



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE DO CAMPUS ARARANGUÁ  
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Lucas Eduardo Schlenert de Oliveira

**Investigando o Uso de Aprendizado Federado na Detecção e Controle de  
Diabetes**

Araranguá  
2024

Lucas Eduardo Schlenert de Oliveira

**Investigando o Uso de Aprendizado Federado na Detecção e Controle de  
Diabetes**

Trabalho de Conclusão de Curso do Curso de Graduação em Engenharia de Computação submetido ao Centro de Ciências, Tecnologias e Saúde do Campus Araranguá da Universidade Federal de Santa Catarina para a obtenção do título de Bacharel em Engenharia de Computação.

Orientadora: Profa. Analúcia Schiaffino Morales, Dra.

Araranguá

2024

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

Oliveira, Lucas Eduardo Schlenert de  
Investigando o Uso de Aprendizado Federado na Detecção e  
Controle de Diabetes / Lucas Eduardo Schlenert de  
Oliveira ; orientadora, Analúcia Schiaffino Morales, 2024.  
30 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Campus Araranguá,  
Graduação em Engenharia de Computação, Araranguá, 2024.

Inclui referências.

1. Engenharia de Computação. 2. Aprendizado Federado. 3.  
Diabetes. I. Morales, Analúcia Schiaffino . II.  
Universidade Federal de Santa Catarina. Graduação em  
Engenharia de Computação. III. Título.

Lucas Eduardo Schlenert de Oliveira

**Investigando o Uso de Aprendizado Federado na Detecção e Controle de Diabetes**

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

Araranguá, 13 de Dezembro de 2024.

---

Prof. Jim Lau, Dr.  
Coordenador do Curso

**Banca Examinadora:**

---

Profa. Analúcia Schiaffino Morales, Dra.  
Orientadora

---

Prof. Jim Lau, Dr.  
Avaliador  
Universidade Federal de Santa Catarina

---

Prof. Alison Roberto Panisson, Dr.  
Avaliador  
Universidade Federal de Santa Catarina

# Investigando o Uso de Aprendizado Federado na Detecção e Controle de Diabetes

Lucas Eduardo Schlenert de Oliveira\*

2024, DEZEMBRO

## Resumo

A redução da mortalidade por doenças crônicas, como a diabetes, é um dos desafios da Agenda 2030 da OMS. O diagnóstico precoce é uma abordagem promissora, pois, em muitos casos, o tratamento na fase inicial da doença pode reverter a situação sem grandes impactos na vida do paciente. Este trabalho investiga o uso de aprendizado federado na detecção e monitoramento de diabetes por biomarcadores fisiológicos. Utilizou-se um conjunto de métodos para seleção de dados, pré-processamento e treinamento de algoritmos, incluindo redes neurais e florestas randômicas, avaliados com métricas de desempenho. Os resultados indicaram alta precisão (0.999) para classificação, demonstrando a eficácia do modelo global, além de comprovada a possibilidade de personalização dos modelos locais. Conclui-se que o aprendizado federado é promissor para aplicações na saúde, apesar da necessidade de dados mais robustos e generalizáveis.

**Palavras-chaves:** Diabetes, Aprendizado Federado.

---

\*engcomp.lucasedu@gmail.com

# Investigating the Use of Federated Learning in Diabetes Detection and Management

Lucas Eduardo Schlenert de Oliveira\*

2024, DECEMBER

## Abstract

The reduction of mortality from chronic diseases, such as diabetes, is one of the challenges outlined in the WHO's 2030 Agenda. Early diagnosis is a promising approach, as in many cases, treatment during the initial stages of the disease can reverse the condition without significantly impacting the patient's life. This study investigates the use of federated learning for the detection and monitoring of diabetes through physiological biomarkers. A range of methods was employed for data selection, preprocessing, and algorithm training, including neural networks and random forests, which were evaluated using performance metrics. The results indicated high accuracy (0.999) in classification, demonstrating the effectiveness of the global model, as well as the feasibility of personalizing local models. It is concluded that federated learning is a promising approach for healthcare applications, despite the need for more robust and generalizable data.

**Key-words:** Diabetes, Federated Learning.

---

\*engcomp.lucasedu@gmail.com

# 1 Introdução

Com a evolução da medicina e da tecnologia, reduziu-se consideravelmente a incidência de algumas doenças, chegando a extingui-las. Contudo, o mesmo avanço acabou propiciando um ambiente favorável para outros tipos de doenças, muitas vezes associado ao desenvolvimento de um estilo de vida mais sedentário. Dessas, destacam-se as doenças crônicas, as quais tendem a afetar a vida do portador por um longo período de tempo e são retratadas como um desafio principal para o desenvolvimento sustentável da Agenda 2030 da Organização Mundial da Saúde (OMS)<sup>1</sup>, sendo elas responsáveis por cerca de 74% dos óbitos anuais, lideradas por, respectivamente, doenças cardiovasculares, câncer, condições respiratórias e diabetes.

Nesse viés, ressalta-se a *diabetes mellitus*, caracterizada pela incapacidade do corpo de produzir ou utilizar adequadamente a insulina, possuindo três grupos principais: diabetes tipo 1, diabetes tipo 2 e diabetes gestacional (RODRIGUEZ-LEÓN et al., 2021). De acordo com o portal da Sociedade Brasileira de Diabetes<sup>2</sup>, os fatores de risco que impulsionam o desenvolvimento da doença incluem: tabagismo, consumo de álcool, obesidade e hipertensão, sendo mais prevalentes em homens e, dentre as maiores consequências da falta de acompanhamento da doença, destacam-se cegueira, necrose e doenças cardiovasculares (BIESSELS et al., 2006). Em 2023, o Brasil registrou cerca de 30 milhões de atendimentos nas Unidades Básicas de Saúde, um número alarmante que reflete a gravidade da situação. Ademais, em 2022, segundo o Sistema de Informações sobre Mortalidade (SIM), foram registrados aproximadamente 80 mil óbitos relacionados à doença no país<sup>3</sup>.

Os impactos na saúde pública elevam os altos custos financeiros globalmente. Em 2022, nos Estados Unidos, os gastos relacionados ao tratamento de diabetes superaram os US\$ 400 bilhões, incluindo custos diretos, como medicamento, consulta e internação, e custos indiretos<sup>4</sup>. No entanto, esse número pode crescer ainda mais nas próximas décadas, impulsionado por fatores como o envelhecimento mundial da população e o crescimento do sedentarismo, o qual tem contribuído significativamente para o aumento da prevalência da diabetes, prevendo-se um impacto financeiro ainda mais severo para sistemas de saúde ao redor do mundo caso tal quadro não seja combatido. Existe a necessidade de elaborar propostas preventivas para evitar a sobrecarga de procedimentos e custos adicionais com o aumento do número de idosos (VERAS, 2023). Além de enfrentar barreiras socioeconômicas, que incluem a idade avançada e a situação financeira, as quais dificultam abordagens não medicamentosas como exercícios físicos e dieta que auxiliam no controle da doença (OLIVEIRA; AMÂNCIO; SILVA, 2024).

A carga financeira e social da doença não é exclusiva dos governos (BREITSCHEIDEL S. STAMENITIS; SCHÖFFSKI, 2010). As famílias também sofrem com os custos relacionados aos medicamentos, as consultas e internações, além de impactos econômicos indiretos, como a sobrecarga tributária. Diante disso, evidencia-se a necessidade de buscar soluções que previnam o agravamento da doença e reduzam os custos associados (BUTT et al., 2024). A prevenção e o controle contínuo são, sem dúvida, mais acessíveis do que o tratamento de complicações severas (NILSON et al., 2020). Neste sentido, smartphones e *wearables* têm o potencial de revolucionar o cuidado de pacientes com diabetes,

<sup>1</sup> <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>

<sup>2</sup> <https://diabetes.org.br/>

<sup>3</sup> <https://www.gov.br/saude/pt-br/assuntos/noticias/2024/junho/dia-nacional-do-diabetes-cerca-de-30-milhoes-de-atendimentos-foram-realizados-em-2023>

<sup>4</sup> <https://diabetes.org/newsroom/press-releases/new-american-diabetes-association-report-finds-annual-costs-diabetes-be>

proporcionando um monitoramento mais preciso e personalizado, além de promover a autonomia e o engajamento dos pacientes no próprio cuidado. A presença de sensores que podem ser utilizados para monitorar parâmetros relevantes para pessoas com diabetes, como glicose no sangue, pressão arterial, calorias, atividade física, condição dos pés e olhos, e níveis de estresse. Esses dispositivos podem realizar esse monitoramento de forma contínua, passiva e não invasiva, sem interferir significativamente na rotina diária dos usuários (RODRIGUEZ-LEÓN et al., 2021).

Associado a esta tecnologia de sensores, a Inteligência Artificial (IA) surge como uma ferramenta promissora, com potencial para auxiliar no monitoramento e a gestão da diabetes. A IA oferece a capacidade de processar grandes volumes de dados, identificar padrões e prever complicações, possibilitando um tratamento mais eficaz (KRAKAUER; MALERBI, 2024). No entanto, a adoção dessa tecnologia no setor de saúde ainda enfrenta barreiras substanciais. Entre elas, estão o tratamento de dados sensíveis, como informações médicas confidenciais, e a dispersão de dados entre diferentes instituições, fenômeno conhecido como "ilhas de dados". Além disso, a volumetria e a diversidade dos dados gerados por dispositivos médicos conectados tornam o processamento em larga escala um desafio técnico e regulatório, especialmente no contexto da Lei Geral de Proteção de Dados (LGPD). Adicionalmente, existe o desafio da personalização dos resultados, cada pessoa com diabetes é única, e as recomendações devem ser individualizadas, de acordo com o tipo de diabetes, a idade, o peso e outras condições de saúde (RODRIGUEZ-LEÓN et al., 2021).

