



CLASSIFICAÇÃO DE FIBRILAÇÃO ATRIAL INTRACARDÍACA UTILIZANDO MACHINE LEARNING

Luís Fillype da Silva*¹, Jonathan Araújo Queiroz¹, Allan Kardec Barros¹

¹UFMA – Universidade Federal do Maranhão

Resumo - O eletrocardiograma é um exame que quantifica a atividade elétrica do coração, permitindo detectar o ritmo cardíaco, o número de batimentos por minuto e o diagnóstico de diversas patologias cardíacas. Este artigo tem como objetivo obter um modelo de classificação baseado nos batimentos de dois grupos de indivíduos: saudáveis e não saudáveis. A metodologia de extração de características foi usada e adaptada para a classificação da fibrilação atrial. As classificações foram realizadas no espaço bidimensional e tridimensional no banco de dados, obtendo uma acurácia de 97% a 100%.

Palavras-Chave - Aprendizado de Máquina, ECG, Estatísticas de Alta Ordem.

CLASSIFICATION OF INTRA CARDIAC ATRIAL FIBRILLATION USING MACHINE LEARNING.

Abstract - The electrocardiogram is an exam that quantifies the electrical activity of the heart, allowing the detection of the heart rhythm, the number of beats per minute and the diagnosis of various cardiac pathologies. This article aims to obtain a classification model based on the beats of two groups of individuals: healthy and unhealthy. The feature extraction methodology was used and adapted for the classification of atrial fibrillation. Classifications were performed in two-dimensional and three-dimensional space in the database, obtaining an accuracy of 97% to 100%.

Keywords - Machine Learning, ECG, High Order Statistics.

I. INTRODUÇÃO

A importância do diagnóstico eficaz das doenças cardíacas tem impulsionado o desenvolvimento de métodos computacionais para a detecção mais precisa dessas doenças cardíacas.

Segundo a Organização Mundial da Saúde [1], as principais causas de morte no mundo hoje estão fortemente relacionadas aos problemas decorrentes do coração. A fibrilação atrial (FA), a arritmia cardíaca sustentada mais comum na prática

clínica, é um problema de saúde pública importante e crescente. Estima-se que cerca de 2,2 milhões de pessoas sejam afetadas nos EUA e quase 4,5 milhões na Europa. Sua prevalência varia com a idade, afetando cerca de 10% dos pacientes com mais de 80 anos. Considerando o fator idade, estudos projetam um aumento de até cinco vezes na prevalência de FA no ano de 2050 [4].

Assim, a análise do eletrocardiograma (ECG) torna-se de suma importância para a detecção prévia de anormalidades, pois é uma das técnicas não invasivas mais utilizadas para auxiliar no diagnóstico da Fibrilação Atrial [4], bem como para evidenciar a propensão ao desenvolvimento ou a existência de outras doenças deste tipo [4].

Trabalhos como o de Queiroz et. al. [3] analisam a variação da tensão que ocorre em um intervalo t de batimento cardíaco usando curtose. No artigo de Kachuee, Fazeli, Sarrafzadeh (2018) [2] é proposto um método baseado em redes neurais convolucionais profundas para a classificação dos batimentos cardíacos capaz de classificar com precisão cinco arritmias diferentes.

Este artigo se propõe a extrair todo o batimento cardíaco de um ECG e agrupar dois grupos usando estatísticas de alta ordem, posteriormente realizando a classificação em dois algoritmos de Aprendizado de Máquina.

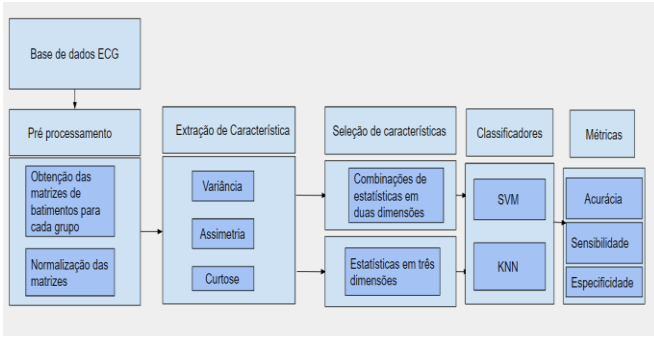
II. MATERIAIS E MÉTODOS

Na Figura 1, é ilustrada a metodologia seguida para a elaboração deste artigo. Primeiramente, foram definidas as bases de dados a serem utilizadas, separando-as em dois conjuntos de dados: indivíduos que apresentam ritmo sinusal normal e indivíduos que apresentam fibrilação atrial.

