



UEPB

UNIVERSIDADE ESTADUAL DA PARAÍBA - UEPB

CAMPUS I

Centro de Ciência e Tecnologia - CCT

Departamento de Computação

Bacharelado em Ciências da Computação

Alice Ferreira da Silva Oliveira

Caio César Silva dos Santos

Isis Julie Assis Gomes.

Orientador: Prof. Dr. Dunfrey Pires Aragão

Sistema C.A.I: Cuidado Assistido de Idosos, um sistema para identificar alterações no estado de saúde dos pacientes.

Campina Grande - PB

2025

1. Introdução

O envelhecimento populacional brasileiro é um fenômeno exponencial: segundo o Censo Demográfico 2022, a população com 65 anos ou mais chega a 10,9 % da população total, com alta de 57,4 % em relação a 2010 (quando era 7,4 %) (IBGE, 2023). Considerando idosos com 60 anos ou mais, são 32 milhões de pessoas, o que representa 15,6 % da população total do país. Diante desse cenário no qual a população idosa é boa parte da população total do país, a tecnologia surge como uma aliada no enfrentamento dos desafios que chegam junto ao aumento dessa população, como a falta de acompanhamento individualizado e contínuo dos pacientes em localizações remotas e a limitações de locomoção dos idosos. Nas últimas décadas, a Inteligência Artificial (IA) tem se consolidado como uma ferramenta fundamental em diversos setores da vida humana, especialmente na área da saúde, por sua capacidade de processar grandes volumes de dados e identificar padrões que muitas vezes passam despercebidos ao olhar humano. Técnicas de aprendizado de máquina já são utilizadas em aplicações como a análise automatizada de exames de imagem, a predição de risco de doenças crônicas e o suporte à tomada de decisão clínica. Sob esse viés, Zhou (2022) destaca que modelos como Random Forest e Gradient Boosting têm obtido alto desempenho na predição de doenças crônicas, como diabetes e problemas cardiovasculares. Ademais, Shickel (2018) demonstra que o uso de aprendizado de máquina aplicado a registros eletrônicos de saúde (EHRs) possui vasto potencial para apoiar a tomada de decisões médicas, auxiliando na personalização de tratamentos. Esses resultados reforçam que tais abordagens já vêm sendo aplicadas em cenários reais no contexto da saúde, servindo de base sólida para projetos que buscam ampliar o monitoramento e a prevenção na população idosa. Na área da medicina assistiva, estes avanços permitem o desenvolvimento de sistemas de monitoramento inteligentes voltados para a população idosa, como destacado por Topol (2019), ao afirmar que a convergência entre inteligência artificial e medicina de alta performance pode ampliar a eficiência dos cuidados de saúde, especialmente na identificação precoce de riscos e na personalização de tratamentos. Esses sistemas e modelos podem atuar de forma preventiva, identificando sinais de risco e antecipando situações críticas, como quedas, arritmias, infartos e acidentes vasculares cerebrais (AVCs). Dessa forma, a IA não apenas amplia a eficiência do cuidado, mas também contribui para a autonomia, a segurança e a qualidade de vida dos idosos, também oferecendo maior tranquilidade às famílias e cuidadores.

A falta de assistência médica em lugares remotos e as dificuldades de locomoção de idosos são fatores que contribuem no aumento da demanda de soluções práticas que integrem prognósticos rápidos, de fácil utilização e que sejam ferramentas automatizadas, habilitadas a dar apoio aos familiares e cuidadores. Nesse contexto, ferramentas que são baseadas em machine learning e Inteligência Artificial podem ampliar o acesso a prognósticos de qualidade, como exposto no estudo recente realizado no âmbito do Brazilian Longitudinal Study of Ageing (ELSI-Brazil), que demonstrou que modelos de aprendizado de máquina, como

Random Forest, Gradient Boosting e XGBoost, alcançaram um bom desempenho ($AUC \approx 0,92$) na predição de mortalidade em brasileiros com 50 anos ou mais, reforçando o potencial dessas abordagens no monitoramento da saúde da população idosa (BARBOSA et al., 2025). Além disso, a utilização de um sistema que apresenta um prognóstico preciso, antecipa os possíveis riscos e aumenta a chance de recuperação.

Apesar dos avanços na área de machine learning e Inteligência Artificial, um dos principais desafios na construção de modelos preditivos em saúde assistiva é garantir a acurácia confiável das previsões dadas, especialmente quando os dados disponíveis são heterogêneos ou limitados. Ao utilizar machine learning, o sistema pode apresentar overfitting, fenômeno no qual o algoritmo se ajusta excessivamente aos dados de treinamento - que foram retirados do dataset - capturando características específicas desse conjunto e comprometendo o desempenho em novos dados, como os fornecidos por usuários do sistema. Nesse sentido, o presente projeto busca solucionar a seguinte questão: Como desenvolver e avaliar um modelo de machine learning que seja preciso, generalizável e interpretável para a predição de riscos de saúde em idosos a partir de dados de monitoramento assistivo?

