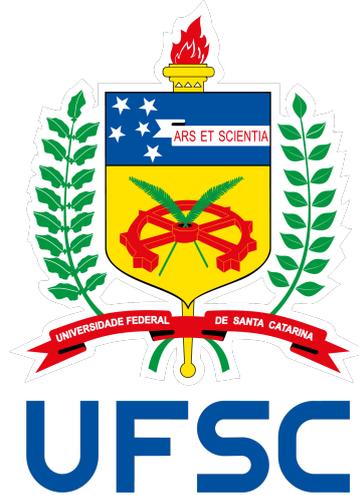


Universidade Federal de Santa Catarina  
Campus Reitor João David Ferreira Lima  
Departamento de Engenharia Elétrica e Eletrônica



Guilherme Alves Nunes de Barros

Estimativa dos Níveis de Glicose Usando Processamento de  
Sinais de Fotopletismografia com Suporte da Variabilidade da  
Frequência Cardíaca e Modelos de Machine Learning

Florianópolis

2024

**Guilherme Alves Nunes de Barros**

**Estimativa dos Níveis de Glicose Usando  
Processamento de Sinais de Fotopletismografia com  
Suporte da Variabilidade da Frequência Cardíaca e  
Modelos de Machine Learning**

Trabalho de Conclusão de Curso apresentado à Universidade Federal de Santa Catarina como parte dos requisitos necessários para a obtenção do título de Bacharel em Engenharia Eletrônica.  
Orientador: Prof. Cesar Ramos Rodrigues, Dr.

Universidade Federal de Santa Catarina  
Campus Reitor João David Ferreira Lima  
Departamento de Engenharia Elétrica e Eletrônica

Florianópolis  
2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.  
Dados inseridos pelo próprio autor.

Barros, Guilherme Alves Nunes de  
Estimativa dos níveis de glicose usando processamento de  
sinais de fotopletismografia com suporte da variabilidade  
da frequência cardíaca e modelos de machine learning /  
Guilherme Alves Nunes de Barros ; orientador, Cesar Ramos  
Rodrigues, 2024.  
86 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Centro Tecnológico,  
Graduação em Engenharia Eletrônica, Florianópolis, 2024.

Inclui referências.

1. Engenharia Eletrônica. 2. Glicose. 3.  
Fotopletismografia. 4. Inteligência Artificial. 5.  
Monitoramento Não Invasivo. I. Rodrigues, Cesar Ramos .  
II. Universidade Federal de Santa Catarina. Graduação em  
Engenharia Eletrônica. III. Título.

# Guilherme Alves Nunes de Barros

## Estimativa dos Níveis de Glicose Usando Processamento de Sinais de Fotopletismografia com Suporte da Variabilidade da Frequência Cardíaca e Modelos de Machine Learning

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “Bacharel em Engenharia Eletrônica” e aprovado em sua forma final pelo Curso de Graduação em Engenharia Eletrônica.

Florianópolis, 19 de dezembro de 2024

---

Profa. Dra. Daniela Ota Hisayasu Suzuki,  
Coordenadora do Curso

### **Banca Examinadora:**

---

Prof. Cesar Ramos Rodrigues, Dr.  
Universidade Federal de Santa Catarina  
Orientador

---

Prof. Jefferson Luiz Brum Marques, Dr.  
Universidade Federal de Santa Catarina  
Avaliador

---

Sofia Pinheiro Klautau, Eng. Biomed.  
Universidade Federal de Santa Catarina  
Avaliadora

Dedico este trabalho à memória da minha querida tia Edna. Sua presença foi uma luz em minha vida, sempre me inspirando com sua sabedoria, bondade e amor incondicional. Mesmo em sua ausência, carrego comigo as lições que você me ensinou e a força que sempre transmitiu. Este trabalho é uma homenagem à sua memória, com o desejo de honrar tudo o que você representou e continua representando para mim. Obrigado por ter sido uma inspiração e um exemplo a ser seguido. Sua ausência é sentida profundamente, mas seu legado vive em meu coração.

# Agradecimentos

Primeiramente a agradeço minha família, que sempre esteve ao meu lado, me apoiando incondicionalmente. Aos meus pais, pela educação, amor e incentivo constantes, e por acreditarem no meu potencial em cada etapa da minha vida. Aos meus irmãos, pelo companheirismo e por serem minha inspiração diária.

Ao meu orientador, Prof. Dr. Cesar Ramos Rodrigues, pela paciência, dedicação e valiosas orientações que foram fundamentais para a realização deste trabalho. Sua confiança e comprometimento com o meu aprendizado foram essenciais para o meu crescimento acadêmico e pessoal.

Aos meus colegas e amigos, que compartilharam dessa caminhada comigo, seja através de conselhos, palavras de encorajamento ou momentos de descontração. A cada um de vocês, o meu muito obrigado por tornarem essa jornada mais leve e significativa.

Gostaria também de expressar minha gratidão aos professores e à instituição de ensino, que proporcionaram o conhecimento e os recursos necessários para que eu pudesse alcançar este objetivo.

Por fim, agradeço a todos que, de alguma forma, contribuíram para a realização deste trabalho, mesmo que indiretamente. A todos, o meu mais sincero obrigado.

A missão da ciência é aliviar o sofrimento humano.  
(Autor desconhecido)

# Resumo

A glicose é a principal fonte de energia do organismo, obtida por meio da alimentação. Seu uso eficiente pelas células depende da regulação adequada dos níveis de glicose no sangue, controlada pela insulina, hormônio produzido pelas células pancreáticas. Na diabetes mellitus, há disfunção na produção ou uso da insulina, resultando em desequilíbrios que podem causar complicações graves, como doenças renais, infartos e acidentes vasculares cerebrais. Segundo a Organização Mundial da Saúde (OMS), a diabetes é a nona maior causa de mortes no mundo, com 1,5 milhão de óbitos anuais. O monitoramento regular dos níveis de glicose no sangue é essencial, especialmente em pessoas com diabetes, permitindo a detecção precoce de complicações. No entanto, os métodos atuais são invasivos, caros e podem ser desconfortáveis, criando a necessidade de alternativas não invasivas e acessíveis. Este estudo propõe uma solução baseada em sinais de fotopletismografia (PPG) e algoritmos de Inteligência Artificial, investigando como a idade e a Variabilidade da Frequência Cardíaca (HRV) podem afetar a precisão das estimativas de glicose. Utilizou-se o conjunto de dados UFSC-BGL, disponibilizado por (DEUS, Luis Felipe de, 2023), para aprimorar estudos prévios e testar novas abordagens. Após o pré-processamento, os sinais PPG foram segmentados em janelas de 10 segundos, sem sobreposição, com extração de 65 características nos domínios da frequência, não linear e morfológico. Os dados foram divididos em treinamento (70%) e teste (30%) para treinar e avaliar modelos de aprendizado de máquina. A análise revelou que a limitação por idade até 50 anos, combinada à Decomposição Empírica de Modos (EMD) e à HRV, apresentou os melhores resultados. O modelo Random Forest (RF) obteve um erro médio absoluto (MAE) de 7,69 mg/dL, erro médio absoluto percentual (MAPE) de 6,27% e 92,2% das estimativas na zona A do gráfico de Clarke, destacando robustez e confiabilidade. O modelo CatBoost também demonstrou bom desempenho, com MAPE de 6,57% e 86,2% das estimativas na zona A. Este estudo traz três contribuições principais: o aprimoramento das análises do conjunto de dados UFSC-BGL, a inclusão de HRV como variável analítica, e o desenvolvimento de modelos precisos que consideram a influência da idade e da dinâmica autonômica na estimativa da glicose, reforçando o potencial de soluções não invasivas para monitoramento contínuo.

**Palavras-Chave:** Glicose, Fotopletismografia, Inteligência Artificial, Monitoramento Não Invasivo.

# Abstract

Glucose is the primary source of energy for the body, obtained through diet. Its efficient utilization by cells depends on the proper regulation of blood glucose levels, controlled by insulin, a hormone produced by pancreatic cells. In diabetes mellitus, there is a dysfunction in insulin production or utilization, resulting in imbalances that can lead to severe complications such as kidney disease, heart attacks, and strokes. According to the World Health Organization (WHO), diabetes is the ninth leading cause of death worldwide, with 1.5 million annual fatalities. Regular monitoring of blood glucose levels is essential, particularly for individuals with diabetes, enabling early detection of complications. However, current methods are invasive, costly, and potentially uncomfortable, highlighting the need for non-invasive and accessible alternatives. This study proposes a solution based on photoplethysmography (PPG) signals and Artificial Intelligence algorithms, investigating how age and Heart Rate Variability (HRV) may impact the accuracy of glucose estimates. The UFSC-BGL dataset, provided by (DEUS, Luis Felipe de, 2023), was utilized to refine previous studies and test new approaches. After preprocessing, PPG signals were segmented into 10-second windows without overlap, with the extraction of 65 features across frequency, non-linear, and morphological domains. The data were split into training (70%) and testing (30%) sets to train and evaluate machine learning models. The analysis revealed that age restriction up to 50 years, combined with Empirical Mode Decomposition (EMD) and HRV, yielded the best results. The Random Forest (RF) model achieved a mean absolute error (MAE) of 7.69 mg/dL, a mean absolute percentage error (MAPE) of 6.27%, and 92.2% of estimates in zone A of the Clarke error grid, demonstrating robustness and reliability. The CatBoost model also showed good performance, with a MAPE of 6.57% and 86.2% of estimates in zone A. This study makes three main contributions: the refinement of analyses using the UFSC-BGL dataset, the inclusion of HRV as an analytical variable, and the development of accurate models that account for the influence of age and autonomic dynamics on glucose estimation, reinforcing the potential of non-invasive solutions for continuous monitoring.

**Keywords:** Glucose, Photoplethysmography, Artificial Intelligence, Non-Invasive Monitoring.

# Lista de figuras

Figura 1 – Forma de onda de ECG e PPG . . . . .	25
Figura 2 – Esquemático para aquisição do sinal de PPG. . . . .	26
Figura 3 – Exemplo de resposta em frequência para os filtros. . . . .	30
Figura 4 – Ilustração da derivação de dados da HRV dos sinais de ECG e PPG, (a) intervalos RR no sinal de ECG e (b) intervalos PP no sinal de PPG. N.U. (Unidades normalizadas). . . . .	33
Figura 5 – Exemplo de gráfico de erro de Clarke. . . . .	37
Figura 6 – Fluxograma detalhado do pré-processamento. . . . .	45
Figura 7 – Fluxograma detalhado do pré-processamento. . . . .	46
Figura 8 – Segmento de 10 s do primeiro (001) @60Hz, glicose de 181 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. . . . .	53
Figura 9 – Domínio da frequência do participante 001: Respectivamente, seg- mento de sinal processado e gráfico de Densidade Espectral de Potência. 54	
Figura 10 – Domínio do tempo do participante 001: Pulsos sobrepostos (em azul) de um segmento de 10 segundos, com o pulso médio realçado em vermelho . . . . .	54
Figura 11 – Extração de características do pulso do sinal de PPG, VPG e APG, destacados em azul, vermelho e verde. Participante 001 pulso re- amostrado a @200 Hz, com glicose de 181 mg/dL. . . . .	55
Figura 12 – Histograma da distribuição de valores de glicose (mg/dL) no conjunto de dados dos pacientes. . . . .	56
Figura 13 – Histograma da distribuição de valores de glicose (mg/dL) nos conjun- tos de treino e teste para os diferentes segmentos de pacientes. . . . .	57
Figura 14 – Gráfico de Clarke das predições de glicose. . . . .	58
Figura 15 – Feature Mais Influentes. . . . .	59
Figura 16 – Gráfico de Clarke das predições de glicose (sem divisão por idade). . .	60
Figura 17 – Gráfico de Clarke das predições de glicose (participantes até 50 anos). 61	
Figura 18 – Gráfico de Clarke das predições de glicose (participantes acima de 50 anos). . . . .	62

Figura 19 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de treino e teste com agrupamento de segmentos. . . . .	63
Figura 20 – Gráfico de Clarke das predições de glicose (agrupamento de segmentos, sem divisão por idade). . . . .	64
Figura 21 – Gráfico de Clarke das predições de glicose (agrupamento de segmentos, até 50 anos). . . . .	65
Figura 22 – Gráfico de Clarke das predições de glicose (agrupamento de segmentos, acima de 50 anos). . . . .	66
Figura 23 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de teste e treino para os segmentos dos pacientes com a incorporação da EMD. . . . .	67
Figura 24 – Gráfico de Clarke das predições de glicose (sem divisão por idade). . . . .	68
Figura 25 – Gráfico de Clarke das predições de glicose (até 50 anos). . . . .	69
Figura 26 – Gráfico de Clarke das predições de glicose (com mais de 50 anos). . . . .	70
Figura 27 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de teste e treino para os segmentos dos pacientes agrupados com a incorporação da EMD. . . . .	71
Figura 28 – Gráfico de Clarke das predições de glicose (sem divisão por idade). . . . .	72
Figura 29 – Gráfico de Clarke das predições de glicose (até 50 anos). . . . .	73
Figura 30 – Gráfico de Clarke das predições de glicose (com mais de 50 anos). . . . .	74

# Lista de tabelas

Tabela 1	– Resultados dos modelos com o conjunto completo. . . . .	58
Tabela 2	– Distribuição das predições nas zonas de precisão do modelo de erro de Clarke. . . . .	58
Tabela 3	– Resultados dos modelos com o conjunto completo (sem divisão por idade). . . . .	60
Tabela 4	– Distribuição das predições nas zonas de precisão do modelo de erro de Clarke. . . . .	60
Tabela 5	– Resultados dos modelos com participantes até 50 anos. . . . .	61
Tabela 6	– Distribuição das predições nas zonas de precisão do modelo de erro de Clarke (até 50 anos). . . . .	61
Tabela 7	– Resultados dos modelos com participantes acima de 50 anos. . . . .	62
Tabela 8	– Distribuição das predições nas zonas de precisão do modelo de erro de Clarke (acima de 50 anos). . . . .	62
Tabela 9	– Resultados dos modelos com agrupamento de segmentos (sem divisão por idade). . . . .	63
Tabela 10	– Distribuição das predições nas zonas de precisão do modelo de erro de Clarke. . . . .	64
Tabela 11	– Resultados dos modelos com agrupamento de segmentos (participantes até 50 anos). . . . .	64
Tabela 12	– Distribuição das predições nas zonas de precisão do modelo de erro de Clarke. . . . .	65
Tabela 13	– Resultados dos modelos com agrupamento de segmentos (participantes acima de 50 anos). . . . .	65
Tabela 14	– Distribuição das predições nas zonas de precisão do modelo de erro de Clarke. . . . .	66
Tabela 15	– Resultados dos modelos com o conjunto completo (sem divisão por idade). . . . .	67
Tabela 16	– Resultado da distribuição das predições utilizando o modelo de erro de Clarke. . . . .	68
Tabela 17	– Resultados dos modelos com participantes com até 50 anos. . . . .	68
Tabela 18	– Resultado da distribuição das predições utilizando o modelo de erro de Clarke. . . . .	69

Tabela 19 – Resultados dos modelos com participantes com mais de 50 anos. . . .	69
Tabela 20 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke. . . . .	70
Tabela 21 – Resultados dos modelos com o conjunto completo (sem divisão por idade). . . . .	71
Tabela 22 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke. . . . .	72
Tabela 23 – Resultados dos modelos com participantes com até 50 anos. . . . .	72
Tabela 24 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke. . . . .	73
Tabela 25 – Resultados dos modelos com participantes com mais de 50 anos. . . .	73
Tabela 26 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke. . . . .	74
Tabela 27 – Resultados dos modelos obtidos por (DEUS, Luis Felipe de, 2023). . .	76
Tabela 28 – Atributos morfológicos extraídos dos pulsos do sinal de PPG. . . . .	84
Tabela 29 – Atributos relacionados à variabilidade da frequência cardíaca (HRV) extraídos dos pulsos do sinal de PPG. . . . .	85
Tabela 30 – Atributos derivados da decomposição em modos empíricos (EMD) extraídos dos pulsos do sinal de PPG. . . . .	85

# Lista de Siglas e Abreviaturas

AGES	<i>Advanced Glycation End-products (Produtos de Glicação Avançada)</i>
APG	<i>Acceleration Plethysmogram (Plethysmograma de Aceleração)</i>
ASRAREA	<i>Área da Representação Analítica do Sinal</i>
ATP	<i>Adenosina Trifosfato</i>
AVC	<i>Acidente Vascular Cerebral</i>
BGL	<i>Blood Glucose Level (Nível de Glicose no Sangue)</i>
CAN	<i>Cardiac Autonomic Neuropathy (Neuropatia Autonômica Cardiovascular)</i>
CAAE	<i>Certificado de Apresentação para Apreciação Ética</i>
CATBOOST	<i>Algoritmo de Aprendizado de Máquina Baseado em Gradient Boosting</i>
CEPSH	<i>Comitê de Ética em Pesquisa com Seres Humanos</i>
EBSERH	<i>Empresa Brasileira de Serviços Hospitalares</i>
ECG	<i>Eletrocardiograma</i>
EMD	<i>Empirical Mode Decomposition (Decomposição Empírica de Modos)</i>
FMI	<i>Função de Modo Intrínseco (Intrinsic Mode Function)</i>
FFT	<i>Fast Fourier Transform (Transformada Rápida de Fourier)</i>
FPG	<i>Fasting Plasma Glucose (Glicose em Jejum)</i>
GCM	<i>Continuous Glucose Monitor (Monitor de Glicose de Uso Contínuo)</i>
HBA1c	<i>Hemoglobina Glicada</i>
HDEM	<i>Método Hilbert Double Envelope</i>
HHT	<i>Transformada de Hilbert-Huang</i>
HRV	<i>Heart Rate Variability (Variabilidade da Frequência Cardíaca)</i>
HU-UFSC	<i>Hospital Universitário Polydoro Ernani de São Thiago</i>
IBIs	<i>Intervalos Interbatimentos (Inter-Beat Intervals)</i>
IMC	<i>Índice de Massa Corporal</i>
IMF	<i>Funções Modais Intrínsecas (Intrinsic Mode Functions)</i>
IPA	<i>Inflection Point Area (Área do Ponto de Inflexão)</i>
IQR-RR	<i>Interquartile Range RR (Intervalo Interquartil dos Intervalos RR)</i>
LED	<i>Light Emitting Diode (Diodo Emissor de Luz)</i>
L1 e L2	<i>Regularizações de penalização utilizadas em modelos de aprendizado de máquina</i>
MAE	<i>Mean Absolute Error (Erro Médio Absoluto)</i>
MAPE	<i>Mean Absolute Percentage Error (Erro Médio Absoluto Percentual)</i>
ME	<i>Mean Error (Erro Médio)</i>

MSE	<i>Mean Squared Error (Erro Quadrático Médio)</i>
N.U.	<i>Unidades normalizadas</i>
OMS	<i>Organização Mundial da Saúde (World Health Organization)</i>
PA	<i>Pressão Arterial</i>
PNN50	<i>Percentual de Intervalos RR com Diferenças Maiores que 50 ms</i>
PPG	<i>Photoplethysmography (Fotoplethysmografia)</i>
PSD	<i>Power Spectral Density (Densidade Espectral de Potência)</i>
PSDBPOW	<i>Potência de Banda da Densidade Espectral</i>
PSDMFREQ	<i>Frequência Média da Densidade Espectral</i>
PSDPKAMP	<i>Amplitude de Pico da Densidade Espectral de Potência</i>
PTP	<i>Peak-to-Peak Amplitude (Amplitude Pico a Pico)</i>
RF	<i>Random Forest (Floresta Aleatória)</i>
RMSE	<i>Root Mean Squared Error (Raiz do Erro Quadrático Médio)</i>
RMSSD	<i>Root Mean Square of Successive Differences (Raiz Quadrada da Média dos Quadrados d</i>
R <sup>2</sup>	<i>Coefficient of Determination (Coeficiente de Determinação)</i>
RBF	<i>Radial Basis Function (Função de Base Radial)</i>
SDNN	<i>Standard Deviation of NN Intervals (Desvio Padrão dos Intervalos NN)</i>
SD	<i>Standard Deviation (Desvio Padrão)</i>
SPO <sub>2</sub>	<i>Saturação de Oxigênio no Sangue</i>
SUS	<i>Sistema Único de Saúde</i>
SVM	<i>Support Vector Machine (Máquina de Vetores de Suporte)</i>
SSQI	<i>Skewness Signal Quality Index (Índice de Qualidade do Sinal Inclinado)</i>
SYS_PEAK	<i>Posição do Pico Sistólico</i>
UFSC	<i>Universidade Federal de Santa Catarina</i>
UFSC-BGL	<i>Base de Dados de Glicose no Sangue da UFSC</i>
WHO	<i>World Health Organization (Organização Mundial da Saúde)</i>
Z-SCORE	<i>Escala de Padronização de Dados</i>

# Sumário

1	INTRODUÇÃO . . . . .	17
1.1	Objetivos . . . . .	21
1.1.1	Objetivos específicos . . . . .	22
1.1.2	Organização do texto . . . . .	22
2	REVISÃO TEÓRICA . . . . .	24
2.1	Fotopletismografia . . . . .	24
2.2	Processamento de Sinais . . . . .	26
2.2.1	Filtro Savitzky-Golay . . . . .	27
2.2.2	Filtro Butterworth . . . . .	28
2.2.3	Normalização Min Max . . . . .	30
2.3	Inteligência Artificial . . . . .	31
2.4	Variabilidade da Frequência Cardíaca . . . . .	32
2.5	Decomposição Empírica por Modos . . . . .	34
2.6	Métricas de Avaliação . . . . .	35
3	REVISÃO BIBLIOGRÁFICA . . . . .	38
4	MÉTODOS E MATERIAIS . . . . .	43
4.1	Conjuntos de Dados . . . . .	43
4.2	Método . . . . .	44
4.2.1	Pré-processamento . . . . .	45
4.2.2	Janelamento . . . . .	46
4.2.3	Avaliação da qualidade do sinal . . . . .	47
4.2.4	Extração de características . . . . .	48
4.2.5	Particionamento de dados . . . . .	50
4.2.6	Aprendizado de Máquina . . . . .	51
5	RESULTADOS . . . . .	53
5.1	Treinamento utilizando todos os segmentos . . . . .	57
5.1.1	Conjunto Completo . . . . .	58
5.1.2	Conjunto Sem Divisão por Idade . . . . .	59
5.1.3	Conjunto Com Participantes Até 50 Anos . . . . .	60

5.1.4	Conjunto Com Participantes Maiores de 50 Anos . . . . .	61
5.2	Treinamento Utilizando Agrupamento de 10 Segmentos. . . .	63
5.2.1	Conjunto Sem Divisão por Idade . . . . .	63
5.2.2	Conjunto Com Participantes Até 50 Anos . . . . .	64
5.2.3	Conjunto Com Participantes Maiores de 50 Anos . . . . .	65
5.3	Treinamento utilizando todos os segmentos com a incorpo- ração da Decomposição Empírica por Modos . . . . .	66
5.3.1	Conjunto Sem Divisão por Idade. . . . .	67
5.3.2	Conjunto para Participantes com até 50 anos . . . . .	68
5.3.3	Conjunto para Participantes com Mais de 50 anos . . . . .	69
5.4	Treinamento Utilizando Agrupamento de 10 Segmentos com a incorporação da EMD . . . . .	70
5.4.1	Conjunto Sem Divisão por Idade . . . . .	71
5.4.2	Conjunto para Participantes com até 50 anos . . . . .	72
5.4.3	Conjunto para Participantes com Mais de 50 anos . . . . .	73
6	DISCUSSÃO . . . . .	75
7	CONCLUSÃO . . . . .	78
	REFERÊNCIAS . . . . .	79
A	APÊNDICE A . . . . .	85

# 1 Introdução

A diabetes mellitus é o termo coletivo para distúrbios metabólicos heterogêneos cuja principal causa é a hiperglicemia crônica, afetando milhões de pessoas ao redor do mundo. Atualmente, não existe uma cura para ela (SCHLEICHER et al., 2022). Essa condição ocorre quando o pâncreas não consegue produzir insulina em quantidade suficiente para atender às necessidades do organismo ou quando as células do corpo apresentam resistência à insulina, reduzindo sua capacidade de utilizar a glicose eficientemente. Esse desequilíbrio resulta em níveis elevados de glicose no sangue, condição conhecida como hiperglicemia, que pode levar a complicações a longo prazo, como danos aos vasos sanguíneos, nervos e órgãos (DEFRONZO; BONADONNA; FERRANNINI, 1992).