Em seguida, foi realizado o pré-processamento dos sinais do banco de dados e organização dos dados para a etapa de extração de recursos. Nessa etapa, são calculados os valores de variância, assimetria e curtose dos conjuntos de dados de cada base. Após essa etapa, foram selecionadas combinações de características, que são representadas pelas estatísticas citadas, e colocadas como entrada para os classificadores, dividindo os dados com a validação cruzada, em treino e teste. Ao final do processo, os valores da métrica de classificação são retornados para avaliação do algoritmo.

*silvaluis_@outlook.com

Figura 1: Metodologia proposta para a realização do trabalho.



A. Base de dados

Os banco de dados utilizados foram o The MIT-BIH Normal Sinus Rhythm Database e o Intracardiac Atrial Fibrillation Database, ambas disponíveis em [4][5]. O banco de pacientes com ritmo sinusal normal contém 18 registros, e destes foram utilizados 13 para esta análise. Já a base de dados de pacientes com fibrilação atrial intracardíaca, tirados de ECG intracardíaco, contém 8 pacientes, sendo destes utilizados todos.

De ambas as bases, foram selecionados os sinais de ECG característicos da derivação DII, a mais utilizada no mundo. Foram selecionados toda a duração do sinal, amostrados a uma frequência de 256 Hz para extração dos batimentos.

B. Pré processamento

Foram selecionadas a derivação DII do ECG de cada paciente, para análise e a posterior extração de características. Após isso, segmentou-se cada sinal selecionados de forma a obter todo o respectivo batimento.

Agrupou-se os batimentos de cada grupo, sendo gerada uma matriz X pela concatenação dos batimentos do grupo saudável, e uma matriz Y de batimentos do grupo com Fibrilação Atrial intracardíaca, como descrito na Equação 1 abaixo.

$$B_{n,m} = [B_{n,a} \ B_{n,b} \ \dots \ B_{n,z} \] \quad (1)$$

Onde:

- n - Número de batimentos.
- m - Número de colunas.
- B - Matriz que contém os batimentos concatenados.

Após o agrupamento, realizou-se a subtração do conjunto obtido pela média, dividindo pela entropia de Shannon do sinal, segundo a Equação 2.

$$B_{n,m} = \frac{B_{n,m} - \frac{1}{N} \sum_1^N B_{n,m}}{-\sum_1^m P_{n,m}(\log_2 \frac{1}{N})} \quad (2)$$

Onde:

- N - Número de batimentos
- P - Probabilidade associada a cada batimento.
- B - Matriz de batimentos da classe

C. Extração de características

Nesta etapa, utilizou a metodologia de extração por meio de estatística de alta ordem, até a quarta, proposta em [3]. Por meio das equações a serem descritas abaixo, obteve-se um vetor para cada uma das estatísticas associadas: variância, curtose, e assimetria. As Equações 3, 4 e 5 representam-nas, respectivamente.

$$\sigma_X^2 = E(X^2) - ((E(X))^2) \quad (3)$$

Onde:

- σ_X^2 - Variância.
- $E(X)$ - Os valores esperados de X.

$$\lambda_X = E[((X - E(X))\sigma^{-1})^3] \quad (4)$$

Onde:

- λ_X - Assimetria.
- $E(X)$ - Os valores esperados de X.

$$\kappa_X = E[((X - E(X))\sigma^{-1})^4] \quad (5)$$

Onde:

- κ_X - Curtose..
- $E(X)$ - Os valores esperados de X.