Assim sendo, torna-se especialmente necessária a aplicação de modelos descentralizados, no intuito de obter dados de múltiplas fontes, tornando o modelo mais robusto e assegurando a privacidade. Nesse contexto, o aprendizado federado (AF) apresenta-se como uma solução viável para muitos desses desafios (DHADE; SHIRKE, 2023). A tecnologia permite que dispositivos vestíveis realizem o treinamento de modelos de IA de forma local, sem a necessidade de transferir todos os dados sensíveis para servidores centralizados. Isso não só preserva a privacidade dos pacientes, conforme as exigências da LGPD, mas também possibilita a escalabilidade do uso da IA no tratamento de doenças crônicas, como a diabetes. Assim, o AF pode ser uma ferramenta crucial no desenvolvimento de uma abordagem mais eficiente e segura para o monitoramento e controle da diabetes.

O objetivo principal do trabalho é investigar o modelo de aprendizado federado e validar a aplicação para identificação da doença através de sinais fisiológicos, ou biomarcadores. Para isso, foram investigados *datasets* e testados alguns modelos de aprendizado de máquinas, incluindo o aprendizado federado, discutindo suas vantagens, desvantagens e possibilidades de personalização. O presente trabalho está organizado da seguinte forma: a Seção 2 apresenta um conjunto de trabalhos relacionados, a Seção 3 descreve os materiais e métodos utilizados no experimento, a Seção 4 apresenta os resultados obtidos, a Seção 5 discute as implicações dos achados. Por fim, são apresentadas as conclusões, trabalhos futuros e as referências utilizadas no desenvolvimento do trabalho.

## 2 Fundamentação Teórica

Diabetes Mellitus<sup>5</sup> é uma doença crônica com três ramificações principais: Tipo 1, Tipo 2 e Gestacional. Embora menos comum, a Tipo 1 é a segunda mais prevalente e está fortemente relacionada à genética. Esta forma da doença é caracterizada pela redução

<sup>5</sup> <https://bvsmms.saude.gov.br/diabetes/>



da produção de insulina pelo corpo, necessitando de injeções regulares do hormônio para regular os níveis de glicose. Em contraste, a Tipo 2 é a mais comum e pode se desenvolver devido a uma combinação de fatores sociais, físicos e emocionais. Neste caso, o corpo começa a desenvolver resistência à insulina e, eventualmente, a produção do hormônio é afetada, requerendo mudanças no estilo de vida do portador. A Diabetes Gestacional, por sua vez, está relacionada à incapacidade do corpo de produzir insulina suficiente para lidar com o aumento dos níveis glicêmicos decorrentes das mudanças hormonais durante a gravidez (ALOTAIBI; ALBALAWI, 2018; CASTRO et al., 2021; GHOZALI, 2024).

Além desses tipos, existe o estado pré-diabético, onde o índice glicêmico está alterado, mas não o suficiente para ser considerado diabetes propriamente dito (STANDARDS . . ., 2014). De acordo com as informações do portal do Ministério da Saúde<sup>6</sup>, em média, um em cada dois pré-diabéticos desenvolve a doença, muitas vezes devido à despreocupação ou falta de monitoramento, sendo esse o maior desafio ao lidar com a diabetes e outras doenças silenciosas. Muitos portadores desconhecem sua condição e, frequentemente, o diagnóstico é tardio, resultando em sequelas inevitáveis. Por isso, há duas frentes distintas quando se aborda soluções para diabetes: diagnóstico e monitoramento (RODAKI et al., 2024).

## 2.1 Diagnóstico vs Monitoramento

Ambas as vertentes são essenciais para o manejo eficaz da doença, contudo, cada uma representa uma perspectiva e abordagem distinta (COMMITTEE, 2024). Diagnóstico é um evento pontual, que define a qual grupo o indivíduo se enquadra, enquanto o monitoramento envolve o acompanhamento contínuo da evolução da doença e do tratamento, fazendo ajustes conforme necessário. Assim como há uma diferença na representatividade, também há nas variáveis coletadas e nas métricas utilizadas (RODAKI et al., 2024).

O diagnóstico tende a ser realizado por intermédio de exames laboratoriais, como hemoglobina glicada (HbA1c), teste oral de tolerância a glicose e glicemia em jejum. Por vezes, esses exames são acompanhados por outros dados, como espessura da pele do tríceps (indício de obesidade), número de gestações, e função pedigree de diabetes (cálculo que considera o histórico familiar de incidência da doença) (GROSS et al., 2002). Esses exames laboratoriais almejam analisar a reação do corpo em diferentes circunstâncias relacionadas à diabetes, mas acabam por não explorar completamente as nuances do corpo no cotidiano (RODAKI et al., 2024).

Em contrapartida, o monitoramento é realizado de maneira contínua, coletando parâmetros fisiológicos e avaliando o que o conjunto representa, geralmente mediante o uso de sensores apropriados (ZOU et al., 2024). Esses sensores podem incluir monitores contínuos de glicose (CGM), que fornecem leituras constantes dos níveis glicêmicos no sangue, sensores de frequência cardíaca e dispositivos que monitoram a pressão arterial e os níveis de oxigênio no sangue (SpO2) (BOLLIMUNTHA; MURUGAN, 2023). A utilização desses dispositivos permite ajustes terapêuticos em tempo real, promovendo uma abordagem proativa na prevenção de complicações relacionadas à diabetes. Além disso, o monitoramento contínuo de parâmetros fisiológicos oferece uma visão detalhada das variações diárias dos indicadores de saúde, permitindo uma gestão mais precisa e personalizada do tratamento, incluindo aspectos como o controle do sono, atividades diárias e do estresse, fornecendo um suporte mais abrangente para pacientes com diabetes (ZOU et al., 2024).

---

<sup>6</sup> <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/diabetes>

## 2.2 Tratamento de dados

Em um cenário ideal, os dados utilizados para treinamentos são balanceados e não ruidosos, porém dados reais não funcionam assim, principalmente em cenários que envolvam pessoas. Por vezes o padrão de coleta não foi seguido a risca, houve falhas na coleta, ou mesmo a coleta foi feita de forma tendenciosa, gerando um desbalanceamento de classes, ou mesmo induzindo a presença de outliers. Por esse motivo, é importante avaliar a qualidade dos dados antes de proceder com o treinamento de um modelo e tratá-los caso necessário. A primeira etapa seria a avaliação de valores nulos e outliers, pontos significativamente distintos dos demais, pois ambos afetam a precisão do modelo (XU; WANG, 2019). Dessa avaliação surgem duas frentes que podem ser exploradas: ajuste dos valores nulos e remoção dos outliers.

Método	Descrição
Média ou Mediana	Substituição dos valores nulos pela média ou mediana dos valores daquela coluna
Valor Fixo	Substituição dos valores nulos por um valor fixo, como zero
Remoção	Remoção da linha que contém o valor nulo, o que pode acabar removendo representatividade dos dados

Tabela 1 – Métodos para Ajuste de Valores Nulos

Método	Descrição
Z-Score	Remoção dos dados externos a $n$ vezes o desvio padrão a partir da média
IQR	Remoção dos dados externos a $n$ vezes a zona interquartil
MAD	Similar ao Z-Score, mas contando a partir da mediana
BADK	Ajuste de Boxplot baseado em k-Nearest Neighbors
MAD-BADK	Proposta híbrida que utiliza o Método MAD ao invés do IQR no ajuste de Boxplot

Tabela 2 – Métodos de Remoção de Outlier

Para a primeira, há três abordagens comuns, substituição por Média/Mediana, por Valor Fixo ou Remoção da Linha, detalhadas na Tabela 1. A estratégia escolhida depende do contexto e do volume de dados nulos e não há uma unanimidade na comunidade científica sobre qual o método utilizar (LIU et al., 2023). Para a segunda, a remoção de outliers pode ser feita de forma manual, por meio de visualização dos dados, ou por meio de técnicas estatísticas. Dentre as técnicas, há certa discordância na literatura sobre os benefícios e limitações de cada uma: (LEYS et al., 2013) aponta uma falha no método Z-Score e no método da Região Interquartil (IQR) ao lidar com datasets pequenos e sugere a utilização de MAD, enquanto (REHMAN; BELHAOUARI, 2021) propõe novas metodologias, como o Ajuste de Boxplot baseado em k-Nearest Neighbors (BADk). A Tabela 2 detalha cada método, incluindo uma proposta híbrida que mescla BADK com MAD.

Quanto à visualização dos dados, a principal limitação é a dimensão do dataset, pois o máximo que conseguimos visualizar são três dimensões, o que pode ser um problema em datasets com muitas características, de modo que seja essencial reduzir a dimensionalidade dos dados. Diversos métodos são citados na literatura, cada qual com suas particularidades, assim como detalhado na Tabela 3.

Após o tratamento da qualidade dos dados, é necessário verificar o equilíbrio das classes. Algumas técnicas comumente utilizadas em cenários de desequilíbrio para reduzir o viés imposto pela classe majoritária incluem Oversampling, Undersampling, SMOTE e SMOTEENN, que podem ser aplicadas para remediar o quadro. No entanto, é essencial ter cuidado para não induzir o modelo ao sobreajuste.