A glicose é um carboidrato simples essencial para a geração de energia no organismo, obtida principalmente por meio de alimentos como frutas, vegetais, grãos e produtos lácteos. Após sua digestão e absorção no trato gastrointestinal, a glicose é liberada na corrente sanguínea, permitindo sua distribuição para as células, onde será metabolizada para produzir energia. Além de sua utilização imediata, a glicose pode ser armazenada no fígado e nos músculos como glicogênio, uma forma de reserva energética mobilizada em situações de alta demanda, como durante exercícios intensos ou períodos prolongados de jejum. Esse sistema de armazenamento e liberação garante uma oferta constante de energia ao organismo, mesmo em condições de privação alimentar ou esforço físico extremo (MARATHE; GAO; CLOSE, 2017).

A insulina é um hormônio peptídico essencial, produzido pelas células  $\beta$  das ilhotas de *Langerhans* no pâncreas, que desempenha um papel central na regulação do equilíbrio glicêmico. Sua principal função é facilitar a captação da glicose do sangue pelas células, especialmente nas células musculares, adiposas e hepáticas, promovendo sua utilização como fonte de energia. Uma vez nas células, a glicose é metabolizada por meio de processos como a glicólise, resultando na produção de ATP (Adenosina Trifosfato).

O ATP é a principal molécula de energia utilizada pelas células para sustentar suas funções vitais. Ele age como a "moeda energética" do organismo, fornecendo a energia necessária para uma ampla gama de processos biológicos, incluindo o transporte ativo de íons e moléculas, a síntese de biomoléculas e a contração muscular. Assim, a ação da insulina não apenas regula os níveis de glicose no sangue, mas também é essencial para a manutenção dos processos metabólicos e energéticos que garantem o funcionamento adequado do organismo.

Além de promover a captação de glicose, a insulina regula seu armazenamento no fígado e nos músculos na forma de glicogênio, o principal reservatório de energia de rápida disponibilidade. Esse processo, denominado glicogênese, é crucial para estabilizar os níveis de glicose no sangue, particularmente após as refeições, quando os níveis de glicose tendem a aumentar devido à absorção de carboidratos (WILCOX, 2005; COMMITTEE; COMMITTEE, 2022).

A diabetes é uma condição metabólica crônica caracterizada pela regulação inadequada dos níveis de glicose no sangue, resultando em hiperglicemia persistente. Essa desregulação é causada por deficiência na produção de insulina, resistência à sua ação, ou uma combinação de ambos. A hiperglicemia crônica pode desencadear uma série de complicações a longo prazo, afetando múltiplos sistemas do organismo.

Entre as complicações mais comuns estão as doenças cardiovasculares, como infarto do miocárdio e acidente vascular cerebral (AVC), as quais são as principais causas de morbidade e mortalidade em pessoas com diabetes. Além disso, a nefropatia diabética, caracterizada por danos progressivos aos rins, é uma das principais causas de insuficiência renal crônica em todo o mundo.

A diabetes também pode causar neuropatia diabética, resultante em danos aos nervos periféricos, levando a dor, formigamento e perda de sensibilidade, especialmente nos membros inferiores. A retinopatia diabética é outra complicação significativa, sendo uma das principais causas de cegueira em adultos em idade produtiva. Por fim, a má circulação sanguínea associada à diabetes, combinada com a neuropatia, pode levar a ulcerações nos pés e, em casos graves, à necessidade de amputações (GONTIJO et al., 2024).

Segundo a (World Health Organization, 2023), a prevalência do diabetes tem aumentado significativamente nas últimas três décadas, tornando-se uma das principais preocupações de saúde pública no século XXI. Esse aumento é atribuído a diversos fatores, como mudanças nos hábitos alimentares, redução da atividade física, crescimento do envelhecimento populacional e urbanização acelerada. A expansão da doença representa um desafio global, exigindo estratégias de prevenção, diagnóstico precoce e manejo eficaz para conter suas consequências e aliviar a carga sobre os sistemas de saúde.

Conforme os dados mais recentes, a diabetes é uma das principais causas de morte e incapacidade no mundo, com cerca de 537 milhões de pessoas vivendo com a doença atualmente, representando 6,1% da população global. Esse número pode ultrapassar 1,3 bilhão até 2050, devido ao crescimento acelerado da prevalência, especialmente do diabetes tipo 2, que compõe 96% dos casos globais. A condição já é uma das dez principais causas de morte, com impacto significativo em doenças como insuficiência renal,

cegueira, AVC e ataques cardíacos. Além disso, a diabetes afeta desproporcionalmente pessoas em países de baixa e média renda, onde mais de 80% dos casos estão concentrados. O índice de massa corporal elevado é o principal fator de risco, seguido por alimentação inadequada, baixa atividade física, tabagismo e consumo de álcool. Prevenir e tratar a doença representa um desafio significativo para os sistemas de saúde em escala global (FEDERATION, 2021; (FIOCRUZ), 2023).

A estimativa precisa dos níveis de glicose no sangue (BGL, do inglês Blood Glucose Levels) é essencial para o controle e o tratamento eficaz da diabetes. Tradicionalmente, o método de referência, conhecido como (*Gold Standard*), é realizado por meio de testes laboratoriais que utilizam amostras de sangue do paciente. Embora altamente precisos, esses métodos são invasivos e podem ser inconvenientes, especialmente para pacientes que necessitam de monitoramento regular (GROSS et al., 2002).

Entre os principais testes laboratoriais amplamente utilizados para avaliar os níveis de glicose no sangue estão o teste de glicose em jejum (FPG, do inglês Fasting Plasma Glucose) e o teste de hemoglobina glicada (HbA1c). O teste de glicose em jejum mede a concentração de glicose no sangue após um período de jejum de pelo menos 8 horas, fornecendo uma estimativa pontual dos níveis glicêmicos no momento do exame. Por outro lado, o teste de HbA1c oferece uma visão mais abrangente, ao medir a média dos níveis de glicose no sangue nos últimos dois a três meses. Isso é possível porque a glicose no sangue se liga à hemoglobina das células vermelhas, e a quantidade dessa ligação reflete os níveis médios de glicose durante o ciclo de vida das hemácias.

O teste de HbA1c é amplamente considerado mais confiável para monitorar o controle glicêmico a longo prazo, ao eliminar as variações diárias que podem ocorrer em medições pontuais, como as do teste de glicose em jejum. Além disso, o HbA1c é menos influenciado por fatores temporários, como o estado de jejum ou ingestão recente de alimentos (SACKS; KIRKMAN; LITTLE, 2024). No entanto, ambos os testes são complementares e desempenham papéis importantes no diagnóstico e na avaliação do controle da diabetes, destacando-se como ferramentas indispensáveis no manejo dessa condição.

No laboratório, a análise da amostra de sangue permite determinar com alta precisão a concentração de glicose presente. Essa maior confiabilidade deve-se ao fato de ambos os testes amplamente utilizados, como o teste de glicose em jejum (FPG) e o teste de hemoglobina glicada (HbA1c), utilizarem sangue venoso em vez de sangue capilar. O uso de sangue venoso proporciona resultados mais consistentes e precisos, sendo um padrão ouro no diagnóstico e monitoramento da diabetes.

Apesar de exigir maior investimento financeiro e de tempo em comparação com dis-

positivos portáteis de monitoramento, esses testes são preferidos em contextos clínicos e de pesquisa devido à sua confiabilidade. Isso os torna essenciais para o diagnóstico e o monitoramento adequado do diabetes, permitindo ajustes precisos no tratamento e contribuindo para a prevenção de complicações associadas. A precisão dos testes laboratoriais também é vital para estabelecer diagnósticos diferenciais, identificar diabetes em estágios iniciais e avaliar a eficácia das intervenções terapêuticas.

Embora os métodos laboratoriais sejam amplamente reconhecidos por sua precisão, há situações em que soluções rápidas são necessárias, tornando o glicosímetro uma ferramenta prática e acessível. Este dispositivo, utilizado com tiras reagentes descartáveis, possibilita a medição rápida dos níveis de glicose no sangue capilar. O procedimento, geralmente realizado pelo próprio usuário, envolve a coleta de uma pequena amostra de sangue obtida por punção com uma lanceta, seguida de análise pelo dispositivo (VASHIST et al., 2011). Além do glicosímetro e das tiras, é essencial o uso de materiais básicos, como álcool e curativos, para prevenir infecções durante o processo.

Devido à sua simplicidade e rapidez, o método tradicional de monitoramento da glicemia é invasivo e pode ser desconfortável, especialmente para pessoas que precisam realizar várias medições diárias. Como alternativa moderna, os monitores de glicose de uso contínuo (GCM), como o FreeStyle Libre, da Abbott®, apresentam uma solução menos invasiva. Esses dispositivos utilizam um sensor fixado na parte posterior do braço, equipado com uma fina agulha que mede os níveis de glicose no fluido intersticial, reduzindo a necessidade de perfurações frequentes e oferecendo dados em tempo real. Essa abordagem melhora o conforto, a conveniência e a adesão ao tratamento, além de impactar positivamente na qualidade de vida dos pacientes.

No entanto, uma desvantagem significativa é o custo elevado associado a esses dispositivos. Diferentemente dos glicosímetros e tiras, disponibilizados pelo Sistema Único de Saúde (SUS) no Brasil, os sensores dos GCMs não são fornecidos pelo sistema público de saúde. Além disso, os sensores têm uma vida útil limitada, geralmente de cerca de 14 dias, o que pode tornar o monitoramento contínuo inacessível para muitos pacientes devido ao custo recorrente.

Os métodos apresentados possuem vantagens e desvantagens distintas, mas compartilham a característica de serem dispositivos invasivos. Contudo, há uma demanda crescente por dispositivos vestíveis que realizem o monitoramento de condições de saúde de forma não invasiva, especialmente com foco no bem-estar e na prevenção (NASIRI, 2019). Esses dispositivos, geralmente, não têm a finalidade de emitir diagnósticos clínicos. Em vez disso, fornecem dados básicos para o acompanhamento de biomarcadores relacionados à saúde.

Entre as principais tecnologias utilizadas em dispositivos vestíveis, o PPG destaca-se como uma ferramenta essencial para a aquisição de biomarcadores de forma não invasiva. O PPG possibilita a extração de informações relacionadas às variações no fluxo sanguíneo, refletindo, indiretamente, a atividade cardíaca. Trata-se de uma técnica de baixo custo e alta acessibilidade, que utiliza luz para iluminar áreas do corpo humanas ricas em vasos sanguíneos, geralmente sob uma fina camada de tecido e gordura, como a ponta dos dedos ou o lóbulo da orelha (TAMURA et al., 2014; LIANG et al., 2018a). Essa técnica aproveita a capacidade da luz de penetrar tecidos biológicos e ser absorvida pelos vasos sanguíneos, fornecendo dados valiosos sobre a dinâmica vascular e a perfusão tecidual (LOH et al., 2022).

O uso da PPG em dispositivos vestíveis tem ganhado destaque devido à sua capacidade de medir uma ampla gama de parâmetros fisiológicos, como frequência cardíaca, saturação de oxigênio no sangue (SpO<sub>2</sub>) e variabilidade da frequência cardíaca (HRV). Esses avanços tornam a PPG uma alternativa promissora para o monitoramento contínuo em cenários clínicos e domiciliares, promovendo uma abordagem mais integrada e preventiva no cuidado à saúde (RAHMAN et al., 2024).

No contexto do monitoramento contínuo da glicose, pesquisas recentes têm buscado aprimorar métodos que integrem técnicas inovadoras, como sensores implantáveis, dispositivos portáteis e algoritmos avançados de previsão. Esses esforços visam proporcionar aos pacientes diabéticos maior controle sobre os níveis de glicose no sangue, reduzindo riscos de complicações e melhorando a gestão da doença de maneira mais prática e eficiente no dia a dia (ZANELLI et al., 2022). Essas iniciativas refletem o potencial transformador das tecnologias de monitoramento em saúde, especialmente quando aplicadas a condições crônicas como diabetes.

## 1.1 Objetivos

O objetivo deste trabalho consiste na implementação de um modelo de aprendizado de máquina (Machine Learning) para estimativa do nível de glicose em termos de miligramas por decilitro (mg/dL) de forma não invasiva e contínua, utilizando o sinal de Fotopletismografia (PPG, do inglês) e a variabilidade da frequência cardíaca (HRV, do inglês).

### 1.1.1 Objetivos específicos

Visando atingir o objetivo principal, foram pensados alguns objetivos específicos expostos a seguir:

1. Identificar um conjunto de dados de participantes com glicose normal e alterada, onde ambas as variáveis desejadas são medidas, o sinal de PPG e o valor de referência de glicose em mg/dL.
2. Estudar de métodos de processamento de sinais, que se concentram na segmentação em ciclos individuais e na filtragem de ruídos provenientes da aquisição dos sinais de PPG.
3. Investigar as características do sinal de PPG que possam ser afetadas pelos níveis de glicose.
4. Pesquisar e implementar métodos relacionados à estimativa da glicose usando o sinal de PPG.
5. Pesquisar e implementar métodos para estimar a HRV por meio do sinal de PPG.
6. Investigar as alterações na forma da onda de PPG relacionadas à idade, avaliando como é seu impacto nas estimativas baseadas em PPG.
7. Desenvolver modelos de inferência para estimativa de glicose com base em PPG, considerando, fatores como VFC e idade.

### 1.1.2 Organização do texto

O presente trabalho visa construir um modelo de machine learning para a previsão não invasiva de glicose, utilizando o sinal de PPG e a análise da HRV. Seu conteúdo está estruturado em sete capítulos.

O Capítulo 2 consiste em uma revisão teórica concisa na qual os conceitos abordados serão explicitados. O Capítulo 3 é uma revisão do estado da arte, revisando a literatura de apoio para o tema do projeto.

No Capítulo 4, a metodologia é detalhada, explicando os componentes do projeto e as abordagens criadas. O capítulo também contém informações sobre os datasets utilizados e a organização dos dados. Os resultados dos experimentos são apresentados no Capítulo 5, com tabelas e gráficos.

A análise e discussão desses resultados são apresentadas no Capítulo 6, evidenciando os resultados encontrados. O Capítulo 7 finaliza o trabalho, revisando se de maneira satisfatória os objetivos propostos foram alcançados e sugerindo aplicações futuras.

## 2 Revisão Teórica

Neste capítulo, será apresentada uma revisão teórica relacionada com tema deste trabalho, para esclarecer os seus principais conceitos. Inicialmente, apresenta-se uma breve introdução sobre a fonte dos dados deste trabalho, o sinal de PPG. Além das técnicas de processamento realizadas para a remoção de artefatos, segmentação do sinal e a obtenção da HRV, será apresentada uma contextualização sobre algoritmos de inteligência artificial, modelos de Machine Learning e as métricas utilizadas para avaliar o desempenho deste trabalho. Além das diferentes técnicas de processamento de sinais utilizadas para remover artefatos deste sinal.

### 2.1 Fotopletismografia

A fotopletismografia (PPG, do inglês) é uma palavra composta que tem sua origem nas palavras “foto”, que significa luz; “pletismo”, que significa volume; e “grafia”, que significa gravação (ALIAN; SHELLEY, 2014). Seus estudos tiveram início com Hertzman, em 1937, onde ele constatou que ao irradiar luz sobre a pele, era possível a detectar por retrodispersão uma alteração significativa de acordo com atividade cardíaca.

Este método, atualmente, utiliza comumente a ponta do dedo como local para emitir uma onda de luz na faixa de 460 a 960 nm. Quando essa luz entra em contato com a área irradiada, ela reflete as variações do fluxo sanguíneo entre as fases sistólica e diastólica do ciclo cardíaco (KYRIACOU; ALLEN, 2021).

O PPG mede a quantidade de luz transmitida ou refletida, que varia conforme a concentração de substâncias no sangue e o percurso óptico, ambos influenciados pela pulsação. A lei de Beer-Lambert explica esse processo, descrevendo como a intensidade da luz é atenuada com base no coeficiente de extinção, na concentração das substâncias e no comprimento do caminho óptico percorrido pela luz ao atravessar o tecido (PANKAJ et al., 2022).

$$A = \varepsilon \cdot c \cdot l \quad (1)$$

Onde:

- $A$  é a absorção;
- $\varepsilon$  é o coeficiente de absorção molar (ou absortividade molar);

- $c$  é a concentração da solução (em mol/L);
- $l$  é o caminho óptico (em cm).

O sinal gerado pelo PPG reflete as variações volumétricas arteriais, composto por duas principais componentes: uma componente pulsátil (AC) e uma componente de variação lenta (DC). A componente AC captura as mudanças no volume sanguíneo microvascular, sincronizadas com cada batimento cardíaco, e está associada ao aumento da atenuação da luz devido ao maior volume de sangue. Esse componente está sobreposto a uma componente quasi-DC muito maior, relacionada aos tecidos e ao volume médio de sangue na amostra, além de incluir oscilações de baixa frequência atribuídas à respiração, atividade do sistema nervoso simpático e termorregulação. Na prática, a forma de onda PPG é frequentemente invertida, sendo sincronizada com o eletrocardiograma (ECG) correspondente (ALLEN, 2007).

Figura 1 – Forma de onda de ECG e PPG



Fonte: Adaptado de (ALLEN, 2007).

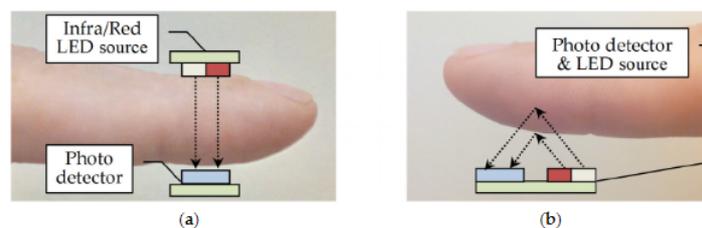
Devido a esses artefatos, a fotopletismografia (PPG) emergiu como um método alternativo, não invasivo, para a extração de sinais fisiológicos do sistema cardiovascular. Esse método utiliza propriedades ópticas, como absorção, dispersão e transmissão da luz, aplicadas em frequências específicas de onda, para detectar variações no volume de sangue nos tecidos. Essas características tornam a PPG uma técnica eficiente para monitorar parâmetros como frequência cardíaca, saturação de oxigênio e até mesmo estimativas de pressão arterial, com aplicações em dispositivos médicos e wearables (PARK et al., 2022).

Um dispositivo PPG consiste em um diodo emissor de luz (LED), responsável por emitir a luz, e um fotodetector, que detecta a luz emitida. O dispositivo pode ser classificado em dois tipos: transmissivo e reflexivo, dependendo da posição relativa entre o LED e o fotodetector. No tipo transmissivo, a luz atravessa o tecido antes de ser captada pelo fotodetector, enquanto no tipo reflexivo, a luz é refletida pelo tecido e, em seguida, detectada.

No tipo reflexivo, o fotodetector está posicionado próximo ao LED e mede a intensidade da luz espalhada após irradiar o tecido da pele. Embora a intensidade da luz captada seja relativamente menor em comparação ao tipo transmissivo, o que pode comprometer a qualidade do sinal, o tipo reflexivo apresenta a vantagem de permitir medições de PPG em regiões além das partes distais do corpo. Ele pode ser utilizado em locais como a testa, pulso, artéria carótida e esôfago, onde a transmissão de luz é mais difícil (BANIK et al., 2020).

O tipo transmissivo mede a intensidade da luz atenuada após esta atravessar os tecidos da pele. Ele é utilizado principalmente para medir PPG em regiões distais do corpo, onde a pele é mais fina, como nos dedos das mãos, dos pés e nos lóbulos das orelhas. Esse tipo de sensor PPG geralmente oferece um desempenho de medição mais estável em comparação ao tipo reflexivo, devido à maior intensidade da luz detectada após passar pelos tecidos (PŘIBIL; PŘIBILOVÁ; FROLLO, 2020a). A Figura 2 ilustra as configurações dos sensores nos modos (a) transmissivo e (b) reflexivo, destacando suas principais características.

Figura 2 – Esquemático para aquisição do sinal de PPG.



Fonte: Adaptado de (PŘIBIL; PŘIBILOVÁ; FROLLO, 2020b).

## 2.2 Processamento de Sinais

Os sinais de PPG são obtidos por meio da incidência de luz, direta ou indireta, sobre o corpo humano. Contudo, fatores como a taxa de gordura corporal, a coloração da pele e a respiração introduzem ruídos significativos, tornando inviável o uso do sinal em sua

forma bruta (raw). Por esse motivo, o pré-processamento é uma etapa fundamental, na qual técnicas de filtragem e correção são aplicadas para remover artefatos e destacar informações relevantes, como apontado por Charlton et al. (CHARLTON et al., 2022) e Allen (ALLEN, 2007), permitindo a extração de características com maior precisão.

Um estudo realizado por (LIANG et al., 2018b) avaliou o impacto e o desempenho de filtros digitais aplicados ao sinal de PPG. A análise considerou nove tipos de filtros, cada um com dez ordens diferentes, resultando em 90 configurações distintas. Os autores concluíram que o filtro Chebyshev II, seguido pelo filtro Butterworth, ambos de 4<sup>a</sup> ordem, apresentaram o melhor desempenho, com base no indicador SSQI (Skewness Signal Quality Index) proposto no estudo. A Eq. 2 ilustra o método de avaliação:

$$SSQI = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \mu_x}{\sigma} \right)^3 \quad (2)$$

Onde "n" é o número de pontos do sinal de PPG, e  $\mu$  e  $\sigma$  são, respectivamente, a média e o desvio padrão de  $x_i$ .

No entanto, nos testes realizados por Luis Felipe de Deus (DEUS, Luis Felipe de, 2023), verificou-se que a melhor técnica para a redução de ruídos nos sinais de PPG foi a combinação dos filtros Savitzky-Golay (SAVITZKY; GOLAY, 1964) e Butterworth (SELESNICK; BURRUS, 1998). A configuração utilizada para o filtro Savitzky-Golay foi de 19 frames e 4<sup>a</sup> ordem, enquanto para o 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

O filtro Savitzky-Golay é um método de suavização de sinais que aplica uma média ponderada dos pontos de dados, utilizando um polinômio para aproximar o valor central em uma janela de dados. Ele é comumente usado para suavizar sinais ruidosos e, ao mesmo tempo, preservar as características importantes do sinal, como picos e vales, sem os distorcer excessivamente (SAVITZKY; GOLAY, 1964).

O filtro Savitzky-Golay é geralmente definido por três parâmetros principais:

- **Tamanho da janela:** Número de pontos que serão usados para calcular a suavização. Geralmente é um número ímpar (por exemplo, 19 pontos), para haver um valor central na janela.
- **Ordem do polinômio:** O grau do polinômio que será ajustado aos pontos da janela. Comumente, são usados polinômios de 2<sup>a</sup> ou 4<sup>a</sup> ordem.

