

# Classificação de Eletrocardiograma com Integração de Modelos de Linguagem Grande e Técnicas de Geração Aumentada de Recuperação para Suporte à Decisão Médica

Ademar Alves Trindade

Departamento de Ciência da Computação  
Universidade do Estado de Mato Grosso (UNEMAT) - Cáceres, MT - Brasil

ademar.alves@unemat.br

**Abstract.** *Cardiovascular disease is one of the world's leading killers, and the most accurate prognosis depends on early detection. Retrieval Augmented Generation (RAG) and Large Scale Language Models (LLM) support diagnostics, improving the detection of heart disease. The system incorporates TensorFlow Keras models to classify ECGs and medical reports, helping healthcare professionals analyze complex data. The combination of automated analysis and clinical history provides a comprehensive and accurate diagnostic tool.*

**Resumo.** *Um dos maiores fatores de morte do planeta são as enfermidades cardiovasculares, e os prognósticos mais precisos dependem da detecção precoce. As Técnicas de Geração Aumentada por Recuperação (RAG) e Modelos de Linguagem de Grande Escala (LLM) oferecem suporte ao diagnóstico, aprimorando a detecção de enfermidades cardíacas. Para classificar ECGs e relatórios médicos, o sistema incorpora modelos TensorFlow Keras, auxiliando profissionais de saúde na análise de dados complexos. A combinação de análises automatizadas com histórico-clínico fornece uma ferramenta diagnóstica abrangente e precisa.*

## 1. Introdução

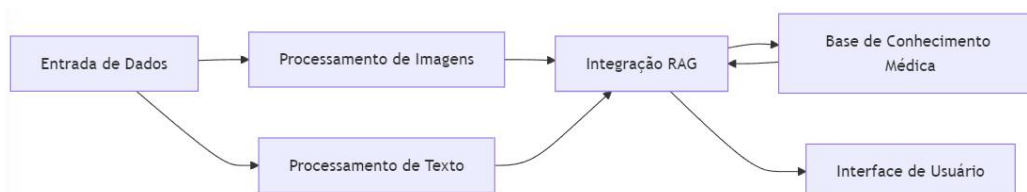
As doenças cardiovasculares representam um grande desafio para a saúde pública, sendo responsáveis por aproximadamente 17,9 milhões de mortes anuais, o que corresponde a 32% das mortes globais (OMS). De acordo com Da Silva et al. (2024), as lesões mais frequentes nessas condições são arritmias, insuficiências cardíacas e ataques cardíacos. Em relação à qualidade de vida, os pacientes e suas famílias sofrem menos do que o enfrentamento de enfermidades de alta prevalência. Bourdette (2024) relata que aos cuidados médicos, hospitalares e tratamentos de longo prazo para enfermidades cardíacas são uma porcentagem significativa dos orçamentos da saúde em diversas nações, sendo necessárias medidas eficazes de prevenção e tratamento. Outro elemento crucial é a complexidade relacionada ao diagnóstico precoce das enfermidades cardíacas. De acordo com Martins (2024), a detecção precisa e rápida é fundamental para aumentar os prognósticos e reduzir a mortalidade associada. Ações médicas oportunas que podem evitar complicações graves e melhorar a qualidade de vida dos pacientes são possibilitadas pela identificação precoce.

Em suma, as doenças cardíacas desempenham um papel crítico na saúde pública, tanto em mortalidade quanto em custos financeiros e sociais. A complexidade dos dados médicos, a necessidade de diagnósticos rápidos e a variabilidade na interpretação são desafios atuais no diagnóstico cardíaco.

Neste artigo, propomos uma abordagem que integra diversas tecnologias para auxiliar no diagnóstico de doenças cardíacas. Na combinação com prontuários eletrônicos, utilizamos uma base de dados de exames cardíacos que abrange o histórico médico de atendimento dos pacientes. A metodologia envolve o desenvolvimento de um assistente de diagnóstico utilizando um Modelos de Linguagem Grande (LLM), cuja integração com os dados do paciente e os resultados dos exames realizados no modelo TensorFlow Keras, conforme descrito por Nazi & Peng (2024). Essa integração é aprimorada usando técnicas de geração aumentada de recuperação (RAG), que permitem enriquecer a interpretação dos dados, promovendo uma análise mais detalhada conforme Iqbal et al. (2024). O objetivo consiste em fornecer uma ferramenta adequada ao trabalho dos especialistas em saúde, a fim de auxiliar na análise de dados complexos e em simplificar o diagnóstico, servindo sempre como um recurso adicional, sem substituir a capacidade e o julgamento clínico do médico.