Método	Descrição
Análise de Componente Principal (PCA)	Utiliza transformações lineares para reduzir a dimensionalidade dos dados, preservando ao máximo a variância em direções principais (PITOYO; KAYO, 2017)
Análise de Componentes de Vizinhança (NCA)	Transforma dados usando uma métrica de Mahalanobis aprendida para maximizar a classificação leave-one-out no k-NN, preservando proximidades relevantes no espaço reduzido (GOLDBERGER et al., 2004)
Análise Discriminante Linear (LDA)	Reduz dimensões projetando os dados em eixos que maximizam a separação entre classes conhecidas, utilizando informações categóricas (PITOYO; KAYO, 2017)
Distância entre os k vizinhos (D-k-NN)	Organiza os dados de acordo com a distância média entre os seus vizinhos, formando uma linha, o que permite a remoção direta dos outliers (REHMAN; BELHAOUARI, 2021)
Embutimento Estocástico de Vizinhança Distribuída (t-SNE)	Reduz dimensões convertendo distâncias em probabilidades de vizinhança, otimizando uma função de custo baseada em distribuições de Student-t para preservar estruturas locais e globais (MAATEN; HINTON, 2008)
Embutimento Linear Local (LLE)	Preserva as relações locais dos dados, projetando-os em um espaço de menor dimensão por meio de reconstruções lineares locais (WANG; AL., 2024)
Mapas Auto-Organizáveis (SOM)	Projeta dados multidimensionais em mapas 2D, preservando a topologia local sem usar informações de categorias, criando representações visuais úteis para análise exploratória (PITOYO; KAYO, 2017)
SOM Relacionados à Classe (CRSOM)	utiliza redes neurais baseadas em funções de base radial para projetar dados multidimensionais em mapas 2D, preservando topologia e contexto, com foco em visualização e classificação (PITOYO; KAYO, 2017)

Tabela 3 – Métodos de Visualização de Dados

Técnica	Descrição	Limitações
Oversampling	Geração de cópias da classe minoritária	Frequentemente leva o modelo ao sobreajuste
Undersampling	Redução de instâncias da classe majoritária	Tende a perder representatividade nos dados
Técnica de Oversampling por Minoria Sintética (SMOTE)	Geração sintética de dados impacta da classe minoritária	Gera clusters de dados, o que modelos como kNN, além de introduzir ruído
SMOTE Editado com Vizinho Próximo (SMOTEENN)	SMOTE com remoção de ruídos após a geração	Por conta do processo extra, é mais lento de processar

Tabela 4 – Técnicas de Balanceamento

## 2.3 Algoritmos de Aprendizado

Não existe uma abordagem universal a ser seguida no que tange ao desenvolvimento de um modelo de IA; contudo, há estudos, diretrizes e evidências na literatura que argumentam e expõem os benefícios e limitações dos algoritmos. Dentre eles, destacam-se Redes Neurais Artificiais (Artificial Neural Network, ANN), Floresta Randômica (Random Forest, RF), Máquina de Vetor de Suporte (Support Vector Machine, SVM), k-Vizinhos mais Próximos (k-Nearest Neighbors, k-NN) e Regressão Logística (Logistic Regression, LR), devido à sua elevada capacidade de classificação e ampla utilização na literatura para problemas de detecção de doenças (CHOUDHURY; GUPTA, 2019).

### 2.3.1 ANN

ANNs são inspiradas na estrutura do cérebro humano, consistindo em camadas de neurônios artificiais que processam informações e aprendem padrões a partir dos dados. As ANNs são amplamente utilizadas para tarefas de classificação e regressão devido à sua capacidade de modelar relações complexas entre entradas e saídas (EL\_JERJAWI; ABU-NASER, 2018).

### 2.3.2 RF

O algoritmo de aprendizado RF faz parte dos métodos ensemble, que aglomeram os resultados de diversas partes antes de apontar o resultado final. No caso do RF, os dados passam por  $n$  Árvores de decisão (AD) e, após cada uma chegar a uma classificação para aquele dado, uma votação por maioria é feita para eleger o resultado final (AHMAD et al., 2022). Além disso, por convenção, em cenários de classificadores binários, sugere-se utilizar números ímpares de árvores, a fim de não possibilitar empates.

### 2.3.3 SVM

O método SVM funciona como um gerador de limites entre as classes, criando um hiperplano que tenta separar as classes em áreas de modo que um novo ponto seja associado à classe em cuja área está inserido (AHMAD et al., 2022). Apesar de robusto, é fortemente impactado pelos outliers, sendo necessário configurar quão permissivo o modelo deve ser no intuito de otimizar a generalização (GÉRON, 2021).

### 2.3.4 k-NN

O algoritmo k-NN classifica um novo ponto com base nos rótulos dos seus  $k$  vizinhos mais próximos (GARCIA-CARRETERO et al., 2020). Este método é simples e eficaz, mas pode ser impactado por dados ruidosos ou desbalanceados. Portanto, recomenda-se atentar-se a eventuais comportamentos anômalos, principalmente em treinamentos que envolvam SMOTE e SMOTEENN, pois esses métodos fortalecem clusters de dados.

### 2.3.5 LR

LR é um método de classificação que modela a probabilidade de um dado pertencer a uma classe como uma função das variáveis de entrada (AHMAD et al., 2022). Apesar de ser mais simples, esse é um modelo que necessita de um ajuste fino de seus parâmetros, pois dados desbalanceados ou em baixa representatividade acabam por minar seu potencial, gerando um sobreajuste do modelo.

## 2.4 Métricas

Após ser treinado, o modelo deve ser avaliado para definir sua aplicabilidade e desempenho na tarefa pretendida e, para tal, diversas métricas podem ser calculadas. Dentre elas, algumas se destacam por serem amplamente usadas em problemas de classificação (ROMANCINI et al., 2022), todas baseadas nos valores coletados da matriz de confusão (relação que permite a análise direta de acertos e erros do modelo, bem como suas respectivas classes).

### 2.4.1 Acurácia

A Acurácia ( $Acc$ ) mede a proporção de previsões corretas entre todas as realizadas pelo modelo, sendo calculada pela Fórmula 1. Apesar de amplamente utilizada, ela pode ser enganosa em datasets desbalanceados, pois não diferencia bem entre classes majoritárias e minoritárias.

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

### 2.4.2 Precisão

A Precisão ( $Prec$ ) avalia a capacidade do modelo de prever corretamente os casos positivos entre todas as suas previsões positivas, conforme a Fórmula 2. É especialmente relevante em cenários onde os falsos positivos têm alto custo, como em diagnósticos médicos.

$$Prec = \frac{VP}{VP + FP} \quad (2)$$

### 2.4.3 Recall

O recall ( $Rec$ ) reflete a habilidade do modelo em identificar corretamente os casos positivos entre todos os verdadeiros positivos existentes, descrito pela Fórmula 3. Essa métrica é crítica em situações onde é importante minimizar falsos negativos, como em sistemas de detecção de fraudes e diagnósticos médicos.

$$Rec = \frac{VP}{VP + FN} \quad (3)$$

#### 2.4.4 Pontuação F1

A Pontuação F1 ( $F1$ ) combina Precisão e Recall em uma única métrica harmônica, oferecendo um equilíbrio entre ambas, conforme definido na Fórmula 4. É ideal para avaliar o desempenho em datasets desbalanceados.

$$F1 = \frac{2 * Acc * Pre}{Acc + Pre} \quad (4)$$

#### 2.4.5 Coeficiente Kappa de Cohen

O coeficiente Kappa ( $\kappa$ ) mede o grau de concordância entre as previsões do modelo e as observações reais, ajustando para acertos aleatórios, conforme a Fórmula 6. Ele é particularmente útil em problemas multiclases ou desbalanceados, pois considera a distribuição de classes.

$$\epsilon = \frac{(VP + FP) * (VP + FN) * (VN + FP) * (VN + FN)}{VN + VP + FN + FP^2} \quad (5)$$

$$\kappa = \frac{Acc - \epsilon}{1 - \epsilon} \quad (6)$$

#### 2.4.6 Coeficiente de Correlação de Mathew

O Coeficiente de Mathew ( $MCC$ ) é uma métrica baseada em correlação que leva em conta todos os elementos da matriz de confusão (VP, FP, VN e FN), como expresso na Fórmula 7. É amplamente reconhecido como uma métrica robusta para datasets desbalanceados, oferecendo uma visão global do desempenho.

$$Mcc = \frac{VP * VN - FP * FN}{\sqrt{(VN + VP) * (FN + FP) * (VN + FP) * (VP + FN)}} \quad (7)$$

### 2.5 Aprendizado Federado

O aprendizado federado é uma configuração de aprendizado de máquinas em que vários clientes treinam um modelo de forma colaborativa. Esse processo é liderado por um coordenador central, e os dados de treinamento permanecem descentralizados (ZHANG et al., 2021), como mostra a Figura 1. De modo resumido, o coordenador central possui um modelo global que é compartilhado com os demais dispositivos que estão participando do treinamento. Cada dispositivo treina o modelo com os dados locais e, periodicamente, envia um vetor com os pesos para o coordenador. Esse coordenador vai calcular os pesos atualizados com base nos vetores com ele compartilhados e irá enviar os novos pesos aos dispositivos, onde o ciclo de treinamento irá continuar com os pesos atualizados. Além de ser uma técnica descentralizada e não permitir o vazamento dos dados na rede por meio das comunicações como coordenador, o modelo local começa a se moldar de acordo com os dados locais, o que permite uma personalização do modelo, de modo que ele se adapte melhor ao usuário que o utiliza. Essa característica de personalização, aliada ao

conceito de treinamento descentralizado, é uma opção ideal para o contexto de medicina personalizada, onde os dados dos pacientes são sensíveis e não podem ser compartilhados, e cada paciente se comporta de forma distinta mediante doenças e tratamentos.

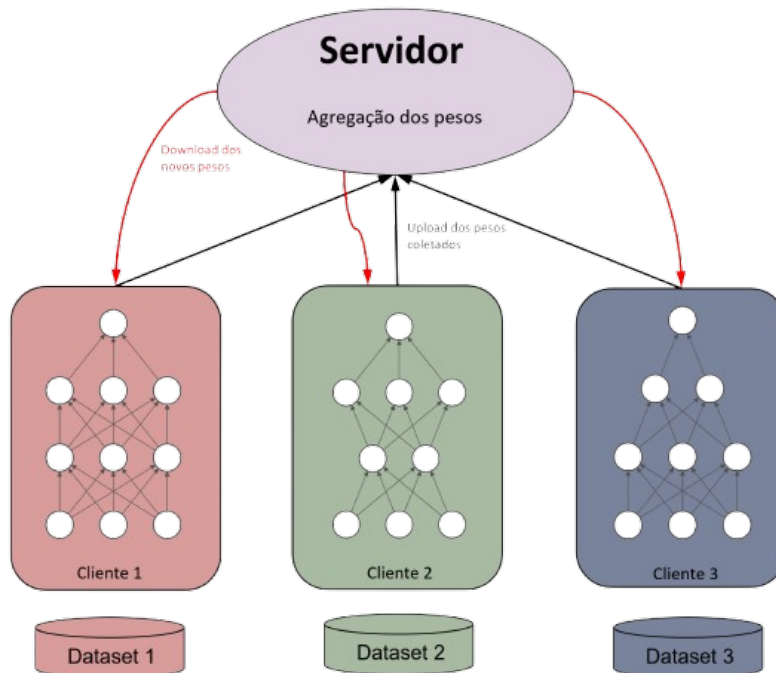


Figura 1 – Exemplo de Servidor Federado

Notoriamente, há diversos estudos na literatura sobre como aproveitar ao máximo essa configuração, sobre qual seria o cálculo ideal para ser feito, considerando que diferentes dispositivos possuem diferentes níveis de poder computacional e protocolos de segurança, haja vista a rigorosidade sobre a proteção dos dados implementada por diversos países. O algoritmo de FedAvg, cálculo da média dos vetores de pesos acumulados pelo coordenador, é um dos algoritmos de cálculo de peso mais utilizados, principalmente pela sua simplicidade e eficiência, mas não é o único, sendo o FedProx e o Protocolo de Compressão Ternária Esparsa (ZHANG et al., 2021) propostos como alternativas.

