de tratamento intensivo. Esta situação pode estar associada a outras comorbidades que afetam o fluxo sanguíneo e, conseqüentemente, o sinal de PPG.

Embora o modelo Random Forest não tenha apresentado os melhores resultados, uma vantagem da sua estrutura é a possibilidade de estimar a importância de cada característica. Para tal, o modelo se baseia no método de *Mean Decrease Impurity*, também conhecido como *Gini importance*. A Fig.16a ilustra que o pico sistólico é uma característica de grande importância e essencial para descrever o modelo. Este atributo contém informações sobre a localização no plano cartesiano onde ocorre o maior fluxo sanguíneo, bem como o tempo entre o início e o fim da fase sistólica. Além

disso, o valor de *Skewness* e a Energia de Kaiser-Teagen também contribuem significativamente, entre outros. No entanto, também é importante mencionar que todas as características possuem alguma relevância para o modelo.

Da mesma forma, para o MIMIC-BGL, observa-se na Fig.16b que o pico sistólico ainda tem grande importância. Entretanto, o valor da largura do pulso em 10% da sua amplitude (*W10*) também se mostra um fator extremamente relevante. Outras características, como o *Rising Time* e a *Kurtosis* do PSD, também contribuem de forma significativa para este modelo. Com base nesses resultados, pode-se concluir que o modelo requer uma dimensionalidade alta, uma vez que todas as características utilizadas têm algum peso na decisão da estimativa de glicose. Além disso, é evidente que o pico sistólico possui a maior importância nesse modelo para ambos os conjuntos de dados, devido a informação contida nessa característica.

Em um segundo experimento, foram obtidos resultados por meio de gráficos que visam demonstrar o comportamento das estimativas para cada participante. Na Fig.17a, é possível observar que as estimativas para o conjunto UFSC-BGL se concentram, em sua maioria, próximas à linha 1:1. Este comportamento indica uma consistência em relação a variação dos valores de referência. Algumas estimativas na faixa de 180 mg/dL a 250 mg/dL são observadas com relativa precisão. No entanto, há um participante com nível de glicose de 350 mg/dL em que o modelo falhou, resultando em uma estimativa de 200 mg/dL.

Da mesma forma, os resultados obtidos utilizando o conjunto MIMIC-BGL, conforme ilustrado na Fig.17b, revelam um padrão em que o modelo foi preciso na faixa de valores acima de 100 mg/dL até aproximadamente 200 mg/dL. Antes e depois dessas faixas, o modelo apresentou instabilidade, com predições bastante dispersas, especialmente na região de valores superiores a 250 mg/dL. Nestes casos, as predições se distanciam frequentemente da linha 1:1. Por fim, quando mesclados os conjuntos de dados não houve grandes variações, o comportamento das predições se manteve similar ao observado quando utilizado o conjunto de dados MIMIC-BGL.

Ao aplicar o gráfico de erro de Clarke às estimativas, conforme ilustrado na Fig.18, foi possível avaliar a acurácia do método utilizando ambos os conjuntos, UFSC-BGL e MIMIC-BGL. No caso do UFSC-BGL, observa-se que 73% dos participantes se concentraram na zona A, que é a região de maior relevância, enquanto 27% estiveram na zona B. Por outro lado, ao utilizar o MIMIC-BGL, houve uma redução na precisão da zona A, que agora abrange apenas 65% das amostras. Enquanto que a zona B continuou a compreender 27% das amostras, os 7% restantes foram classificados na zona D, onde a precisão já não é mais tolerável. Além disso, ao utilizar o UFSC+MIMIC, verificou-se um leve aumento na precisão, com 66% das amostras na zona A. No entanto, também foi observado que um pequeno número de erros dispersos, representando 1% das predições, agora estão localizados na zona C do gráfico.

Contudo, um último experimento foi conduzido para avaliar a precisão das previsões ao longo do tempo para um grupo de participantes. Essa análise foi possível devido à disponibilidade de múltiplas referências de glicose para cada indivíduo no conjunto de dados MIMIC-BGL. A Fig.19 apresenta os resultados para o primeiro participante, identificado como ID 12122. Este participante possui mais de 60 referências de glicose com variação entre 60 mg/dL e 250 mg/dL. Ao examinar as estimativas observa-se que em alguns momentos o modelo não foi capaz de acompanhar os valores reais, especialmente entre a amostra 10 e a amostra 30. No entanto, em certos pontos, o modelo demonstrou alta precisão e manteve o padrão da série temporal, resultando em um coeficiente de similaridade de 0,85 para este indivíduo. Por outro lado, os participantes seguintes, identificados como IDs 18846 e 16590, apresentaram um menor número de amostras, cerca de 30. Estes, exibiram um padrão de série temporal mais suave, com variações ocorrendo de forma progressiva. Nestes casos, o desempenho do modelo foi mais preciso visualmente, no entanto em momentos de pico glicêmico houve discrepâncias, resultando em coeficientes de similaridade de 0,77 e 0,78, respectivamente, para IDs 18846 e 16590.

Por último, os resultados desta investigação são comparados com outros autores que exploraram a estimativa não invasiva da glicose utilizando sinais de fotopletismografia. A Tab.11 descreve os resultados para este trabalho, bem como para cada um dos autores selecionados.

Tabela 11 – Comparação entre resultados deste trabalho e outros autores.

Autor	Método	MAE (mg/dL)	r2 (p)	Zona A (%)
(YEN <i>et al.</i> , 2022)	Características morfológicas e demográficas	5,09	0,99	100
(HABBU <i>et al.</i> , 2019)	MFCC	-	0,90	85
(ALONSO-SILVERIO <i>et al.</i> , 2021)	MFCC	11,62	-	92
(ISLAM <i>et al.</i> , 2021)	Características morfológicas	17,02	-	-
(GUZMAN; CAZARES; MARTINEZ-TORTEYA, 2021)	HRV	16,45	-0,18	-
<b>Este trabalho</b>	Características morfológicas, frequência e não-lineares	17,96	0,64	73

Fonte: Autor.



Conforme descrito na Tab.11, este estudo obteve resultados semelhantes aos de (GUZMAN; CAZARES; MARTINEZ-TORTEYA, 2021) e (ISLAM *et al.*, 2021), com um erro médio absoluto entre 16 e 18 mg/dL. No entanto, os autores desses estudos utilizaram um número reduzido de participantes se comparado a este trabalho, sendo 5, 22 e 80 indivíduos, respectivamente. O número de amostras impacta significativamente os resultados em modelos de treinamento supervisionado, o que pode ter influenciado nos resultados dos autores.