Nesse sentido, evidencia-se a necessidade de desenvolver e avaliar sistemas capazes de integrar tecnologia e eficiência, não apenas na detecção de alterações no estado de saúde dos pacientes, mas também na geração de prognósticos precisos que auxiliem na prevenção de complicações. O sistema deve apresentar um desempenho satisfatório em termos de acurácia e ser capaz de generalizar para diferentes perfis de indivíduos idosos. Além disso, busca-se ter incorporado no sistema e interface, mecanismos de fácil interpretação, possibilitando que os cuidadores e profissionais da área da saúde possam compreender a previsão gerada.

O presente trabalho justifica-se pelo potencial impacto benéfico na qualidade de vida da população idosa que um sistema de machine learning pode proporcionar, ao fornecer prognósticos precisos e acessíveis com base em dados fisiológicos. A proposição de um sistema que integre acurácia, prognóstico preciso e de fácil interpretabilidade contribui para uma necessidade real da população idosa. Ademais, o atual projeto contribui para a comunidade científica por possuir procedimento replicável e de código aberto, podendo contribuir para futuros avanços na área da tecnologia integrada à medicina.

Nesse sentido, o presente projeto propõe um sistema baseado em um modelo baseado em gradient boosting para classificar o risco que o usuário possui com base em dados vitais específicos. Resultados preliminares apresentaram uma acurácia de 99,8% na classificação dos riscos apresentados das entradas testadas, que foram retirados e tratados do datasetVitalsigns. Tais resultados mostram que a iniciativa pode contribuir significativamente para o ramo do cuidado assistivo de idosos, o protótipo desenvolvido tem potencial de mostrar-se útil para o combate a problemática anteriormente mencionada.

2.0 Justificativa

Considerando o aumento significativo do índice de envelhecimento no Brasil, que chegou a 24,7% no último censo demográfico do IBGE, o cenário atual revela a necessidade da criação de soluções práticas e eficazes para o cuidado preventivo da população idosa. Mesmo com o grande avanço tecnológico, ainda são poucas as pesquisas que tratem diretamente de um sistema que busque utilizar machine learning como meio de prevenção para situações de risco e fornecimento de prognósticos. A acessibilidade a recursos de saúde é um desafio significativo para a população idosa que geralmente possui limitações de locomoção, especialmente em regiões remotas. Assim, a utilização desse recurso pode contribuir para o monitoramento dessa parcela da população, atuando na prevenção de acidentes e doenças, otimização de recursos em situações de risco e, conseqüentemente, para a promoção do bem-estar e segurança desses indivíduos. Ao utilizar machine learning e Inteligência Artificial, o sistema oferece uma metodologia acessível e com vasto potencial de aplicação, também possibilitando sua posterior integração em dispositivos wearables.

3.0 Objetivos

3.1 Objetivo geral

Desenvolver um sistema inteligente de monitoramento da saúde de idosos, utilizando técnicas de inteligência artificial e machine learning, com o intuito de prevenir riscos de acidentes e eventos médicos críticos na população idosa.

3.2 Objetivos específicos

- Tratar o banco de dados Vitalsigns Dataset, utilizando as informações favoráveis;
- Implementar algoritmos de machine learning para analisar os dados;
- Validar a eficácia do sistema e verificar se houve overfitting e comparar com trabalhos existentes;
- Ligar o sistema com uma interface simples para implementar um protótipo de funcionamento.

4.0 Métodos e Materiais

Esse projeto configura-se como uma pesquisa quantitativa, com caráter experimental, com objetivo de formular um sistema que gere prognósticos precisos. O foco da pesquisa é alcançar uma acurácia robusta, focando na prevenção de eventos críticos, como quedas, infartos e AVCs.

4.1 Base e tratamento de dados

O Human Vitalsigns Dataset, disponível na plataforma [Kaggle](#), é uma base de dados composta por registros de sinais vitais de indivíduos. Cada linha inclui medições de frequência cardíaca (BPM), taxa de respiração (RESP), saturação de oxigênio (SpO₂), temperatura corporal (°C) e um rótulo indicando se os sinais vitais estão normais ou anormais. A quantidade exata de registros não é especificada na documentação oficial, porém ao analisar constatamos 200028 linhas de registros no

documento csv, o dataset é amplamente utilizado em estudos de aprendizado de máquina para classificação de sinais vitais humanos. Pesquisas anteriores, como a de Mihirette (2025), utilizaram versões deste dataset para desenvolver modelos de redes neurais recorrentes (RNNs) com o objetivo de identificar estados de saúde normais ou anormais com alta precisão.