- **Valor de suavização:** Determina o grau de suavização do sinal, com janelas maiores proporcionando uma suavização mais forte, mas também podendo eliminar mais detalhes.

Nos sinais biomédicos, o ruído de alta frequência pode surgir de diversos fatores, como interferências eletrônicas, movimentos do paciente, instabilidades nos equipamentos ou outros artefatos externos. Esses ruídos prejudicam a qualidade do sinal captado, tornando a interpretação e análise dos dados menos precisas. A aplicação de filtros para atenuar o ruído de alta frequência é, portanto, essencial e encontra ampla utilidade em contextos como a análise de sinais de PPG, ECG, espectros de sinais, entre outros.

A equação matemática que define o filtro Savitzky-Golay é baseada na expansão de um polinômio local, e o processo de suavização é realizado aplicando-se esse polinômio em uma janela deslizante sobre os dados. Nesse contexto, o tamanho da janela é dado por  $2m + 1$ , onde  $m$  representa o número de pontos adjacentes considerados em cada direção a partir do ponto central. O filtro pode ser expresso da seguinte forma para um sinal  $y_i$  em um ponto  $i$ , com uma janela de tamanho  $2m + 1$  e um polinômio de ordem  $k$ :

$$y_i^{SG} = \frac{1}{2m + 1} \sum_{j=-m}^m C_j y_{i+j} \quad (3)$$

Onde  $C_j$  são os coeficientes que dependem da ordem do polinômio e do tamanho da janela.

## 2.2.2 Filtro Butterworth

O filtro Butterworth é um tipo de filtro eletrônico, amplamente utilizado em processamento de sinais devido à sua resposta em frequência suave e sem ondulações. Ele foi proposto por Stephen Butterworth em 1930, sendo caracterizado por uma resposta de magnitude que, no intervalo passante, é plana e sem distorções. O principal objetivo do filtro Butterworth é maximizar a suavidade da resposta na região passante, ao mesmo tempo que minimiza a atenuação na região de rejeição (BUTTERWORTH et al., 1930).

Características principais:

- **Resposta de magnitude:** O filtro Butterworth é caracterizado por uma resposta de magnitude plana na faixa passante (sem ondulações), o que significa que ele não introduz distorção significativa no sinal na região de frequências baixas.

- **Faixa de transição:** A transição entre a faixa passante e a faixa de rejeição ocorre suavemente, sem picos ou dips, mas a atenuação fora da faixa passante aumenta conforme a ordem do filtro.
- **Ordem do filtro:** A ordem do filtro determina a suavidade da transição. Filtros de maior ordem têm transições mais abruptas entre a faixa passante e a faixa de rejeição, mas exigem mais componentes em circuitos analógicos ou mais cálculos em implementações digitais.
- **Função de transferência:** A função de transferência de um filtro Butterworth pode ser expressa como:

$$H(s) = \frac{1}{\prod_{k=1}^n (s + \omega_k)} \quad (4)$$

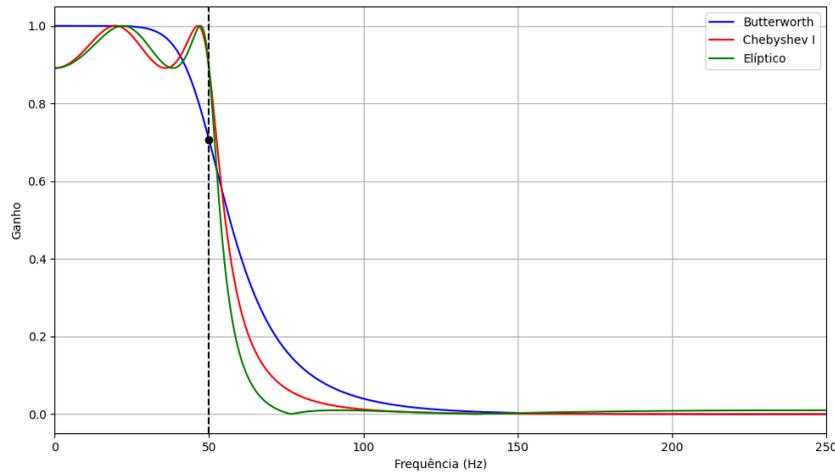
Onde  $s$  é a variável complexa,  $\omega_k$  são as raízes dos polos do filtro (distribuídas em uma circunferência no plano complexo), e  $n$  é a ordem do filtro.

A resposta em frequência de um filtro Butterworth é dada por:

$$|H(j\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}} \quad (5)$$

Onde  $\omega_c$  é a frequência de corte e  $n$  é a ordem do filtro. A resposta em frequência é suave e sem ondulações na faixa passante, e a atenuação cresce rapidamente após a frequência de corte (BUTTERWORTH et al., 1930). A Figura 3, 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 de sinais biomédicos (BOGDAN; PANU, 2015).

Figura 3 – Exemplo de resposta em frequência para os filtros.



Fonte: Adaptado de (BOGDAN; PANU, 2015).

### 2.2.3 Normalização Min Max

A última etapa de pré-processamento aplicada no sinal de PPG (Fotopleletismografia) é a normalização. Essa técnica é fundamental para evitar que diferentes magnitudes em amplitude influenciem desigualmente as análises subsequentes. O sinal de PPG foi normalizado utilizando a técnica de normalização **Min-Max**, que reescala os valores do sinal para um intervalo entre 0 e 1. Como o nome sugere, essa técnica utiliza o valor mínimo e máximo de amplitude do sinal como parâmetros, conforme descrito pela equação 6:

$$PPGn[i] = \frac{PPGf[i] - \min(PPGf)}{\max(PPGf) - \min(PPGf)} \quad (6)$$

Onde:

- $PPGn$  é o PPG normalizado,
- $PPGf$  é o PPG filtrado,
- $\min(PPGf)$  é o valor mínimo do sinal filtrado,
- $\max(PPGf)$  é o valor máximo do sinal filtrado.

A normalização é essencial para garantir que as variações de amplitude do sinal de PPG não causem distorções nas análises, especialmente em modelos de Machine Learning. Essa técnica assegura que todas as variáveis sejam comparáveis e o modelo não seja tendencioso devido a diferenças de escala entre as variáveis.

## 2.3 Inteligência Artificial

A Inteligência Artificial (IA) aplicada à área biomédica tem se destacado como uma ferramenta poderosa para transformar o diagnóstico, tratamento e prevenção de doenças. A crescente quantidade de dados gerados no setor de saúde, provenientes de exames laboratoriais, imagens médicas e registros eletrônicos de saúde, impulsiona o uso de algoritmos de aprendizado de máquina e técnicas de IA para melhorar a precisão e a eficiência dos cuidados médicos.

De acordo com um estudo de Rajpurkar, "o uso de IA para diagnosticar doenças a partir de imagens médicas tem mostrado resultados promissores, muitas vezes alcançando ou até superando a performance dos médicos humanos em determinadas áreas, como na detecção de câncer de pulmão" (RAJPURKAR et al., 2022). Além disso, os modelos de IA são utilizados para prever a progressão de doenças, como diabetes e doenças cardiovasculares, com base em dados históricos de pacientes. Estes modelos auxiliam os médicos a tomar decisões mais informadas e a personalizar o tratamento para cada paciente.

A IA também desempenha um papel crucial na medicina de precisão, onde os tratamentos são adaptados com base nas características genéticas e clínicas individuais de um paciente. Conforme apontado por Topol, "as tecnologias de IA, quando combinadas com a genômica, oferecem uma nova abordagem para tratamentos personalizados, otimizando os resultados para os pacientes" (TOPOL, 2019).

Outro aspecto importante é a integração de sistemas de IA com dispositivos vestíveis, como monitores de frequência cardíaca e sensores de glicemia, que permitem o monitoramento contínuo da saúde do paciente. Isso possibilita a detecção precoce de anomalias e a intervenção imediata, reduzindo complicações e melhorando a qualidade de vida. Como menciona (NENSA; DEMIRCIOGLU; RISCHPLER, 2019), "os avanços na IA e no monitoramento em tempo real são fundamentais para a evolução da medicina preventiva e da gestão de doenças crônicas".

Além disso, os algoritmos de IA desempenham um papel fundamental na área biomédica, trazendo avanços significativos e promissores. Eles são aplicados em diversas áreas, como diagnóstico médico, descoberta de medicamentos, monitoramento de saúde, análise de imagens médicas e previsão de doenças. No campo do diagnóstico médico, por exemplo, algoritmos de IA conseguem analisar grandes volumes de dados clínicos e históricos dos pacientes, auxiliando os médicos na tomada de decisões rápidas e precisas (MAGRABI et al., 2019).

A IA também é amplamente utilizada na análise de imagens médicas, permitindo a

detecção precoce de doenças e o auxílio na interpretação de exames radiológicos (DUNG et al., 2019). Essas tecnologias têm o potencial de melhorar a eficiência dos sistemas de saúde, personalizar tratamentos e otimizar os resultados para os pacientes. Com a constante evolução da IA, espera-se que esses algoritmos desempenhem um papel cada vez mais importante no avanço da medicina e na melhoria da qualidade de vida.

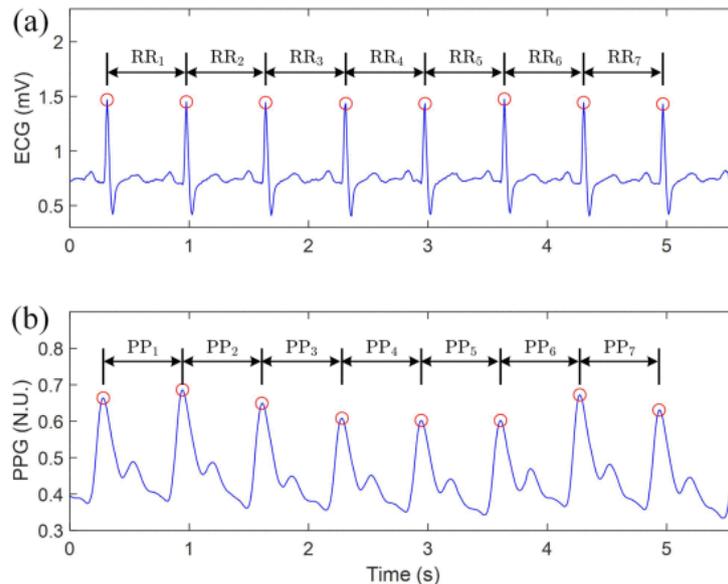
Embora a Inteligência Artificial, de maneira geral, seja definida como a ciência capaz de imitar as habilidades humanas, o aprendizado de máquina (Machine Learning) é uma vertente específica da IA que treina as máquinas para aprender a partir dos dados. Na atualidade, não há recurso mais abundante do que dados estruturados e não estruturados. Na segunda metade do século XX, o aprendizado de máquina se consolidou como uma subárea da IA, com algoritmos de autoaprendizado capazes de extrair conhecimento dos dados para fazer previsões futuras (RASCHKA; MIRJALILI, 2019).

Ao invés de programadores escreverem rotinas manualmente, o aprendizado de máquina oferece uma alternativa mais eficiente para adquirir conhecimento a partir dos dados, permitindo a melhoria gradual da performance das predições. Neste contexto, a IA está transformando a medicina, trazendo benefícios significativos tanto na prática clínica quanto na pesquisa biomédica. Neste trabalho, utilizou-se a linguagem de programação Python, versão 3.11, para o desenvolvimento, assim como o pacote Scikit-Learn, versão 1.3.2 (PEDREGOSA et al., 2011), para a implementação dos modelos de inteligência artificial.

## 2.4 Variabilidade da Frequência Cardíaca

A variabilidade da frequência cardíaca (HRV, do inglês) é amplamente reconhecida como um indicador da atividade do sistema nervoso autônomo (SNA, do inglês), refletindo o equilíbrio entre as divisões simpática e parassimpática. Recentemente, a extração da HRV a partir do sinal de fotopleletismografia (PPG) tem ganhado destaque devido à praticidade e à ampla aplicabilidade em dispositivos vestíveis e no monitoramento remoto. A PPG é obtida por meio de variações na absorção de luz pela pele, que refletem mudanças no volume sanguíneo nos tecidos. Esse método não invasivo permite estimar a variabilidade dos intervalos interbatimentos (IBIs), similares aos intervalos entre os picos das ondas R obtidos pelo eletrocardiograma (ECG). No sinal de PPG, os IBIs são derivados dos picos ou vales do sinal pulsátil, denominados intervalos P-P.

Figura 4 – Ilustração da derivação de dados da HRV dos sinais de ECG e PPG, (a) intervalos RR no sinal de ECG e (b) intervalos PP no sinal de PPG. N.U. (Unidades normalizadas).



Fonte: Adaptado de (LU et al., 2023).

A simplicidade e portabilidade da PPG tornam-na uma solução viável para dispositivos vestíveis, como smartwatches e pulseiras fitness, possibilitando o monitoramento contínuo da HRV em ambientes não clínicos. Estudos demonstram que a PPG apresenta alta correlação com o ECG em condições estáticas e dinâmicas controladas, evidenciando sua robustez como alternativa ao ECG (SHAFFER; GINSBERG, 2017; LU et al., 2023). Entre os parâmetros derivados da HRV, destacam-se os do domínio do tempo, índices amplamente utilizados incluem o valor médio dos intervalos normais (meanNN), o desvio padrão dos intervalos normais (SDNN), o desvio padrão das diferenças entre intervalos adjacentes (SDSD), o percentual de diferenças entre intervalos superiores a 50 ms (pNN50) e o índice triangular de HRV. Essas medidas são úteis para diferentes finalidades: o SDNN e o índice triangular são indicados para avaliar a variabilidade global, enquanto o RMSSD é mais apropriado para analisar componentes de curto prazo. No domínio de frequência, métricas como LF (Low Frequency) e HF (High Frequency), que quantificam as influências simpática e parassimpática, podem ser calculadas utilizando algoritmos baseados em transformada rápida de Fourier (FFT) ou wavelets.

Apesar de suas vantagens, o uso da PPG enfrenta desafios, como artefatos de movimento, condições de iluminação e características individuais da pele, que podem comprometer a qualidade do sinal. Estudos apontam que técnicas de filtragem e aprendizado

de máquina são fundamentais para reduzir os erros na derivação de IBIs a partir da PPG (LU et al., 2009). Ainda assim, a PPG tem mostrado grande potencial em aplicações clínicas e esportivas.

## 2.5 Decomposição Empírica por Modos

A Decomposição Empírica por Modos (EMD) é uma técnica poderosa para a análise de sinais não estacionários e não lineares, introduzida por (HUANG et al., 1998). Diferentemente de métodos tradicionais como a Transformada de Fourier e a Transformada Wavelet, a EMD não requer uma função base pré-definida, adaptando-se diretamente às características do sinal analisado. O objetivo principal da EMD é decompor um sinal em uma soma de funções oscilatórias denominadas Modos Intrínsecos de Função (IMFs - *Intrinsic Mode Functions*), além de um resíduo monotônico. Esses IMFs representam as diferentes frequências do sinal, organizadas da mais alta para a mais baixa, possibilitando a análise de componentes específicas de frequência adaptativamente.

O processo de decomposição é iterativo e baseado em dois critérios fundamentais: (1) cada IMF deve possuir apenas uma inversão de sinal entre máximos e mínimos (*zero-crossing*); e (2) o valor médio dos envelopes superior e inferior, definidos pelos máximos e mínimos locais do sinal, deve ser aproximadamente zero. A decomposição segue os seguintes passos: inicialmente, os máximos e mínimos locais do sinal são identificados, e os envelopes superior e inferior são construídos utilizando interpolação cúbica. Em seguida, a média dos envelopes superior ( $E_{\text{superior}}$ ) e inferior ( $E_{\text{inferior}}$ ) é calculada, e o componente médio ( $m(t)$ ) é subtraído do sinal original ( $S(t)$ ):

$$m(t) = \frac{E_{\text{superior}} + E_{\text{inferior}}}{2} \quad (7)$$

$$h(t) = S(t) - m(t) \quad (8)$$

O resultado ( $h(t)$ ) é então verificado para atender aos critérios de IMF. Caso não satisfaça, o processo (chamado de *sifting*) é repetido até que um IMF válido seja obtido. Após a extração de um IMF ( $IMF_i(t)$ ), ele é subtraído do sinal original, gerando um resíduo ( $R(t)$ ), sobre o qual o mesmo processo é aplicado para extrair os IMFs subsequentes. O sinal pode, assim, ser representado como:

$$S(t) = \sum_{i=1}^n IMF_i(t) + R(t) \quad (9)$$

Onde  $S(t)$  é o sinal original,  $IMF_i(t)$  são os Modos Intrínsecos de Função, e  $R(t)$  é o resíduo monotônico.

A EMD apresenta vantagens significativas em relação a outros métodos de análise de sinais. Por ser adaptativa, a técnica é altamente eficaz na decomposição de sinais complexos, especialmente aqueles que apresentam comportamentos não estacionários. Isso a torna ideal para aplicações em áreas como análise de sinais biomédicos (ECG, PPG, EEG), processamento de sinais de engenharia (detecção de falhas e análise de vibrações), meteorologia (séries temporais climáticas) e economia (identificação de tendências e ciclos econômicos) (HUANG et al., 1998; SANTOS, 2020).

Entretanto, a EMD também possui limitações. A sensibilidade a ruídos pode comprometer a qualidade dos IMFs, e a interpretação dos modos extraídos nem sempre é trivial. Além disso, a técnica não possui uma fundamentação matemática rigorosa, sendo considerada um método heurístico (RILLING et al., 2003). Apesar dessas limitações, a EMD permanece uma ferramenta valiosa e amplamente utilizada em contextos onde os métodos convencionais falham em capturar a natureza dinâmica e não linear dos sinais.

Portanto, a EMD oferece uma abordagem flexível e intuitiva para a análise de sinais complexos, permitindo insights significativos sobre os componentes oscilatórios subjacentes a fenômenos não estacionários. Seu impacto em diversas áreas demonstra seu valor como uma técnica inovadora e amplamente aplicável.

## 2.6 Métricas de Avaliação

Apesar do modelo utilizado para a estimativa da glicose, é fundamental empregar métricas adequadas para avaliar seu desempenho. Essas métricas fornecem uma base objetiva para medir a eficácia e a qualidade dos resultados obtidos. Ao aplicar essas métricas, é possível comparar diferentes modelos, identificar suas limitações e implementar melhorias. Além disso, as métricas auxiliam na tomada de decisões fundamentadas, facilitando a escolha do modelo mais adequado para atingir os objetivos desejados.

Portanto, o uso de métricas de avaliação é um componente essencial em qualquer análise de desempenho de modelos, contribuindo para melhorar a eficiência e a confiabilidade dos resultados. Neste trabalho, foram adotadas as mesmas métricas de avaliação utilizadas no estudo de Luis, visando comparar os resultados obtidos.

- **Mean Error (ME):** O erro médio descreve o valor médio entre a diferença do valor real com o valor estimado (Eq. 10).

$$ME = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (10)$$

- **Mean Absolute Error (MAE):** O erro médio absoluto descreve o valor médio entre a diferença absoluta do valor real com o valor estimado (Eq. 11) (WILLMOTT; MATSUURA, 2005).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

- **Mean Absolute Percentage Error (MAPE):** O erro absoluto percentual descreve o erro em termos de uma porcentagem em relação ao valor de referência (Eq. 12).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100 \quad (12)$$

- **Root Mean Squared Error (RMSE):** A raiz do erro quadrático médio é uma medida quadrática que visa realçar erros extremos que seriam suavizados pela média em uma amostra grande (Eq. 13) (WILLMOTT; MATSUURA, 2005).

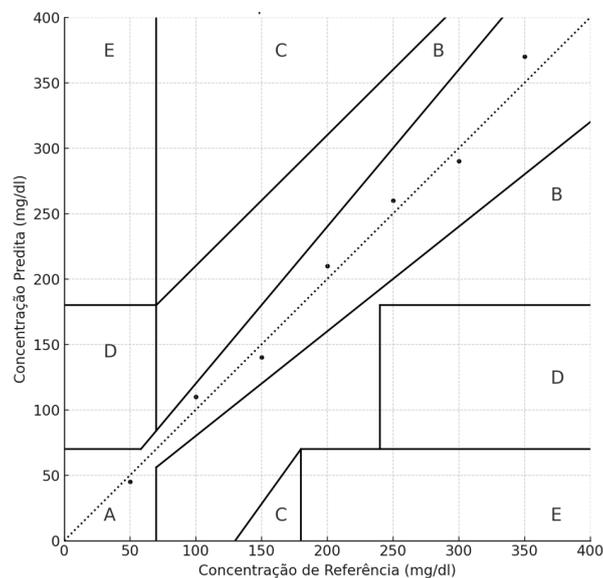
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

- **Coefficient of determination ( $r^2$ ):** 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 consegue explicar a variabilidade dos dados e 1 indica que o modelo se ajusta perfeitamente aos dados (NETER; WASSERMAN; KUTNER, 1983).
- **O Clarke Error Grid** é um método utilizado para avaliar a precisão de sistemas de monitoramento contínuo de glicose (CLARKE et al., 1987). Ele classifica as estimativas de glicose com base na discrepância em relação aos valores de referência. A origem desse gráfico está na comparação entre os valores medidos pelo glicosímetro e os valores de referência, geralmente obtidos por métodos laboratoriais de maior precisão. Essa análise permite avaliar a precisão e a confiabilidade do glicosímetro, identificando discrepâncias que podem impactar o monitoramento e o controle glicêmico. O gráfico é dividido em cinco zonas de precisão, cada uma com implicações clínicas específicas:

- **Zona A:** A faixa de maior precisão, com discrepância de  $\pm 20\%$  dos valores de referência. As estimativas nesta zona indicam um ajuste adequado e não exigem alterações nas ações clínicas.
- **Zona B:** Discrepância entre  $\pm 20\%$  e  $\pm 40\%$  em relação aos valores de referência. Embora ainda aceitável, sugere uma precisão inferior e pode exigir modificações na medicação, mas sem alterar o diagnóstico.
- **Zona C:** Discrepância entre  $\pm 40\%$  e  $\pm 60\%$ . Indica uma precisão moderada e pode levar a alterações no tratamento, como prescrição de medicamentos, e até alterações no diagnóstico.
- **Zona D:** Discrepância entre  $\pm 60\%$  e  $\pm 80\%$ . A precisão é baixa, e alterações significativas podem ser necessárias no tratamento, com risco clínico associado.
- **Zona E:** Discrepância maior que  $\pm 80\%$ . A imprecisão é significativa e pode levar a ações clínicas perigosas, com riscos para a saúde do paciente.

Esse método é importante para garantir a eficácia e segurança de sistemas de monitoramento contínuo de glicose, especialmente em pacientes com diabetes, para as decisões clínicas serem baseadas em estimativas confiáveis.

Figura 5 – Exemplo de gráfico de erro de Clarke.



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

### 3 Revisão Bibliográfica

Nesta seção, será apresentada uma revisão da literatura relacionada ao tema deste trabalho. O objetivo é fornecer uma visão geral das pesquisas anteriores e enfatizar as contribuições e descobertas relevantes neste domínio.