Por outro lado, (HABBU *et al.*, 2019) e (ALONSO-SILVERIO *et al.*, 2021) ambos utilizaram coeficientes cepstrais obtidos do sinal de PPG como fonte de informação para o treinamento. Estes autores alcançaram, respectivamente, 85% e 92% das estimativas na zona A do CEG, enquanto este estudo obteve 73%. No entanto, ao adotar a mesma abordagem, este trabalho não obteve resultados semelhantes. A utilização de dados públicos poderia ter ajudado a solucionar esta incógnita. Assim, teria permitido a recriação e teste da mesma abordagem, com os mesmos dados com os quais os resultados foram originalmente publicados.

Por fim, (YEN *et al.*, 2022) demonstraram os melhores resultados ao combinar informações provenientes de várias fontes. Para tal, foram utilizados, o sinal de PPG de dois comprimentos de onda, impedância bioelétrica, além de características fisiológicas e demográficas do paciente. Os autores alcançaram um erro médio absoluto de 5,09 mg/dL e 100% das estimativas na zona A do CEG. Todavia, é mencionado pelos autores, e do entendimento deste trabalho, que os valores de hemoglobina (g/L) também foram usados no treinamento. Estes valores poderiam ser valores do histórico de hemoglobina glicada do paciente. Em ambos os casos, a utilização destas características, bem como outros sinais vitais, foge do escopo deste trabalho.

## 7 CONCLUSÃO

Este estudo propôs a pesquisa e desenvolvimento de um modelo computacional para a estimativa não invasiva dos níveis de glicose utilizando o sinal de Fotopletismografia. Durante o período da pesquisa, foi proposto um novo conjunto de dados que inclui tanto os sinais de fotopletismografia quanto as referências de glicose. Este conjunto de dados, denominado UFSC-BGL, engloba 80 participantes que foram recebidos no Hospital Universitário Polydoro de São Thiago da Universidade Federal de Santa Catarina (HU-UFSC/EBSERH).

Além disso, um segundo conjunto de dados, denominado MIMIC-BGL, foi utilizado neste projeto com o objetivo de ampliar o escopo da pesquisa. O MIMIC-BGL não foi coletado especificamente para esta pesquisa, mas foi criado a partir da combinação de outros dois conjuntos de dados: MIMIC-III e Curated-BGL. O conjunto de dados MIMIC-III contém os sinais de fotopletismografia relacionados aos pacientes disponíveis no conjunto de dados Curated-BGL, que inclui medições de glicose. O recém desenvolvido MIMIC-BGL conta com 2.954 e mais de 98.000 registros. Entretanto apenas uma fração destes foram utilizados nesta pesquisa devido à complexidade e dimensionalidade dos modelos, como também o custo computacional.

Foram realizados experimentos com o objetivo de investigar a hipótese inicial deste trabalho, que é a de que o sinal de fotopletismografia contém informações sobre as variações nos níveis de glicose. Dentre as abordagens desenvolvidas e testadas, destacam-se o uso dos coeficientes cepstrais obtidos por meio de bancos de filtros de Mel e a utilização de imagens de espectrograma e escalograma. Além disso, características morfológicas do sinal de fotopletismografia também foram avaliadas.

Contudo, este trabalho optou por um método desenvolvido através da utilização de características morfológicas, não lineares e no domínio da frequência. Um total de 51 atributos foram extraídos de segmentos de 10 segundos, sem sobreposição, dos sinais de fotopletismografia. Os dados foram divididos em dois conjuntos, treinamento (67%) e teste (33%), com base no identificador do participante. Modelos de aprendizado de máquina foram treinados e avaliados utilizando métricas de desempenho no conjunto de teste.

Os resultados obtidos revelaram que o modelo CatBoost obteve a melhor performance ao utilizar o conjunto de dados UFSC-BGL. O modelo alcançou um erro médio absoluto de 17,96 mg/dL, erro médio absoluto percentual de 11,91% e 73% das estimativas dos participantes estavam localizadas na zona A do gráfico de Clarke. Ao utilizar o conjunto de dados MIMIC-BGL, os resultados demonstraram que o modelo Support Vector Machine apresentou o melhor desempenho. Foi observado um erro médio absoluto de 31,02 mg/dL, erro médio absoluto percentual de 20,64% e 65% das estimativas dos participantes estavam na zona A do gráfico de Clarke. Por fim, quando

ambos os conjuntos de dados foram mesclados, o modelo Support Vector Machine continuou com o melhor desempenho. O modelo alcançou um erro médio absoluto de 31,11 mg/dL, erro médio absoluto percentual de 20,75% e 66% das estimativas dos participantes estavam na zona A do gráfico de Clarke.

Este trabalho contribui para o avanço das pesquisas de métodos de estimativa não invasiva de glicose através da concepção de dois novos conjuntos dados. Além disso, demonstrou experimentos práticos e resultados sólidos de um método inicial para a estimativa da glicose através de sinais de fotopletismografia. Dentre as perspectivas de trabalhos futuros é possível citar a utilização por completo do conjunto de dados MIMIC-BGL e explorar o método de decomposição de modos intrínsecos. Além disso, explorar outras características que podem refinar e aumentar a robustez deste modelo.

## REFERÊNCIAS

(WHO), World Health Organisation. **Global Report on Diabetes**. [S./], abr. 2016. Disponível em: <https://ideas.repec.org/p/ess/wpaper/id10553.html>.

AGUIAR, Luís Francisco; SOARES, Maria Joana. **The Continuous Wavelet Transform: A Primer**. [S./], 2011. Disponível em: <https://ideas.repec.org/p/nip/nipewp/16-2011.html>.

ALLEN, John. Photoplethysmography and its application in clinical physiological measurement. **Physiological Measurement**, v. 28, n. 3, r1, jan. 2007.

ALONSO-SILVERIO, Gustavo A.; FRANCISCO-GARCÍA, Víctor; GUZMÁN-GUZMÁN, Iris P.; VENTURA-MOLINA, Elías; ALARCÓN-PAREDES, Antonio. Toward non-invasive estimation of blood glucose concentration: A comparative performance. **Mathematics**, MDPI, v. 9, 20 out. 2021. ISSN 22277390.

ARÉVALO, Aldo Robles *et al.* Data-driven curation process for describing the blood glucose management in the intensive care unit. **Scientific Data**, Nature Research, v. 8, 1 dez. 2021. ISSN 20524463.

CLARKE, William L; COX, Daniel; GONDER-FREDERICK, Linda A; CARTER, William; POHL, Stephen L. Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. **Diabetes Care**, v. 10, n. 5, p. 622–628, set. 1987. ISSN 0149-5992. eprint: <https://diabetesjournals.org/care/article-pdf/10/5/622/515265/10-5-622.pdf>.

CLARKE, William L. The original Clarke Error Grid Analysis (EGA). **Diabetes technology & therapeutics**, v. 7 5, p. 776–9, 2005.