## 2.6 Trabalhos Correlatos

A procura por métodos mais eficientes e socialmente econômicos para a detecção e controle de diabetes não é exatamente uma novidade, desde 2010 essa pauta é abordada pela comunidade científica (BARAKAT; BRADLEY; BARAKAT, 2010), testando novos métodos, implementações, sistemas ou *frameworks* para alavancar o desenvolvimento de soluções ante tal doença atualmente incurável.

Objetivando desenvolver um modelo capaz de auxiliar médicos no diagnóstico de diabetes, (EL\_JERJAWI; ABU-NASER, 2018) fizeram um levantamento de possíveis fatores correlacionados ao desenvolvimento da doença (tipo I, II e pré-diabetes) e, defendendo o ponto de que regressões não se adaptam adequadamente a problemas com múltiplas variáveis independentes, desenvolveram uma RNA utilizando-se de resultados de exames laboratoriais e medidas corporais, alcançando uma acurácia de 87.3%. Com um objetivo semelhante, (BARAKAT; BRADLEY; BARAKAT, 2010) desenvolveu uma Máquina de



Vetores de Suporte, utilizando-se de subsampling e clusterização de K-means para lidar com desequilíbrio entre as classes, obtendo uma acurácia de 94% e sensibilidade de 93%.

Buscando a seleção ideal dos parâmetros, (SNEHA; GANGIL, 2019) utilizou-se de cálculos correlacionais para filtrar os inputs, removendo cinco colunas pouco relacionadas ao diagnóstico de diabetes. Após, aplicou o dataset tratado a cinco modelos, sendo eles SVM, RF, Nãive Bayes (NB), Árvore de Decisão (AD) e K-NN, com os quais obteve uma acurácia média de 72.56% e destacou a especificidade de RF e AD e a precisão de NB (82.30%) após segunda avaliação dos inputs.

Em uma temática similar, (AWONIRAN et al., 2022) explorou técnicas de pré-processamento, incluindo análise de componentes principais (PCA) para redução de dimensionalidade, a fim de melhorar o desempenho dos modelos na predição de diabetes. Utilizando um dataset de diabetes do Kaggle (ddK), foram testados três algoritmos principais: SVM, RF e Redes Neurais Profundas (DNN). A pesquisa destacou a acurácia de 89% do DNN, comparada com 76% do SVM e 77% do RF. Os autores demonstraram a ligação entre a eficácia e a qualidade dos dados pré-processados, sugerindo que melhorias nessa etapa podem gerar avanços ainda mais expressivos em futuras implementações.

Expandindo o uso das ANNs, (UPAMANYU et al., 2024) propuseram um modelo com uma camada inicial de Memória de Longo e Curto Prazo, visando contornar o problema de gradiente desaparecer de redes com retropropagação. Após utilizar uma versão modificada para LSTM do ddK, os autores expuseram a qualidade do modelo proposto ao comparar seus resultados aos de outros modelos já conhecidos, com acurácia, precisão, recall e F1-Score superiores a 97%.

Visando uma robustez adicional aos métodos tradicionais, que apresentavam baixa estabilidade e precisão, (XU; WANG, 2019) explorou o uso de múltiplos modelos para gerar o diagnóstico, utilizando-se de uma combinação entre RF e XGBoost, aliados a Intervalo Interquartil (IQR) para a remoção de valores anômalos. Por fim, foi feita uma validação cruzada K-fold, e o modelo final apresentou acurácia de 93.75% contra os 76%-88% dos modelos tradicionais apresentados.

Levando em consideração as dificuldades causadas pela heterogeneidade dos dados, devido a, por exemplo, diferentes padrões abordados por dispositivos ou instituições, (N et al., 2024) buscaram criar um framework que empregasse abordagens de AF personalizadas no nível dos dispositivos, visando a implementação de modelos ajustados às necessidades específicas de cada aplicação. Para validar o novo framework, os autores fizeram alguns estudos de casos, envolvendo datasets relacionados a aplicações de Internet das Coisas Médicas (IoMT), incluindo predição de diabetes, e mostraram como a proposta obteve resultados superiores a técnicas centralizadas e ao FedAvg (algoritmo simples para agregar atualizações de modelos locais, calculando a média ponderada dessas atualizações) em todos os casos estudados.

Nota-se que há um desenvolvimento em campo no que diz respeito a estudos na predição de diabetes, contudo, não foi encontrado na literatura um trabalho cujo foco seja semelhante ao proposto no presente trabalho, pois, enquanto as pesquisas existentes focam em apoiar o diagnóstico do médico ou realizar um diagnóstico com base nos dados de exames laboratoriais, seja utilizando uma estrutura federada ou centralizada, o objetivo é acompanhar pacientes com diabetes em um cenário cotidiano, tomando como base biomarcadores que possam ser coletados de maneiras não intrusivas e sem processos complexos por sensores.



### 3 Materiais e Métodos

A metodologia adotada neste estudo seguiu uma abordagem estruturada para investigar o aprendizado federado na detecção de diabetes. O processo foi dividido em quatro etapas principais:

1. **Seleção e Preparação dos Dados:** foram avaliados e tratados datasets, com foco em garantir a qualidade e a representatividade dos dados.
2. **Preprocessamento:** incluiu avaliação de valores nulos, detecção e eliminação de outliers e balanceamento das classes para corrigir o desbalanceamento do dataset.
3. **Treinamento e Validação dos Modelos:** diferentes algoritmos de aprendizado de máquina foram testados, e suas métricas comparadas para selecionar o modelo mais eficiente.
4. **Implementação do AF:** o modelo escolhido foi implantado em dispositivos ESP32, onde foi realizada a simulação de treinamento federado para validar a personalização e a escalabilidade da solução.

Os segmentos a seguir detalham cada uma dessas etapas, explicando os métodos utilizados e justificando as escolhas feitas, focadas na busca por resultados confiáveis e aplicáveis, alinhados aos objetivos propostos.

#### 3.1 Dataset

No intuito de selecionar um *dataset* condizente com a proposta da pesquisa, que é servir de protótipo para um modelo acoplado a sensores vestíveis, foram avaliados *datasets* utilizados nos trabalhos correlatos e *datasets* provenientes de repositórios de bases de dados, como o Kaggle, o Google Dataset Search, UCI Machine Learning Repository, Physionet e Harvard Dataverse. O objetivo da busca é selecionar bases de dados com dados fisiológicos que possam ser utilizados na pesquisa, semelhantes a dados que poderiam ser coletados por sensores vestíveis não invasivos. Dessa forma, o *dataset* que serviria para treinar o modelo não poderia ser composto somente por dados laboratoriais ou informações coletadas de questionários, eliminando a vasta maioria dos *datasets* encontrados, incluindo alguns dos recorrentemente utilizados, como o PIMA (*dataset* indiano voltado para o diagnóstico de diabetes). Outros *datasets* apresentavam apenas dados de concentração de glicose no sangue, úteis para previsão de glicose, mas não exatamente para a detecção de diabetes, haja vista o interesse em investigar outros biomarcadores.

Após a extensiva pesquisa, três *datasets* foram avaliados como elegíveis para a pesquisa: Ohio T1DM (dados de monitoramento contínuo de pacientes com diabetes tipo 1) (MARLING; BUNESCU, 2020), DiaHealth (dataset bengalês para predição de diabetes tipo II) (PRAMA et al., 2024) e um *dataset* sobre Glicose e Biomarcadores Superficiais (GBS) (JVALE; DESAI, 2021).

Apesar de ser utilizado em algumas literaturas e ter dados parcialmente relevantes para esse estudo, o Ohio T1DM apresenta frequente injeção de insulina nos pacientes e possui uma única classe de pacientes (diabetes tipo I). Tais fatores podem gerar uma visão enviesada do modelo. Ademais, não é de uso público e requer um acordo de uso de dados para o acesso, sendo descartado por esses motivos.

O DiaHealth oferece uma visão mais ampla, contendo dados de classes diabéticas e não-diabéticas, embora levemente desbalanceado, pendendo para a classe majoritária não-diabética. Contudo, a concentração de glicose coletada é referente à concentração na urina, e não no sangue, de modo que seria necessário um aparato mais robusto do que o vestível desejado pelo escopo do projeto e, como a concentração de glicose é uma medida crucial para a detecção, esse *dataset* também foi descartado.

Apresentando um desequilíbrio para a classe majoritária diabética, o GBS dispõe de dez colunas, sendo sete delas, valores coletáveis por sensores, duas, não coletáveis e uma, com o diagnóstico binário, mostrando-se apto a ser utilizado para o treinamento do modelo. A única ressalva é a proporção de 50:1 entre as classes, necessitando de um pré-processamento antes de ser utilizado efetivamente no modelo. A possibilidade de juntar múltiplos *datasets* foi considerada para reduzir tal desproporção, porém, a escassez desse tipo de *dataset* mostrou-se um obstáculo, anulando a possibilidade de junção.

### 3.2 Pré-processamento dos Dados

Originalmente, o GBS possui dez colunas, sendo elas Idade, Índice Glicêmico (IG), Pressão Arterial Diastólica (PAD), Pressão Arterial Sistólica (PAS), Frequência Cardíaca (FC), Temperatura Corporal (TC), Saturação de Oxigênio (SO<sub>2</sub>), Suor, Tremor e Diabético (D). Visto que Suor e Tremor eram tidas como entradas binárias (presente/ausente), ambas as colunas foram removidas. Além disso, no intuito de simplificar o processo de treinamento no vestível, aqui representado por um ESP32, as colunas PAD e PAS foram simplificadas para Pressão Sanguínea (PS). Por fim, a coluna TC teve seus valores convertidos de °F para °C, mantendo-se duas unidades decimais, e a coluna D teve seus dados convertidos de *Sim/Não* para saída binária.