Realizando uma busca na *PubMed*, utilizando as palavras-chave glucose, PPG e HRV foi retornado 1 resultado, entretanto a pesquisa está fora do escopo deste trabalho, por se tratar de um estudo sobre como a galantamina alivia inflamação e resistência à insulina em pacientes com síndrome metabólica em um estudo randomizado. Alterando as palavras-chave para BGL, PPG e HRV não houve resultado retornado. É de compreensão do autor os poucos estudos realizados sobre estimativa não invasiva dos níveis de glicose com o auxílio da variabilidade cardíaca até o momento. Além disso, a maioria dos estudos publicados utilizou dados que não são acessíveis ao público. Para avançar nessa área e desenvolver abordagens mais fortes e generalizáveis, essa carência de pesquisas e a falta de dados abertos são um obstáculo. As abordagens utilizadas para estimativa dos níveis de glicose com o auxílio da variabilidade cardíaca variam desde a utilização exclusiva do sinal de fotopletismografia até a combinação desse sinal com outras informações biomédicas.

Shin et al. (SHI et al., 2023) buscou uma solução de inteligência artificial para reconhecer níveis elevados de glicose no sangue ( $\geq 7,8$  mmol/L) de forma não invasiva e avaliar o risco de diabetes com base em medições repetidas. O estudo foi realizado no KK Women's and Children's Hospital, em Cingapura, e 500 participantes foram recrutados (idade média 38,73, desvio padrão 10,61 anos; IMC médio 24,4, desvio padrão 5,1 kg/m<sup>2</sup>). Os BGL da maioria dos participantes foram medidos antes e após consumir 75 g de bebidas açucaradas usando um glicosímetro convencional (Accu-Chek Performa) e um dispositivo portátil de pulso. Os resultados obtidos no glicosímetro foram usados como medidas de verdade. Os autores realizaram uma extensa engenharia de recursos nos dados do sensor de PPG, o que possibilitou a identificação de características sensíveis às alterações nos níveis de glicose.

Alguns modelos de aprendizado de máquina foram treinados e avaliados com validação cruzada de 10 vezes, usando dados demográficos dos participantes e características críticas extraídas das medições de PPG como preditores. Uma máquina de vetores de suporte com kernel de função de base radial teve o melhor desempenho de detecção, com precisão média de 84,7%, sensibilidade de 81,05%, especificidade de 88,3%, precisão

de 87,51%, média geométrica de 84,54% e Pontuação F de 84,03%. Estas descobertas sugerem que as medidas de PPG podem ser usadas para identificar participantes com medidas elevadas de glicose no sangue e auxiliar na triagem de participantes quanto ao risco de diabetes.

Vargová et al. (VARGOVÁ; NĚMCOVÁ; NOVÁKOVÁ, 2023) trouxeram como objetivo principal de pesquisa propor métodos de classificação glicêmica em dois grupos distintos, glicemia baixa e alta, e prever valores específicos de glicemia usando técnicas de aprendizado de máquina. Dois conjuntos de dados foram criados medindo sinais PPG de 16 voluntários, 11 mulheres e 5 homens com idade entre 22 e 79 anos, dentre eles 5 tinham diabetes ou pré-diabetes. Os conjuntos de dados foram criados por meio de dois dispositivos. O primeiro conjunto de dados veio da pulseira *Empatica E4* (Empatica Inc., Estados Unidos), o sinal PPG derivado do *Empatica* é gerado usando um algoritmo patenteado (EP-2948043-A1, AU-2014208391-B2 e AU 2017264977) que combina luz verde e vermelha. Os sinais são amostrados a uma frequência de 64 Hz. O segundo foi medido usando um smartphone *Samsung Galaxy Note 20* (*Samsung Electronics, Vietnam*). A resolução do vídeo foi definida para 1080×1920 píxel e a taxa de quadros foi de 30 quadros por segundo.

Cada medição durou 15 minutos. Durante toda a medição, o sinal PPG foi registrado continuamente usando a pulseira *Empatica E4*. Além disso, os sinais PPG de um minuto foram medidos três vezes usando o smartphone durante toda a medição. Os participantes colocaram o dedo na câmera do smartphone, cobrindo tanto a lente quanto a fonte de luz LED. Do vídeo do dedo, o componente vermelho médio correspondente ao sinal PPG bruto foi extraído. Simultaneamente, o BGL de referência foi medido três vezes utilizando um glicosímetro certificado *Fora Diamond Mini*.

Para aumentar a quantidade de dados obtidos de indivíduos com maior BGL, os participantes diabéticos foram submetidos a todo o processo de medição duas vezes. O BGL medido variou de 4,3 a 13,2 mmol/l. Houve um pré-processamento nos sinais, resultando em segmentações de 10 segundos. Dos sinais segmentados, 27 características diferentes foram extraídas, algumas dessas características são baseadas na análise da HRV, devido à redução da atividade parassimpática observada em diabéticos. Recursos adicionais são derivados da morfologia da forma de onda PPG, influenciada pela variação da absorção de luz do sangue com diferentes níveis de glicose. Os valores discrepantes foram substituídos, todos os atributos foram normalizados usando o *Z-score* e uma análise estatística dos recursos foi realizada. Para reduzir ainda mais o número de recursos, um método de seleção denominado “Máxima Relevância e Redundância Mínima” (mRMR) foi aplicado ao conjunto restante de 23 características, reduzindo para 10 características.

Diversos modelos de aprendizado de máquina foram testados. *Random Forest* (RF) e *Support Vector Machine* (SVM) com o kernel de função de base radial (RBF, do inglês) tiveram melhor desempenho na classificação de sinais PPG em dois grupos. Esses modelos alcançaram uma precisão de 76% (SVM) e 75% (RF) no conjunto de dados de teste de pulseira *Empatica E4*. A funcionalidade dos modelos propostos foi então verificada no conjunto de dados de teste do smartphone, onde ambos os modelos alcançaram precisão semelhante: 74% (SVM) e 75% (RF). Para prever valores específicos de glicemia, o FR teve melhor desempenho. O erro absoluto médio (MAE, do inglês) foi de 1,25 mmol/l no conjunto de dados de teste de pulseira *Empatica E4* e 1,37 mmol/l no conjunto de dados de teste do smartphone.

O estudo apresentado por (SATTER et al., 2024) se concentrou na previsão de glicose no sangue usando sinais PPG de pulso e explorando diversos recursos baseados em formas de onda PPG, incluindo relação AC para DC (AC/DC) e recursos baseados em função de modo intrínseco (FMI, do inglês) derivados da decomposição de modo empírico (EMD, do inglês). Os dados foram coletados de um total de 34 participantes. Todo o processo de coleta de dados foi supervisionado e aprovado pelo conselho de revisão institucional (IRB, do inglês) da Universidade Kookmin, Seul, Coreia Sul, e o número do protocolo do IRB é KMU-202111-BR-286. A proporção de sujeitos entre homens e mulheres é de 50:50, media de idade de 55 anos, com a mínima de 23 e máxima de 86. O medidor de glicose no sangue CareSens II foi usado para medir o nível real de glicose no sangue correspondente e um sinal PPG simultâneo foi registrado por três minutos, a taxa de amostragem para dados PPG foi de 24 Hz. Além disso, a seleção baseada na importância do recurso também foi usada para melhorar a precisão do modelo proposto.

Entre os quatro algoritmos de aprendizado de máquina considerados neste estudo, *CatBoost* superou consistentemente *XGBoost*, *LightGBM* e RF em um amplo número de recursos. O modelo de melhor desempenho, *CatBoost*, alcançou *Pearson's r* de 0,96, MSE 0,08,  $R^2$  score 0,92 e MAE 8,01 ao considerar os 50 principais recursos selecionados tanto dos recursos baseados em formas de onda PPG quanto dos recursos baseados no FMI. O *p-values* para todos os modelos foram  $<0,001$ , indicando correlações estatisticamente significativas.

Neste projeto não serão propostos novos métodos de análise sobre HRV, pois os métodos existentes até o momento já demonstram ampla utilização e eficácia nas áreas de pesquisa. O método da transformada de Hilbert (HHT, do inglês): Método Hilbert Double Envelope (HDEM), consiste na aplicação da transformada de Hilbert e na detecção dos picos resultantes. Esta técnica vem sendo amplamente utilizada desde a década de 1990 em áreas como processamento de sinais biomédicos, análise de vibrações mecânicas

e engenharia elétrica, consolidando-se como uma ferramenta robusta e confiável (HUANG et al., 1998; CLIFFORD et al., 2006). Este método obteve uma grande aceitação pela comunidade, devido aos resultados alcançados por (ESGALHADO et al., 2022), utilizando este algoritmo nos sinais de eletrocardiografia (ECG) e PPG, comprovaram uma alta correlação entre eles, e resultados de 99,07% e 99,45% respectivamente.

O estudo de (WU et al., 2021) sobre a classificação da pressão arterial (PA) a partir da fotopletismografia (PPG) destaca que alterações na pressão arterial afetam o volume sanguíneo nos capilares, influenciando diretamente as medições do sinal de PPG. Esse fenômeno sugere que, se é possível estimar a pressão arterial a partir do PPG, a pressão arterial pode interferir na medida da glicose, uma vez que tanto a PA quanto os níveis de glicose estão relacionados à dinâmica hemodinâmica e à elasticidade vascular. Níveis elevados de glicose, típicos em condições de hiperglicemia, alteram a viscosidade sanguínea e a rigidez arterial, fatores que afetam a precisão das leituras do PPG.

O impacto da glicose na Variabilidade da Frequência Cardíaca (VFC) ocorre devido a níveis elevados de glicose, comuns em condições de hiperglicemia crônica, que contribuem diretamente para alterações na dinâmica cardiovascular e no tônus autonômico. A hiperglicemia compromete a função endotelial, reduzindo a elasticidade dos vasos sanguíneos e favorecendo a vasoconstrição, interferindo na modulação parassimpática e simpática da frequência cardíaca. Esse processo leva a uma redução na complexidade da VFC, como demonstrado no estudo de (COSSUL et al., 2023), que relaciona a progressão da neuropatia autonômica cardiovascular (CAN) com padrões específicos nos sinais de HRV. Além disso, a alteração da viscosidade sanguínea causada pela hiperglicemia, conforme discutido por (BARRETT et al., 2017), compromete a perfusão tecidual e a dinâmica do fluxo sanguíneo, intensificando o impacto nos sinais cardíacos. Dessa forma, os estudos existentes reforçam que níveis elevados de glicose estão associados à disfunção autonômica, corroborando a hipótese de que a VFC é diretamente afetada pelo metabolismo glicêmico desregulado, resultando em mudanças na regulação cardiovascular.

O trabalho de (ELGENDI, 2012) destaca o efeito da idade na forma da onda do PPG, evidenciando que o envelhecimento está associado a um aumento da rigidez arterial e a uma maior velocidade de propagação da onda de pulso. Essas mudanças afetam diretamente o contorno da onda PPG, com a redução do intervalo entre os picos sistólico e diastólico e a atenuação da incisura dicrótica, característica de artérias saudáveis em indivíduos mais jovens. À medida que a idade avança, a capacidade das artérias de se expandirem e contraírem de maneira eficiente diminui, refletindo a deterioração da função vascular. Essas alterações relacionadas à idade interferem na dinâmica do fluxo sanguíneo e, conseqüentemente, na interpretação dos sinais PPG. Como a medição da

glicose baseada no PPG depende das variações volumétricas de sangue nos capilares, a rigidez arterial e a redução da elasticidade vascular com o envelhecimento podem impactar negativamente a precisão dessas estimativas. Portanto, a idade representa um fator relevante que deve ser considerado em modelos de inferência de glicose baseados no PPG.

## 4 Métodos e Materiais

Este capítulo descreve a metodologia utilizada no desenvolvimento deste projeto. Inicialmente, são apresentados os conjuntos de dados empregados, seguidos por uma visão geral do método proposto. Cada um dos principais componentes desse método será detalhado, com ênfase no objetivo principal: a estimativa dos níveis de glicose.

### 4.1 Conjuntos de Dados

A obtenção de um volume adequado de dados é essencial para o treinamento de modelos de inteligência artificial baseados em aprendizado supervisionado. Para este trabalho, foi selecionado o conjunto de dados denominado UFSC-BGL, coletado durante a pesquisa descrita em (DEUS, Luis Felipe de, 2023) em parceria com a Universidade Federal de Santa Catarina (UFSC) e o Hospital Universitário Polydoro Ernani de São Thiago (HU-UFSC/EBSERH).

O UFSC-BGL foi desenvolvido visando reunir dados abrangentes de participantes com níveis normais e alterados de glicose em um contexto hospitalar. Na época da concepção de seu estudo, não havia conjuntos de dados publicamente disponíveis na literatura que atendessem a essas necessidades. Assim, a criação do UFSC-BGL não apenas preencheu essa lacuna, mas também contribuiu significativamente para a pesquisa na área.

O projeto recebeu aprovação do Comitê de Ética em Pesquisa com Seres Humanos (CEPSH-UFSC) em 10 de outubro de 2022 e está registrado sob o identificador CAAE 60111422.5.0000.0121.

O estudo incluiu 92 participantes, com idades entre 18 e 75 anos. A composição do conjunto de dados é de 45 participantes com glicose considerada normal e 47 participantes com glicose alterada. A classificação dos grupos foi realizada com base nas diretrizes brasileiras de diabetes para testes aleatórios (RODACKI et al., 2024).

Os critérios de inclusão foram os seguintes:

- Grupo diabético: participantes com níveis de glicose maiores ou iguais a 120 mg/dL em jejum, ou maiores, ou iguais a 140 mg/dL sem jejum.
- Grupo controle (saudáveis): Participantes com níveis de glicose menores que 120 mg/dL em jejum ou menores que 140 mg/dL sem jejum.

Esses critérios foram adotados para assegurar a distinção clara entre os dois grupos e garantir a relevância dos dados para o objetivo do estudo.

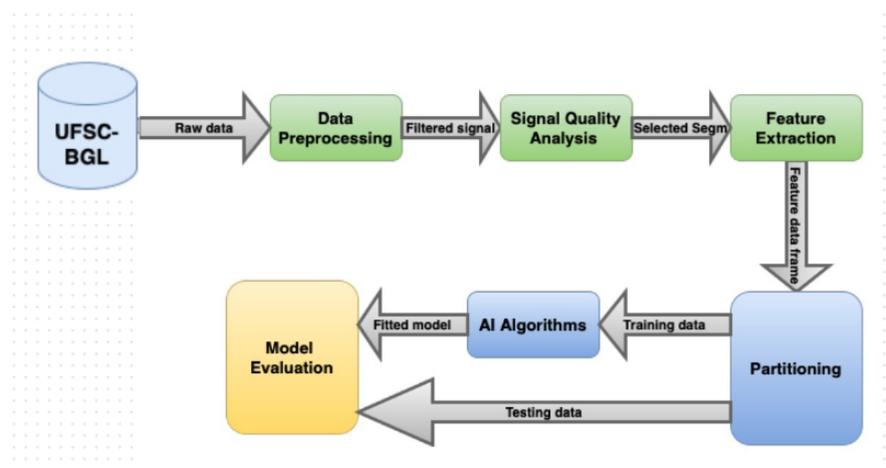
## 4.2 Método

Este trabalho procura aprimorar os resultados obtidos na pesquisa de mestrado conduzida por Luis Felipe de Deus, explorando abordagens que não puderam ser implementadas durante o desenvolvimento do estudo original. O método proposto é estruturado em seis etapas principais:

- Pré-processamento: Utilização dos mesmos procedimentos aplicados na pesquisa original para garantir consistência e comparabilidade.
- Análise da qualidade do sinal: Emprego das técnicas previamente utilizadas por (DEUS, Luis Felipe de, 2023) para avaliar a confiabilidade dos dados.
- Extração de características: Identificação e seleção de atributos relevantes para a modelagem.
- Particionamento dos dados: Divisão dos conjuntos de dados para treinamento e teste, assegurando uma avaliação robusta.
- Treinamento de algoritmos de inteligência artificial: Desenvolvimento e ajuste de modelos utilizando abordagens avançadas de IA.
- Avaliação do método: Comparação rigorosa do desempenho do novo método com os resultados da pesquisa original, validando a eficácia das abordagens propostas.

Essa abordagem busca superar as limitações anteriores, contribuindo para a evolução dos métodos e resultados obtidos no campo de estudo.

Figura 6 – Fluxograma detalhado do pré-processamento.



Fonte: Adaptado de (DEUS, Luis Felipe de, 2023).

### 4.2.1 Pré-processamento

Sinais de PPG estão frequentemente sujeitos a ruídos de alta e baixa frequência, bem como artefatos devido à perda de contato do sensor ou movimentação, o que torna necessário aplicar técnicas de pré-processamento para melhorar a qualidade dos dados (LI; HUANG, 2020).

O pré-processamento do sinal envolve diversas etapas essenciais para aprimorar a qualidade dos dados e garantir sua adequação às análises subsequentes. Inicialmente, foi realizada uma segmentação do sinal por meio de janelamento, etapa fundamental para dividir o sinal contínuo em intervalos menores e facilitar o processamento. Esse procedimento é particularmente importante para lidar com variações temporais no sinal e garantir que as etapas seguintes sejam aplicadas consistentemente em todas as partes do dado.

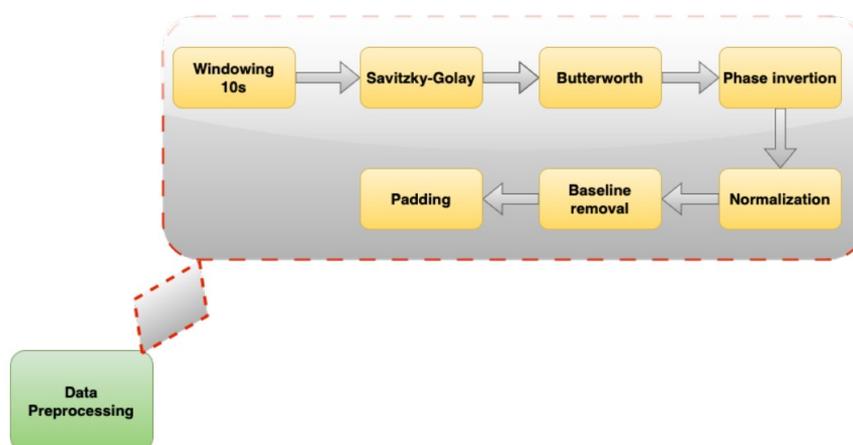
Após o janelamento, foram aplicados os filtros Savitzky-Golay e Butterworth. O filtro Savitzky-Golay (SAVITZKY; GOLAY, 1964) foi utilizado para suavizar o sinal (smoothing), ajustando-o com base em padrões identificados na série temporal e reduzindo o ruído sem comprometer as características essenciais do dado. Em seguida, o filtro passa-banda Butterworth (SELESNICK; BURRUS, 1998) foi empregado para remover frequências indesejadas, como aquelas associadas a movimentos respiratórios ou artefatos gerados pelo dispositivo de coleta. Para corrigir a inversão de fase inerente ao filtro Butterworth, foi realizada uma inversão do segmento do sinal, preservando sua integridade e alinhamento.

Após a aplicação dos filtros, o sinal foi submetido a uma normalização de amplitude utilizando o método Min-Max, que ajusta os valores para uma escala predefinida, favorecendo a uniformidade e o processamento subsequente. Por fim, a correção da linha de base foi realizada com a técnica de interpolação cubic spline, eliminando desvios e proporcionando maior estabilidade ao sinal processado.

Essa abordagem foi inspirada na metodologia originalmente desenvolvida por Luis Felipe de Deus em seu trabalho, mantendo os mesmos princípios para assegurar consistência nos resultados e facilitar a comparação entre as abordagens.

A Figura 7 apresenta o fluxograma detalhado das etapas que compõem o pré-processamento, destacando a sequência e as interações entre os diferentes blocos do processo.

Figura 7 – Fluxograma detalhado do pré-processamento.



Fonte: Adaptado de (DEUS, Luis Felipe de, 2023).

### 4.2.2 Janelamento

Devido à baixa quantidade de participantes no dataset e à presença de sinais de PPG com duração de 5 minutos, optou-se pela segmentação dos sinais, a fim de gerar segmentos menores para cada paciente, aumentando, assim, a quantidade de amostras disponíveis para cada nível de glicose. Esse processo foi fundamental para a extração das características a serem utilizadas no treinamento do modelo.

A técnica de janelamento, amplamente empregada, consiste em dividir um sinal de longa duração em segmentos menores. Neste trabalho, adotaram-se os mesmos critérios de janelamento utilizados no estudo de (DEUS, Luis Felipe de, 2023), onde foi realizado o janelamento do sinal original de 5 minutos em janelas de 10 segundos. Esse inter-

valo de tempo relativamente curto permite que o sinal original contínuo seja dividido em 30 segmentos menores, possibilitando previsões rápidas. Os segmentos podem ser processados individualmente, evitando a exclusão de um sinal inteiro caso algum trecho contenha artefatos ou interferência.

Os primeiros e os últimos pulsos do sinal de PPG de cada segmento foram descartados, visando eliminar inconsistências decorrentes da divisão inicial dos segmentos e de introduzir um preenchimento de janela (padding), garantindo maior precisão na análise e no treinamento do modelo. Embora a aplicação de funções matemáticas para modificar o sinal em cada segmento seja uma prática comum, neste estudo optou-se por não as utilizar, seguindo o método empregado por (DEUS, Luis Felipe de, 2023).

### 4.2.3 Avaliação da qualidade do sinal

Mesmo com o uso de técnicas avançadas de processamento de sinais, a presença de ruídos e artefatos é inevitável. Por isso, torna-se crucial incorporar um componente que avalie e quantifique a qualidade de cada segmento do sinal, removendo partes que não atendam aos critérios estabelecidos. Para este projeto, foi adotado o sistema de avaliação da qualidade do sinal desenvolvido por (DEUS, Luis Felipe de, 2023), cuja eficiência foi comprovada em seu estudo original, garantindo uma base sólida para a análise e processamento subsequentes.

Esse sistema é projetado para identificar e descartar segmentos espúrios, garantindo que dados comprometidos não sejam enviados aos algoritmos de inteligência artificial. A estratégia consiste na delimitação exata da duração de cada pulso com base no valor mínimo entre dois picos sistólicos, permitindo identificar com exatidão o início de cada pulso.

Após a segmentação, cada pulso é avaliado individualmente utilizando o Índice de Qualidade do Sinal (SSQI), conforme descrito por Li e Huang (LI; HUANG, 2020). Esse método atribui a cada pulso uma pontuação de qualidade que varia de 0 a 1, em que 1 indica uma qualidade perfeita. Esse critério quantitativo assegura que apenas os pulsos mais confiáveis sejam utilizados nas etapas subsequentes de análise.

Um filtro baseado em características morfológicas foi implementado para complementar a avaliação da qualidade do sinal, analisando pulsos individuais extraídos de janelas de 10 segundos. Este filtro identifica atributos como tempo de subida e descida, amplitude, relação entre tempos e presença de incisura dicrótica, além de características específicas como o pico sistólico, sendo o ponto de maior amplitude do pulso de PPG e, conforme (ALLEN, 2007), deve ocorrer tipicamente no segundo quadrante, sendo deslo-

camentos significativos indicativos de artefatos ou anomalias. Outras métricas incluem a assimetria do pulso, avaliada pela skewness, e a Área do Ponto de Inflexão (IPA), que corresponde à razão entre as áreas anterior e posterior ao pico diastólico, permitindo uma análise detalhada das alterações morfológicas. Esses atributos são comparados a valores-limite estabelecidos com base em sinais fisiologicamente plausíveis, conforme descrito na literatura por (CHARLTON et al., 2022). Caso uma janela apresente uma proporção de pulsos válidos abaixo de um limiar predefinido, ela é considerada não confiável.

Para validar um segmento, pelo menos 9 de 10 pulsos nele contidos devem atender aos seguintes critérios:

- $SSQI \geq 0,9$ , conforme o método de Li e Huang (LI; HUANG, 2020), garantindo alta qualidade do sinal;
- O pico sistólico deve ocorrer antes de 750 ms do tempo total do pulso;
- O valor de skewness deve estar entre 0 e 2 (valor adimensional);
- O valor da IPA deve ser maior que 1,5 (valor adimensional).