CONTEC HEALTH. **Oxímetro de pulso CMS50E**. [S./: s.n.], 2022. Acessado em 04 setembro 2022. Disponível em: <https://contechealth.com>.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.

DEFRONZO, Ralph A; BONADONNA, Riccardo C; FERRANNINI, Eleuterio.

**Pathogenesis of NIDDM A Balanced Overview.** [S.l.]. Disponível em: <http://diabetesjournals.org/care/article-pdf/15/3/318/440690/15-3-318.pdf>.

DOROGUSH, Anna Veronika; ERSHOV, Vasily; GULIN, Andrey. CatBoost: gradient boosting with categorical features support. **ArXiv**, abs/1810.11363, 2018.

DUONG, Michael Tran; RAUSCHECKER, Andreas M; RUDIE, Jeffrey D; CHEN, Po-Hao; COOK, Tessa S; BRYAN, R Nick; MOHAN, Suyash. Artificial intelligence for precision education in radiology. en. **Br. J. Radiol.**, British Institute of Radiology, v. 92, n. 1103, p. 20190389, nov. 2019.

ELGENDI, M.; LIANG, Y.; CHEN, Z.; WARD, R. Analysis: An optimal filter for short photoplethysmogram signals. **Scientific Data**, 2018.

ERDFELDER, E.; FAUL, F.; BUCHNER, A. GPOWER: A general power analysis program. **Behavior Research Methods, Instruments and Computers**, v. 28, n. 1, p. 1–11, 1996.

GORE, John C. Artificial intelligence in medical imaging. **Magnetic Resonance Imaging**, v. 68, a1–a4, 2020. ISSN 0730-725X.

GTECH. **Glicosímetro digital.** [S.l.: s.n.], 2022. Acessado em 04 setembro 2022. Disponível em: <https://accumed.com.br/smart/>.

GUZMAN, Luis; CAZARES, Andres Marcelo Garza Fox; MARTINEZ-TORTEYA, Antonio. Model for Glycemic Level Detection using Heart Rate Variability in a Mexican Sample. **Proceedings - 2020 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2020**, Institute of Electrical e Electronics Engineers Inc., p. 505–510, mar. 2021.

HABBU, Shraddha K.; JOSHI, Shrikant; DALE, Manisha; GHONGADE, Rajesh Bhaskar. Noninvasive Blood Glucose Estimation Using Pulse Based Cepstral Coefficients. **2019 2nd International Conference on Signal Processing and Information Security, ICSPIS 2019**, Institute of Electrical e Electronics Engineers Inc., out. 2019.

HOSNY, Ahmed; PARMAR, Chintan; QUACKENBUSH, John; SCHWARTZ, Lawrence H; AERTS, Hugo J W L. Artificial intelligence in radiology. en. **Nat. Rev. Cancer**, v. 18, n. 8, p. 500–510, ago. 2018.

ISLAM, Tanvir Tazul; AHMED, Md Sajid; HASSANUZZAMAN, Md; AMIR, Syed Athar Bin; RAHMAN, Tanzilur. Blood glucose level regression for smartphone ppg signals using machine learning. **Applied Sciences (Switzerland)**, MDPI AG, v. 11, p. 1–20, 2 jan. 2021. ISSN 20763417.

JIANG, Fei *et al.* Artificial intelligence in healthcare: past, present and future. **Stroke and Vascular Neurology**, BMJ Specialist Journals, v. 2, n. 4, p. 230–243, 2017. ISSN 2059-8688. eprint: <https://svn.bmj.com/content/2/4/230.full.pdf>.

JOHNSON, Alistair E.W. *et al.* MIMIC-III, a freely accessible critical care database. **Scientific Data**, Nature Publishing Groups, v. 3, mai. 2016. ISSN 20524463.

KAVSAOGLU, A.R.; POLAT, K.; BOZKURT, M.R. An innovative peak detection algorithm for photoplethysmography signals: an adaptive segmentation method. **Turkish Journal of Electrical Engineering and Computer Sciences**, 2016.

LI, Hongxing; HUANG, Shizhao. A High-Efficiency and Real-Time Method for Quality Evaluation of PPG Signals. **IOP Conference Series: Materials Science and Engineering**, IOP Publishing, v. 711, n. 1, p. 012100, jan. 2020.

LIANG, Yongbo; ELGENDI, Mohamed; CHEN, Zhencheng; WARD, Rabab. Analysis: An optimal filter for short photoplethysmogram signals. **Scientific Data**, Nature Publishing Groups, v. 5, mai. 2018. ISSN 20524463.

LIAW, Andy; WIENER, Matthew. Classification and Regression by randomForest. **R News**, v. 2, n. 3, p. 18–22, 2002.

LOUPPE, Gilles. **Understanding Random Forests: From Theory to Practice**. Out. 2014. Tese (Doutorado).

MAGRABI, Farah *et al.* Artificial intelligence in clinical decision support: Challenges for evaluating AI and practical implications. en. **Yearb. Med. Inform.**, Georg Thieme Verlag KG, v. 28, n. 01, p. 128–134, ago. 2019.

MARATHE, Payal H.; GAO, Helen X.; CLOSE, Kelly L. American Diabetes Association Standards of Medical Care in Diabetes 2017. **Journal of Diabetes**, v. 9, n. 4, p. 320–324, 2017. eprint:

<https://onlinelibrary.wiley.com/doi/pdf/10.1111/1753-0407.12524>.

MIHAI, Bogdan; MIHAI, Panu. LabVIEW Modeling and Simulation of The Digital Filters, 2015.

NASIRI, Noushin. **Introductory Chapter: Wearable Technologies for Healthcare Monitoring**. [S.l.]: IntechOpen, dez. 2019.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

RASCHKA, S.; MIRJALILI, V. **Python Machine Learning**. [S.l.]: Packt Publishing, 2019. P. 742.

SACKS, David B. A1C versus glucose testing: a comparison. en. **Diabetes Care**, United States, v. 34, n. 2, p. 518–523, fev. 2011.

SANTANA, Sílvio Heverton Campelo de. **Implementação de uma arquitetura microcontrolada para análise do contorno da curva fotopletismográfica**. 2017. F. 45. Monografia (Trabalho de Conclusão de Curso) – Curso de Graduação em Meteorologia, Universidade Federal de Santa Maria, Recife.

SAVITZKY, Abraham.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, 1964.

SCHAFER, R. W. What Is a Savitzky-Golay Filter? **IEEE Signal Processing Magazine**, v. 28, n. 4, p. 111–117, 2011.

SELESNICK, I.W.; BURRUS, C.S. Generalized digital Butterworth filter design. **IEEE Transactions on Signal Processing**, v. 46, n. 6, p. 1688–1694, 1998.