Inicialmente para o tratamento dos dados, foram retirados dados repetidos, eliminando as linhas onde houveram as ocorrências, para conseguir maior precisão. Em seguida, as linhas que possuíam dados faltosos também foram retiradas, visando não atrapalhar o modelo a fazer correlações de maneira errônea, bem como as colunas desnecessárias no contexto do trabalho (Patient ID, Timestamp, Gender). Tendo em vista que o sistema será direcionado ao público idoso, também foram retiradas as linhas de informações onde o paciente possuía idade menor ou igual a 59 anos. Ao final do tratamento do banco de dados, permaneceram 16.626 linhas para o treinamento do sistema.

4.2 Treinamento do modelo

No processo de treinamento do modelo foram consideradas oito variáveis, sendo sete utilizadas como preditoras (features) e uma como variável-alvo (target). A seleção das variáveis independentes foi fundamentada em evidências da literatura que demonstram sua relevância clínica e influência direta sobre o estado geral de saúde e o risco associado ao indivíduo. As variáveis incluídas foram: frequência cardíaca, marcador consolidado de estresse cardiovascular e preditor de morbimortalidade (Fox et al., 2007); temperatura corporal, relacionada à homeostase e ao sistema termorregulatório (Mackowiak, 1997); saturação de oxigênio, parâmetro fundamental para a avaliação da função respiratória (Bickler et al., 2005); pressão arterial sistólica e diastólica, amplamente utilizadas na estratificação de risco cardiovascular (Whelton et al., 2018); além de idade, peso, altura e índice de massa corporal (IMC), este último reconhecido como um dos principais indicadores de risco metabólico e de doenças crônicas, como diabetes tipo 2 e hipertensão (WHO, 2020).

A variável-alvo adotada, denominada Coluna de Risco, classifica os registros do conjunto de dados em uma escala binária, em que 0 representa baixo risco e 1 indica alto risco. Essa estrutura possibilita que o modelo de aprendizado de máquina aprenda padrões fisiológicos associados a diferentes níveis de vulnerabilidade clínica.

O pré-processamento dos dados foi realizado em Python, utilizando bibliotecas amplamente difundidas na comunidade científica, como pandas e NumPy para manipulação e normalização dos dados, e scikit-learn para operações auxiliares de divisão do conjunto em treino e teste, bem como para métricas de avaliação. O algoritmo de aprendizado escolhido foi o Extreme Gradient Boosting (XGBoost), implementado por meio da biblioteca xgboost, devido ao seu desempenho superior

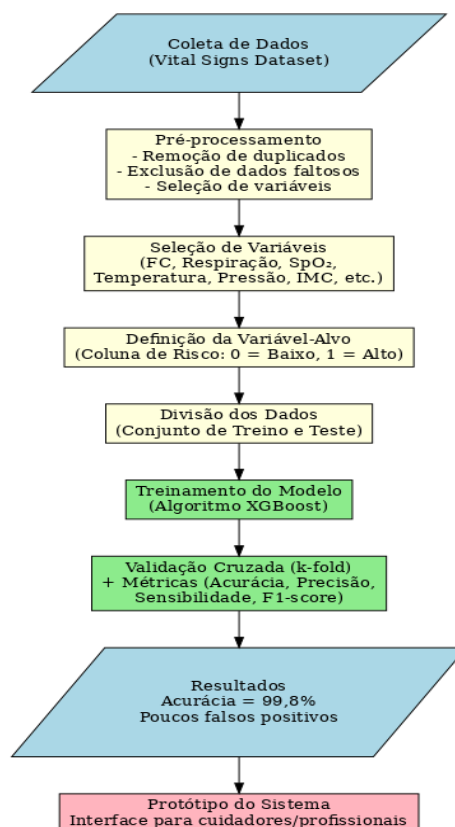
em tarefas de classificação tabular, capacidade de lidar com interações complexas entre variáveis e robustez contra overfitting.

O modelo foi treinado utilizando validação cruzada (*k*-fold cross-validation) para garantir maior generalização e confiabilidade dos resultados. A métrica principal de desempenho foi a acurácia, complementada por indicadores adicionais como precisão, sensibilidade e F1-score, obtidos por meio da geração da matriz de confusão.

A escolha das variáveis foi guiada por sua representatividade de diferentes sistemas fisiológicos — cardiovascular, respiratório, metabólico e termorregulatório — permitindo ao modelo uma visão abrangente do estado de saúde. Além disso, considerou-se a viabilidade de coleta em ambientes clínicos e domiciliares, reforçando o caráter prático da proposta. O monitoramento desses indicadores é reconhecido como fundamental em processos de triagem clínica e acompanhamento de saúde, sobretudo em populações idosas, em que tais medidas apresentam maior impacto na capacidade fisiológica de resposta (Gonik, Chester & Rudolph, 2010).