Após esse remanejamento nas colunas, tomou-se cuidado com os valores nulos e as anomalias. Como evidenciado na Tabela 5, não havia valores ausentes nas colunas, eliminando a necessidade de substituir esses valores, avançando direto para a etapa de avaliação e remoção de anomalias para o melhor usufruto dos dados.

	Idade	IG	PS	FC	TC	SPO2	D
Frequência	16701	16701	16701	16701	16701	16701	16701
Valores Nulos	0	0	0	0	0	0	0
Média	31.16	96.00	12.11	91.61	36.31	97.38	0.98
$\sigma$	25.60	43.28	0.77	10.45	0.45	0.85	0.14
Mínimo	9.00	50.00	9.60	78.00	35.56	93.00	0.00
25%	9.00	68.00	11.70	84.00	35.93	97.00	1.00
50%	14.00	84.00	11.80	89.00	36.29	98.00	1.00
75%	55.00	108.00	12.70	95.00	36.64	98.00	1.00
Máximo	77.00	250.00	14.90	130.00	37.22	99.00	1.00

Tabela 5 – Estatísticas dos dados presentes no Dataset GBS

Após análise das vantagens, desvantagens e limitações dos métodos de redução de dimensionalidade, foram selecionados o t-SNE, NCA, D-k-NN e CRSOM para os testes iniciais, com foco na visualização de outliers e validação de suas remoções. Esses métodos foram escolhidos por sua habilidade de preservar a topologia dos dados, a eficiência em espaços de alta dimensionalidade e a capacidade de manter as relações locais, permitindo uma análise mais detalhada e precisa dos dados.

Para a detecção e remoção de outliers, optou-se por testar o uso de três abordagens: o BADk, o MAD e uma versão híbrida que substitui o uso de IQR do BADk pelo MAD, permitindo uma análise robusta aos desvios em dados mais assimétricos. Os resultados obtidos com cada método serão discutidos e comparados na seção 4, possibilitando uma avaliação clara do impacto de cada técnica na qualidade dos dados.

Contudo, ainda há o ponto da desproporção das classes no GBS. Para remediar essa falha, foram aplicados os métodos de balanceamento SMOTE, SMOTEENN e undersampling. O método de oversampling foi descartado devido à constante tendência de gerar o sobreajuste nos modelos testados. Assim sendo, foram produzidas 16 variações do GBS (três técnicas de remoção de outlier mais a original, por três técnicas de balanceamento mais a original).

### 3.3 Algoritmos de Aprendizado

Para a construção do modelo de classificação de diabetes a ser utilizado pelo AF, foram testados os algoritmos de aprendizado ANN, RF, SVM, kNN e RL. As implementações e treinamentos desses modelos foram realizados utilizando o software Orange Data Mining, que oferece uma plataforma interativa para configurar e otimizar os parâmetros dos algoritmos. Almejando o melhor modelo com os dados disponíveis, foram feitas avaliações com todas as 16 variações geradas do dataset GBS em cada tipo de modelo, totalizando 80 modelos, considerando apenas as definições finais dos hiperparâmetros, cujos valores foram selecionados para otimizar a capacidade preditiva.

Para o ANN, foram seguidas algumas diretrizes e recomendações (Géron, 2021), como a utilização da função de ativação *sigmoid* para classificações binárias, e a utilização de duas camadas ocultas, com ReLu como função de ativação, com 10 neurônios cada. Para todas as variações do GBS, essa configuração oferecia resultados satisfatórios.

Já para o RF, foi considerada a convenção de número ímpar de árvores, além de ter sido avaliado o número ideal de árvores para cada dataset, buscando o equilíbrio entre métricas e quantidade de árvores, pois, embora mais árvores tendam a prover resultados melhores, também acarretam um aumento no custo computacional, fator relevante, principalmente em vista das limitações da utilização do modelo no ESP32.

Para o desenvolvimento do modelo desejado com SVM, apenas dois hiperparâmetros precisaram ser ajustados, o  $C$  e o  $\gamma$ . Contudo, assim como no caso do RF, os valores mudavam com cada variação do GBS, buscando sempre o melhor conjunto de métricas.

A busca das melhores métricas com o kNN resultou em um comportamento semelhante ao SVM e RF, onde cada variação requeria um valor diferente de vizinhos. Contudo, como mencionado anteriormente, o kNN requer uma atenção especial ao ser treinado com SMOTE e SMOTEENN, portanto, todos os resultados foram rigorosamente estudados para avaliar sua validade.

Por fim, considerando que as variações do GBS contemplam dados desbalanceados (sem técnica de balanceamento aplicada), sintéticos (SMOTE e SMOTEENN) ou em baixa quantidade (undersampling), preferiu-se utilizar a regularização *Ridge (L2)* aliada a um parâmetro  $C$  adequado à variação utilizada em cada iteração, tanto para preservar informações dos dados, quanto para evitar o sobreajuste.

### 3.4 Aprendizado Federado

A estrutura federada criada, ilustrada na Figura 2, dispõe de quatro requisições, sendo elas a coleta dos pesos dos ESPs, o download do modelo inicial, a coleta do gradiente de alteração dos pesos e a inscrição dos ESP na estrutura. Assim que um ESP se conectar à Api do servidor federado, é aberta uma conexão websocket (WS) para envio de alertas, como heartbeat e finalização do cálculo de agregação do gradiente, e o dispositivo baixa o modelo inicial do servidor, salvando-o localmente. Foram utilizados dois ESP32 Wemos D1 R32 (módulo com Bluetooth e Wi-Fi) para simular dois clientes, um diabético com 19 anos e um saudável com 30 anos. Os dados utilizados por cada um dos clientes foram gerados sinteticamente com base nos dados coletados do GBS e armazenados em datasets distintos. O intuito é simular os biomarcadores coletados dos usuários em tempo real e avaliar em qual classe tais biomarcadores enquadram o usuário.

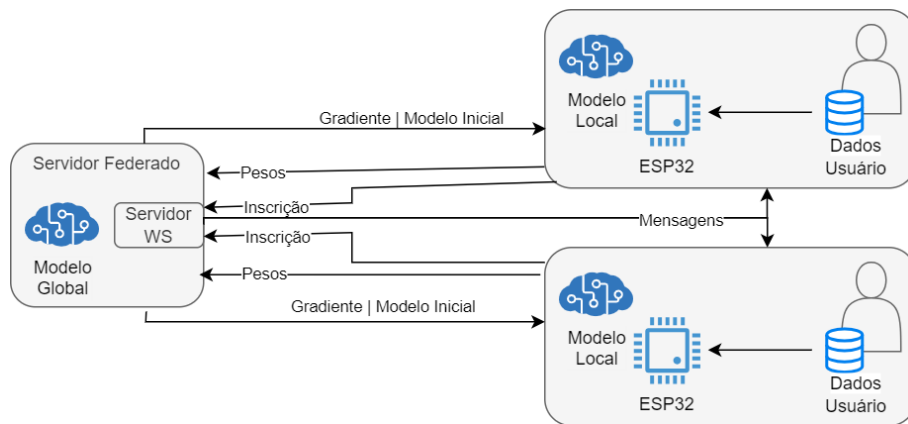


Figura 2 – Representação da Rede Federada

O fluxo básico do funcionamento desta estrutura é:

1. Um dispositivo se conecta à Api e ao serviço WS;
2. O servidor dá-lhe um id, salva a conexão na lista de dispositivos ativos;
3. O dispositivo requisita o modelo, que é enviado por requisição HTTP no formato JSON;
4. Em intervalos regulares, o dispositivo coleta um dado aleatório do dataset correspondente, salvando-o em um vetor;
5. Assim que houver dez coletas, o dispositivo treina o modelo local e atualiza os pesos localmente quando erra;
6. Finalizada a inferência, o dispositivo coleta os pesos do modelo e envia para o servidor;
7. Assim que o servidor obtém os vetores de todos os dispositivos conectados, ele inicia o processo de agregação;
8. Calculado o gradiente de mudança dos pesos, o servidor envia uma mensagem pelo WS, avisando todas as conexões ativas;
9. Os dispositivos coletam os gradientes e atualizam seus pesos;
10. O ciclo de coleta e atualização se reinicia.

Visando a necessidade da personalização dos modelos locais aos seus usuários, escolheu-se utilizar o gradiente, calculado com o algoritmo de FedAvg, ao invés dos pesos brutos, concedendo a liberdade necessária para o processo. Vale ressaltar que, enquanto o intuito é seguir padrões de uma arquitetura federada, não foram aplicadas camadas de segurança propriamente ditas à rede ou à Api, pois o escopo da pesquisa é avaliar a personalização e eficácia no diagnóstico de diabetes. Além disso, apesar da noção clássica do uso de ANN para o AF, essa não é uma limitação, de modo que outros modelos possam ser utilizados nesta estrutura de aprendizado, o diferencial seria a dificuldade de aplicar o modelo ao dispositivo, sendo o ANN o mais comum, tendo auxílio de bibliotecas como TensorFlow.

## 4 Resultados

Dos métodos de visualização dos dados baseados em redução de dimensionalidade, o NCA foi o que demonstrou melhor capacidade de expor os clusters de dados desse dataset. Enquanto CRSOM apresentou certas inconsistências, apontando zero clusters da classe não diabética, D-k-NN trouxe todos os dados em uma linha, como era esperado, porém, essa não foi uma perspectiva vantajosa para a análise dos dados. Por fim, t-SNE conseguiu trazer uma visão razoavelmente útil, contudo, menos visivelmente informativa que o NCA e, portanto, a análise dos métodos de remoção de outliers foi feita com base no NCA.