SENGUPTA, Sohini; HANDOO, Anil; HAQ, Inaamul; DAHIYA, Karamvir; MEHTA, Sanjay; KAUSHIK, Mradul. Clarke Error Grid Analysis for Performance Evaluation of Glucometers in a Tertiary Care Referral Hospital. **Indian Journal of Clinical Biochemistry**, v. 37, p. 199–205, mar. 2021.

SOCIEDADE BRASILEIRA DE DIABETES. **Diretriz da Sociedade Brasileira de Diabetes 2023**. [S.l.: s.n.], 2023. Acessado em 11 Junho 2023. Disponível em: <https://diretriz.diabetes.org.br>.

TAMURA, Toshiyo; MAEDA, Yuka; SEKINE, Masaki; YOSHIDA, Masaki. Wearable Photoplethysmographic Sensors—Past and Present. **Electronics**, v. 3, n. 2, p. 282–302, 2014. ISSN 2079-9292.

VASHIST, Sandeep Kumar; ZHENG, Dan; AL-RUBEAN, Khalid; LUONG, John H.T.; SHEU, Fwu Shan. **Technology behind commercial devices for blood glucose monitoring in diabetes management: A review**. v. 703. [S.l.: s.n.], out. 2011. P. 124–136.

WILCOX, Gisela. Insulin and Insulin Resistance. **Clinical biochemist Reviews**, v. 26, p. 19–39, 2005.

WORLD HEALTH ORGANIZATION. **World Health Statistics**. [S.l.: s.n.], 2020. Acessado em 04 setembro 2022. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/hypertension>.

YEN, Chih Ta; CHEN, Un Hung; WANG, Guo Chang; CHEN, Zong Xian. Non-Invasive Blood Glucose Estimation System Based on a Neural Network with Dual-Wavelength Photoplethysmography and Bioelectrical Impedance Measuring. **Sensors**, MDPI, v. 22, p. 2–12, 12 jun. 2022. ISSN 14248220.

ZHANG, Yuwei; ZHANG, Yuan; SIDDIQUI, Sarah Ali; KOS, Anton. **Non-invasive Blood-Glucose Estimation Using Smartphone PPG Signals and Subspace KNN Classifier**. v. 86. [S.l.], 2019. P. 68–74.



## APÊNDICE A – RESULTADOS AVALIANDO SEGMENTOS DE 10 SEGUNDOS

Tabela 12 – Resultados dos modelos utilizando o conjunto UFSC-BGL para predições de segmentos de 10 segundos no conjunto de teste.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
<b>SVM</b>	6.85	21.82	39.74	0.50	13.81
<b>CatBoost</b>	3.41	19.71	35.83	0.59	12.58
<b>RF</b>	3.33	20.19	37.51	0.55	12.78

Fonte: Autor.

Tabela 13 – Resultados dos modelos utilizando o conjunto MIMIC-BGL para predições de segmentos de 10 segundos no conjunto de teste.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
<b>SVM</b>	6.12	30.58	52.01	0.48	20.28
<b>CatBoost</b>	1.68	32.34	51.89	0.49	22.83
<b>RF</b>	1.00	38.35	57.73	0.36	27.50

Fonte: Autor.

Tabela 14 – Resultados dos modelos utilizando o conjunto UFSC+MIMIC para predições de segmentos de 10 segundos no conjunto de teste.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
<b>SVM</b>	4.05	30.52	49.33	0.43	20.63
<b>CatBoost</b>	-2.66	30.26	45.99	0.51	22.79
<b>RF</b>	-2.03	36.53	53.72	0.34	27.03

Fonte: Autor.

Tabela 15 – Resultado da distribuição das predições de cada segmento no modelo de erro de Clarke em cada um dos conjuntos.

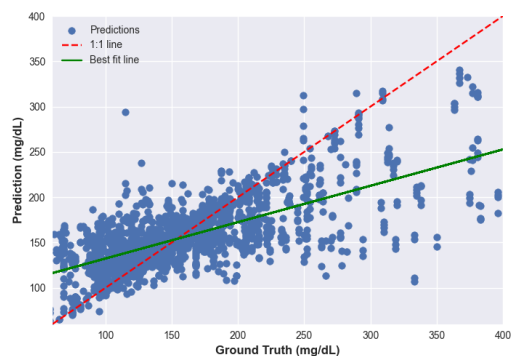
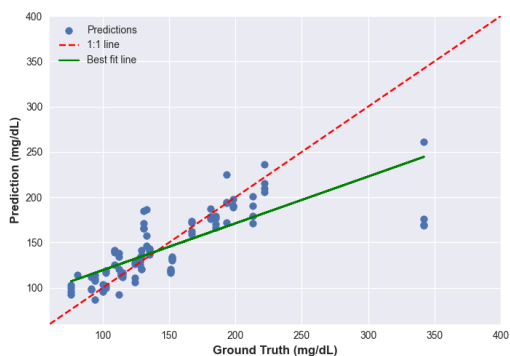
Conjunto	Zona A (%)	Zona B (%)	Zona C (%)	Zona D (%)	Zona E (%)
<b>UFSC-BGL</b>	74	23	0	3	0
<b>MIMIC-BGL</b>	67	26	0	7	0
<b>UFSC+MIMIC</b>	66	27	1	6	0

Fonte: Autor.

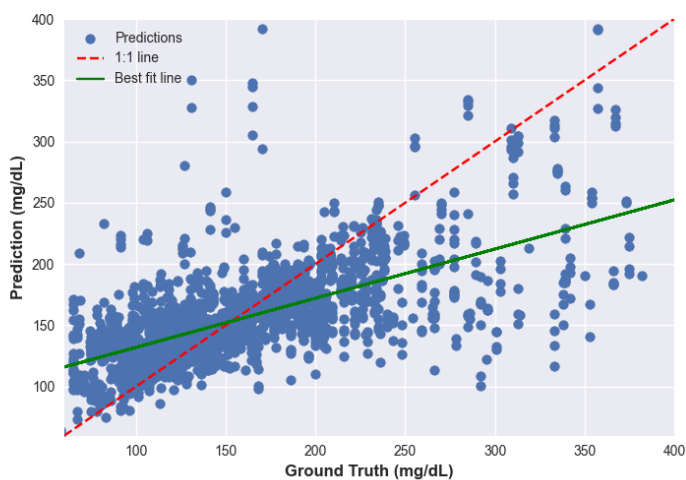
Figura 20 – Gráfico de dispersão dos valores preditos de glicose para cada segmento de 10 segundos, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL. Linha 1:1 e Linha de melhor ajuste realçadas respectivamente em vermelho tracejado e verde sólido.