Essa abordagem, originalmente empregada por (DEUS, Luis Felipe de, 2023), garante a integridade e a qualidade dos dados antes de serem enviados para as etapas de extração de características e processamento subsequente. Ao assegurar que apenas segmentos de sinal válidos e de alta qualidade sejam utilizados, é possível otimizar os resultados nas fases seguintes do processamento e análise

#### 4.2.4 Extração de características

A extração de características de sinais biomédicos, especialmente do sinal de fotopleto-mografia (PPG), é essencial para a análise e interpretação dos dados. A forma de onda do sinal nem sempre revela, de maneira clara e coesa, as correlações entre o sinal biomédico e as condições clínicas subjacentes. Características como amplitude, deslocamento e largura dos pulsos do sinal de PPG são frequentemente utilizadas para quantificar e descrever as propriedades fisiológicas que o sinal representa.

Através da detecção e medição dessas características, é possível obter informações valiosas sobre a função cardiovascular e outras condições de saúde relacionadas. Entre as características mais relevantes, destaca-se a variabilidade da frequência cardíaca (VFC), que fornece informações importantes sobre o equilíbrio autonômico do sistema nervoso

e o estado do sistema cardiovascular. A HRV é frequentemente utilizada para avaliar o risco de doenças cardiovasculares e monitorar o estresse e a recuperação física.

Além disso, a decomposição empírica do modo (EMD) tem se mostrado uma técnica eficaz para extrair e analisar componentes de diferentes escalas temporais do sinal PPG. A EMD permite separar o sinal em componentes oscilatórios chamadas funções de modo intrínseco (IMFs), oferecendo uma visão mais detalhada das diferentes frequências presentes no sinal, o que pode ser crucial para detectar anomalias e padrões específicos.

Através da detecção e medição dessas características, é possível obter informações valiosas sobre a função cardiovascular e outras condições de saúde relacionadas. Essas medidas podem ser usadas para identificar padrões específicos da onda, avaliar a integridade do sistema circulatório ou até mesmo detectar alterações e anomalias.

Portanto, a extração de características morfológicas e a aplicação de métodos avançados como HRV e a EMD desempenham um papel crucial na compreensão do funcionamento fisiológico, contribuindo para uma análise mais detalhada e precisa, e tendo grande aplicação em monitoramento da saúde e diagnóstico clínico.

Neste estudo, utilizando a mesma abordagem aplicada por Luis Felipe, cada segmento de 10 segundos de sinal PPG, que passou pelo filtro de qualidade, foi utilizado para a extração de características. Um segmento de 10 segundos de sinal contém normalmente entre 10 e 15 pulsos, no entanto, nem todos os pulsos em um segmento apresentam qualidade ideal. Para evitar a extração de características a partir de pulsos ruidosos, foi adotada a técnica de média de todos os pulsos presentes no segmento. Essa abordagem assegura uma representação mais confiável e robusta das características do sinal de PPG, minimizando a influência de pulsos de baixa qualidade e aumentando a precisão na análise das propriedades do sinal.

As principais características extraídas do sinal de PPG incluem:

- SYS\_PEAK: Posição do pico sistólico.
- DIA\_PEAK: Posição do pico diastólico.
- RISING\_TIME: Tempo de subida entre o início do pulso e o pico sistólico.
- NOTCH: Posição do notch no pulso de PPG.

Além dessas características morfológicas, também são analisadas as componentes extraídas utilizando técnicas como a Decomposição Modal Empírica (EMD), que fornecem as seguintes informações:

- Mean (IMFmean): O valor médio da Componente Modal Intrínseca (IMF).

- Variance (IMFvar): Medida da dispersão dos valores da IMF em torno da média.
- Standard Deviation (IMFstd): A raiz quadrada da variância, que indica a extensão da dispersão dos valores da IMF.
- Peak-to-Peak Amplitude (PTP): Diferença entre os valores máximo e mínimo da IMF.
- Skewness (IMFskew): Medida da assimetria da distribuição de probabilidade da IMF.
- Kurtosis (IMFkurt): Medida da "cauda" da distribuição de probabilidade da IMF, indicando a presença de valores extremos.

Adicionalmente, são extraídas características relacionadas à variabilidade da frequência cardíaca (HRV), fundamentais para a análise da flutuação dos intervalos entre os batimentos cardíacos:

- SDNN: Desvio padrão dos intervalos RR, uma medida global da variabilidade cardíaca.
- RMSSD: Raiz quadrada da média dos quadrados das diferenças entre intervalos RR sucessivos, refletindo a variabilidade de alta frequência.
- PNN50: Percentual de intervalos RR consecutivos com uma diferença superior a 50 ms, utilizado para medir a variabilidade cardíaca em baixas frequências.
- IQR\_RR: Intervalo interquartil dos intervalos RR, uma medida de dispersão dos intervalos RR, útil para avaliar a variabilidade dos batimentos cardíacos.

### 4.2.5 Particionamento de dados

O particionamento de dados em conjuntos de treinamento e teste é uma prática essencial em algoritmos de aprendizado de máquina. Esse processo envolve dividir o conjunto de dados disponível em duas partes distintas: um conjunto de treinamento, utilizado para treinar o modelo, e um conjunto de teste, usado para avaliar o desempenho do modelo em dados previamente não vistos. Essa divisão é crucial para testar a capacidade do modelo de generalizar o conhecimento adquirido durante o treinamento, verificando sua precisão ao lidar com dados desconhecidos.

Um particionamento adequado dos dados assegura a validade e a confiabilidade dos resultados obtidos pelo modelo de aprendizado de máquina. Sem essa separação, o

modelo pode se ajustar excessivamente aos dados de treinamento (overfitting), o que comprometeria sua capacidade de desempenho em novos dados. Assim, essa prática contribui diretamente para a construção de modelos mais robustos e eficientes.

No presente estudo, o particionamento dos dados foi realizado com base nos segmentos extraídos e nas informações adquiridas a partir deles, utilizando um embaralhamento aleatório para a divisão dos conjuntos. Inicialmente, os pacientes foram distribuídos de forma que 70% dos pacientes foram selecionados para compor o conjunto de treinamento, enquanto os 30% restantes foram destinados ao conjunto de teste, abrangendo todos os segmentos de dados. Após os testes com esse primeiro grupo, os segmentos foram agrupados por paciente em intervalos de 10 amostras, e o particionamento foi realizado novamente, com 70% dos dados destinados ao treinamento e 30% ao teste.

As variações na utilização desses dois grupos e a metodologia de particionamento são detalhadas no Capítulo 5.

#### 4.2.6 Aprendizado de Máquina

Para comparar os resultados obtidos por Luis, foram utilizados os mesmos modelos de aprendizado de máquina empregados em seu estudo: Random Forest e CatBoost. Esses modelos foram treinados utilizando os dados segmentados, com o conjunto de treinamento sendo utilizado para ajustar os parâmetros do modelo e o conjunto de teste para avaliar sua capacidade de generalização.

A escolha desses modelos foi baseada nos resultados positivos reportados por (DEUS, Luis Felipe de, 2023), que demonstraram alta precisão e confiabilidade na estimativa de glicose em seu estudo. Em particular, foram avaliadas métricas como o erro médio absoluto (Mean Absolute Error - MAE) e o coeficiente de correlação ( $R^2$ ), que indicaram uma boa correspondência entre os valores estimados e os valores de referência de glicose. Esses resultados foram utilizados como parâmetro para comparar o desempenho no presente trabalho, permitindo verificar tanto a consistência quanto possíveis melhorias na estimativa de glicose em relação ao método original.

O Random Forest (RF), descrito por Breiman (BREIMAN, 2001), é um algoritmo de aprendizado de máquina baseado em um conjunto de árvores de decisão, sendo amplamente utilizado para tarefas de classificação e regressão. O método constrói diversas árvores de decisão de maneira independente, utilizando amostras aleatórias do conjunto de dados de treinamento, mediante um processo conhecido como bootstrap (LOUPPE, 2014). Durante a construção de cada árvore, são selecionados aleatoriamente subconjuntos das características, aumentando a diversidade entre as árvores e melhora a gene-

realização do modelo.

A principal vantagem do Random Forest está na combinação de previsões: a partir da votação das árvores de decisão ou da média de suas previsões, o algoritmo gera uma estimativa final mais robusta. Isso contribui para reduzir a variância do modelo e diminuir o risco de overfitting, ao passo que melhora sua precisão, especialmente em conjuntos de dados complexos. Dessa forma, o Random Forest é uma técnica poderosa e amplamente utilizada em diversas áreas, como análise de dados biomédicos, reconhecimento de padrões, e previsão em áreas como saúde, finanças e marketing

O CatBoost (DOROGUSH; ERSHOV; GULIN, 2018) é um algoritmo de gradient boosting que visa combinar múltiplos modelos fracos (geralmente árvores de decisão) para criar um modelo mais forte e preciso. A técnica de boosting utiliza uma abordagem sequencial, onde cada árvore de decisão sucessiva tenta corrigir os erros das árvores anteriores. O CatBoost se destaca por sua regularização embutida, que ajuda a evitar o overfitting, utilizando uma combinação de regularizações L1, L2 e regularização por desvio. Essas regularizações controlam a complexidade do modelo e reduzem o risco de ajustes excessivos aos dados de treinamento.

Uma característica notável do CatBoost é a sua detecção automática de melhorias significativas durante o processo de treinamento. Utilizando métodos estatísticos, o CatBoost determina quando é necessário adicionar uma nova árvore à sequência, garantindo que o treinamento continue de forma eficiente e eficaz. Essa abordagem não só contribui para melhorar a precisão do modelo, mas também otimiza o tempo de treinamento, evitando a criação de árvores desnecessárias e, assim, acelerando o processo de construção do modelo.

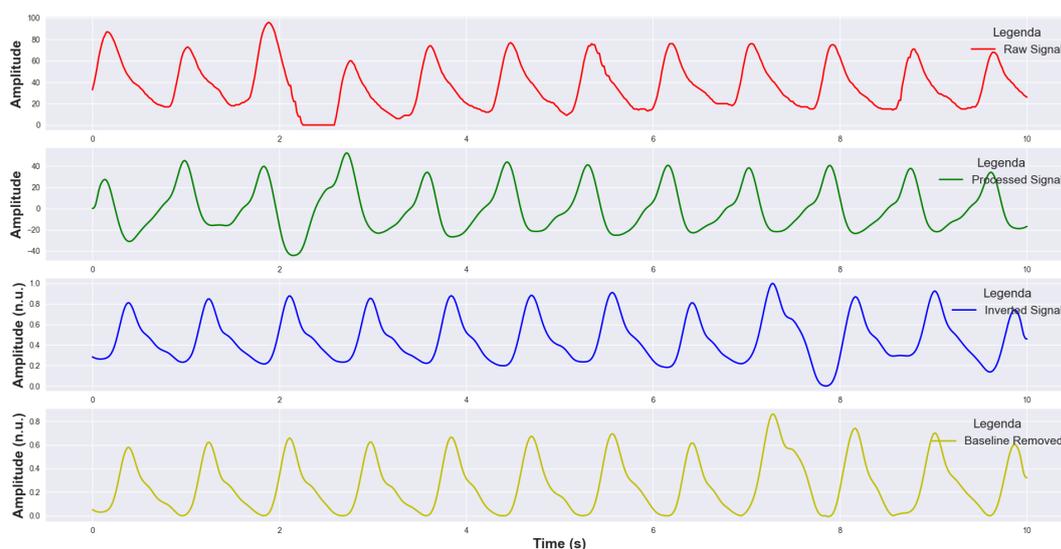
O CatBoost é particularmente eficaz em problemas com dados categóricos e pode lidar com essas variáveis sem a necessidade de pré-processamento complexo, como codificação explícita. Sua robustez e capacidade de lidar com grandes volumes de dados, com a sua eficiência computacional, fazem dele uma excelente opção em tarefas de aprendizado supervisionado.

## 5 Resultados

Este capítulo planeja apresentar os resultados obtidos neste trabalho. Primeiramente, vale ressaltar a preocupação central deste estudo na quantidade de sinais coletados e sua qualidade, considerando as diversas interferências que invalidaram diversos dos seus segmentos, tendo como suas possíveis causas a taxa de gordura corporal ou circulação de sangue na área onde foi realizada a coleta do sinal. Além disso, foram discutidos possíveis impactos da qualidade do sinal obtido por técnicas ópticas.

A figura 8 é um exemplo do sinal obtido pela medição e todas as suas etapas de pré-processamento.

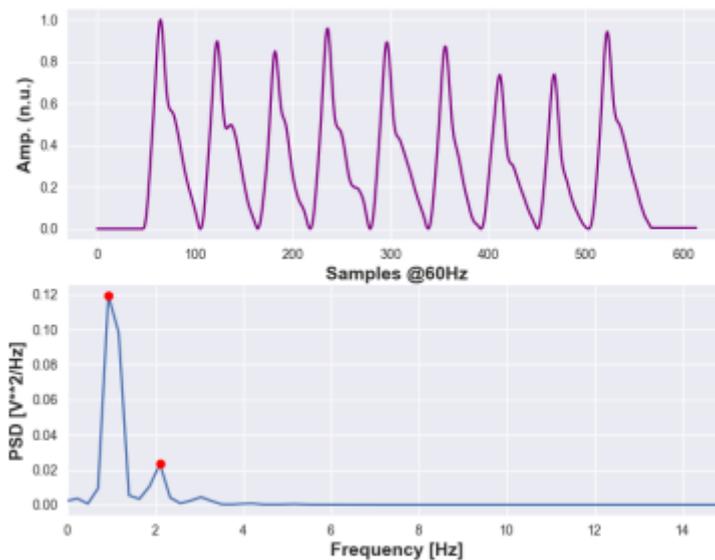
Figura 8 – Segmento de 10 s do primeiro (001) @60Hz, glicose de 181 mg/dL. a) sinal raw, b) sinal filtrado com Savitzky-Golay + Butterworth, c) sinal normalizado + inversão de fase, d) sinal com remoção de linha base.



Fonte: Autor.

Após a etapa de pré-processamento, foram extraídas características do domínio da frequência a partir de segmentos de 10 s. Por meio do gráfico de Densidade Espectral, identificaram-se atributos como a frequência das primeiras e segundas harmônicas, além da potência máxima nesses pontos específicos. A Figura 9 ilustra esse processo, destacando os pontos correspondentes a essas harmônicas.

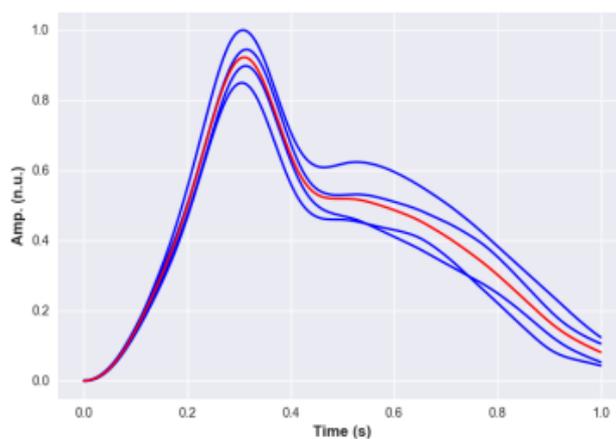
Figura 9 – Domínio da frequência do participante 001: Respectivamente, segmento de sinal processado e gráfico de Densidade Espectral de Potência.



Fonte: Autor.

Após a separação dos pulsos no segmento do sinal, foi realizada uma média para obter um único pulso representativo do segmento. Esse processo está ilustrado na Figura 10. Esse pulso médio serviu como base para a extração de características morfológicas relevantes para a análise.

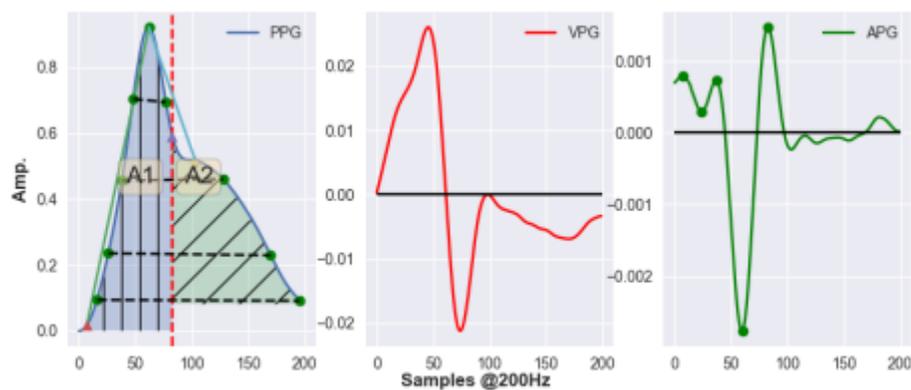
Figura 10 – Domínio do tempo do participante 001: Pulsos sobrepostos (em azul) de um segmento de 10 segundos, com o pulso médio realçado em vermelho



Fonte: Autor.

O pulso médio foi então utilizado para a obtenção de outras características que, posteriormente, foram concatenadas e empregadas nos modelos de inteligência artificial (IA). Um exemplo desse processamento de características morfológicas é mostrado na Figura 11, onde são calculadas a primeira e a segunda derivada do sinal. A primeira derivada é utilizada para identificar variações rápidas, como picos e vales, enquanto a segunda derivada fornece informações sobre a curvatura do sinal, permitindo a análise de padrões mais detalhados, como pontos de inflexão. Todas as características descritas foram extraídas de cada pulso e armazenadas em um objeto do tipo DataFrame para facilitar o processamento e a integração com os modelos de IA.

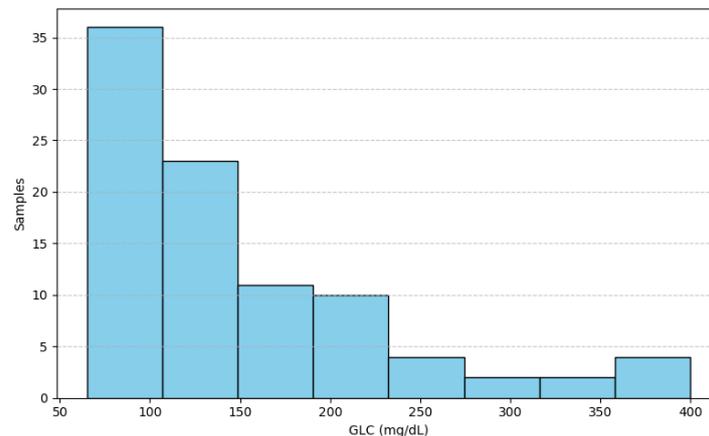
Figura 11 – Extração de características do pulso do sinal de PPG, VPG e APG, destacados em azul, vermelho e verde. Participante 001 pulso re-amostrado a @200 Hz, com glicose de 181 mg/dL.



Fonte: Autor.

O histograma fornece uma descrição do conjunto de dados com base no nível de glicose de cada paciente (Figura 12).

Figura 12 – Histograma da distribuição de valores de glicose (mg/dL) no conjunto de dados dos pacientes.



Fonte: Autor.

Com as características extraídas, foram realizados os treinamentos para identificar como idade, HRV e a EMD influenciam na estimativa do nível de glicose.

Tendo como hipótese que a separação por grupos de idade poderia melhorar a estimativa do nível de glicose, ela baseia-se em alguns fundamentos fisiológicos. A idade é um fator determinante que afeta diversos processos metabólicos e fisiológicos, incluindo a variabilidade da frequência cardíaca (HRV) e os níveis de glicose no sangue (ELECTROPHYSIOLOGY, 1996). Com o envelhecimento, ocorrem alterações significativas na resposta autonômica, na sensibilidade à insulina e na dinâmica cardiovascular, que podem impactar diretamente a relação entre essas variáveis.

No caso da HRV, evidências científicas indicam que ela tende a diminuir com o avanço da idade, devido a uma redução na atividade parassimpática e na capacidade de adaptação do sistema nervoso autônomo (SHAFFER; MCCRATY; ZERR, 2014; ELGENDI, 2012). Essas alterações podem modificar os padrões de variação capturados pela Decomposição Empírica por Modos (EMD).

A separação dos dados em grupos etários permite capturar diferenças intrínsecas nos padrões fisiológicos, o que pode aumentar a acurácia dos modelos de estimativa. Essa abordagem reconhece que os parâmetros extraídos possuem pesos e significados distintos para cada faixa etária. Em indivíduos mais jovens, por exemplo, a maior variabilidade autonômica pode fornecer sinais mais robustos para correlacionar HRV com os níveis de glicose (CATAI et al., 2020; WU et al., 2021). Por outro lado, em indivíduos mais velhos, fatores como a redução da elasticidade vascular, o declínio da sensibilidade autonômica e alterações no metabolismo basal podem influenciar significativamente os

padrões fisiológicos. Esses fatores demandam ajustes específicos nos métodos de análise para garantir que os modelos de estimativa sejam igualmente precisos nessa faixa etária, considerando as mudanças graduais que ocorrem com o envelhecimento.

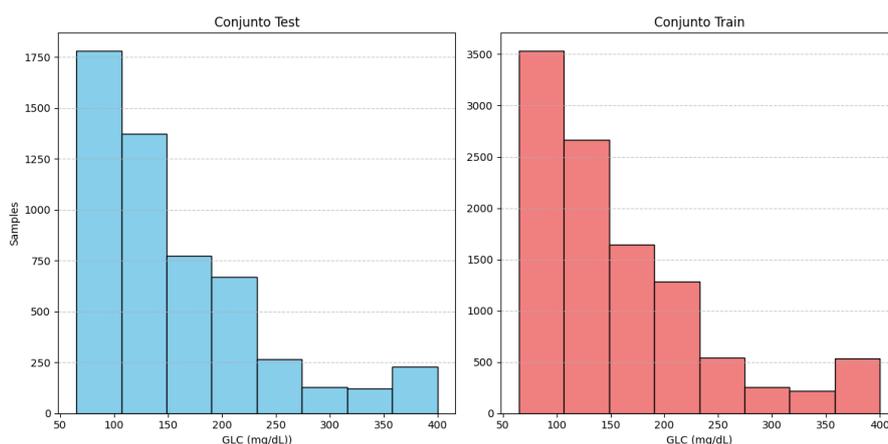
Dessa forma, a estratégia de segmentação etária fundamenta-se na ideia de que modelos adaptados por faixa etária conseguem captar nuances que podem ser diluídas em análises gerais, resultando em predições mais precisas e alinhadas às características fisiológicas de cada grupo.

Os treinamentos realizados incluíram duas abordagens principais: (1) utilização de todos os segmentos extraídos e (2) criação de conjuntos formados por segmentos de 10 segundos com cálculo de suas médias aritméticas. Para cada uma dessas abordagens, foram analisados dois cenários: (a) sem divisão por idade e (b) com divisão em dois grupos etários, de até 50 anos (25 participantes) e maiores de 50 anos (67 participantes). A divisão dos grupos de dados foi feita considerando 70% para treino e 30% para teste em cada caso, de modo a garantir a consistência e a robustez dos modelos desenvolvidos.

## 5.1 Treinamento utilizando todos os segmentos

Nas próximas seções, são apresentados os resultados do treinamento realizado com o conjunto de dados UFSC-BGL, considerando diferentes abordagens de segmentação por idade: sem divisão por faixa etária, participantes com até 50 anos e participantes com mais de 50 anos. A Figura 13 ilustra a distribuição das amostras utilizadas nos conjuntos de treino e teste.