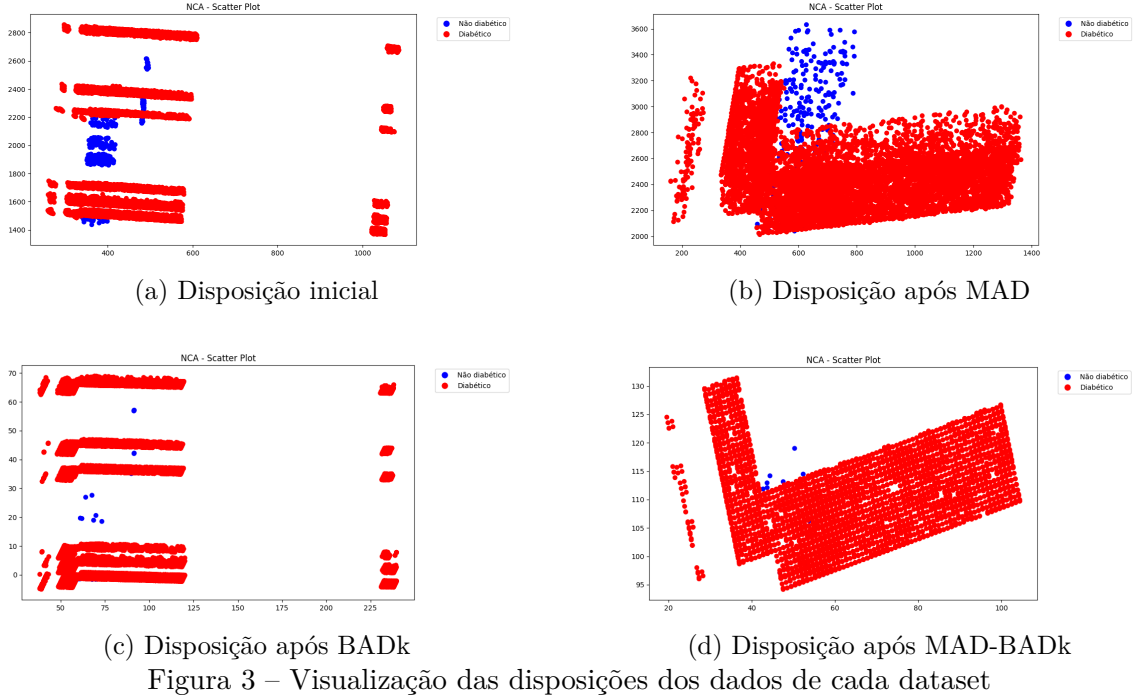
A Figura 3 apresenta uma comparação entre o dataset inicial e os métodos de remoção de outliers (MAD, BADK, MAD-BADK). No original, há a presença de clusters vagamente definidos e espalhados, bem como uma zona de sobreposição das classes. O método MAD tornou mais evidente a forma dos clusters, tanto sozinho quanto quando aplicado ao método BADk, e, apesar de não ter separado completamente as classes, o que pode indicar a presença de outliers não removidos, demonstra eficácia equilibrada, sem remover muitos valores dúbios. Contudo, o mesmo não pode ser afirmado sobre o método BADk, o qual demonstra certa agressividade nas remoções, eliminando dados cruciais aos clusters, sendo tal comportamento observado tanto sozinho quanto aliado ao método MAD.

Portanto, há fortes indicativos de que o método MAD é o mais efetivo para o tratamento do dataset GBS. Porém, resolveu-se manter as três versões tratadas do GBS, cada qual correspondendo à seleção filtrada por um método e, em seguida, as quatro versões passaram pelos métodos de balanceamento Undersampling, SMOTE e SMOTEENN, gerando assim as 16 variações do dataset inicial. A Tabela 5 aborda as principais informações sobre as variantes finais.

Embora o método MAD tenha demonstrado ser o mais eficaz para a remoção de outliers, todas as 16 versões do dataset foram utilizadas na avaliação dos modelos. O objetivo foi analisar de forma abrangente o impacto das remoções no desempenho de aprendizagem dos modelos e identificar qual deles seria o mais adequado para o diagnóstico de diabetes.

As avaliações foram realizadas utilizando as seis métricas de desempenho nos 80 modelos, os quais alcançaram resultados variados, oscilando entre ruim e excelente. O gráfico da Figura 4 ilustra a média das métricas em cada algoritmo de aprendizado testado. Já a Tabela 7 mostra o melhor resultado de cada algoritmo, destacando as métricas alcançadas e a versão do dataset utilizado. Esse conjunto permite uma visualização clara das comparações, enquanto fornece um panorama do desempenho de cada um.

Como pode ser observado na Tabela 7, o melhor resultado dentre os modelos



	Remoção	Balanco	Diabéticos	Não diabéticos	Total	Razão
1	-	-	16373	328	16701	49.92
2	-	SMOTE	16311	16251	32562	1.00
3	-	SMOTEENN	16168	16009	32177	1.01
4	-	Undersampling	656	328	984	2.00
5	MAD	-	8841	233	9074	37.94
6	MAD	SMOTE	8841	8793	17634	1.01
7	MAD	SMOTEENN	8765	8499	17264	1.03
8	MAD	Undersampling	466	233	699	2.00
9	BADk	-	15823	59	15882	268.19
10	BADk	SMOTE	15823	15761	31584	1.00
11	BADk	SMOTEENN	15818	15392	31210	1.03
12	BADk	Undersampling	118	59	177	2.00
13	MAD-BADk	-	8518	45	8563	189.29
14	MAD-BADk	SMOTE	8518	8470	16988	1.01
15	MAD-BADk	SMOTEENN	8497	8087	16584	1.05
16	MAD-BADk	Undersampling	90	45	135	2.00

Tabela 6 – Estatísticas das versões do dataset

testados foi obtido pelo algoritmo k-NN, seguido pelo RF. Além disso, contrário aos indícios iniciais, MAD não foi o método que gerou os melhores modelos, mas sim BADk e sua proposta híbrida, ambos servidos com dados sintéticos. Porém, como o k-NN é muito sensível a inserções de dados sintéticos, a probabilidade de estar enviesado é elevada, fator esse evidenciado pela indiferença em utilizar valores superiores a dois para o  $k$ . Portanto, o modelo RF treinado com o dataset SMOTEENN MAD-BADk foi escolhido para ser utilizado como modelo inicial do servidor da arquitetura federada.

Porém, como o algoritmo RF não é diretamente replicável para o código do ESP e o ESP utilizado possuía claras limitações de RAM, algumas modificações foram necessárias:

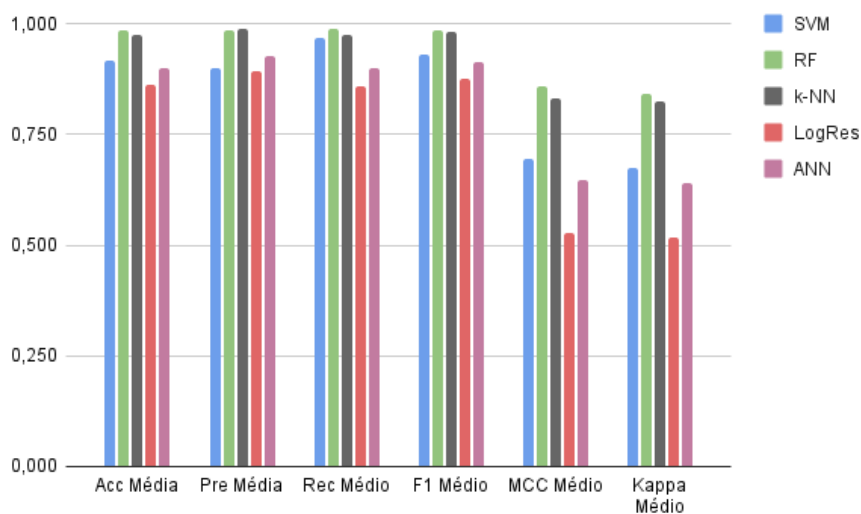


Figura 4 – Média das métricas de desempenho

Algoritmo	Dataset	Acc	Prec	Recall	F1	MCC	Kappa
ANN	Smoteenn BADk	0.986	<b>1.000</b>	0.972	0.986	0.973	0.973
k-NN	Smoteenn BADk	<i>1.000</i>	<b>1.000</b>	<i>0.999</i>	<i>1.000</i>	<i>0.999</i>	<i>0.999</i>
RF	Smoteenn MAD-BADk	<b>0.997</b>	<b>1.000</b>	<b>0.994</b>	<b>0.997</b>	<b>0.994</b>	<b>0.994</b>
RL	Under MAD	0.904	0.893	0.979	0.934	0.771	0.760
SVM	Smoteenn MAD-BADk	0.974	0.957	0.990	0.973	0.949	0.948

Tabela 7 – Melhores métricas de cada algoritmo de aprendizado

- Replicar o modelo treinado com base em árvores binárias, onde cada nó era representado no formato da estrutura Node:

```
// Estrutura criada para representar um nó da árvore
struct Node{
    int feature_index;
    float threshold;
    int left_index;
    int right_index;
    float value;
};
```

- Separar o modelo em 25 arquivos JSON menores, cada qual representando uma das árvores criada pelo modelo inicial;
- Criar um função de inferência e atualização dos pesos locais a cada inferência errada, mudando somente o pesos dos nodos pelos quais o dado passou na árvore;
- Criar uma função para coletar todos os pesos de todos os 25 arquivos, para enviar os valores para o servidor calcular a agregação;



- Criar uma função para agregar o gradiente enviado do servidor aos nós das árvores correspondentes.

Por fim, após um período da estrutura trabalhando com os dados locais, tornou-se evidente a diferença nos pesos entre os dois ESPs. Isso se deve por conta das atualizações locais dos pesos a cada inferência errada, bem como à atualização via gradientes, o que demonstra fortemente a possibilidade de personalização da estratégia federada, haja vista a tendência do modelo se adequar aos dados por ele processados. A Figura 5 contém os pesos coletados dos ESPs no final do período de testes. Vale ressaltar que esse teste nos ESPs foi feita com o intuito de avaliar a personalização e, como os dados utilizados eram, em sua maioria, sintéticas, não foram avaliadas as métricas das versões locais do modelo, sendo as detalhes anteriormente nessa seção referentes aos modelos iniciais.