(a) UFSC-BGL

(b) MIMIC-BGL



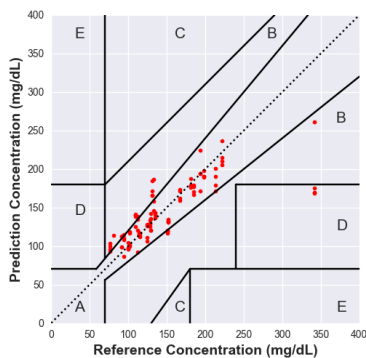
(c) UFSC+MIMIC



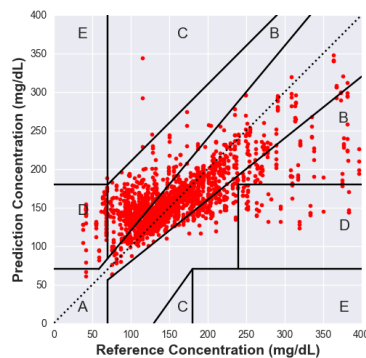
Fonte: Autor.

Figura 21 – Gráfico de Clarke dos valores preditos de glicose para cada segmento de 10 segundos, em relação ao seu respectivo valor real. Eixo das abcissas demonstra o valor real de glicose enquanto o eixo das ordenadas o valor predito pelo modelo, ambos em termos de mg/dL.

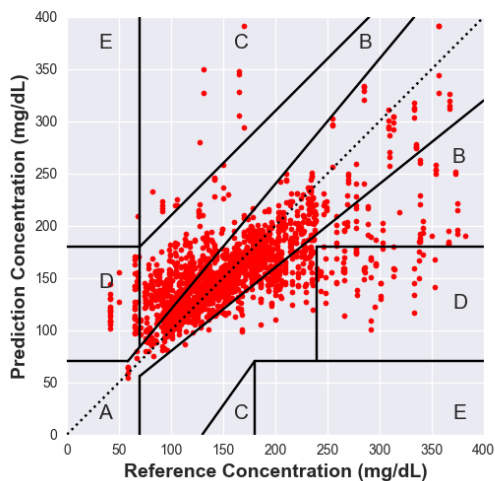
(a) UFSC-BGL



(b) MIMIC-BGL

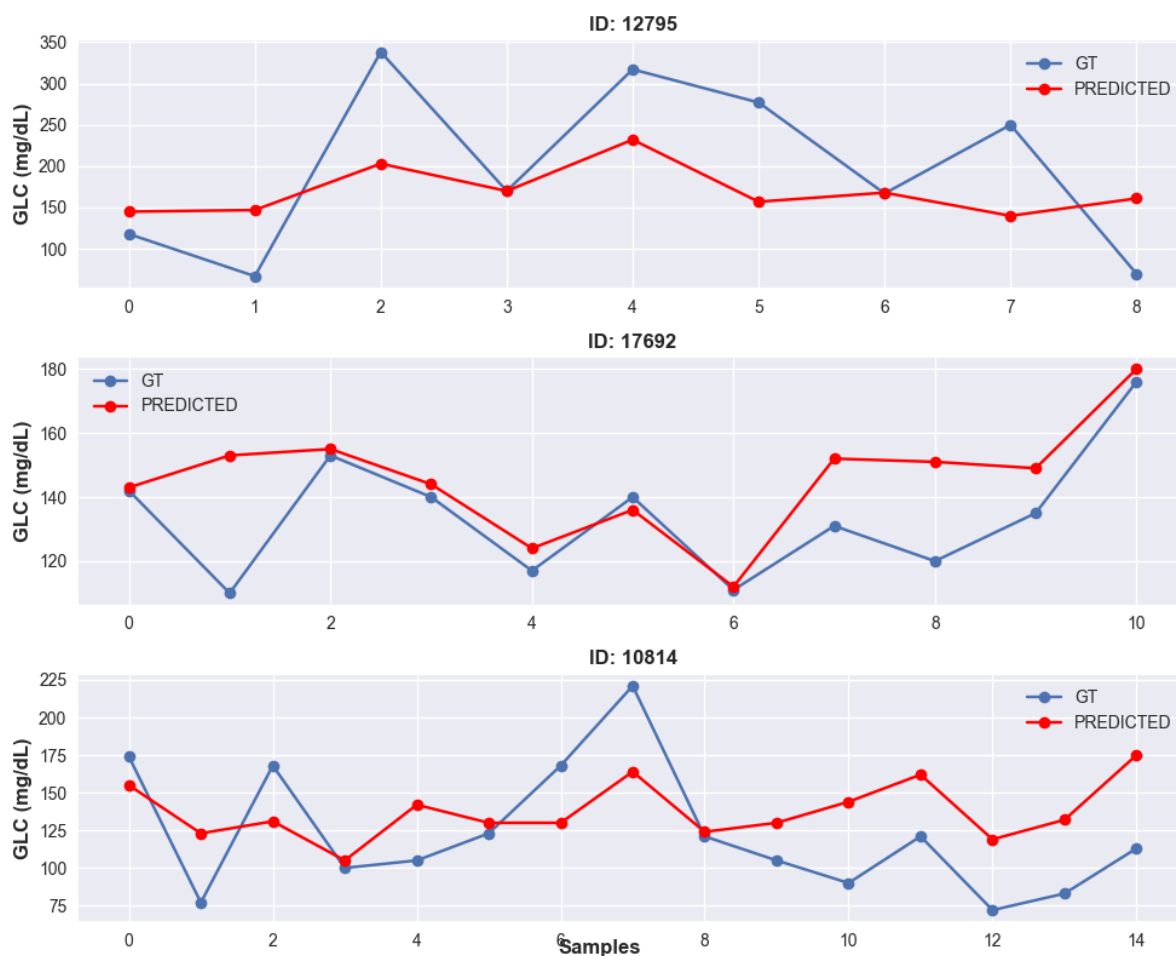


(c) UFSC+MIMIC



Fonte: Autor.

Figura 22 – Gráfico de segmento dos valores reais e preditos pelo modelo para três participantes diferentes. ID: 12795, 17692 e 10814



Fonte: Autor.

## APÊNDICE B – CARACTERÍSTICAS EXTRAÍDAS DO SINAL DE FOTOPLETISMOGRAFIA

Tabela 16 – Características extraídas do segmento de 10 segundos do sinal de PPG e posteriormente utilizadas nos modelos de aprendizado de máquina.