D. Métricas de avaliação

Foram usadas como métrica de avaliação dos modelos dos classificadores os valores de acurácia, sensibilidade e especificidade, definidos pelas Equação 6, Equação 7 e Equação 8.

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \times 100 \quad (6)$$

$$Sensibilidade = \frac{VP}{VP + FN} \times 100 \quad (7)$$

$$Especificidade = \frac{VN}{VN + FP} \times 100 \quad (8)$$

Onde:

- VP - verdadeiros positivos.

- VN - verdadeiros negativos.
- FP - falso positivos.
- FN - falso negativos.

E. Validação cruzada

A validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo, a partir de um conjunto de dados. Neste artigo, a divisão dos dados é usando o método holdout, que consiste em dividir os dados em 70 e 30 de maneira aleatória. Utilizou-se 70% dos dados para treino e 30% para teste, a fim de evitar valores discrepantes.

III. RESULTADOS

Este trabalho analisou os batimentos extraídos do ECGs de vários pacientes saudáveis com fibrilação atrial, a fim de classificá-los.

Para a etapa de classificação, foram geradas matrizes, onde cada coluna é representada por variância, assimetria e curtose, respectivamente. Tais matrizes foram as entradas dos classificadores K-mais próximos Neighbours (KNN) e Support Vector Machine (SVM) para verificar quais momentos estatísticos apresentam maior acurácia, sensibilidade e especificidade.

Além disso, foi apresentado um conjunto de estatísticas de alta ordem para cada sinal a fim de obter qual combinação proporciona maior eficiência no processo de classificação, onde σ_x^2 representa a variância dos batimentos dos indivíduos na base de dados, κ_x a curtose da base e λ_x a assimetria dos dados. A seguir, nas Figura 2 e Figura 3, são apresentados os batimentos de um sinal de cada banco de dados utilizado. Na Figura 2 são apresentados os batimentos de um indivíduo com batimentos saudáveis e, na Figura 3 os batimentos de um indivíduo pertencente ao grupo de fibrilação atrial intracardíaca.

Figura 2: Batimentos saudáveis

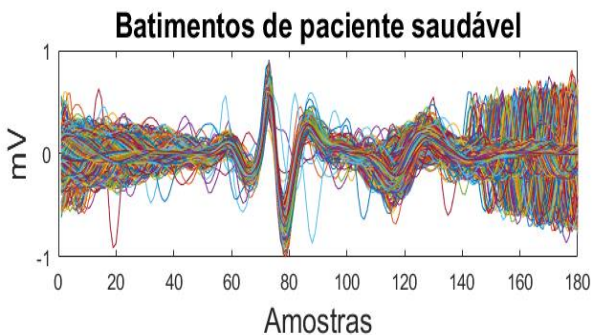
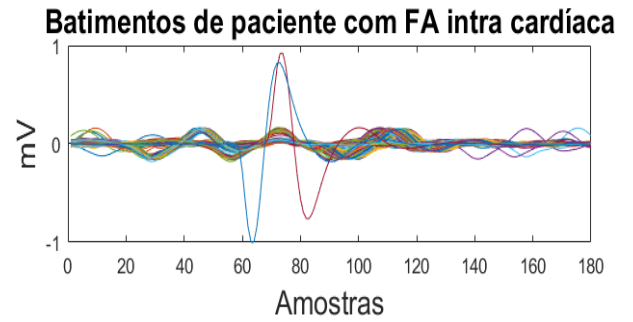


Figura 3: Batimentos com Fibrilação Atrial Intra cardíaca..



Os resultados estão expostos na Tabela I, Tabela II e Tabela III.

Tabela 1: Métricas de avaliação com Variância x Curtose

Estatísticas	KNN	SVM
Acurácia	100%.	100%.
Sensibilidade	100%.	100%.
Especificidade	87%.	90%.

Tabela 2: Métricas de avaliação com Variância x Assimetria x Curtose

Estatísticas	KNN	SVM
Acurácia	100%.	97%.
Sensibilidade	100%.	96.5%.
Especificidade	85%.	87%.