## 2. Materiais e Métodos

Nesta seção, detalhamos a metodologia empregada para desenvolver o assistente de diagnóstico de doenças cardíacas, aproveitando relatórios médicos e imagens. A metodologia é dividida em várias etapas, desde a coleta e pré-processamento dos dados até a integração dos modelos de classificação com ferramentas de processamento de linguagem natural. Para fornecer uma compreensão clara das ferramentas e procedimentos adotados, cada etapa é explicada.



Figural - Visão geral da proposta

### 2.1 Conjunto de Dados

Adotou-se o conjunto de dados de Arritmia MIT-BIH, foi desenvolvido pelo Massachusetts Institute of Technology (MIT) em colaboração com o Beth Israel Hospital (BIH) Moody (2001). O conjunto de dados foi anotado no nível de batimento cardíaco por dois ou mais cardiologistas de forma independente. 14 tipos de batimento cardíaco originais são consolidados em 5 grupos de acordo com a recomendação da Associação para o Avanço da Instrumentação Médica. Conforme citado por Kachuee (2018), a variedade e amplitude dos tipos de músculo cardíaco encontrados na base do MIT-BIH, faz com que este conjunto de dados seja um dos mais utilizado em estudos de análise automatizada de ECG conforme relatado por Luiz et al. (2011), permitem que os modelos treinados generalizem bem para uma variedade de cenários clínicos.

### 2.2 Descrição dos Dados Utilizados

No total, 2.350 imagens de ECG distribuídas entre cinco tipos de batimentos cardíacos (470 imagens por categoria) foram usadas para treinar um algoritmo de classificação,

aprendendo as características específicas de cada tipo. Um conjunto de teste 50 imagens por categoria (totalizando 250 imagens) avaliou o desempenho final do modelo, garantindo generalização e robustez.

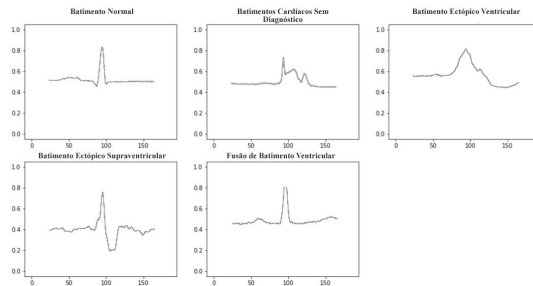


Figura 2 - Sinais de ECG para diferentes tipos de batimentos cardíacos

Para preparar as imagens para o modelo de rede neural, elas foram convertidas para formato RGB, redimensionadas para 224x224 pixels e normalizadas para o intervalo [-1, 1], melhorando o desempenho do treinamento. As imagens pré-processadas foram transformadas em arrays NumPy para predição.

### 2.4 Modelo de Classificação com TensorFlow Keras

Foi desenvolvido uma Rede Neural Convolutiva (CNN) usando TensorFlow Keras. Inicialmente treinada com o Teachable Machine (Colak, 2024), aprimoramos a flexibilidade e eficiência do modelo implementando a camada personalizada CustomDepthwiseConv2D (segundo Sari, 2024), substituindo a camada padrão do TensorFlow e garantindo compatibilidade com o modelo treinado.

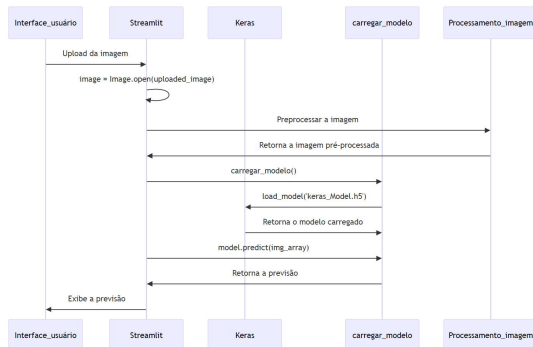


Figura 3 - Fluxo de interações no processo de classificação

A Figura 3 apresenta as etapas de preparação, análise e exibição do sistema, onde o TensorFlow Keras realiza o pré-processamento da imagem de ECG, fazendo os ajustes necessários para o modelo de classificação. Em seguida, o Keras é novamente empregado para carregar o modelo previamente treinado, que analisa a imagem pré-processada e identifica padrões de batimentos cardíacos. Após a classificação, o Streamlit exibe os resultados na interface do usuário, facilitando o acesso e a interpretação dos dados processados. Esse fluxo contínuo destaca como cada componente do sistema colabora para processar a imagem, carregar o modelo de aprendizado de máquina e gerar uma previsão.

### 3. Resultados e Discussões

Modelos de Linguagem Grande (LLMs) e técnicas de Recuperação Aumentada por Geração (RAG) combinam previsões com relatórios médicos textuais, oferecendo interpretação completa. O RAG interpreta resultados e informações dos prontuários, fornecendo insights para profissionais de saúde. A integração facilita a comunicação entre dados dos prontuários e classificações do TensorFlow Keras, permitindo considerar análises automatizadas e histórico clínico, proporcionando visão abrangente do estado de saúde via dashboard interativo.