<b>Característica</b>	<b>Descrição</b>
<b>MAX_POWER_1st</b>	Valor da potência de pico da primeira harmônica do segmento de PPG.
<b>IDX_MAX_1st</b>	Índice da posição do pico da primeira harmônica do segmento de PPG.
<b>F_MAX_1st</b>	Frequência onde ocorre o pico da primeira harmônica do segmento de PPG.
<b>MAX_POWER_2nd</b>	Valor da potência de pico da segunda harmônica do segmento de PPG.
<b>IDX_MAX_2nd</b>	Índice da posição do pico da segunda harmônica do segmento de PPG.
<b>F_MAX_2nd</b>	Frequência onde ocorre o pico da segunda harmônica do segmento de PPG.
<b>MEAN_PSD</b>	Valor médio da densidade de potência espectral do segmento de PPG.
<b>STD_PSD</b>	Desvio padrão da densidade de potência espectral do segmento de PPG.
<b>VAR_PSD</b>	Variância da densidade de potência espectral do segmento de PPG.
<b>KUR_PSD</b>	Kurtose da densidade de potência espectral do segmento de PPG.
<b>KTE</b>	Energia de Kaiser-Teagen.
<b>QZE</b>	Energia de Qi-Zheng.
<b>TOTAL_ENERGY</b>	Energia total do segmento de PPG.
<b>RMS</b>	Valor RMS da energia do segmento de PPG.

Fonte: Autor.

Tabela 17 – Características extraídas dos pulsos do sinal de PPG e posteriormente utilizadas nos modelos de aprendizado de máquina.

<b>Característica</b>	<b>Descrição</b>
<b>SYS_VALUE</b>	Valor da amplitude do pico sistólico.
<b>SYS_PEAK</b>	Posição do pico sistólico.
<b>RISING_TIME</b>	Tempo de duração, em amostras, entre o início do pulso e o pico sistólico.
<b>SKEWNESS</b>	Assimetria do pulso de PPG.
<b>KURTOSIS_F</b>	Kurtose do pulso de PPG pelo critério de Fisher.
<b>KURTOSIS_P</b>	Kurtose do pulso de PPG pelo critério de Pearson.
<b>VARIANCE</b>	Variancia do pulso de PPG.
<b>W75</b>	Largura do pulso de PPG em 75% de amplitude.
<b>W50</b>	Largura do pulso de PPG em 50% de amplitude.
<b>W25</b>	Largura do pulso de PPG em 25% de amplitude.
<b>W10</b>	Largura do pulso de PPG em 10% de amplitude.
<b>NOTCH</b>	Posição do notch do pulso de PPG.
<b>DIA_VALUE</b>	Valor da amplitude do pico diastólico.
<b>DIA_PEAK</b>	Posição do pico diastólico.
<b>DIA_SLOPE</b>	Tempo de duração entre o pico diastólico e o final do pulso.
<b>AREA</b>	Valor da área sobre a curva do pulso de PPG extraído utilizando a regra trapezoidal.
<b>AREA_1</b>	Valor da área sobre a curva entre o início do pulso até o pico diastólico.
<b>AREA_2</b>	Valor da área sobre a curva entre o pico diastólico e o final do pulso.
<b>IPA</b>	Área do Ponto de inflexão. Definido como a razão entre AREA_2 e AREA_1.
<b>A</b>	Ponto "A" da segunda derivada do pulso de PPG.
<b>B</b>	Ponto "B" da segunda derivada do pulso de PPG.
<b>C</b>	Ponto "C" da segunda derivada do pulso de PPG.
<b>D</b>	Ponto "D" da segunda derivada do pulso de PPG.
<b>E</b>	Ponto "E" da segunda derivada do pulso de PPG.
<b>B/A</b>	Razão entre pontos B e A da segunda derivada do pulso de PPG.
<b>B/A_t</b>	Razão entre pontos B e A da segunda derivada do pulso de PPG em termos de segundos.
<b>C/A</b>	Razão entre pontos C e A da segunda derivada do pulso de PPG.
<b>C/A_t</b>	Razão entre pontos C e A da segunda derivada do pulso de PPG em termos de segundos.
<b>D/A</b>	Razão entre pontos D e A da segunda derivada do pulso de PPG.
<b>D/A_t</b>	Razão entre pontos D e A da segunda derivada do pulso de PPG em termos de segundos.
<b>E/A</b>	Razão entre pontos E e A da segunda derivada do pulso de PPG.
<b>E/A_t</b>	Razão entre pontos E e A da segunda derivada do pulso de PPG em termos de segundos.
<b>X</b>	Diferença entre o valor do pico sistólico e o último ponto do pulso de PPG.
<b>Y</b>	Diferença entre o valor do pico diastólico e o último ponto do pulso de PPG.
<b>AI</b>	<i>Augmentation Index</i> . Valor da razão entre a diferença entre X e Y, e X.
<b>RI</b>	<i>Reflection Index</i> . Valor da razão entre Y e X.
<b>DT</b>	Valor absoluto da diferença entre a posição do pico sistólico e diastólico.
<b>DT_t</b>	Valor absoluto da diferença entre a posição do pico sistólico e diastólico em termos de segundos.

Fonte: Autor.

**APÊNDICE C – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)**

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Você está sendo convidado(a) como voluntário(a) a participar da pesquisa “**Estimativa dos níveis de glicemia através do processamento de sinais de PPG e modelos de inteligência artificial**”, que tem como objetivo a utilização do sinal de Fotopletismografia (PPG) para implementação de uma técnica de estimativa dos níveis de glicemia, de forma não invasiva e contínua. Normalmente, o método convencional é preciso fazer um pequeno furo, geralmente no dedo, para verificar a glicose. Este trabalho busca uma alternativa em que não é necessário amostra de sangue para ter uma estimativa confiável dos mesmos valores. Esta estimativa é um valor próximo ao valor real que seria obtido pela amostra de sangue.

Esta pesquisa está associada ao projeto de Dissertação de Mestrado de Luis Felipe de Deus, aluno do Programa de Pós-Graduação em Engenharia Elétrica e ao Trabalho de Conclusão de Curso de Bruno Castro Valle, aluno de graduação do curso de Engenharia Elétrica, ambos da Universidade Federal de Santa Catarina. Este documento que você está lendo é chamado de Termo de Consentimento Livre e Esclarecido (TCLE). Ele contém explicações sobre o estudo que você está sendo convidado a participar. Antes de decidir se deseja participar de livre e espontânea vontade você deverá ler e compreender todo o conteúdo.

Acreditamos que a pesquisa seja importante, porque visa a implementação de uma técnica que ajude, principalmente para pessoas que sofrem com diabetes e precisam diariamente observar a sua glicose através de equipamentos que requerem amostra de sangue (invasivos), se tornando assim um grande passo para melhorar a qualidade de vida e a segurança.

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina (CEPSH-UFSC). O CEPSH é um órgão colegiado interdisciplinar, deliberativo, consultivo e educativo, vinculado à Universidade Federal de Santa Catarina, mas independente na tomada de decisões, criado para defender os interesses dos participantes da pesquisa em sua integridade e dignidade e para contribuir no desenvolvimento da pesquisa dentro de padrões éticos.