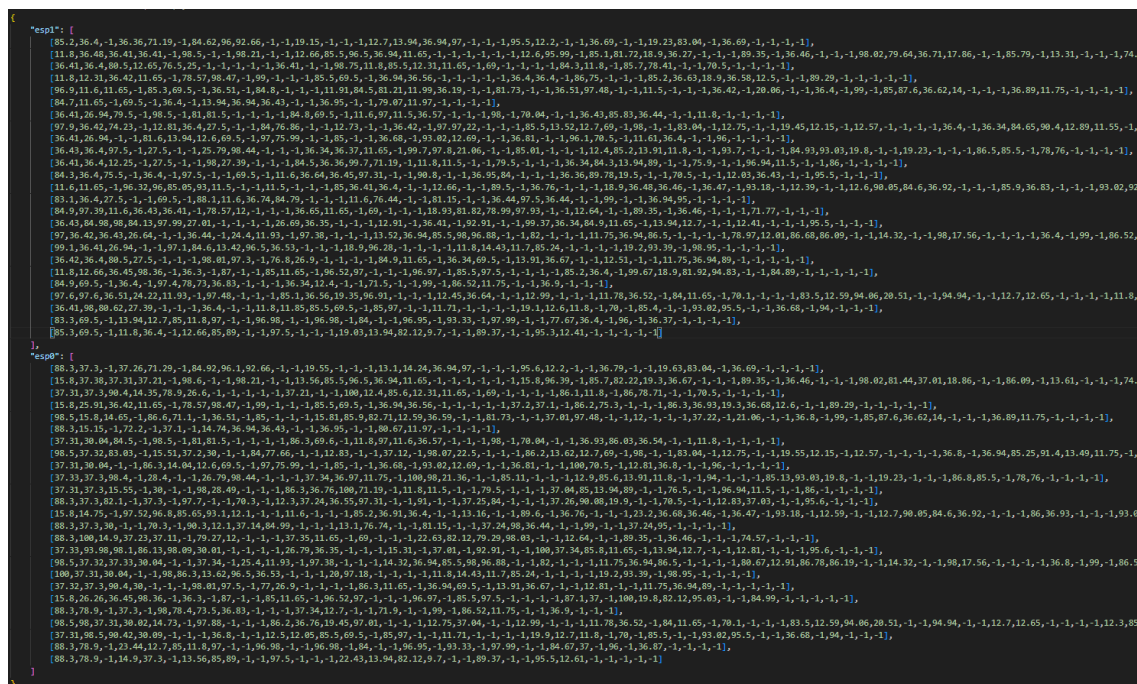


Figura 5 – Comparação Entre os Pesos dos ESPs

## 5 Discussão

O presente estudo explorou diferentes modelos de aprendizado de máquina para detecção de diabetes a serem aplicados a uma rede federada, identificando resultados promissores, mas também ressaltando limitações significativas que precisam ser abordadas para garantir a aplicabilidade prática e a confiabilidade do sistema. Dentre as limitações, destacam-se o desequilíbrio de classes, limitações dos datasets, desafios de implementação e aspectos éticos relacionados à privacidade dos dados.

### 5.1 Impacto do Desequilíbrio de Classes

Apesar de alguns modelos terem apresentado métricas excelentes, quase utópicas, existem ressalvas a serem feitas sobre algumas limitações observadas, a começar pelo desequilíbrio das classes. Modelos desequilibrados tendem a carregar uma visão enviesada, dificultando a convergência correta. Mesmo com técnicas de balanceamento, como o



SMOTEENN, estas se revelam apenas medidas paliativas que reforçam relações já existentes no dataset. Embora baseados em dados reais, dados sintéticos não capturam toda a complexidade do mundo real e podem impactar negativamente o desempenho de alguns algoritmos, como o k-NN.

O conceito do k-NN é categorizar um ponto com base na classe dos seus vizinhos. Embora robusto, a introdução de dados sintéticos alterou significativamente sua dinâmica. A geração de dados sintéticos fortalece clusters no dataset, criando vizinhos artificiais da mesma classe muito próximos. Essa interferência foi evidenciada ao encontrar o valor ideal de ( $k = 2$ ), o que viola o princípio do desempate e ilustra o impacto do balanceamento artificial na qualidade do modelo.

## 5.2 Limitações dos Datasets

Não obstante, os datasets utilizados apresentam limitações significativas. A convergência satisfatória de um modelo não implica, necessariamente, em sua adequação à tarefa desejada. A exemplo, embora o dataset GBS tenha sido o melhor encontrado, ele não é ideal para o propósito desta pesquisa. Apesar de conter as variáveis relevantes (biomarcadores fisiológicos), os dados foram coletados em ambientes controlados, comprometendo sua capacidade de representar situações do cotidiano.

Analisando a Tabela 5, identificam-se limitações significativas:

- **FC:** Variando entre 78 e 130 bpm, o intervalo exclui condições comuns como bradicardia ou usuário adormecido (abaixo de 60 bpm) e elevações acima de 150 bpm em atividades intensas.
- **PS:** Valores entre 9.6 e 14.9 representam a normalidade, ignorando casos de hipertensão (acima de 17 mmHg) e hipotensão (abaixo de 9.6 mmHg).
- **SP02:** Com intervalos entre 95% e 100%, não inclui condições como hipoxemia (abaixo de 95%) e hiperóxia (acima de 100%).

Essas discrepâncias ressaltam a necessidade de um novo dataset que contemple uma maior variabilidade fisiológica, abrangendo condições reais, extremas e menos usuais. Isso fortaleceria a capacidade do modelo de lidar com dados do cotidiano, promovendo maior generalização e robustez na detecção e monitoramento de diabetes.

## 5.3 Desafios de Implementação no ESP

A implementação do modelo no ESP revelou limitações relacionadas à capacidade limitada de armazenamento do dispositivo. O modelo inicial ocupava cerca de 80% espaço disponível, sendo necessário cindir o modelo em árvores binárias salvas como arquivos .json via SPIFFS. Isso permitiu a alocação momentânea apenas da árvore em processamento. Embora essa abordagem tenha reduzido a velocidade de processamento devido às alocações sequenciais, o intervalo entre leituras e requisições é suficientemente longo para que o impacto seja mínimo.

## 5.4 Conformidade com a LGPD

Embora o AF minimize o compartilhamento direto de informações sensíveis ao manter os dados localmente, ainda deve-se cumprir as diretrizes da LGPD. É essencial que

os dados sejam anonimizados e que dispositivos e servidores sigam normas de segurança robustas. Neste estudo, não foi criado um protocolo específico para garantir a conformidade, pois o foco era avaliar a aplicabilidade do AF. Entretanto, em aplicações reais, é recomendável implementar protocolos de segurança para proteger a privacidade dos dados dos usuários.

## 5.5 Relevância dos Marcadores Fisiológicos

Os marcadores fisiológicos analisados desempenham papel central na detecção e no monitoramento da diabetes:

- **FC:** Pode indicar variações metabólicas associadas à glicemia ou respostas anormais ao esforço físico.
- **PS:** Relaciona-se diretamente a complicações cardiovasculares comuns em diabéticos.
- **IG:** É a base do diagnóstico e controle da doença.
- **SP02:** Permite identificar possíveis complicações respiratórias associadas à diabetes.

Esses marcadores, em associação com o modelo binário de classificação, permitem o monitoramento em tempo real, identificando mudanças críticas antes que se tornem emergências médicas.

## 5.6 Aplicação do Modelo em Cenário Real

O modelo desenvolvido tem potencial para ser utilizado em dispositivos vestíveis para monitoramento de indivíduos em suas atividades diárias. Por exemplo, uma pessoa não diagnosticada como diabética poderia usar um dispositivo integrado ao modelo que, durante suas atividades cotidianas, detectaria frequências cardíacas elevadas e outros marcadores fora do padrão. O modelo classificaria o indivíduo como potencialmente diabético, emitindo um alerta para procurar orientação médica. Essa abordagem pode evitar complicações futuras, oferecendo um sistema preventivo que desafoga os sistemas de saúde ao tratar problemas precocemente.

## 6 Conclusão e Trabalhos Futuros

Este estudo validou a aplicação do aprendizado federado na detecção de diabetes utilizando biomarcadores fisiológicos. Os objetivos foram atingidos ao demonstrar que o modelo proposto pode alcançar métricas promissoras, como a acurácia de 0.999, além de evidenciar a viabilidade de implementação em dispositivos de ponta. Os resultados reforçam o potencial dessa abordagem para soluções de monitoramento não invasivas e personalizáveis na área da saúde, embora testes mais aprofundados com dados reais sejam necessários para validar completamente a eficácia de tal personalização.

Apesar do sucesso obtido, algumas limitações foram identificadas, principalmente relacionadas ao dataset utilizado, que apresentava baixa variabilidade e dados coletados em condições controladas. Isso restringe a generalização do modelo para cenários do dia a dia, onde as condições são naturalmente mais dinâmicas. Com base nisso, futuras pesquisas devem focar na criação de datasets mais representativos, que capturem maior diversidade de condições fisiológicas e comportamentos reais dos usuários, incluindo marcadores como:

- Frequência Cardíaca
- Taxa de Sudorese
- Pressão Diastólica
- Pressão Sistólica
- Condutância da Pele
- Frequência Respiratória
- Hidratação Corporal
- Temperatura Corporal
- Saturação de Oxigênio
- Saturação de Dióxido de Carbono

Além disso, a continuidade dos estudos pode explorar formas de melhorar a eficiência computacional dos modelos implantados em dispositivos com restrições de hardware, como os utilizados neste trabalho. Conclui-se que o aprendizado federado oferece um caminho promissor para o monitoramento de doenças crônicas, como a Diabetes, contribuindo para o desenvolvimento de soluções mais acessíveis, eficientes e adaptáveis no diagnóstico e acompanhamento de pacientes.

## Referências

AHMAD, G. N. et al. Mixed machine learning approach for efficient prediction of human heart disease by identifying the numerical and categorical features. *Applied Sciences*, v. 12, n. 15, 2022. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/12/15/7449>>. Citado (2) vezes nas páginas [12 e 13].

ALOTAIBI, M.; ALBALAWI, M. A mobile gestational diabetes management and educational system for gulf countries: System architecture. In: *2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC)*. [S.l.: s.n.], 2018. p. 193–196. Citado na página [9].

AWONIRAN, O. et al. A machine learning technique for detection of diabetes mellitus. In: *2022 5th Information Technology for Education and Development (ITED)*. [S.l.: s.n.], 2022. p. 1–6. Citado na página [16].