Figura 13 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de treino e teste para os diferentes segmentos de pacientes.



Fonte: Autor.

### 5.1.1 Conjunto Completo

No primeiro experimento, foi utilizado o conjunto de dados completo, sem segmentação por faixa etária. Os resultados indicaram que o modelo o Random Forest (RF) teve um melhor desempenho em relação ao CatBoost, conforme detalhado na Tabela 1. Adicionalmente, a Tabela 2 apresenta a porcentagem de previsões em cada zona de precisão do modelo de erro de Clarke, amplamente utilizado para avaliar a acurácia em estimativas de glicose.

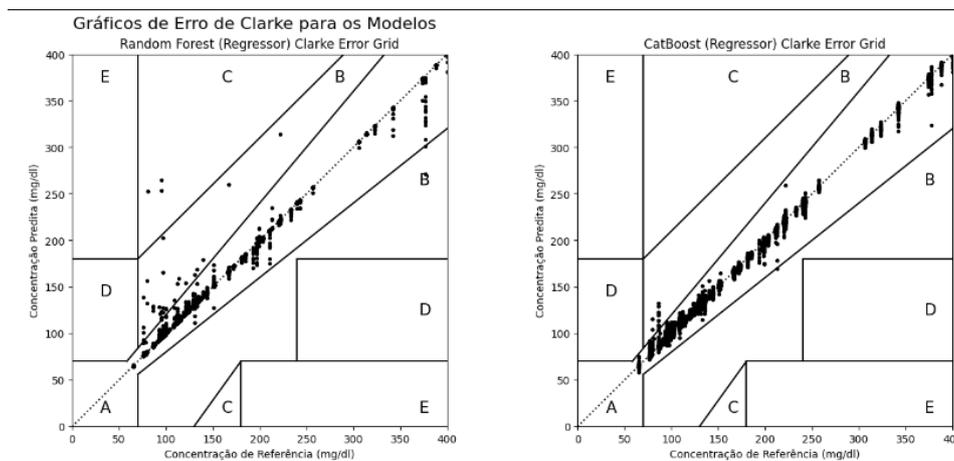
Tabela 1 – Resultados dos modelos com o conjunto completo.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	0.228922	0.821784	6.072357	0.993822	0.626443
CatBoost	0.079765	2.089387	3.692846	0.997715	1.580574

Fonte: Autor.

A Figura 14 apresenta o gráfico de erro de Clarke, que permite uma análise detalhada da precisão dos modelos, destacando a frequência de previsões em diferentes zonas de erro.

Figura 14 – Gráfico de Clarke das previsões de glicose.



Fonte: Autor.

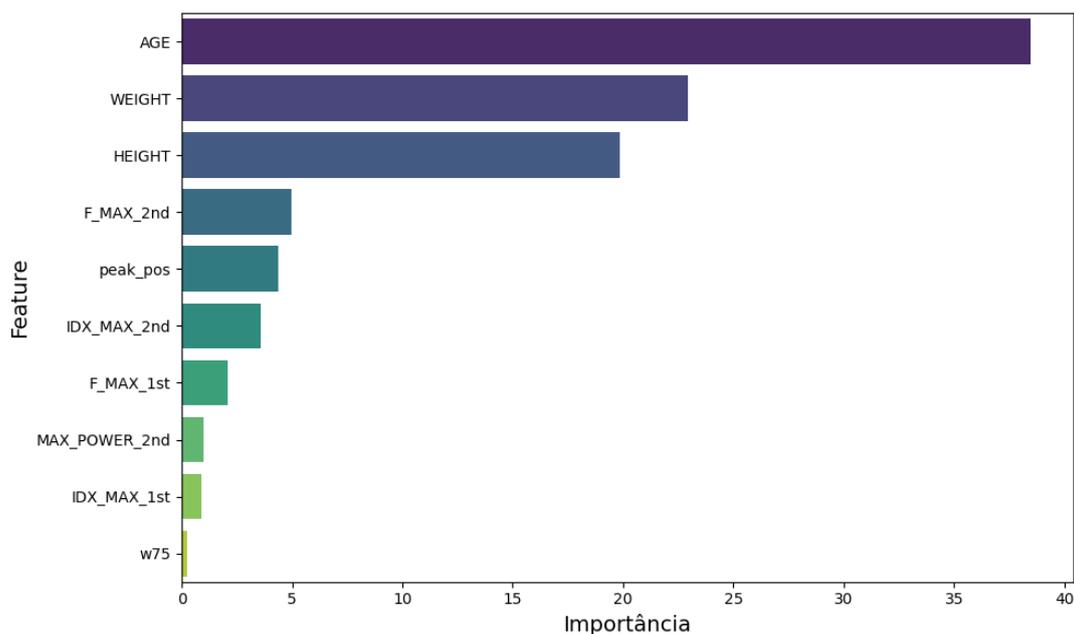
Tabela 2 – Distribuição das previsões nas zonas de precisão do modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	5300	38	5	0	0
CatBoost	5326	17	0	0	0

Fonte: Autor.

Ao analisar os atributos que influenciavam o treinamento do modelo, como ilustrado na Figura 15, observou-se que as três mais relevantes era idade, altura e peso. Para modelos que visam analisar a variação de amostras, é importante considerar que features invariantes para diferentes segmentos de amostras de um mesmo indivíduo podem impactar como o modelo prevê a variável-alvo, neste caso, o nível de glicose. Desta forma, as features idade, altura e peso foram descartadas do dataset, os resultados treinamentos seguintes foram feitos sem elas.

Figura 15 – Feature Mais Influentes.



Fonte: Autor.

### 5.1.2 Conjunto Sem Divisão por Idade

Os resultados indicaram que o modelo CatBoost teve um melhor desempenho em relação ao Random Forest (RF), conforme detalhado na Tabela 3. Adicionalmente, a Tabela 4 apresenta a porcentagem de predições em cada zona de precisão do modelo de erro de Clarke, amplamente utilizado para avaliar a acurácia em estimativas de glicose. É possível comprovar que sem as features idade, peso e altura o desempenho dos modelos cai em comparação quando elas estavam no dataset.

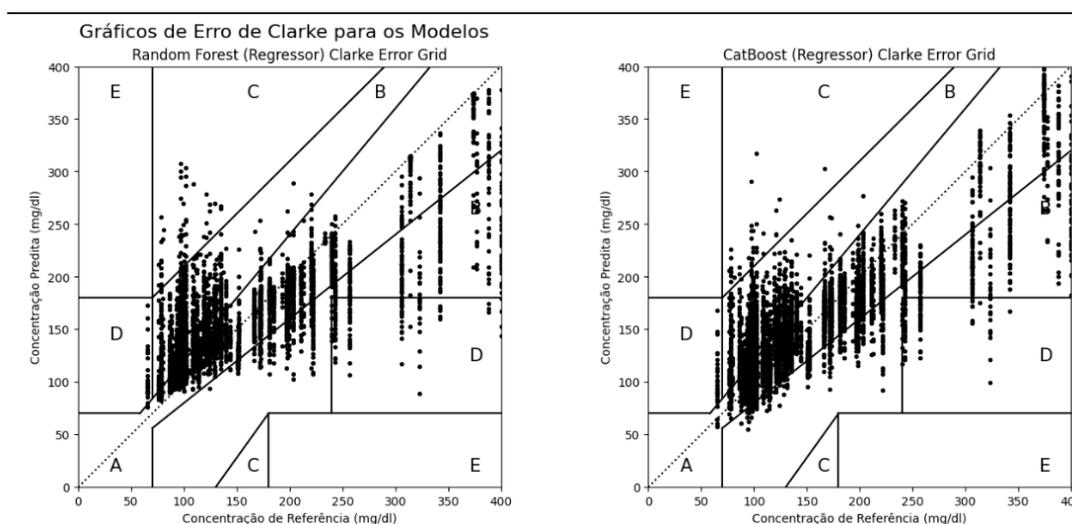
Tabela 3 – Resultados dos modelos com o conjunto completo (sem divisão por idade).

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	2.474907	30.829760	47.508965	0.621856	21.895065
CatBoost	1.113285	30.589405	43.201912	0.687311	21.732137

Fonte: Autor.

A Figura 16 apresenta o gráfico de erro de Clarke, que permite uma análise detalhada da precisão dos modelos, destacando a frequência de previsões em diferentes zonas de erro.

Figura 16 – Gráfico de Clarke das previsões de glicose (sem divisão por idade).



Fonte: Autor.

Tabela 4 – Distribuição das previsões nas zonas de precisão do modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	2785	1150	1408	0	0
CatBoost	2557	1370	1416	0	0

Fonte: Autor.

### 5.1.3 Conjunto Com Participantes Até 50 Anos

No segundo experimento, foram analisados apenas os participantes com idade até 50 anos. Conforme mostrado na Tabela 5, os erros dos modelos CatBoost e RF diminuem em comparação ao conjunto completo. A Figura 17 exibe o gráfico de erro de Clarke para essa análise. Neste experimento, conseguimos observar que o erro dos modelos RF

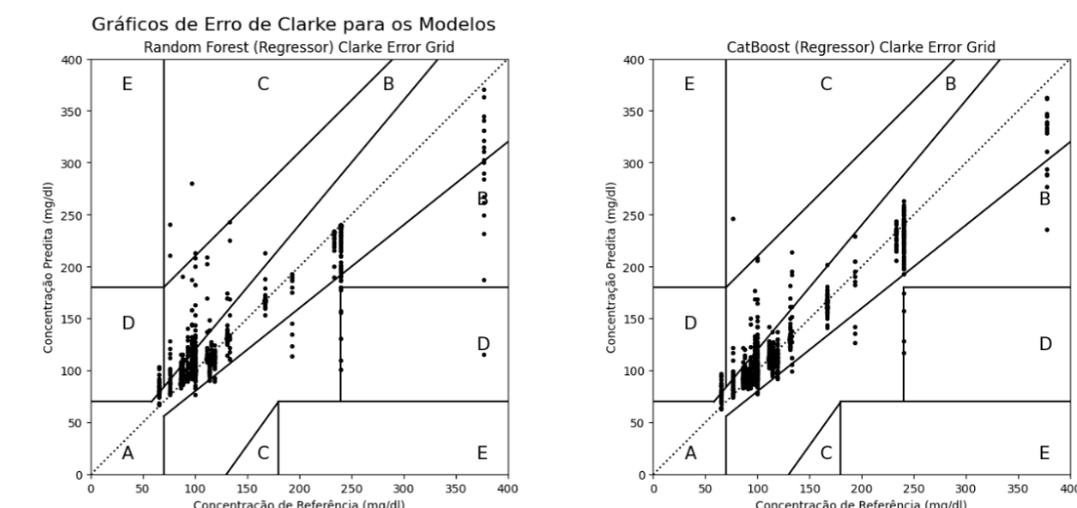
e CatBoost tiveram um leve aumento. Na Fig. 17 é exposto o gráfico de erro de Clarke. A Tabela 6, demonstra o resultado da porcentagem de estimativas em cada zona de precisão.

Tabela 5 – Resultados dos modelos com participantes até 50 anos.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	1.031602	8.614607	21.622522	0.834094	7.114916
CatBoost	0.669883	8.764018	16.572184	0.902544	7.424740

Fonte: Autor.

Figura 17 – Gráfico de Clarke das predições de glicose (participantes até 50 anos).



Fonte: Autor.

Tabela 6 – Distribuição das predições nas zonas de precisão do modelo de erro de Clarke (até 50 anos).

Modelo	A	B	C	D	E
RF	1160	86	65	0	0
CatBoost	1181	88	42	0	0

Fonte: Autor.

### 5.1.4 Conjunto Com Participantes Maiores de 50 Anos

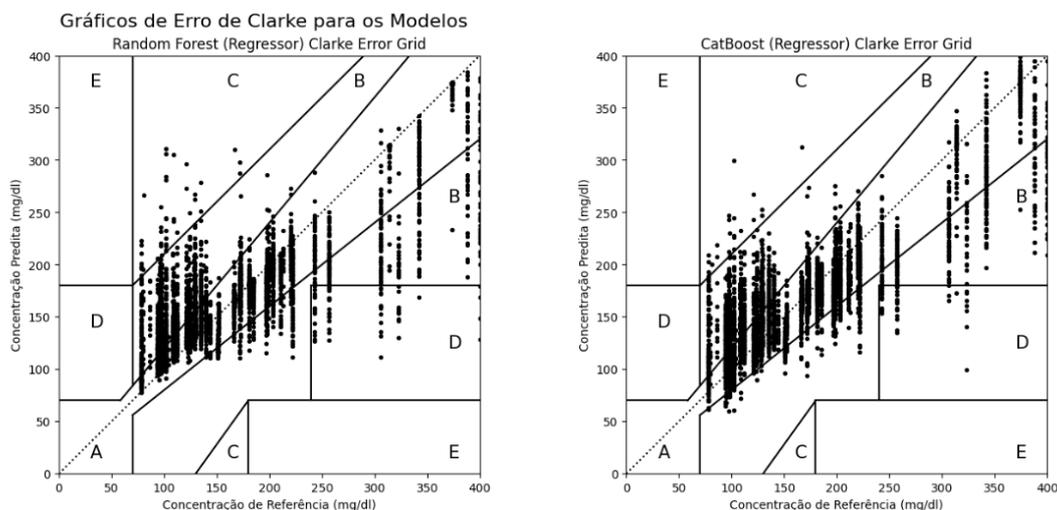
O último experimento considerou apenas os participantes com mais de 50 anos. A Tabela 7 apresenta os resultados, que indicam um aumento leve nos erros para os modelos CatBoost e RF. O gráfico de erro de Clarke, exibido na Figura 18, complementa a análise.

Tabela 7 – Resultados dos modelos com participantes acima de 50 anos.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	2.177866	30.025756	46.120529	0.669603	20.45623
CatBoost	1.312959	29.741054	41.509647	0.732363	20.19762

Fonte: Autor.

Figura 18 – Gráfico de Clarke das predições de glicose (participantes acima de 50 anos).



Fonte: Autor.

Tabela 8 – Distribuição das predições nas zonas de precisão do modelo de erro de Clarke (acima de 50 anos).

Modelo	A	B	C	D	E
RF	2098	917	1017	0	0
CatBoost	1972	1025	1035	0	0

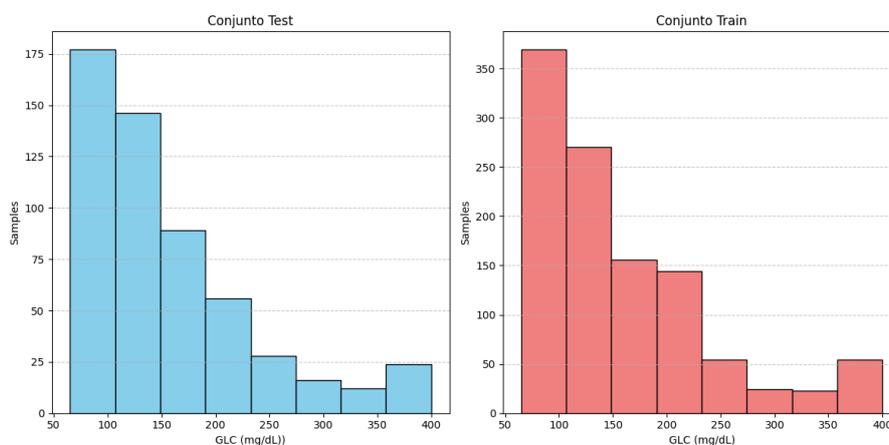
Fonte: Autor.

Os resultados reforçam a eficiência dos modelos CatBoost e RF, especialmente para populações mais jovens. A avaliação contínua desses modelos pode contribuir significativamente para a precisão em estimativas de glicose.

## 5.2 Treinamento Utilizando Agrupamento de 10 Segmentos.

Nesta seção, são apresentados os resultados do treinamento com dados extraídos dos sinais de PPG de cada paciente, utilizando o agrupamento de 10 segmentos e a média aritmética dos mesmos. Os testes foram realizados para três cenários: sem segmentação por faixa etária, com participantes até 50 anos e com participantes acima de 50 anos. A Figura 19 ilustra a distribuição das amostras para os conjuntos de treino e teste.

Figura 19 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de treino e teste com agrupamento de segmentos.



Fonte: Autor.

### 5.2.1 Conjunto Sem Divisão por Idade

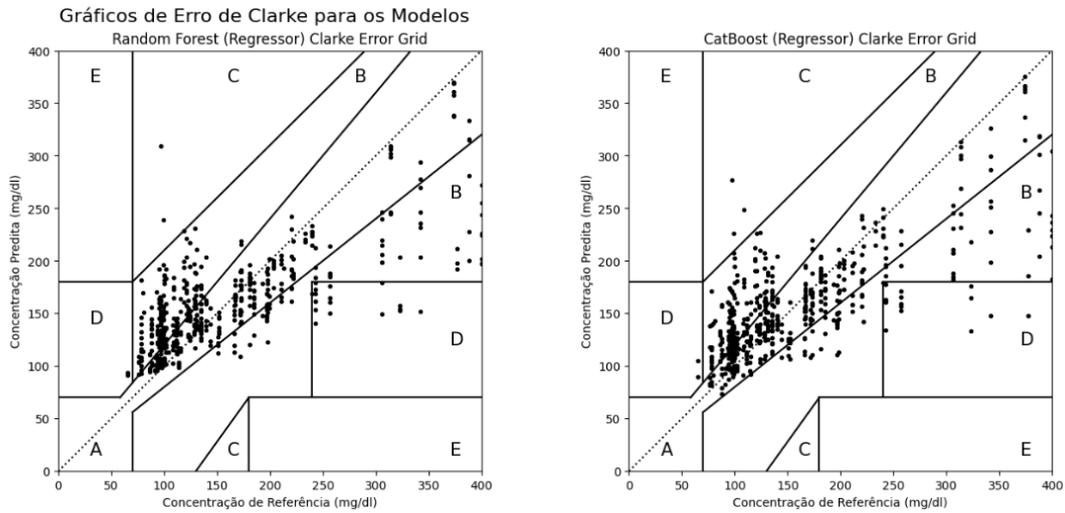
No primeiro experimento, foi utilizado o conjunto completo, sem segmentação por faixa etária. Os resultados indicaram que o modelo CatBoost teve um melhor desempenho em relação ao RF, conforme mostrado na Tabela 9. A Figura 20 exibe o gráfico de erro de Clarke.

Tabela 9 – Resultados dos modelos com agrupamento de segmentos (sem divisão por idade).

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	0.408788	36.431864	52.769395	0.537956	24.679454
CatBoost	-2.182566	35.560691	51.540153	0.559231	23.663726

Fonte: Autor.

Figura 20 – Gráfico de Clarke das predições de glicose (agrupamento de segmentos, sem divisão por idade).



Fonte: Autor.

Tabela 10 – Distribuição das predições nas zonas de precisão do modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	221	151	176	0	0
CatBoost	237	147	164	0	0

Fonte: Autor.

### 5.2.2 Conjunto Com Participantes Até 50 Anos

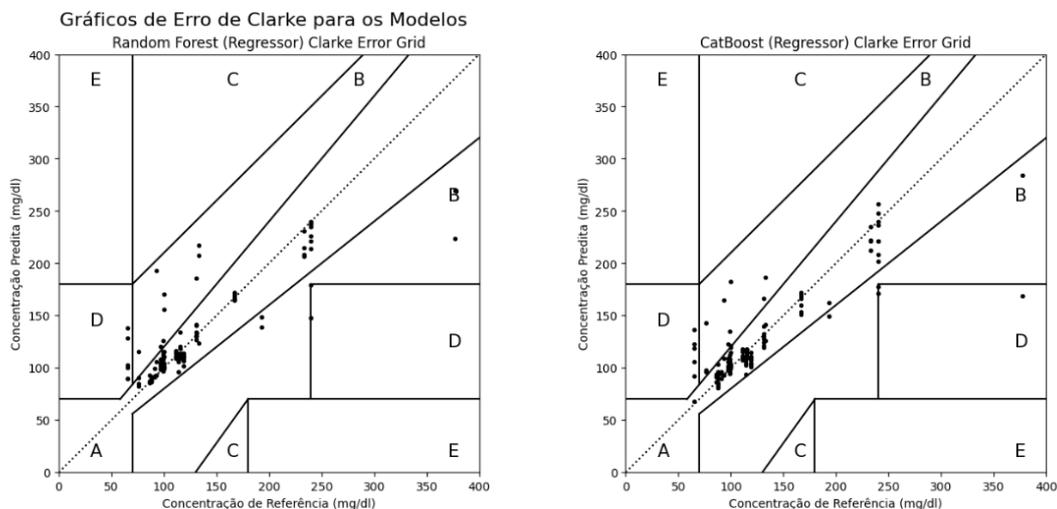
O segundo experimento considerou apenas os participantes com até 50 anos. Conforme mostrado na Tabela 11, os erros dos modelos RF e CatBoost apresentaram uma leve melhora em relação ao conjunto completo com agrupamento de segmentos. A Figura 21 exibe o gráfico de erro de Clarke.

Tabela 11 – Resultados dos modelos com agrupamento de segmentos (participantes até 50 anos).

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	1.580584	14.921168	29.599122	0.711654	12.193978
CatBoost	0.348430	14.511542	28.567599	0.731402	12.149967

Fonte: Autor.

Figura 21 – Gráfico de Clarke das predições de glicose (agrupamento de segmentos, até 50 anos).



Fonte: Autor.

Tabela 12 – Distribuição das predições nas zonas de precisão do modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	113	10	14	0	0
CatBoost	112	12	13	0	0

Fonte: Autor.

### 5.2.3 Conjunto Com Participantes Maiores de 50 Anos

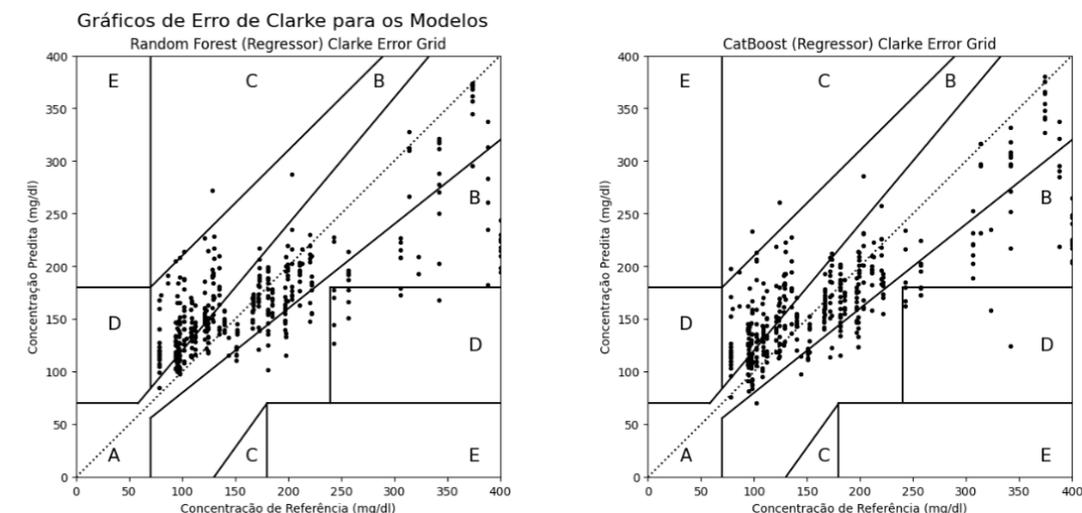
O terceiro experimento avaliou os participantes acima de 50 anos. Como observado na Tabela 13, os modelos CatBoost e RF apresentaram um leve aumento no erro em relação ao conjunto completo com agrupamento de segmentos. A Figura 22 apresenta o gráfico de erro de Clarke para esse grupo.