### **PARTICIPAÇÃO NO ESTUDO**

A participação nesta pesquisa é voluntária, o que significa que você não é obrigado a aceitar. Caso decida participar, a pesquisa consistirá dos seguintes procedimentos: Primeiramente você irá responder um questionário solicitando as informações como idade, etnia, sexo e algumas informações pertinentes para a pesquisa. Em seguida, será aplicado um questionário referente à avaliação de doenças e condições existentes, como medicações e tratamentos que o participante de pesquisa faz uso.

Será colocado um dispositivo do tipo grampo (clip) no dedo indicador da mão direita do participante de pesquisa. Este dispositivo irá coletar sinais de Fotopletismografia (PPG) por 5 (cinco) minutos e não tem nenhum desconforto. Em seguida, será coletado os níveis de glicemia do participante de pesquisa utilizando um glicosímetro, onde será feito um pequeno furo no dedo indicador da mão esquerda para coletar uma gota de sangue. Vale ressaltar que durante todos os procedimentos, isto é, desde a aplicação do questionário até a aquisição dos níveis de glicemia você estará sempre acompanhado(a) ou pelo pesquisador, ou médico associado ou profissionais da saúde, que lhe prestarão toda a assistência e esclarecimentos necessários.



A duração da sua participação é estimada em 10 minutos. Os procedimentos serão realizados no laboratório de Endocrinologia e Metabologia do Hospital Universitário Polydoro de São Thiago da Universidade Federal de Santa Catarina (HU-UFSC/EBSERH), no estado de Santa Catarina, Brasil.

## RISCOS E BENEFÍCIOS

É importante destacar que você não deve esperar qualquer benefício pessoal proveniente da pesquisa e que toda pesquisa envolve riscos. O protocolo do presente estudo envolve duas etapas:

1. O procedimento de aferição dos níveis de glicemia do participante é feito através de um medidor convencional que requer uma amostra de sangue obtida de um dos dedos da mão.
2. A aquisição do sinal de Fotopletismografia (PPG), que será feito através de um oxímetro de pulso do tipo grampo, colocado no dedo indicador da mão direita do participante.

Dentre os riscos que prevemos existe o risco de contaminação do participante através da exposição invasiva, ainda que mínima do sangue no dedo. Como também a contaminação pelo vírus da COVID-19. Estes riscos são considerados médios, devido à pequena área de exposição invasiva, bem como todos os membros da pesquisa terem completado o esquema vacinal com pelo menos duas doses ou a vacina de dose única.

Para minimizar estes riscos, os pesquisadores irão utilizar luvas descartáveis e máscaras em todo o decorrer do protocolo. Ainda, estarão presentes no laboratório apenas os responsáveis pela pesquisa, quando necessário. Existe também o risco de quebra de sigilo do participante da pesquisa, sendo este risco considerado baixo. Para minimizar este risco, os pesquisadores tomarão todas as medidas possíveis para manter o sigilo do participante da pesquisa, identificando-os apenas através de códigos, mantendo a correlação entre os participantes de pesquisa e códigos em local sigiloso e separado dos dados e não publicando nenhuma informação que possa identificar o participante da pesquisa. Isto constitui um risco baixo. Pela composição dos riscos, a participação nesta pesquisa pode ser considerada como **de médio risco**. Caso ocorram danos causados pela participação na pesquisa, devido aos riscos previstos ou não previstos, o participante da pesquisa e seus acompanhantes terão direito a acompanhamento, tratamento médico, indenização e reparo.

Este projeto visa promover benefícios coletivos à sociedade, em particular a as pessoas que sofrem de diabetes e necessitam aferir os níveis de glicemia periodicamente, algumas vezes por mais de duas vezes ao dia. Contudo, no ponto de vista do participante, nenhum benefício será obtido por meio de sua participação, posto que a metodologia proposta não equivale como procedimentos realizados no tratamento da sua condição de saúde por um profissional médico. Será assegurado o direito do participante de não arcar com qualquer despesa relacionada com a sua participação direta nesta pesquisa. Logo, caso o participante tenha vindo apenas para o estudo, independente do seu grupo, será oferecido alimentação no formato de duas barras de cereal 25g e restituição do valor de duas passagens de ônibus coletivo, para o participante e seu acompanhante, se for aplicável. Além disso, não será oferecido nenhum bônus financeiro pela participação, almeja-se que o participante da pesquisa não obtenha lucros ou qualquer tipo de prejuízo financeiro decorrentes da sua participação nesta pesquisa.

## **SIGILO E PRIVACIDADE**

Tendo em vista que nenhuma identificação do participante de pesquisa será coletada, o risco de quebra de sigilo é mínimo, onde os dados do participante de pesquisa serão identificados por códigos numéricos, escolhidos de forma randômica. Os resultados deste trabalho poderão ser apresentados em encontros ou revistas científicas, contudo, mostrarão apenas os resultados obtidos como um todo, sem revelar data, instituição ou qualquer outra informação relacionada à sua privacidade.

## **RESSARCIMENTO E INDENIZAÇÃO**

A legislação brasileira não permite que você tenha qualquer compensação financeira pela sua participação na pesquisa, mas você ou seu acompanhante não terão nenhuma despesa advinda da sua participação na pesquisa. Caso alguma despesa extraordinária associada à pesquisa venha a ocorrer, você e seu acompanhante serão ressarcidos nos termos da lei, ou caso você tenha algum prejuízo material ou imaterial em decorrência da pesquisa poderá solicitar indenização, de acordo com a legislação vigente e amplamente consubstanciada.

## **ASSISTÊNCIA**

Você e o seu acompanhante receberão assistência integral e imediata, de forma gratuita, pelo tempo que for necessário em caso de danos decorrentes da pesquisa.

## **AUTONOMIA**

Antes de assinar, faça perguntas sobre tudo o que não tiver entendido bem. Sua participação é voluntária, portanto, sinta-se absolutamente à vontade em deixar de participar da pesquisa a qualquer momento, sem ter que apresentar qualquer tipo de justificativa. A desistência da participação da pesquisa não lhe trará nenhum tipo de prejuízo ou penalidade, bastará você entrar em contato com um dos pesquisadores, retirando seu consentimento.

