

Um modelo explicável para classificação de arritmias cardíacas utilizando a rede LSTM

Amanda Souza Zírpolo¹, Evandro Tinoco Mesquita², Taiane Coelho Ramos¹

¹ Instituto de Computação – Universidade Federal Fluminense (UFF)
Niterói, Rio de Janeiro, Brasil

²Hospital Universitário Antônio Pedro – Universidade Federal Fluminense (UFF)
Niterói, Rio de Janeiro, Brasil

amandazirpolo@id.uff.br, etmesquita@gmail.com, taiane-ramos@id.uff.br

Resumo. Arritmias cardíacas são anomalias nos impulsos elétricos do coração que resultam em um ritmo alterado, podendo indicar doenças cardíacas potencialmente fatais. No Brasil, essas doenças são a principal causa de morte e, globalmente, representam 16% dos óbitos. O diagnóstico precoce é essencial para aumentar as chances de tratamento e reduzir a mortalidade. O principal exame utilizado é o eletrocardiograma (ECG), um método de baixo custo e rápida obtenção de resultados. No entanto, a identificação de arritmias enfrenta desafios, como a necessidade de análise criteriosa por profissionais especializados, possíveis divergências nas interpretações, interferências nos sinais do ECG e a baixa sensibilidade na detecção de algumas doenças. Para superar essas limitações, sistemas automatizados baseados em aprendizado profundo demonstram um enorme potencial, proporcionando uma análise mais precisa dos sinais cardíacos e aumentando a assertividade do diagnóstico. Dentre as abordagens mais promissoras, redes neurais profundas recorrentes (RNNs) têm demonstrado um excelente desempenho na análise de séries temporais de ECG. Este trabalho obteve 92% de acurácia no conjunto teste classificando 5 classes. Aplicamos o framework SHAP para validar as respostas da rede, fazendo com que a explicabilidade aprimorasse a confiabilidade na identificação e na classificação de arritmias cardíacas. Os pontos destacados pelo modelo como importantes para o diagnóstico da rede foram coerentes com a análise feita pelo especialista da área.

1. Caracterização do problema

Arritmias cardíacas são anormalidades nos impulsos elétricos do coração que alteram seu ritmo e podem indicar doenças graves, muitas vezes fatais. Os principais sintomas incluem palpitações, tonturas e síncope. Segundo a Organização Mundial da Saúde (OMS), as doenças cardiovasculares foram responsáveis por 16% das mortes globais entre 2000 e 2019 [World Health Organization 2019], sendo a principal causa de óbitos no Brasil, com 185.192 mortes registradas em 2016 [Marinho et. al 2018].

Apesar da gravidade, as arritmias podem ser tratadas se diagnosticadas precocemente. O tratamento envolve o uso de medicamentos antiarrítmicos, ablação por cateter e dispositivos elétricos, como marcapassos e desfibriladores implantáveis (CDIs)

[Costa et. al 2024]. Para o diagnóstico, diversos exames são utilizados, como eletrocardiograma (ECG), Holter, ecocardiograma, testes de esforço e ressonância magnética cardíaca [Andalaft 2012], [Cabanelas et. al 2012], [Costa et. al 2024]. Dentre esses, o ECG é amplamente utilizado por ser um exame acessível, de baixo custo e rápida obtenção de resultados. Ele fornece informações detalhadas sobre frequência cardíaca, formato das ondas e possíveis anomalias na condução elétrica do coração [Ansari et. al 2023]. No entanto, a interpretação do ECG apresenta desafios, como baixa sensibilidade na detecção de algumas doenças, suscetibilidade a ruídos e interferências e a necessidade de análise criteriosa por especialistas, o que pode gerar discordâncias diagnósticas.

2. Motivação

O aprendizado de máquina, um ramo da IA, tem sido aplicado para aprimorar o diagnóstico automatizado de arritmias, reduzindo custos com exames e aumentando a precisão do diagnóstico médico [Monard, Baranauskas 2003]. O uso de aprendizado profundo se destaca pela capacidade de extrair informações diretamente dos sinais de ECG, eliminando a necessidade de seleção manual de características. Essas técnicas vêm demonstrando vantagens significativas, como maior confiabilidade na análise dos sinais, monitoramento contínuo e aprimoramento da detecção de arritmias [Ansari et. al 2023].

A correta identificação das classes arrítmicas é essencial para um tratamento adequado conforme a condição cardíaca. No entanto, muitos estudos possuem limitações, como a dificuldade de detectar múltiplas classes de arritmia e contaminação de dados, comprometendo o auxílio aos diagnósticos médicos.