Tabela 3: Métricas de avaliação com Variância x Assimetria

Estatísticas	KNN	SVM
Acurácia	87%.	74%.
Sensibilidade	92%.	90%.
Especificidade	83%.	81%.

IV. DISCUSSÕES

Este trabalho apresentou análise das características em duas e três dimensões. Na Figura 4, a análise em duas dimensões mostra a separação dos dados por variância e curtose, conforme mostrado na Tabela. I, II e III, e também reafirmando os resultados discutidos em Queiroz et. al. [3] e em Lucena et. al. [6], que apontam que a curtose pode ser uma abordagem adequada para medir sinais esparsos, como o ECG.

Na Figura 5, a combinação de curtose e assimetria não alcançou uma separação muito evidente entre os grupos. Isto porque no plano cartesiano, os valores de assimetria das duas classes ficam bem próximos entre si, resultando no não agrupamento. Na Figura 6, usando a variância e assimetria dos dados, uma melhor separação pode ser obtida.

Figura 4: Variância x Curtose.

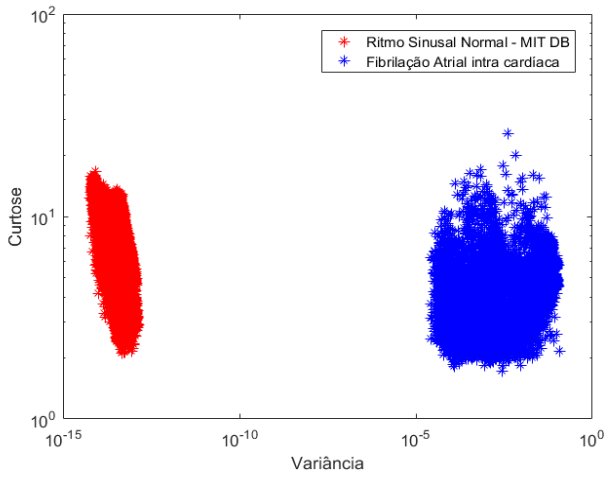


Figura 5: Curtose x Assimetria.

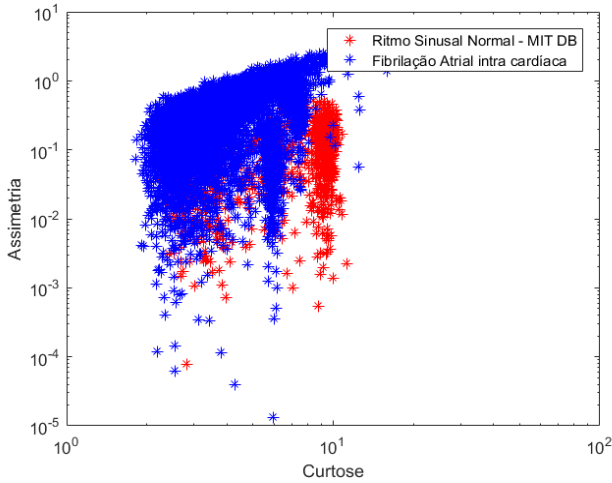
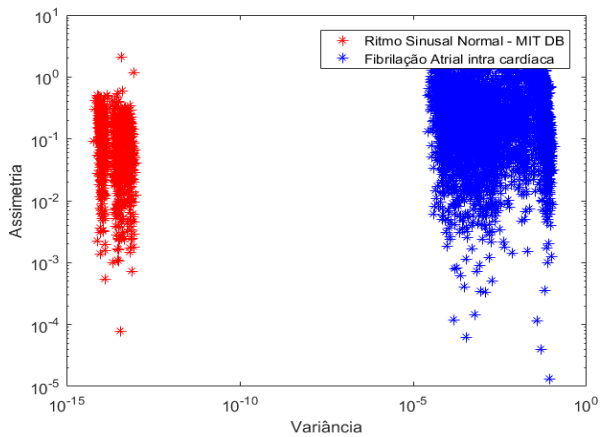
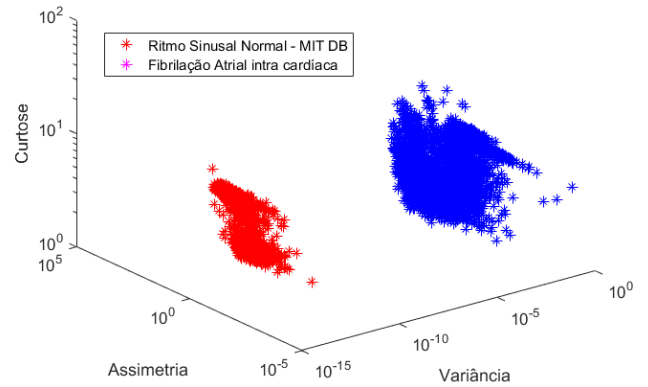