## **CONTATO**

A equipe deste estudo responderá às suas dúvidas a qualquer momento antes, durante e após o estudo. Caso tenha alguma dúvida sobre os procedimentos ou sobre o projeto, você poderá entrar em contato com os pesquisadores envolvidos com o projeto que são Jefferson Luiz Brum Marques, Cesar Ramos Rodrigues, Luis Felipe de Deus, Bruno Castro Valle, todos vinculados ao Instituto de Engenharia Biomédica da Universidade Federal de Santa Catarina (IEB UFSC), no qual, poderá manter contato pelos endereços pessoais ou profissionais, dos pesquisadores. Pesquisador Cesar Ramos Rodrigues, telefone: (48) 998671778 ou pelo e-mail [cesar@ieee.org](mailto:cesar@ieee.org). Pesquisador Jefferson Luiz Brum Marques, telefone: (48) 37218686, ou pelo e-mail [jmarques@ieb.ufsc.br](mailto:jmarques@ieb.ufsc.br). Pesquisador Luis Felipe de Deus, telefone: (55) 999136926 ou pelo e-mail: [luis.fd@posgrad.ufsc.br](mailto:luis.fd@posgrad.ufsc.br). Pesquisador Bruno Castro Valle, telefone (48) 991094275 ou pelo email: [bruno.castro.valle@gmail.com](mailto:bruno.castro.valle@gmail.com). Ou também pelo endereço profissional dos pesquisadores Instituto de Engenharia Biomédica (IEB) localizado na Rua Desembargador Vitor Lima S/N, Trindade, Florianópolis. Horário de funcionamento: 2ª a 6ª feira – 8:00 às 12:00 e das 13:30 às 17:30min. Telefone para contato: (48) 3721-868.

Dúvidas sobre a pesquisa envolvendo princípios éticos poderão ser questionadas ao Comitê de Ética em Pesquisa com Seres Humanos da UFSC (CEPSH-UFSC), com atendimento no prédio Reitoria II, 7º andar, sala 701, localizado na Rua Desembargador Vitor Lima, nº 222, Trindade, Florianópolis. Telefone para contato: (48) 3721-6094 e

e-mail cep.propesq@contato.ufsc.br. Horário de funcionamento: 2ª a 6ª feira – 10:00h às 12:00h e das 16:00 às 18:00. Reclamações e/ou insatisfações relacionadas à participação na pesquisa também poderão ser comunicadas por escrito à Secretaria do CEPESH-UFSC.

### **DECLARAÇÃO**

Este documento foi elaborado em duas vias, datadas, com páginas numeradas, uma das vias será destinada a você e outra ao pesquisador responsável. Tanto você quanto o pesquisador responsável devem assinar ambas as vias. Todas as páginas deverão ser rubricadas tanto por você, como pelo pesquisador responsável. Guarde cuidadosamente a sua via, pois é um documento que traz informações importantes de contato e garante os seus direitos como participante desta pesquisa.

O pesquisador responsável, que também assina esse documento, compromete-se a conduzir a pesquisa de acordo com o que preconiza a Resolução 466/12 de 12/06/2012, que trata dos preceitos éticos e da proteção aos participantes da pesquisa.

Pesquisador responsável: \_\_\_\_\_

Assinatura: \_\_\_\_\_

**Eu, \_\_\_\_\_,**  
**Concordo com a participação por livre e espontânea vontade da pesquisa**  
**“Estimativa dos níveis de glicemia através do processamento de sinais de PPG e**  
**modelos de inteligência artificial”.**

Florianópolis, \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_.

\_\_\_\_\_  
Assinatura do participante da pesquisa

## **APÊNDICE D – QUESTIONÁRIOS APLICADOS DURANTE A COLETA DE DADOS**

## QUESTIONÁRIO 1

**Informativo:** Este questionário faz parte da pesquisa “*Estimativa dos níveis de glicemia através do processamento de sinais de PPG e modelos de inteligência artificial*” cujo Pesquisadores Responsáveis são o Prof. Dr. Cesar Ramos Rodrigues, Prof. Dr. Jefferson Luiz Brum Marques, Eng. Luis Felipe de Deus e Bruno Castro Valle, vinculados ao Instituto de Engenharia Biomédica da Universidade Federal de Santa Catarina.

O questionário somente poderá ser respondido pelos participantes que estão de acordo e assinaram devidamente o Termo de Consentimento Livre e Esclarecido (TCLE).

Todas as informações a serem preenchidas nos itens a seguir devem ser realizadas somente pelo participante da pesquisa.

### Dados Pessoais

Idade: \_\_\_\_ anos

Peso aproximado: \_\_\_\_\_ kg

Altura aproximada: \_\_\_\_\_ cm

Etnia: ( ) Branca ( ) Parda ( ) Preta ( ) Amarela ( ) Indígena

Sexo: ( ) Masculino ( ) Feminino ( ) Outro.

---

**Participante da pesquisa**

---

**Pesquisador responsável**

## QUESTIONÁRIO 2

**Informativo:** Este questionário faz parte da pesquisa “*Estimativa dos níveis de glicemia através do processamento de sinais de PPG e modelos de inteligência artificial*” cujo Pesquisadores Responsáveis são o Prof. Dr. Cesar Ramos Rodrigues, Prof. Dr. Jefferson Luiz Brum Marques, Eng. Luis Felipe de Deus e Bruno Castro Valle, vinculados ao Instituto de Engenharia Biomédica da Universidade Federal de Santa Catarina.

O questionário somente poderá ser respondido pelos participantes que estão de acordo e assinaram devidamente o Termo de Consentimento Livre e Esclarecido (TCLE).

Todas as informações a serem preenchidas nos itens a seguir devem ser realizadas somente pelo participante da pesquisa.

### INFORMAÇÕES RELEVANTES

- a) **Você ingeriu algum alimento ou bebida nas últimas 8h?**  
 Sim  Não
- b) **Se sim, qual a predominância do alimento, por exemplo: Proteínas, Carboidratos, Lipídios. Se preferir pode descrever a refeição.**  
\_\_\_\_\_
- c) **Você faz uso de algum medicamento de uso contínuo?**  
 Sim  Não  
Se sim, qual? \_\_\_\_\_
- d) **Você faz uso de algum tratamento contínuo?**  
 Sim  Não  
Se sim, qual? \_\_\_\_\_
- e) **Se disponível, qual o último resultado do exame de Hemoglobina Glicada (HbA1C) ?**  
R. \_\_\_\_\_
- f) **Você é diagnosticado como portador de diabetes ?**  
 Sim  Não
- g) **Você possui algum aparelho medidor de níveis de glicemia (Glicosímetro) em sua casa?**  
 Sim  Não
- h) **Com qual frequência você faz o acompanhamento dos níveis de glicemia?**  
 Diariamente  Semanalmente  Mensalmente  Não faz
- i) **Se diariamente ou semanalmente, quantas vezes em média ?**  
Diariamente \_\_\_\_\_ vezes por dia.  
Semanalmente \_\_\_\_\_ vezes por semana.

- j) **Com que frequência você sofre com altos níveis de glicemia ?**  
( ) Diariamente ( ) Semanalmente ( ) Mensalmente ( ) Não sei dizer

---

**Participante da pesquisa**

---

**Pesquisador responsável**