Tabela 13 – Resultados dos modelos com agrupamento de segmentos (participantes acima de 50 anos).

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	-0.907609	38.172665	54.210209	0.5805903	24.850944
CatBoost	-2.284349	36.722419	51.689053	0.618694	23.647710

Fonte: Autor.

Figura 22 – Gráfico de Clarke das predições de glicose (agrupamento de segmentos, acima de 50 anos).



Fonte: Autor.

Tabela 14 – Distribuição das predições nas zonas de precisão do modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	154	119	138	0	0
CatBoost	165	114	132	0	0

Fonte: Autor.

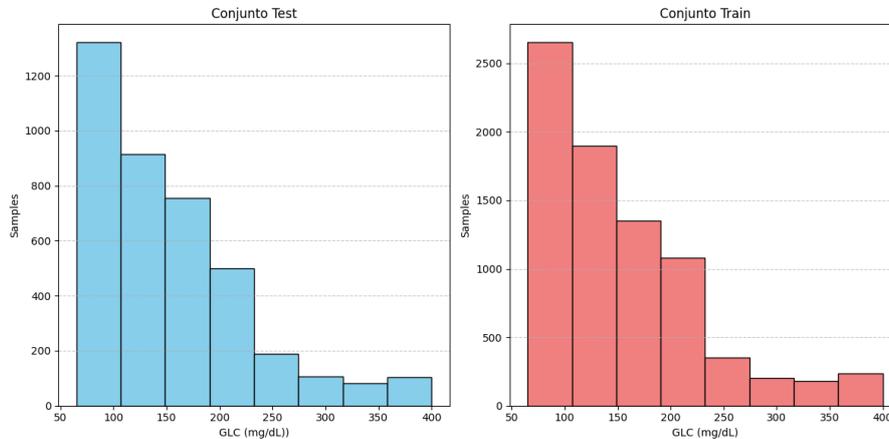
Os resultados indicam que o agrupamento de segmentos e a média aritmética podem influenciar o desempenho dos modelos, especialmente em cenários com divisão por faixa etária. Os modelos CatBoost e RF continuam apresentando resultados mais consistentes, especialmente para participantes até 50 anos.

### 5.3 Treinamento utilizando todos os segmentos com a incorporação da Decomposição Empírica por Modos

Nesta seção, apresentamos os resultados do treinamento utilizando segmentos extraídos dos sinais de PPG de cada paciente com a incorporação da Decomposição Empírica por Modos (EMD). Os dados foram analisados em três cenários: sem divisão por idade,

para participantes com até 50 anos, e para participantes com mais de 50 anos. Na Figura 23, está ilustrada a distribuição das amostras nos conjuntos de treino e teste.

Figura 23 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de teste e treino para os segmentos dos pacientes com a incorporação da EMD.



Fonte: Autor.

### 5.3.1 Conjunto Sem Divisão por Idade.

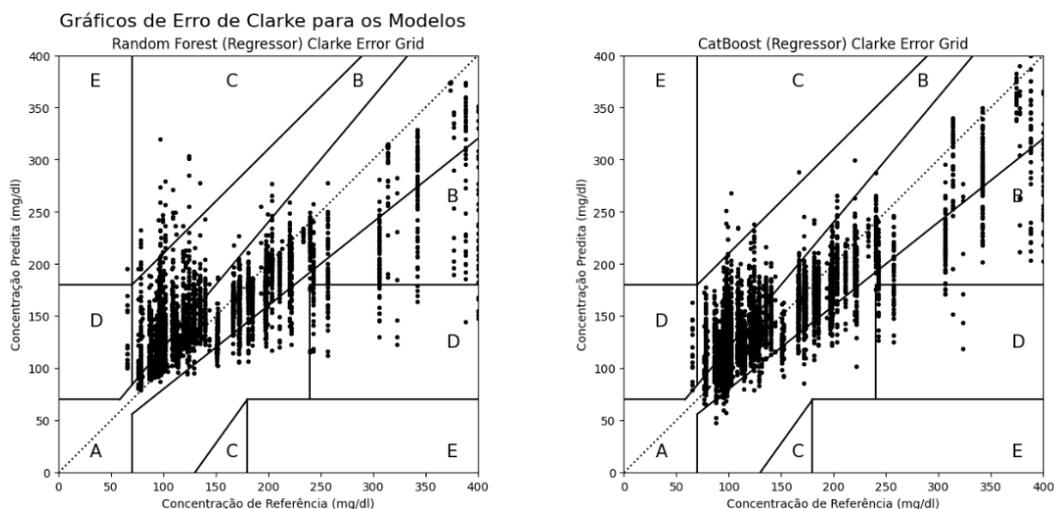
No primeiro teste, utilizando o conjunto de dados completo, o modelo CatBoost teve um melhor desempenho em relação ao RF. Os resultados detalhados estão na Tabela 15, enquanto a Tabela 16 apresenta a distribuição das estimativas nas diferentes zonas do modelo de erro de Clarke.

Tabela 15 – Resultados dos modelos com o conjunto completo (sem divisão por idade).

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	1.890092	28.666398	44.433976	0.615085	20.325538
CatBoost	0.490598	27.798770	39.034219	0.702953	19.778590

Fonte: Autor.

Figura 24 – Gráfico de Clarke das predições de glicose (sem divisão por idade).



Fonte: Autor.

Tabela 16 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	2120	899	954	0	0
CatBoost	2002	1060	911	0	0

Fonte: Autor.

### 5.3.2 Conjunto para Participantes com até 50 anos

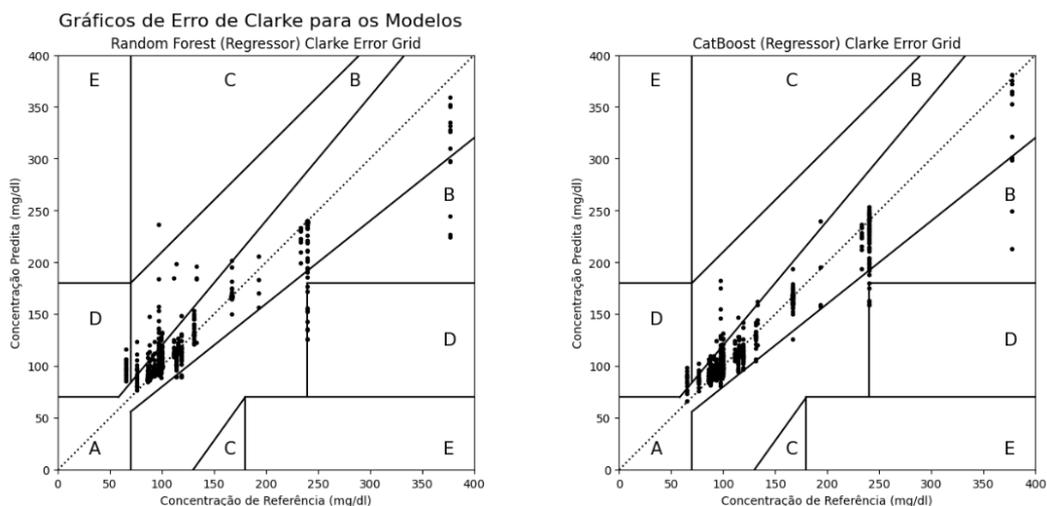
No segundo teste, os modelos CatBoost e RF mantiveram a melhora no desempenho como nos outros conjuntos de dados de participantes até 50 anos em relação ao conjunto completo e acima de 50 anos. Os resultados são exemplificados na Tabela 17 e a distribuição das estimativas apresentada na Tabela 18.

Tabela 17 – Resultados dos modelos com participantes com até 50 anos.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	0.592639	8.170284	19.809259	0.861991	6.574956
CatBoost	-0.175254	7.691031	14.887585	0.922049	6.267183

Fonte: Autor.

Figura 25 – Gráfico de Clarke das predições de glicose (até 50 anos).



Fonte: Autor.

Tabela 18 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	810	64	43	0	0
CatBoost	542	51	24	0	0

Fonte: Autor.

### 5.3.3 Conjunto para Participantes com Mais de 50 anos

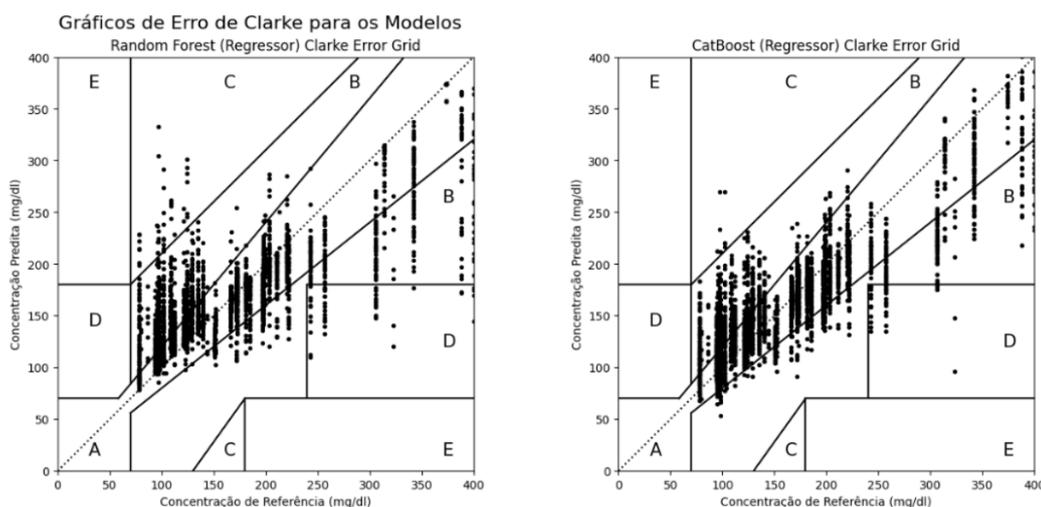
No terceiro teste, os modelos RF e CatBoost continuaram com desempenho ligeiramente superior ao conjunto sem divisão por idade. Os resultados estão detalhados na Tabela 19, enquanto a Tabela 20 apresenta a distribuição das estimativas.

Tabela 19 – Resultados dos modelos com participantes com mais de 50 anos.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	1.206903	29.841420	45.014684	0.621293	20.193271
CatBoost	-0.653477	26.942306	37.659775	0.734936	18.078713

Fonte: Autor.

Figura 26 – Gráfico de Clarke das predições de glicose (com mais de 50 anos).



Fonte: Autor.

Tabela 20 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	1587	715	754	0	0
CatBoost	1483	826	647	0	0

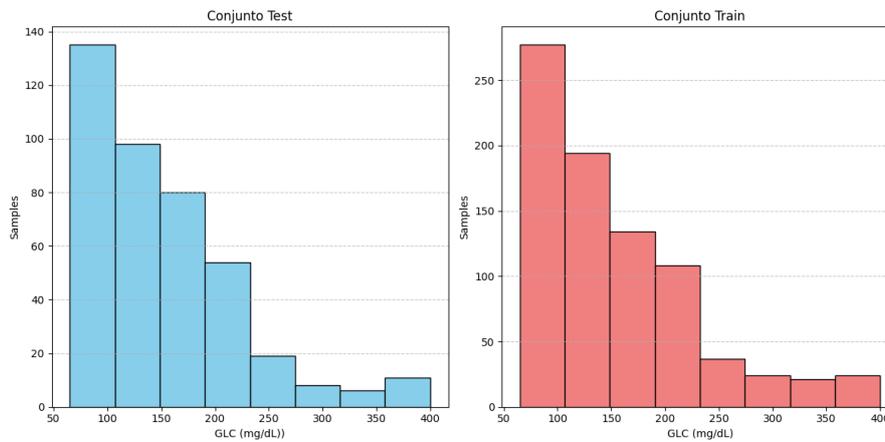
Fonte: Autor.

Os resultados confirmam que a aplicação da EMD contribui para melhorar a precisão dos modelos, especialmente o CatBoost e o RF, em todos os cenários. Esses modelos apresentaram maior estabilidade e precisão, especialmente para participantes até 50 anos.

## 5.4 Treinamento Utilizando Agrupamento de 10 Segmentos com a incorporação da EMD

Nesta seção, apresentamos os resultados do treinamento utilizando agrupamento de 10 segmentos extraídos dos sinais de PPG de cada paciente com a incorporação da EMD. Na Figura 27, está ilustrada a distribuição das amostras nos conjuntos de treino e teste.

Figura 27 – Histograma da distribuição de valores de glicose (mg/dL) nos conjuntos de teste e treino para os segmentos dos pacientes agrupados com a incorporação da EMD.



Fonte: Autor.

#### 5.4.1 Conjunto Sem Divisão por Idade

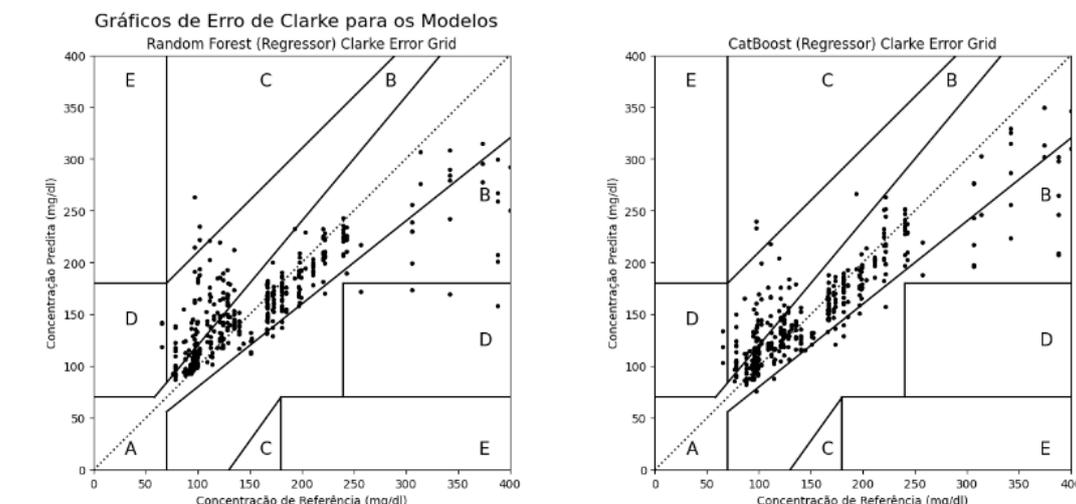
No primeiro teste, utilizando o conjunto de dados completo, o modelo CatBoost teve um melhor desempenho em relação ao RF. Os resultados detalhados estão na Tabela 21, enquanto a Tabela 22 apresenta a distribuição das estimativas nas diferentes zonas do modelo de erro de Clarke.

Tabela 21 – Resultados dos modelos com o conjunto completo (sem divisão por idade).

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	3.557397	25.455791	40.155514	0.662157	18.435115
CatBoost	0.360500	21.692200	33.902684	0.759180	15.588881

Fonte: Autor.

Figura 28 – Gráfico de Clarke das previsões de glicose (sem divisão por idade).



Fonte: Autor.

Tabela 22 – Resultado da distribuição das previsões utilizando o modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	246	89	76	0	0
CatBoost	256	96	59	0	0

Fonte: Autor.

### 5.4.2 Conjunto para Participantes com até 50 anos

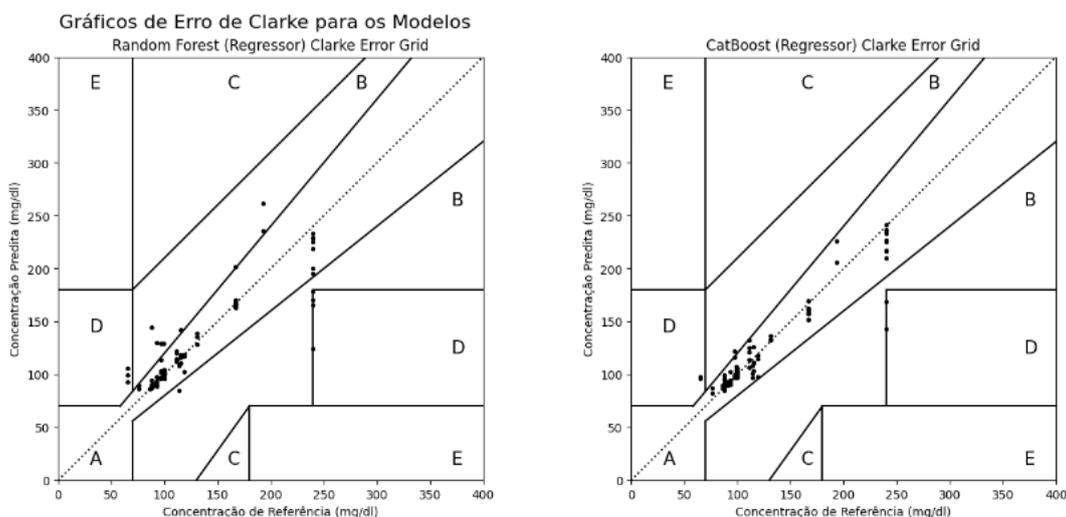
No segundo teste, o desempenho dos modelos continuaram melhores em relação ao conjunto completo, mas sendo inferiores ao resultados dos conjuntos sem agrupamento de segmentos. Os resultados estão na Tabela 23, com a distribuição das estimativas apresentada na Tabela 24.

Tabela 23 – Resultados dos modelos com participantes com até 50 anos.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	-0.520297	10.911584	23.066138	0.777048	8.062696
CatBoost	-1.016935	8.548553	16.084784	0.891584	6.776666

Fonte: Autor.

Figura 29 – Gráfico de Clarke das predições de glicose (até 50 anos).



Fonte: Autor.

Tabela 24 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	82	10	9	0	0
CatBoost	89	10	2	0	0

Fonte: Autor.

### 5.4.3 Conjunto para Participantes com Mais de 50 anos

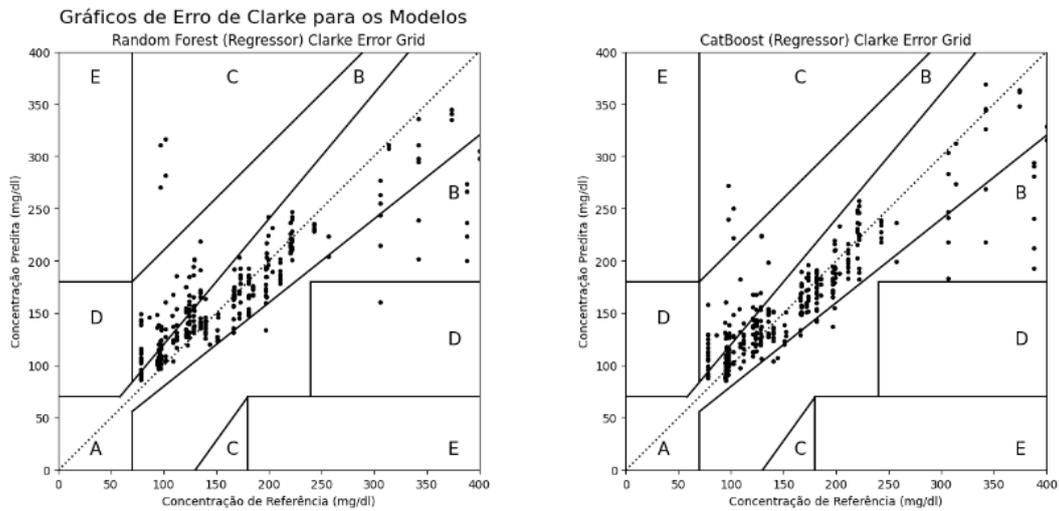
No terceiro teste, os modelos obtiveram um desempenho melhor em relação ao conjunto sem a divisão de idade, mas inferior ao desempenho dos modelos sem agrupamento de segmentos. Os resultados estão detalhados na Tabela 25, enquanto a Tabela 26 apresenta a distribuição das estimativas.

Tabela 25 – Resultados dos modelos com participantes com mais de 50 anos.

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	2.179581	25.957387	42.670204	0.647408	17.949883
CatBoost	0.734466	22.893983	37.040374	0.734936	15.805179

Fonte: Autor.

Figura 30 – Gráfico de Clarke das predições de glicose (com mais de 50 anos).



Fonte: Autor.

Tabela 26 – Resultado da distribuição das predições utilizando o modelo de erro de Clarke.

Modelo	A	B	C	D	E
RF	177	86	47	0	0
CatBoost	202	67	41	0	0

Fonte: Autor.

O treinamento utilizando agrupamento de 10 segmentos com EMD revelou a consistência dos modelos CatBoost e RF em todos os cenários analisados. Ambos apresentaram alta precisão, especialmente no conjunto sem divisão por idade e no grupo de participantes com menos de 50 anos. O uso da EMD reforça a robustez dos modelos CatBoost e RF para predições mais confiáveis.

## 6 Discussão

O Capítulo de Discussão visa apresentar uma análise crítica dos resultados obtidos neste estudo, conforme descrito no capítulo anterior. O estudo utilizou o conjunto de dados UFSC-BGL, fornecido por (DEUS, Luis Felipe de, 2023) para análise comparativa em sua tese de mestrado.

O conjunto de dados UFSC-BGL inclui 92 indivíduos, dos quais 45 apresentam níveis de glicose alterados e 47 com níveis considerados normais. Ao analisar os resultados, destaca-se que 22 participantes utilizam insulina para controle glicêmico, sendo geralmente a opção para casos mais severos. Outro ponto preocupante foi que 5 participantes com níveis de glicose alterados não estavam cientes de sua possível condição de diabetes. A falta de conscientização pode resultar em sérios riscos, visto que a diabetes necessita de monitoramento e controle constantes.

A conscientização sobre a condição de glicemia alterada é crucial para o tratamento eficaz e a prevenção de complicações associadas à diabetes. A análise dos dados deve considerar não apenas os níveis de glicose, mas também fatores clínicos e comportamentais, oferecendo uma abordagem mais completa no controle da doença. Este estudo reforça a importância da detecção precoce e do monitoramento contínuo da glicose, aliados à educação sobre os riscos da doença e as opções de tratamento disponíveis.

Após as etapas de processamento e filtragem, foi realizada a extração das características do sinal, abrangendo aspectos morfológicos e informações nos domínios da frequência e não linear. Na extração morfológica, foi adotada a soma dos pulsos de um segmento e o pulso médio como representação, visando estabilidade ao modelo. Como ilustrado na Figura 10, observam-se pequenas variações em cada pulso de PPG, e essas diferenças podem tornar os modelos de inteligência artificial suscetíveis a flutuações mínimas. Assim, a média dos pulsos foi utilizada para suavizar essas variações e melhorar a robustez do modelo. Características nos domínios da frequência e não linear também foram extraídas de segmentos de 10 segundos, com o intuito de adicionar informações complementares ao modelo, além do domínio do tempo métodos explorados por Luis.

Além disso, a aplicação de métodos avançados, como HRV e a EMD, desempenha um papel fundamental na compreensão do funcionamento fisiológico, permitindo uma análise mais precisa. Esses métodos têm grande aplicação em monitoramento da saúde e diagnóstico clínico, contribuindo significativamente para a avaliação e o acompanhamento de condições de saúde.

Com base nas características extraídas Apêndice (A), foram realizados testes iniciais com os dois modelos de aprendizado de máquina utilizados por (DEUS, Luis Felipe de, 2023): CatBoost e Random Forest.