3. Objetivos

Este trabalho tem como objetivo identificar e classificar quatro classes de arritmia e a classe normal, garantindo a não contaminação e a generabilidade dos resultados em uma base de dados pública. Além disso, utilizaremos técnicas de explicabilidade de sinal para melhor compreensão dos resultados da rede a fim de colaborar com a clínica médica.

4. Trabalhos Relacionados

Na literatura, vários modelos de *machine learning* (ML) e de *deep learning* (DL) foram utilizados para realizar a identificação e a classificação automática de arritmias cardíacas em dados de ECG. O estudo de Khan and Kim (2020) utilizou uma rede LSTM para classificar arritmias cardíacas em ECG. Os dados, extraídos do *UCI Machine Learning*, incluem 452 amostras. O conjunto contém 16 classes distintas (0 a 15) e foi dividido em treino (70%), validação (20%) e teste (10%). O modelo alcançou 93% de acurácia na validação e 88% no teste utilizando todas as classes disponíveis na base de dados. Por outro lado, Sraitih et al. (2021) testaram três métodos clássicos de ML: *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) e *Random Forest* (RF). Após normalização e segmentação dos dados e ajuste de hiperparâmetros do modelo, os modelos foram avaliados na base MIT-BIH, alcançando acurácias entre 80% e 84% na validação e até 99% no teste, indicando boa capacidade preditiva.

Embora esses estudos apresentem resultados promissores, existem limitações na abordagem utilizada para o tratamento dos dados. Caso a segmentação e o aumento de dados não sejam realizados adequadamente, é possível que amostras do mesmo paciente

estejam presentes nos conjuntos de treino, validação e teste, o que pode levar a rede a aprender a reconhecer características individuais do paciente em vez de aprender a característica de interesse no sinal, inflando a métrica de acurácia. Esta pesquisa se propõe a desenvolver um modelo de predição sem uso de recursos como técnicas de *data augmentation* e segmentação de dados, garantindo que todos os dados de um paciente estejam em apenas um conjunto (treino, validação ou teste), possibilitando uma melhor avaliação na capacidade de generalização do modelo para dados nunca vistos.

5. Métodos de pesquisa

5.1. Base de dados utilizada

O conjunto de dados utilizado é o *Chapman-Shaoxing* (CUSPH) [Zheng et. al 2020], que contém registros de ECG de 10646 pacientes, sendo 5733 homens e 4514 mulheres, com idades entre 4 e 89 anos. A idade média é de 60 anos e o desvio padrão é de 17,02 anos. Os sinais foram coletados a uma taxa de 500 Hz por 10 segundos utilizando o sistema MUSE da *General Electric* (GE). O conjunto de dados inclui 12 derivações de ECG e abrange 11 classes de arritmia, além de 67 condições cardiovasculares. No presente trabalho, utilizamos apenas 4 classes de arritmia (bradicardia sinusal, taquicardia sinusal, fibrilação atrial e taquicardia supraventricular) e a classe normal.

5.2. Pré-processamento

Reduzimos a taxa de amostragem dos dados de 500Hz para 50Hz considerando 1 a cada 10 pontos do dado. Foram aplicados os filtros *Butterworth Low Pass* e *Butterworth High Pass*, para eliminar frequências fora do intervalo 0,5Hz e 150Hz que impactassem a forma do sinal cardíaco e o filtro *Notch* para eliminar a frequência da linha de energia em 50Hz. Selecionamos para classificação apenas os eletrodos II, V1 e V6. Optamos por manter o conjunto de dados desbalanceado, sendo composto por dados de 4813 pacientes, dos quais 2002 possuem bradicardia sinusal, 644 possuem taquicardia sinusal, 422 possuem fibrilação atrial, 379 possuem taquicardia supraventricular e 1366 não possuem arritmia. Os dados foram divididos em treino, validação, e teste com a proporção de 70%, 20% e 10%.

5.3. Modelo Long Short Term Memory (LSTM)

A arquitetura do modelo possui três camadas: duas de LSTM, uma com 128 e outra com 64 neurônios respectivamente para processar os dados temporais, seguida de uma camada densa de 5 neurônios para a classificação. As saídas da camada LSTM passam pela função de ativação *Tanh* para capturar padrões complexos. O modelo foi treinado por 87 épocas com o *batch size* de 150 e foi selecionada o modelo considerando um equilíbrio entre a melhor acurácia e o menor *loss* na validação.