A acurácia do classificador, avaliada por matriz de confusão, mostrou bom desempenho na identificação de batimentos normais e ectópicos supraventriculares, mas confusão na classificação de batimentos ventriculares. Isso sugere acurácia aceitável, mas necessidade de aprimoramento para reduzir erros em categorias específicas.

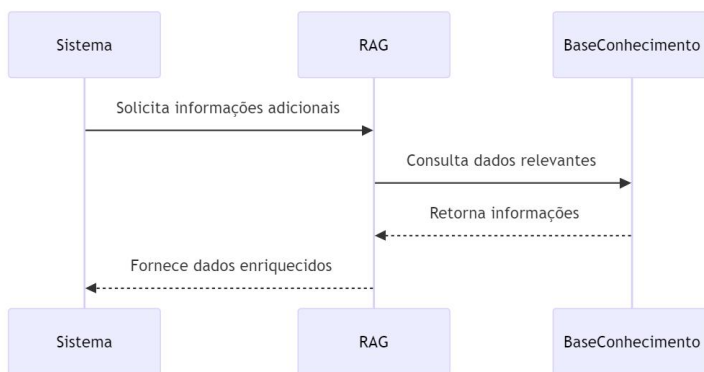


Figura 4 - Diagrama de sequência para consulta RAG

O diagrama acima ilustra como a interface obtém dados do RAG via a base de conhecimento, resultando em análise clínica mais completa para suporte ao diagnóstico. Combinando relatórios médicos com resultados do TensorFlow Keras, é fornecida uma visão abrangente do estado de saúde do paciente, permitindo recomendações diagnósticas mais precisas.

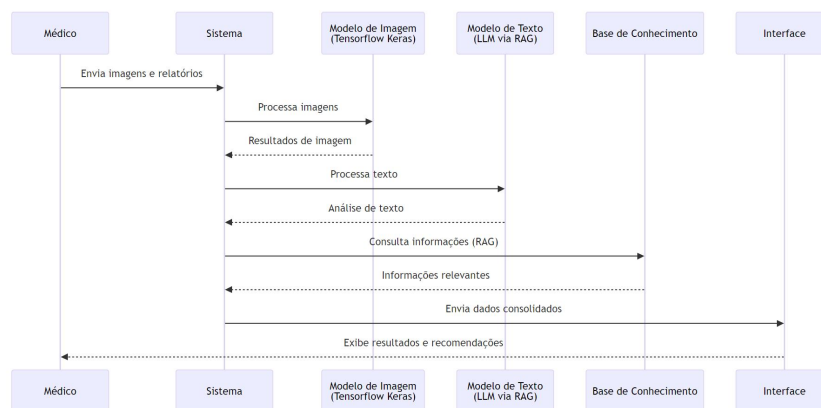


Figura 5 - Fluxo de dados detalhado

A Figura 5 detalha o fluxo de dados, onde, após a classificação, os resultados são integrados ao histórico médico de atendimento do paciente utilizando técnicas de Recuperação Aumentada por Geração. Finalmente, os resultados combinados são

apresentados ao usuário em uma interface interativa desenvolvida em Streamlit, proporcionando uma visualização intuitiva e rápida das informações analisadas.

Para a aplicação final, foi criada uma interface visual interativa utilizando o framework Streamlit. A escolha do Streamlit se deu à sua facilidade de uso e rápida implementação, permitindo a criação de uma interface voltada para a classificação de ECG com integração de LLMs e RAG. À esquerda, a interface permite o upload de arquivos para análise; ao centro, exibe o tipo de batimento cardíaco e a confiança da predição; e à direita, apresenta a "Resposta do Classificador" com uma interpretação básica e outra seção com informações de relatório médico. A interface intuitiva facilita o acesso a dados claros e contextualizados, apoiando o Suporte à Decisão Médica.