Seus resultados obtidos nos modelos treinados são apresentados na Tabela 27, mostraram um erro médio absoluto variando entre 18 e 20 mg/dL e um erro percentual absoluto médio entre 12% e 14%. É importante observar que esses valores de erro são ampliados devido à ampla faixa de valores de glicose nos dados, variando entre 60 mg/dL e 350 mg/dL. Além disso, a distribuição das amostras não é uniforme em todas as faixas de glicose, o que pode impactar a precisão do modelo. Por exemplo, uma diferença de 100 mg/dL entre um valor de referência de 350 mg/dL e um valor estimado de 250 mg/dL pode ser considerada semelhante, dado a falta de amostras suficientes em determinadas faixas.

Tabela 27 – Resultados dos modelos obtidos por (DEUS, Luis Felipe de, 2023).

Modelo	ME (mg/dL)	MAE (mg/dL)	RMSE (mg/dL)	r2 (p)	MAPE (%)
RF	1.23	20.00	36.98	0.57	13.66
CatBoost	2.04	17.96	33.85	0.64	11.91

Fonte: (DEUS, Luis Felipe de, 2023).

O método utilizado para treinar esses modelos foi a média aritmética de cada segmento de paciente, visando reduzir a dispersão do sinal. No entanto, essa abordagem limitou o número de amostras disponíveis para treinamento, o que pode ter impactado a capacidade de generalização dos modelos. Isso sugere que uma estratégia alternativa, que considere melhor a variabilidade dos dados, pode ser necessária para aprimorar os resultados.

Com base nestes resultados, o presente trabalho apresentou quatro abordagens complementares aos estudos anteriores para melhorar os resultados utilizando o conjunto de dados UFSC-BGL. Devido à limitação no número de amostras, em função da quantidade reduzida de participantes, foi adotada uma estratégia diferente na construção das features nos modelos. Em vez de calcular a média dos segmentos dos pacientes, consideraram-se todos os segmentos de cada paciente. Além disso, os dados foram divididos aleatoriamente em 30% para o conjunto de teste e 70% para o conjunto de treino. Essa abordagem visou aumentar a quantidade de dados utilizados no treinamento, explorando melhor a variabilidade do sinal e melhorando, potencialmente, a performance dos modelos.

Na primeira abordagem, os resultados (Tabela 3) mostraram desempenho ligeiramente inferior ao de (DEUS, Luis Felipe de, 2023) no MAE, com valores variando entre

30,4% a 30,9%, e no MAPE, que ficou entre 21,7% e 21,9%. Contudo, o coeficiente de correlação de Pearson (R) demonstrou uma melhora significativa no modelo de CatBoost, com incremento de 0,11.

Ao dividir os dados por faixa etária, criando dois subconjuntos até 50 anos e acima de 50 anos, observou-se que os resultados para pessoas com mais de 50 anos foram semelhantes aos anteriores em relação ao MAPE e RMSE, com uma leve melhora no coeficiente de correlação de Pearson (R).

Por outro lado, para pacientes com até 50 anos, os resultados corroboraram a hipótese sobre a influência da idade nas curvas do sinal de PPG para estimativa do nível de glicose. Conforme apresentado na Tabela 5, os modelos CatBoost e Random Forest mostraram desempenho superior nas métricas avaliadas em comparação aos resultados de Luis, destacando-se o modelo Random Forest com um MAPE de 7,11% e o modelo CatBoost com coeficiente R de Pearson de 0,902.

Para analisar as respostas dos modelos à redução da quantidade de amostras, foi realizado um agrupamento de 10 amostras por média aritmética, resultando em 20 segmentos para cada paciente. Para o conjunto completo, os resultados não mostraram uma melhora ou piora significativa em relação aos obtidos anteriormente. Para pacientes com mais de 50 anos (Tabela 11), houve um leve aumento nos valores de MAE e MAPE, com uma pequena redução no coeficiente de correlação de Pearson (R). Já para pacientes com até 50 anos, houve uma leve piora em comparação aos resultados anteriores, embora ainda superassem os valores de Luis.

Uma abordagem adicional testada foi a incorporação da Decomposição Empírica por Modos (EMD), técnica utilizada para analisar sinais não estacionários e não lineares. Essa metodologia decompõe o sinal em uma série de componentes chamados Modos Intrínsecos de Função (IMFs), extraídos de cada segmento dos pacientes e utilizados no treinamento dos modelos.

Para o conjunto completo, houve uma leve melhora nos resultados, conforme apresentado na Tabela 15, com destaque para o coeficiente de correlação de Pearson (R). Já no subconjunto de pacientes com mais de 50 anos (Tabela 19), observou-se uma melhora significativa em comparação aos resultados anteriores, com valores de MAPE e R superiores aos obtidos por (DEUS, Luis Felipe de, 2023).

Os melhores resultados foram obtidos para pacientes com até 50 anos, com destaque para o modelo CatBoost, que apresentou MAE de 7,69%, MAPE de 6,28% e coeficiente de correlação de Pearson (R) de 0,922.

## 7 Conclusão

Este estudo apresentou uma abordagem inovadora para a estimativa não invasiva dos níveis de glicose, utilizando sinais de Fotopletismografia (PPG) combinados com técnicas avançadas de aprendizado de máquina. A pesquisa foi conduzida utilizando o novo conjunto de dados **UFSC-BGL**, desenvolvido especificamente para este projeto, composto por 92 participantes que contribuíram para uma análise detalhada e robusta.

Os resultados evidenciaram que o modelo **CatBoost** apresentou o melhor desempenho, alcançando um erro médio absoluto (MAE) de 7,69 mg/dL, um percentual absoluto médio (MAPE) de 6,27%, e 73% das estimativas localizadas na zona A do gráfico de Clarke. Esses números destacam a eficácia do modelo na interpretação dos sinais PPG para a estimativa glicêmica, demonstrando seu potencial para aplicações práticas.

Além disso, a pesquisa enfatizou a importância de características extraídas dos sinais PPG, incluindo informações morfológicas, não lineares e no domínio da frequência, que se mostraram fundamentais para o desempenho dos modelos. Estratégias complementares, como a segmentação etária e a análise detalhada de variabilidades nos sinais, mostraram-se eficazes, especialmente para o grupo de pacientes com até 50 anos, onde os melhores resultados foram obtidos.

A análise crítica dos resultados revelou desafios relevantes, como a variabilidade dos sinais entre indivíduos e a necessidade de amostras mais balanceadas em diferentes faixas de glicose. Apesar disso, a abordagem adotada demonstrou ser robusta, com resultados que indicam grande aplicabilidade no contexto clínico. Este trabalho estabelece uma base sólida para o desenvolvimento de métodos não invasivos para o monitoramento contínuo de glicose, contribuindo para a evolução de soluções tecnológicas voltadas ao diagnóstico e acompanhamento da diabetes.

Sugere-se, como perspectivas futuras, ampliar a coleta e diversificação de dados para abranger uma maior variabilidade de participantes, explorar novas técnicas de extração de características, incluindo métodos avançados de análise nos domínios da frequência e não linear, realizar validações práticas em cenários clínicos para garantir a aplicabilidade e eficácia dos modelos propostos, além de investir na integração do método com dispositivos de monitoramento portáteis, visando sua utilização em larga escala.

Os avanços obtidos por este estudo reforçam o potencial da fotopletismografia como ferramenta promissora para estimativa glicêmica não invasiva, contribuindo para um cuidado acessível e confortável para pacientes que necessitam de monitoramento contínuo.

# Referências

- ALIAN, A. A.; SHELLEY, K. H. Photoplethysmography. Best Practice & Research Clinical Anaesthesiology, Elsevier, v. 28, n. 4, p. 395–406, 2014. 24
- ALLEN, J. Photoplethysmography and its application in clinical physiological measurement. Physiological measurement, IoP Publishing, v. 28, n. 3, p. R1, 2007. 25, 27, 47
- BANIK, P. P. et al. Development of a wearable reflection-type pulse oximeter system to acquire clean ppg signals and measure pulse rate and spo2 with and without finger motion. Electronics, MDPI, v. 9, n. 11, p. 1905, 2020. 26
- BARRETT, E. J. et al. Diabetic microvascular disease: an endocrine society scientific statement. The Journal of Clinical Endocrinology & Metabolism, Endocrine Society Washington, DC, v. 102, n. 12, p. 4343–4410, 2017. 41
- BOGDAN, M.; PANU, M. Labview modeling and simulation, of the low-pass and high-pass analog filters. In: IEEE. 2015 13th International Conference on Engineering of Modern Electric Systems (EMES). [S.l.], 2015. p. 1–4. 29, 30
- BREIMAN, L. Random forests. Machine learning, Springer, v. 45, p. 5–32, 2001. 51
- BUTTERWORTH, S. et al. On the theory of filter amplifiers. Wireless Engineer, v. 7, n. 6, p. 536–541, 1930. 28, 29
- CATAI, A. M. et al. Heart rate variability: are you using it properly? standardisation checklist of procedures. Brazilian journal of physical therapy, Elsevier, v. 24, n. 2, p. 91–102, 2020. 56
- CHARLTON, P. H. et al. Detecting beats in the photoplethysmogram: benchmarking open-source algorithms. Physiological Measurement, IOP Publishing, v. 43, n. 8, p. 085007, 2022. 27, 48
- CLARKE, W. L. et al. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes care, Am Diabetes Assoc, v. 10, n. 5, p. 622–628, 1987. 36, 37
- CLIFFORD, G. D. et al. Advanced methods and tools for ECG data analysis. [S.l.]: Artech house Boston, 2006. v. 10. 41
- COMMITTEE, A. D. A. P. P.; COMMITTEE, A. D. A. P. P. 2. classification and diagnosis of diabetes: Standards of medical care in diabetes—2022. Diabetes care, Am Diabetes Assoc, v. 45, n. Supplement\_1, p. S17–S38, 2022. 18
- COSSUL, S. et al. The use of empirical mode decomposition on heart rate variability signals to assess autonomic neuropathy progression in type 2 diabetes. Applied Sciences, MDPI, v. 13, n. 13, p. 7824, 2023. 41

- DEFRONZO, R. A.; BONADONNA, R. C.; FERRANNINI, E. Pathogenesis of niddm: a balanced overview. Diabetes care, Am Diabetes Assoc, v. 15, n. 3, p. 318–368, 1992. 17
- DEUS, Luis Felipe de. Estimativa dos níveis de glicose através do processamento de sinais de fotopletismografia (PPG) e modelos de inteligência artificial. Dissertação (Dissertação (Mestrado)) — Universidade Federal de Santa Catarina, Florianópolis, 2023. Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica. 7, 8, 12, 27, 43, 44, 45, 46, 47, 48, 51, 75, 76, 77
- DOROGUSH, A. V.; ERSHOV, V.; GULIN, A. Catboost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363, 2018. 52
- DUONG, M. T. et al. Artificial intelligence for precision education in radiology. The British journal of radiology, The British Institute of Radiology., v. 92, n. 1103, p. 20190389, 2019. 32
- ELECTROPHYSIOLOGY, T. F. o. t. E. S. o. C. t. N. A. S. o. P. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation, Am Heart Assoc, v. 93, n. 5, p. 1043–1065, 1996. 56
- ELGENDI, M. On the analysis of fingertip photoplethysmogram signals. Current cardiology reviews, Bentham science publishers, v. 8, n. 1, p. 14–25, 2012. 41, 56
- ESGALHADO, F. et al. Peak detection and hrv feature evaluation on ecg and ppg signals. Symmetry, MDPI, v. 14, n. 6, p. 1139, 2022. 41
- FEDERATION, I. D. Diabetes Atlas - 10th Edition. 2021. Accessed December 1, 2024. Disponível em: <<https://diabetesatlas.org>>. 19
- (FIOCRUZ), F. O. C. Diabetes: prevalência e desafios globais. 2023. Accessed December 1, 2024. Disponível em: <<https://portal.fiocruz.br/diabetes>>. 19
- GONTIJO, A. P. V. P. V. et al. Nefropatia diabética: fisiopatologia, diagnóstico e tratamento. Brazilian Journal of Health Review, v. 7, n. 4, p. e72220–e72220, 2024. 18
- GROSS, J. L. et al. Diabetes melito: diagnóstico, classificação e avaliação do controle glicêmico. Arquivos Brasileiros de Endocrinologia & Metabologia, SciELO Brasil, v. 46, p. 16–26, 2002. 19
- HUANG, N. E. et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, The Royal Society, v. 454, n. 1971, p. 903–995, 1998. 34, 35, 41
- KYRIACOU, P. A.; ALLEN, J. Photoplethysmography: technology, signal analysis and applications. [S.l.]: Academic Press, 2021. 24

- LI, H.; HUANG, S. A high-efficiency and real-time method for quality evaluation of ppg signals. In: IOP PUBLISHING. IOP Conference Series: Materials Science and Engineering. [S.l.], 2020. v. 711, n. 1, p. 012100. 45, 47, 48
- LIANG, Y. et al. An optimal filter for short photoplethysmogram signals. Scientific data, Nature Publishing Group, v. 5, n. 1, p. 1–12, 2018. 21
- LIANG, Y. et al. An optimal filter for short photoplethysmogram signals. Scientific data, Nature Publishing Group, v. 5, n. 1, p. 1–12, 2018. 27
- LOH, H. W. et al. Application of photoplethysmography signals for healthcare systems: An in-depth review. Computer Methods and Programs in Biomedicine, Elsevier, v. 216, p. 106677, 2022. 21
- LOUPPE, G. Understanding random forests: From theory to practice. arXiv preprint arXiv:1407.7502, 2014. 51
- LU, G. et al. A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects. Journal of medical engineering & technology, Taylor & Francis, v. 33, n. 8, p. 634–641, 2009. 34
- LU, L. et al. Uncertainties in the analysis of heart rate variability: A systematic review. IEEE Reviews in Biomedical Engineering, IEEE, v. 17, p. 180–196, 2023. 33
- MAGRABI, F. et al. Artificial intelligence in clinical decision support: challenges for evaluating ai and practical implications. Yearbook of medical informatics, Georg Thieme Verlag KG, v. 28, n. 01, p. 128–134, 2019. 31
- MARATHE, P. H.; GAO, H. X.; CLOSE, K. L. American Diabetes Association Standards of Medical Care in Diabetes 2017. [S.l.]: Wiley Online Library, 2017. 17
- NASIRI, N. Introductory chapter: Wearable technologies for healthcare monitoring. Wearable Devices Big Wave Innovation, IntechOpen: London, p. 3–10, 2019. 20
- NENSA, F.; DEMIRCIOGLU, A.; RISCHPLER, C. Artificial intelligence in nuclear medicine. Journal of Nuclear Medicine, Soc Nuclear Med, v. 60, n. Supplement 2, p. 29S–37S, 2019. 31
- NETER, J.; WASSERMAN, W.; KUTNER, M. H. Applied linear regression models. [S.l.]: Richard D. Irwin, 1983. 36
- PANKAJ et al. A review on computation methods used in photoplethysmography signal analysis for heart rate estimation. Archives of Computational Methods in Engineering, Springer, v. 29, n. 2, p. 921–940, 2022. 24
- PARK, J. et al. Photoplethysmogram analysis and applications: an integrative review. Frontiers in Physiology, Frontiers Media SA, v. 12, p. 808451, 2022. 25

- PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. the Journal of machine Learning research, JMLR. org, v. 12, p. 2825–2830, 2011. 32
- PŘIBIL, J.; PŘIBILOVÁ, A.; FROLLO, I. Comparative measurement of the ppg signal on different human body positions by sensors working in reflection and transmission modes. Engineering proceedings, MDPI, v. 2, n. 1, p. 69, 2020. 26
- PŘIBIL, J.; PŘIBILOVÁ, A.; FROLLO, I. Comparative measurement of the ppg signal on different human body positions by sensors working in reflection and transmission modes. Engineering proceedings, MDPI, v. 2, n. 1, p. 69, 2020. 26
- RAHMAN, M. H. et al. Predictive health analysis in industry 5.0: A scientometric and systematic review of motion capture in construction. Digital Engineering, Elsevier, p. 100002, 2024. 21
- RAJPURKAR, P. et al. Ai in health and medicine. Nature medicine, Nature Publishing Group US New York, v. 28, n. 1, p. 31–38, 2022. 31
- RASCHKA, S.; MIRJALILI, V. Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. [S.l.]: Packt publishing ltd, 2019. 32
- RILLING, G. et al. On empirical mode decomposition and its algorithms. In: GRADO: IEEE. IEEE-EURASIP workshop on nonlinear signal and image processing. [S.l.], 2003. v. 3, n. 3, p. 8–11. 35
- RODAKCI, M. et al. Diagnóstico de diabetes mellitus. Diretriz Oficial da Sociedade Brasileira de Diabetes (2024). [S.l.]: Sociedade Brasileira de Diabetes, 2024. ISBN 978-65-272-0704-7. 43
- SACKS, D. B.; KIRKMAN, M. S.; LITTLE, R. R. Point-of-care hba1c in clinical practice: Caveats and considerations for optimal use. Diabetes Care, Am Diabetes Assoc, p. dci230040, 2024. 19
- SANTOS, M. Classificação de complexo QRS em Sinais Cardíacos empregando abordagens baseadas na Decomposição Empírica de Modo. Dissertação (Mestrado), 2020. 35
- SATTER, S. et al. Emd-based noninvasive blood glucose estimation from ppg signals using machine learning algorithms. Applied Sciences, MDPI, v. 14, n. 4, p. 1406, 2024. 40
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, American Chemical Society, v. 36, n. 8, p. 1627–1639, 1964. 27, 45
- SCHLEICHER, E. et al. Definition, classification and diagnosis of diabetes mellitus. Experimental and Clinical Endocrinology & Diabetes, Georg Thieme Verlag KG, v. 130, n. S 01, p. S1–S8, 2022. 17

- SELESNICK, I. W.; BURRUS, C. S. Generalized digital butterworth filter design. IEEE Transactions on Signal Processing, IEEE, v. 46, n. 6, p. 1688–1694, 1998. 27, 45
- SHAFFER, F.; GINSBERG, J. P. An overview of heart rate variability metrics and norms. Frontiers in public health, Frontiers Media SA, v. 5, p. 258, 2017. 33
- SHAFFER, F.; MCCRATY, R.; ZERR, C. L. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Frontiers in psychology, Frontiers Media SA, v. 5, p. 1040, 2014. 56
- SHI, B. et al. Assessing elevated blood glucose levels through blood glucose evaluation and monitoring using machine learning and wearable photoplethysmography sensors: Algorithm development and validation. JMIR AI, JMIR Publications Inc., Toronto, Canada, v. 2, n. 1, p. e48340, 2023. 38
- TAMURA, T. et al. Wearable photoplethysmographic sensors—past and present. Electronics, MDPI, v. 3, n. 2, p. 282–302, 2014. 21
- TOPOL, E. Deep medicine: how artificial intelligence can make healthcare human again. [S.l.]: Hachette UK, 2019. 31
- VARGOVÁ, E.; NĚMCOVÁ, A.; NOVÁKOVÁ, Z. Non-invasive ppg-based estimation of blood glucose level. Lékař a technika-Clinician and Technology, v. 53, n. 1, p. 19–24, 2023. 39
- VASHIST, S. K. et al. Technology behind commercial devices for blood glucose monitoring in diabetes management: A review. Analytica chimica acta, Elsevier, v. 703, n. 2, p. 124–136, 2011. 20
- WILCOX, G. Insulin and insulin resistance. Clinical biochemist reviews, Australasian Association for Clinical Biochemistry and Laboratory Medicine, v. 26, n. 2, p. 19, 2005. 18
- WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. Climate research, v. 30, n. 1, p. 79–82, 2005. 36
- World Health Organization. Diabetes. 2023. Acesso em: 9 jun. 2024. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/diabetes>>. 18
- WU, J. et al. Improving the accuracy in classification of blood pressure from photoplethysmography using continuous wavelet transform and deep learning. International journal of hypertension, Wiley Online Library, v. 2021, n. 1, p. 9938584, 2021. 41, 56
- ZANELLI, S. et al. Diabetes detection and management through photoplethysmographic and electrocardiographic signals analysis: A systematic review. Sensors, MDPI, v. 22, n. 13, p. 4890, 2022. 21

# A Apêndice A

Tabela 28 – Atributos morfológicos extraídos dos pulsos do sinal de PPG.

Atributo	Descrição
peak_value	Valor da amplitude do pico.
peak_pos	Posição do pico no tempo.
rising_time	Tempo de subida do pulso.
skewness	Assimetria do pulso.
kurtosis_f	Kurtose pelo critério Fisher.
kurtosis_p	Kurtose pelo critério Pearson.
variance	Variância do pulso.
w75	Largura do pulso em 75% da amplitude.
w50	Largura do pulso em 50% da amplitude.
w25	Largura do pulso em 25% da amplitude.
w10	Largura do pulso em 10% da amplitude.
notch_y	Ponto de inflexão no pulso.
diastolic_peak_y	Posição do pico diastólico.
diastolic_slope	Inclinação da descida diastólica.
area	Área total do pulso.
area_1	Área antes do pico sistólico.
area_2	Área após o pico diastólico.
IPA	Índice de Proporção de Área ( $\text{area}_2 / \text{area}_1$ ).
A, B, C, D, E	Pontos derivados do pulso.
B/A	Razão entre os pontos B e A.
B/A_t	Razão temporal entre B e A.
C/A	Razão entre os pontos C e A.
C/A_t	Razão temporal entre C e A.
D/A	Razão entre os pontos D e A.
D/A_t	Razão temporal entre D e A.
E/A	Razão entre os pontos E e A.
E/A_t	Razão temporal entre E e A.
X	Diferença entre o pico sistólico e o final do pulso.
Y	Diferença entre o pico sistólico e o início do pulso.
AI	Augmentation Index: Razão entre X e Y.
RI	Reflection Index: Outra métrica de X e Y.
DT	Diferença absoluta entre os picos sistólico e diastólico.
DTt	Diferença temporal entre os picos sistólico e diastólico.
AGE	Idade do indivíduo.
GLC	Glicose do indivíduo.
ID	Identificação do indivíduo.
IP	Índice de pulso.
EP	Energia do pulso.

Tabela 29 – Atributos relacionados à variabilidade da frequência cardíaca (HRV) extraídos dos pulsos do sinal de PPG.

<b>Atributo</b>	<b>Descrição</b>
MAX_POWER_1st	Máxima potência na primeira frequência dominante.
IDX_MAX_1st	Índice da máxima potência na primeira frequência.
F_MAX_1st	Primeira frequência dominante.
MAX_POWER_2nd	Máxima potência na segunda frequência dominante.
IDX_MAX_2nd	Índice da máxima potência na segunda frequência.
F_MAX_2nd	Segunda frequência dominante.

Tabela 30 – Atributos derivados da decomposição em modos empíricos (EMD) extraídos dos pulsos do sinal de PPG.

<b>Atributo</b>	<b>Descrição</b>
IMF_1_mean	Média do sinal IMF 1.
IMF_1_var	Variância do sinal IMF 1.
IMF_1_std	Desvio padrão do sinal IMF 1.
IMF_1_ptp	Diferença pico a pico do IMF 1.
IMF_1_skewness	Assimetria do IMF 1.
IMF_1_kurtosis	Kurtose do IMF 1.
IMF_1_dom_freq	Frequência dominante do IMF 1.
IMF_1_total_power	Potência total do IMF 1.
IMF_1_ae_mean	Média da amplitude instantânea do IMF 1.
IMF_1_if_mean	Média da frequência instantânea do IMF 1.
IMF_1_zcr	Taxa de cruzamento por zero do IMF 1.
IMF_1_extrema	Número de extremos no IMF 1.
IMF_1_psd_mean	Média da densidade espectral do IMF 1.
IMF_1_psd_var	Variância da densidade espectral do IMF 1.
IMF_1_sc	Centroide espectral do IMF 1.
IMF_1_se	Energia espectral do IMF 1.
IMF_1_sf	Frequência espectral do IMF 1.