Figura 6: Variância x Assimetria.



Observou-se que as combinações que utilizam variância e curtose em suas representações das características obtiveram maior separação entre os dois grupos. Isso ocorre porque o ECG de indivíduos com fibrilação atrial apresenta a maior variância, enquanto indivíduos com ritmo sinusal normal apresentam menor variância nos dados. Na Figura 7, é mostrada uma representação tridimensional, levando em consideração a variância, assimetria e curtose do conjunto de batimentos de cada grupo. Percebe-se que embora uma representação 3D proporcione uma melhor visualização dos dados, a acurácia de usar apenas duas características teve um resultado maior nos classificadores utilizados.

Figura 7: Variância x Assimetria x Curtose.



V. CONCLUSÕES

Este artigo traz um método que reforça outros estudos que mostram a eficácia do uso de estatísticas de alta ordem para extrair características e classificar as cardiopatias, como a fibrilação atrial. Além disso, os resultados obtidos pelo método apresentado neste artigo podem ser utilizados para auxiliar na tomada de decisão em processos clínicos, além de detectar arritmias de forma autônoma.

Em trabalhos futuros, diferentes doenças cardiovasculares e técnicas podem ser utilizadas para melhorar o pré-processamento, bem como para aplicar outros classificadores para avaliar métricas.

AGRADECIMENTOS

Agradecemos o laboratório de Processamento da Informação Biológica da Universidade Federal do Maranhão pelo apoio à pesquisa realizada, e ao CNPQ pelo fomento disponibilizado.

REFERÊNCIAS

- [1] World Health Organization (2020). Cardiovascular diseases. Acedido em 17 de Setembro de 2020, em: <https://www.who.int/health-topics/cardiovascular-diseases>.
- [2] M. Kachuee, S. Fazeli and M. Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation," 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, 2018, pp. 443-444. Acedido em 18 de Setembro de 2020, em: <https://arxiv.org/pdf/1805.00794.pdf>.
- [3] Q. J. A. J. A. L. F. B. AK; "Diagnostic decision support systems for atrial fibrillation based on a novel electrocardiogram approach," Journal of electrocardiology. 2017. Acedido em 18 de Setembro de 2020, em <https://pubmed.ncbi.nlm.nih.gov/29187299/>.
- [4] G. A. A. L. G. L. H. J. I. P. C. MR; "Intracardiac Atrial Fibrillation Database". PhysioBank, PhysioToolkit, and PhysioNet. Acedido em 14 de Setembro de 2020, em: <https://archive.physionet.org/physiobank/database/itafdb/>
- [5] G. A. A. L. G. L. H. J. I. P. C. MR; "The MIT-BIH Normal Sinus Rythm Database". PhysioBank, PhysioToolkit, and PhysioNet. Acedido em 14 de Setembro de 2020, em: <https://archive.physionet.org/physiobank/database/nsrdb/>
- [6] L. F. A. K. J. P. O. N. "Statistical coding and decoding of heartbeat intervals". PLoS One. 2011. Acedido em 19 de Setembro de 2020, em: https://www.researchgate.net/publication/51239927_Statistical_Coding_and_Decoding_of_Heartbeat_Intervals.