Configuramos a rede para treinar ao longo de 100 épocas com *batch* de 150 amostras, aplicando a técnica de *Early Stopping* para evitar *overfitting*, fazendo com que a rede treinasse por 87 épocas. A Tabela 1 representa os hiperparâmetros utilizados na rede.

O desempenho da rede foi calculado através da acurácia geral e das métricas de precisão, revocação e *F1-score* para cada classe.

Hiperparâmetros da rede LSTM	
Neurônios na primeira camada de LSTM	128
Neurônios na segunda camada de LSTM	64
Função de ativação das camadas de LSTM	Função Tanh
Neurônios na camada densa	5
Função de ativação da camada densa	Função Softmax
Dropout após a segunda camada de LSTM	0,5
Função de perda	Categorical Crossentropy
Otimizador de parâmetro	Adam

Tabela 1. Hiperparâmetros utilizados na rede LSTM.

5.4. Explicabilidade

Aplicamos o uso de uma técnica de explicabilidade para entender como a rede faz a classificação a partir das características do sinal, permitindo assim a validação dos resultados fornecidos pela rede e a compreensão de como esses resultados podem ser úteis na clínica médica.

Para isso, utilizamos o *framework* de explicabilidade SHAP (*SHapley Additive ex-Planations*, disponível na linguagem *Python* [Scott Lundberg 2018]. O SHAP é baseado na teoria dos valores de *Shapley*, um conceito da teoria dos jogos que distribui de maneira justa a importância de cada variável de entrada na predição de um modelo. O SHAP avalia o impacto de cada entrada ao considerar todas as possíveis combinações de características do sinal interpretado pela rede e calcula a diferença que cada uma faz na predição final.

Os valores de SHAP são calculados usando uma porção dos dados de treino e o explicador utiliza uma porção dos dados de teste. Foram utilizados 2 pacientes de cada classe, totalizando 10 pacientes do treino para o cálculo dos valores de SHAP e 1 paciente de cada classe, totalizando 5 pacientes do teste para o explicador. Utilizamos 5 pacientes para melhor avaliação dos resultados em um diagnóstico único na classe. Uma limitação desse trabalho é o pouco número de amostras utilizados na explicabilidade por motivos de poder computacional.

6. Resultados

Nossa proposta consiste em identificar e classificar 4 classes de arritmia e a classe normal. O modelo apresentou 98% de acurácia na validação. O melhor modelo obtido foi aplicado para classificar o conjunto teste, obtendo 92% de acurácia. A Tabela 2 apresenta os resultados para cada classe no teste e a Figura 1 apresenta a matriz de confusão dos dados de teste. É possível perceber que as classes 0 e 1 apresentaram melhor performance comparadas as classes 3 e 4, devido o menor número de amostras das classes 3 e 4.

O cálculo da importância média permite ao explicador determinar se uma característica teve importância positiva, negativa ou neutra. A importância positiva ocorre quando um ponto do sinal possui valor SHAP positivo, aumentando a probabilidade da rede acertar a classe, como a irregularidade do complexo QRS na fibrilação atrial. Já a negativa ocorre quando o valor SHAP é negativo, aumentando a probabilidade da rede errar a classe, como a variação da onda R, que pode confundir a rede ao se assemelhar a outras arritmias. A importância neutra surge quando o valor SHAP é próximo de zero, indicando irrelevância na decisão do modelo, como o segmento ST, que não impacta a identificação da fibrilação atrial.

Tabela 2. Resultados das métricas de avaliação do teste por classe.

Classe	Precisão	Revocação	F1-Score	Suporte
Normal (0)	0,93	0,89	0,91	274
Bradicardia Sinusal (1)	0,95	0,98	0,97	441
Taquicardia Sinusal (2)	0,92	0,90	0,91	130
Fibrilação Atrial (3)	0,71	0,75	0,73	85
Taquicardia Supraventricular (4)	0,89	0,84	0,87	77
Acurácia	0,92			
Média Macro	0,88	0,87	0,88	1007
Média Ponderada	0,92	0,92	0,92	1007

Matriz de Confusão

0	244	12	9	9	0
1	5	433	0	3	0
2	6	0	117	2	5
3	8	9	1	64	3
4	0	0	0	12	65
	0	1	2	3	4

Verdadeiro

Predito

Figura 1. Matriz de confusão por classes do conjunto de teste.