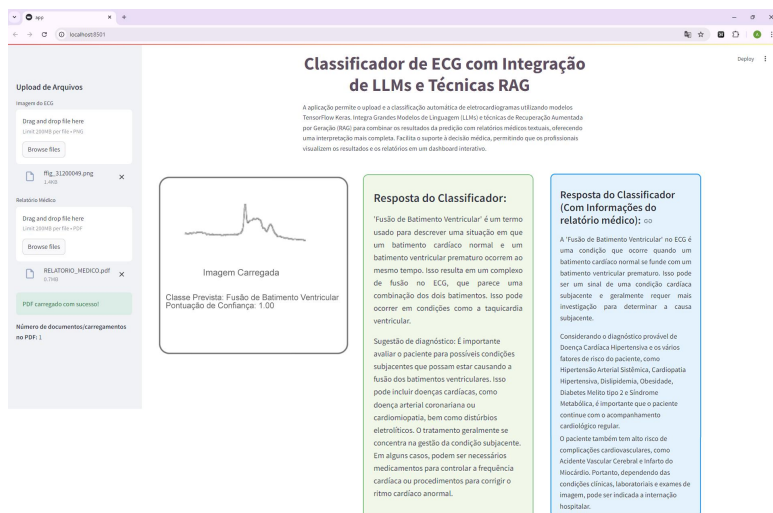


Figura 6 - Interface front-end da aplicação criada utilizando o framework Streamlit

#### 4. Considerações Finais

A integração de modelos de aprendizado profundo com técnicas de Recuperação Aumentada por Geração (RAG) mostrou-se eficaz na análise de imagens de ECG e relatórios médicos textuais. Utilizando um modelo TensorFlow Keras para classificar batimentos cardíacos, integrado a LLMs para interpretação contextual, a ferramenta oferece suporte poderoso à análise clínica, fornecendo aos profissionais de saúde uma visão mais abrangente dos resultados, sem substituir a avaliação clínica especializada.

No entanto, LLMs ainda não são reconhecidos para uso clínico por entidades regulatórias devido a preocupações com precisão e confiabilidade (Nazi, 2024). Além disso, o uso desses modelos com dados sensíveis de pacientes apresenta riscos à privacidade, podendo comprometer a confidencialidade mesmo quando anonimizados (Hacker, 2023). Nesse contexto, a Lei Geral de Proteção de Dados Pessoais (LGPD) pode ser aplicada.

#### 5. Referências Bibliográficas

BOURDETTE, Marcelo Segalerba. A brucelose humana no Brasil sob a perspectiva da saúde única. 2024.

- COLAK, Ertugrul et al. Análise de desempenho de modelos de rede neural convolucional pré-treinados para classificação de doenças oftalmológicas. *Arquivos Brasileiros de Oftalmologia*, v. 87, n. 5, p. 1-10, 2024.
- DA SILVA MACHADO, Robert Douglas; ARAÚJO, Mirelia Rodrigues; CAVALCANTE, Katiane Serra. Complicações em pacientes cardiopatas no pós-COVID-19: revisão de literatura. *Brazilian Journal of Health Review*, v. 7, n. 2, p. e69170-e69170, 2024.
- GAO, Y. et al. Retrieval-augmented generation for large language models: A survey. Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, 2023.
- HACKER, Philipp; ENGEL, Andreas; MAUER, Marco. Regulating ChatGPT and other large generative AI models. In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 2023. p. 1112-1123.
- IQBAL, JL Mazher et al. A novel deep learning approach for early detection of cardiovascular diseases from ECG signals. *Medical Engineering & Physics*, v. 125, p. 104111, 2024.
- KACHUEE, Mohammad; FAZELI, Shayan; SARRAFZADEH, Majid. Ecg heartbeat classification: A deep transferable representation. In: *2018 IEEE international conference on healthcare informatics (ICHI)*. IEEE, 2018. p. 443-444.
- LUZ, Eduardo; MENOTTI, David. How the choice of samples for building arrhythmia classifiers impact their performances. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011. p. 4988-4991.
- MARTINS, Georgiton Carvalho et al. TAMPONAMENTO CARDÍACO: ABORDAGEM DIAGNÓSTICA E TERAPÊUTICA NA MEDICINA INTENSIVA. *Brazilian Journal of Implantology and Health Sciences*, v. 6, n. 8, p. 4030-4037, 2024.
- MOODY, George B.; MARK, Roger G. The impact of the MIT-BIH arrhythmia database. *IEEE engineering in medicine and biology magazine*, v. 20, n. 3, p. 45-50, 2001.
- NAZI, Zahir Al; PENG, Wei. Large language models in healthcare and medical domain: A review. In: *Informatics*. MDPI, 2024. p. 57.
- SARI, Dewi Mutiara et al. Classification of Pupil Turbidity Based on Convolutional Neural Network (CNN) as an Early Detection of Cataract Step Using a Smartphone. In: *International Conference on Applied Science and Technology on Engineering Science 2023 (iCAST-ES 2023)*. Atlantis Press, 2024. p. 287-301.
- SOARES, Ronan Pardo. QUERYAUGMENT-RAG: APRIMORANDO A RECUPERAÇÃO E GERAÇÃO DE INFORMAÇÕES EM MODELOS DE LINGUAGEM DE GRANDE ESCALA. *Revista Contemporânea*, v. 4, n. 6, p. e4470-e4470, 2024.