O funcionamento do sistema desenvolvido pode ser visualizado de forma resumida no fluxograma apresentado na Figura 1, que descreve desde a coleta e tratamento dos dados até a obtenção dos resultados e o protótipo final.

Figura 1- Fluxograma do funcionamento do sistema C.A.I:



4.3 Resultados obtidos

Após o treinamento do modelo, os testes retornaram bons parâmetros, o modelo apresentou em Acurácia 99,8% de eficiência, com o parâmetro de avaliação sendo a matriz de confusão o modelo se mostrou robusto e de 16626 casos testados, o teste apontou apenas 32 ocasiões onde o modelo fez o registro de falsos positivos para alto risco. Por tratar-se de um protótipo para este projeto de pesquisa, ainda não foram realizados testes em situações cotidianas e/ou clínicas. Os testes foram realizados em uma CPU formada por um processador core i7 de 4ª Geração de 2.0GHz e dual core, com 8gb de ram. Garantindo a replicabilidade e acessibilidade de uso em sistemas não tão potentes.

5.0 Conclusão

O sistema desenvolvido apresentou resultados promissores no ramo de previsões do estado de saúde de pessoas idosas, alcançando acurácia de 99,8% na predição de riscos a partir de sinais vitais. Esses resultados indicam o potencial da proposta como ferramenta de apoio preventivo e assistivo. No entanto, ainda se faz necessário validar o modelo em contextos reais e comparar com os modelos já utilizados no cotidiano nacional, garantindo maior generalização e impacto prático no cuidado à população idosa. Apesar dos resultados positivos, o sistema ainda se encontra em estágio inicial, restrito a testes em ambiente controlado. Para validações futuras, torna-se essencial a aplicação em cenários reais, considerando variáveis contextuais como diferenças individuais de pacientes e condições clínicas específicas. Ademais, também se faz necessário expandir os testes para outros conjuntos de dados, para garantir maior generalização do modelo. Outrossim, trabalhos futuros podem trazer o sistema atual integrado a dispositivos wearables, como smartwatches, tornando o controle de saúde ainda mais prático ao coletar os dados fisiológicos automaticamente.

REFERÊNCIAS:

IBGE – Instituto Brasileiro de Geografia e Estatística. Censo 2022: número de pessoas com 65 anos ou mais de idade cresceu **57,4 % em 12 anos**. *Portal Agência de Notícias IBGE*, 27 out. 2023. Disponível em: agenciadenoticias.ibge.gov.br/agencia-noticias/noticias/38186-censo-2022-numero-d-e-pessoas-com-65-anos-ou-mais-de-idade-cresceu-57-4-em-12-anos. Acesso em: 23 ago. 2025.

IBGE – Instituto Brasileiro de Geografia e Estatística. Censo 2022: número de pessoas com 60 anos ou mais de idade chegou a 32.113.490 (15,6 % da população), incremento de 56 % em relação a 2010. *Portal Agência de Notícias IBGE*, 27 out. 2023. Disponível em: agenciadenoticias.ibge.gov.br/agencia-noticias/noticias/38186-censo-2022-numero-d-e-pessoas-com-65-anos-ou-mais-de-idade-cresceu-57-4-em-12-anos. Acesso em: 23 ago. 2025.

IBGE – Instituto Brasileiro de Geografia e Estatística. Censo 2022: índice de envelhecimento — razão entre idosos (≥ 65 anos) e crianças (0–14 anos) — saltou de 30,7 em 2010 para **55,2 em 2022**. *Portal Agência de Notícias IBGE*, 27 out. 2023. Disponível em: agenciadenoticias.ibge.gov.br/agencia-noticias/noticias/38186-censo-2022-numero-d-e-pessoas-com-65-anos-ou-mais-de-idade-cresceu-57-4-em-12-anos. Acesso em: 23 ago. 2025.

BARBOSA, Vinícius dos Santos et al. Predicting all-cause mortality with machine learning among Brazilians aged 50 and over: results from the Brazilian Longitudinal Study of Ageing (ELSI-Brazil). *BMC Medical Research Methodology*, v. 25, n. 1, p. 1-14, 2025. DOI: <https://doi.org/10.1186/s12874-024-02442-2>.

AYUB, Nasir. *Human Vital Sign Dataset*. Disponível em: <https://www.kaggle.com/datasets/nasirayub2/human-vital-sign-dataset>. Acesso em: 23 ago. 2025.

MIHIRETTE, S. *Intent Recognition Using Recurrent Neural Networks on Vital Sign Data: A Machine Learning Approach*. 2025. Disponível em: https://www.researchgate.net/publication/373470912_Intent_Recognition_Using_Recurrent_Neural_Networks_on_Vital_Sign_Data_A_Machine_Learning_Approach. Acesso em: 23 ago. 2025.