BARAKAT, N.; BRADLEY, A. P.; BARAKAT, M. N. H. Intelligible support vector machines for diagnosis of diabetes mellitus. *IEEE Transactions on Information Technology in Biomedicine*, v. 14, n. 4, p. 1114–1120, 2010. Citado na página [15].

BIESSELS, G. J. et al. Risk of dementia in diabetes mellitus: a systematic review. *The Lancet Neurology*, Elsevier, v. 5, n. 1, p. 64–74, 2006. Citado na página [7].

BOLLIMUNTHA, M.; MURUGAN, K. Design and implementation of an interoperable iot based health monitoring system for diabetes. In: *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*. [S.l.: s.n.], 2023. p. 386–390. Citado na página [9].

BREITSCHEIDEL S. STAMENITIS, F.-W. D. L.; SCHÖFFSKI, O. Economic impact of compliance to treatment with antidiabetes medication in type 2 diabetes mellitus: a review paper. *Journal of Medical Economics*, Taylor & Francis, v. 13, n. 1, p. 8–15, 2010. PMID: 19947905. Citado na página [7].

BUTT, M. D. et al. A systematic review of the economic burden of diabetes mellitus: contrasting perspectives from high and low middle-income countries. *Journal of Pharmaceutical Policy and Practice*, Taylor & Francis, v. 17, n. 1, p. 2322107, 2024. Citado na página [7].

CASTRO, R. M. F. d. et al. Diabetes mellitus e suas complicações - uma revisão sistemática e informativa/ diabetes mellitus and its complications - a systematic and informative review. *Brazilian Journal of Health Review*, v. 4, n. 1, p. 3349–3391, 2021. Disponível em: <<https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/24958>>. Citado na página [9].

CHOUDHURY, A.; GUPTA, D. A survey on medical diagnosis of diabetes using machine learning techniques. In: KALITA, J. et al. (Ed.). *Recent Developments in Machine Learning and Data Analytics*. Singapore: Springer Singapore, 2019. p. 67–78. ISBN 978-981-13-1280-9. Citado na página [12].

COMMITTEE, A. D. A. P. P. Diagnosis and classification of diabetes: Standards of care in diabetes-2024. *Diabetes Care*, v. 47, p. S20–S42, 2024. Citado na página [9].

DHADE, P.; SHIRKE, P. Federated learning for healthcare: A comprehensive review. *Engineering Proceedings*, v. 59, n. 1, 2023. ISSN 2673-4591. Disponível em: <<https://www.mdpi.com/2673-4591/59/1/230>>. Citado na página [8].

EL\_JERJAWI, N. S.; ABU-NASER, S. S. Diabetes prediction using artificial neural network. *International Journal of Advanced Science and Technology*, v. 121, p. 54–64, 2018. Disponível em: <<https://doi.org/10.14257/ijast.2018.121.05>>. Citado (2) vezes nas páginas [12 e 15].

GARCIA-CARRETERO, R. et al. Use of a k-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. *Medical & Biological Engineering & Computing*, v. 58, n. 5, p. 991–1002, 2020. ISSN 1741-0444. Disponível em: <<https://doi.org/10.1007/s11517-020-02132-w>>. Citado na página [13].

GHOZALI, M. T. Improving self-management of type 2 diabetes: Evaluating the effectiveness of a mobile app-based patient education approach. In: *2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*. [S.l.: s.n.], 2024. p. 1–5. Citado na página [9].

GOLDBERGER, J. et al. Neighbourhood components analysis. In: SAUL, L.; WEISS, Y.; BOTTOU, L. (Ed.). *Advances in Neural Information Processing Systems*. MIT Press, 2004. v. 17. Disponível em: <[https://proceedings.neurips.cc/paper/\\_files/paper/2004/file/42fe880812925e520249e808937738d2-Paper.pdf](https://proceedings.neurips.cc/paper/_files/paper/2004/file/42fe880812925e520249e808937738d2-Paper.pdf)>. Citado na página [11].

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. Citado na página [9].

GÉRON, A. *Maos à Obra: Aprendizado de Máquina com Scikit-Learn & Tensorflow*. 2ª edição. ed. Rio de Janeiro: Alta Books, 2021. ISBN 9788550815480. Citado (2) vezes nas páginas [12 e 19].

JAVALE, D.; DESAI, S. *Dataset for People for their Blood Glucose Level with their Superficial body feature readings*. IEEE Dataport, 2021. Disponível em: <<https://dx.doi.org/10.21227/c4pp-6347>>. Citado na página [17].

KRAKAUER, M.; MALERBI, F. *Inteligência Artificial Revoluciona o Tratamento do Diabetes*. 2024. Disponível em: <<https://diretriz.diabetes.org.br/inteligencia-artificial-revoluciona-o-tratamento-do-diabetes>>. Citado na página [8].

LEYS, C. et al. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, v. 49, n. 4, p. 764–766, 2013. ISSN 0022-1031. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022103113000668>>. Citado na página [10].

LIU, M. et al. Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques. *Artificial Intelligence in Medicine*, v. 142, p. 102587, 2023. ISSN 0933-3657. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S093336572300101X>>. Citado na página [10].

MAATEN, L. Van der; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. 11, 2008. Citado na página [11].

MARLING, C.; BUNESCU, R. The ohiot1dm dataset for blood glucose level prediction: Update 2020. *CEUR Workshop Proceedings*, v. 2675, p. 71–74, 2020. Citado na página [17].

N, S. D. et al. Fedcure: A heterogeneity-aware personalized federated learning framework for intelligent healthcare applications in iomt environments. *IEEE Access*, v. 12, p. 15867–15883, 2024. Citado na página [16].

NILSON, E. A. F. et al. Custos atribuíveis a obesidade, hipertensão e diabetes no sistema Único de saúde, brasil, 2018. *Revista Panamericana de Salud Pública*, v. 44, p. e32, 2020. ISSN 1680-5348. Disponível em: <<https://doi.org/10.26633/RPSP.2020.32>>. Citado na página [7].

OLIVEIRA, G. P. M.; AMÂNCIO, N. d. F. G.; SILVA, J. L. d. A relação dos fatores socioeconômicos no desenvolvimento e tratamento do diabetes mellitus tipo 2. *Brazilian Journal of Implantology and Health Sciences*, v. 6, n. 2, p. 1873–1887, 2024. Disponível em: <<https://bjih.emnuvens.com.br/bjih/article/view/1454>>. Citado na página [7].

PITOYO, H.; KAYO, O. Intuitive analysis by visualizing context relevant e-learning data. v. 3, n. 2, p. 20–27, 2017. ISSN 2188-4234. Disponível em: <<http://id.nii.ac.jp/1001/00182183>>. Citado na página [11].

PRAMA, T. T. et al. *DiaHealth: A Bangladeshi Dataset for Type 2 Diabetes Prediction*. [S.l.]: Mendeley Data, 2024. Citado na página [17].

- REHMAN, A. ur; BELHAOUARI, S. B. Unsupervised outlier detection in multidimensional data. *Journal of Big Data*, v. 8, n. 1, p. 80, 2021. ISSN 2196-1115. Disponível em: <<https://doi.org/10.1186/s40537-021-00469-z>>. Citado (2) vezes nas páginas [10 e 11].
- RODAKCI, M. et al. *Diagnóstico de diabetes mellitus. Diretriz Oficial da Sociedade Brasileira de Diabetes (2024)*. [S.l.]: Sociedade Brasileira de Diabetes, 2024. Citado na página [9].
- RODRIGUEZ-LEÓN, C. et al. Mobile and wearable technology for the monitoring of diabetes-related parameters: Systematic review. *JMIR Mhealth Uhealth*, v. 9, n. 6, p. e25138, 2021. ISSN 2291-5222. Citado (2) vezes nas páginas [7 e 8].
- ROMANCINI, E. M. R. et al. Desagregação de cargas residenciais: detecção de eventos e classificação em ambiente real. 2022. Citado na página [13].
- SNEHA, N.; GANGIL, T. Analysis of diabetes mellitus for early prediction using optimal features selection. *Journal of Big Data*, v. 6, n. 1, p. 13, 2019. ISSN 2196-1115. Disponível em: <<https://doi.org/10.1186/s40537-019-0175-6>>. Citado na página [16].
- STANDARDS of Medical Care in Diabetes—2015: Summary of Revisions. *Diabetes Care*, v. 38, n. Supplement\_1, p. S4–S4, 2014. ISSN 0149-5992. Disponível em: <<https://doi.org/10.2337/dc15-S003>>. Citado na página [9].
- UPAMANYU, M. S. et al. Early prediction of type- ii diabetes mellitus in young adults using lstm. In: *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. [S.l.: s.n.], 2024. p. 1–6. Citado na página [16].
- VERAS, R. P. Doenças crônicas e longevidade: desafios futuros. *Revista Brasileira de Geriatria e Gerontologia*, Universidade do Estado do Rio Janeiro, v. 26, p. e230233, 2023. ISSN 1809-9823. Disponível em: <<https://doi.org/10.1590/1981-22562023026.230233.pt>>. Citado na página [7].
- WANG, H.; AL. et. A classification algorithm based on improved locally linear embedding. *International Journal of Computational Intelligence and Applications (IJCINI)*, v. 18, n. 1, p. 1–9, 2024. Disponível em: <<https://doi.org/10.4018/IJCINI.344020>>. Citado na página [11].
- XU, Z.; WANG, Z. A risk prediction model for type 2 diabetes based on weighted feature selection of random forest and xgboost ensemble classifier. In: *2019 Eleventh International Conference on Advanced Computational Intelligence (ICACI)*. [S.l.: s.n.], 2019. p. 278–283. Citado (2) vezes nas páginas [10 e 16].
- ZHANG, C. et al. A survey on federated learning. *Knowledge-Based Systems*, v. 216, p. 106775, 2021. ISSN 0950-7051. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950705121000381>>. Citado (2) vezes nas páginas [14 e 15].
- ZOU, Y. et al. Research progress and prospects of intelligent diabetes monitoring systems: A review. *IEEE Sensors Journal*, v. 24, n. 15, p. 23401–23435, 2024. Citado na página [9].