A Figura 2 ilustra a resposta SHAP para o diagnóstico de fibrilação atrial. Nos eletrodos II e V6, a primeira característica de fibrilação atrial é perceptível, descrita pelo intervalo RR irregular, ou seja, as distâncias entre os picos da onda R são diferentes entre um complexo QRS e outro. Outro fator importante é a ausência da onda P bem definida, indicando a contração atrial desorganizada, responsável por causar irregularidade nos intervalos RR. Os pontos em vermelho são indicados pelo explicador como positivos, ocorrendo frequentemente onde deveria existir uma onda P bem definida, inferindo que a rede foi capaz de aprender as características de um diagnóstico de fibrilação atrial. Os pontos em azul são indicados pelo explicador como negativos, como a variação da onda R, e os pontos em branco são indicados pelo explicador como nulo, ou seja, não apresentaram impacto na análise feita.

O resultado da explicabilidade é coerente com o observado no dia a dia nos hospitais, inferindo a boa capacidade da rede LSTM em detectar arritmias em dados nunca vistos. A aplicação de sistemas de detecção automáticos tem potencial em cenários de emergências hospitalares que não possuem cardiologistas 24h, auxiliando o profissional na tomada de decisões.

7. Conclusão

Este trabalho se propôs a identificar 5 classes de arritmias utilizando um modelo LSTM em dados públicos de ECG. Obtivemos uma acurácia de 98% na validação e 92% no teste, indicando confiabilidade na classificação das arritmias avaliadas. Além disso,

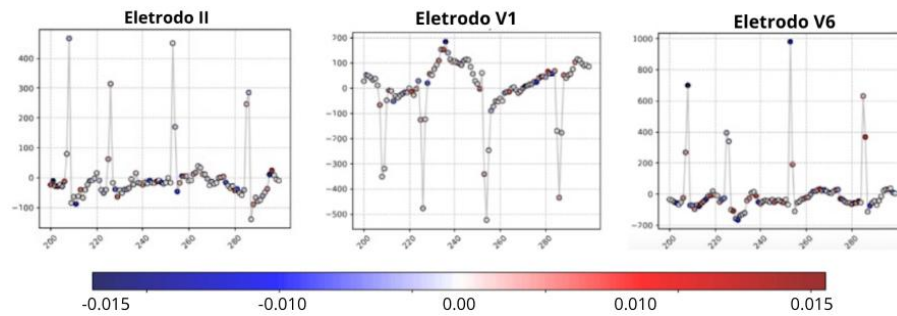


Figura 2. Trecho do ECG do paciente JS0947 com diagnóstico de fibrilação atrial, exibindo a análise de explicabilidade do SHAP. Os painéis da imagem representam trechos de 100 pontos dos eletrodos II, V1 e V6, destacando as regiões mais influentes para a predição do modelo.

utilizamos o framework *SHAP* para a explicabilidade dos dados de ECG a fim de validar os resultados da rede, que se mostrou coerente com as características observadas no ECG.

8. Agradecimentos

Ao PIBIC (Programa Institucional de Bolsas de Iniciação Científica) e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), pela oferta da bolsa de iniciação científica da aluna.

Referências

- Andalaft (2012). The use of non-invasive diagnostic methods for arrhythmias in childhood. *Journal of Cardiac Arrhythmias*.
- Ansari et. al (2023). Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology*.
- Cabanelas et. al (2012). Added value of cardiac magnetic resonance in etiological diagnosis of ventricular arrhythmias. *Portuguese Journal of Cardiology*.
- Costa et. al (2024). Arritmias Cardíacas: Diagnóstico, Tratamento e Prevenção. *Brazilian Journal of Implantology and Health Sciences*.
- Khan and Kim (2020). Cardiac Arrhythmia Disease Classification Using LSTM Deep Learning Approach. *Tech Science Press*.
- Marinho et. al (2018). Burden of disease in Brazil, 1990–2016: a systematic subnational analysis for the Global Burden of Disease Study 2016. *Elsevier*.
- Monard, Baranauskas (2003). Conceitos sobre Aprendizado de Máquina. Manole.
- Scott Lundberg (2018). SHAP documentation.
- Sraitih et. al (2021). An Automated System for ECG Arrhythmia Detection Using Machine Learning Techniques. *Journal of Clinical Medicine*.
- World Health Organization (2019). Leading causes of death and disability: A visual summary of global and regional trends 2000-2019. Acesso em abril, 2024.
- Zheng et. al (2020). A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Nature*.