



ConBRepro

X CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO



EVENTO
ON-LINE

02 a 04
de dezembro 2020

Classificação de Fibrilação Atrial e Fibrilação Atrial Intra Cardíaca utilizando Estatísticas de Alta Ordem e Aprendizado de Máquina

Luís Fillype da Silva

Departamento de Engenharia de Eletricidade - UFMA

Jonathan Araújo Queiroz

Departamento de Engenharia de Eletricidade - UFMA

Allan Kardec Barros

Departamento de Engenharia de Eletricidade - UFMA

Resumo: O eletrocardiograma (ECG) é um exame que apresenta uma representação gráfica da atividade elétrica do coração. Por meio dele, é possível observar o ritmo dos batimentos cardíacos, o número de batimentos por minuto, além de possibilitar o diagnóstico de várias arritmias. Este artigo tem como objetivo desenvolver um modelo de classificação baseado nos batimentos de dois grupos de indivíduos: com Fibrilação Atrial e Fibrilação Atrial intra cardíaca. A metodologia de extração de características baseada e adaptada para classificar a Fibrilação Atrial e seu subtipo. As classificações foram realizadas no espaço tridimensional em duas etapas: com aplicação de Análise dos Componentes Principais (PCA) e sem aplicação do mesmo, por meio de Redes Neurais (RNA), Support Vector Machines (SVM) e K-nearest Neighbors (KNN), obtendo acurácia de 95% a 100%.

Palavras-chave: ECG, Aprendizado de Máquina, Estatística de Alta Ordem.

Classification of Atrial Fibrillation and Intracardiac Atrial Fibrillation using High Order Statistics and Machine Learning

Abstract: The electrocardiogram (ECG) is an exam that provides a graphical representation of the electrical activity of the heart. Through it, it is possible to observe the heartbeat rhythm, the number of beats per minute, in addition to enabling the diagnosis of various arrhythmias. This article aims to develop a classification model based on the beats of two groups of individuals: with Atrial Fibrillation and Intracardiac Atrial Fibrillation. The methodology of extraction of characteristics based and adapted to classify Atrial Fibrillation and its subtype. The classifications were performed in three-dimensional space using Neural Networks (RNA), Support Vector Machines (SVM) and K-nearest Neighbors (KNN), obtaining an accuracy of 95% to 100%.

Keywords: ECG, Machine Learning, High-Order Statistics.

1. Introdução

Pesquisas realizadas pela Organização Mundial da Saúde (2020), ressaltam que 17,3 milhões de pessoas em todo o mundo são vítimas de doenças cardíacas a cada ano. No Brasil, deste total, 950 pessoas morrem diariamente por doenças cardiovasculares. Assim, o diagnóstico eficaz das doenças cardíacas tem impulsionado o desenvolvimento de métodos autônomos que auxiliam na detecção de cardiopatias. Um exame que quantifica a atividade elétrica do coração, possibilitando detectar a frequência cardíaca e o número de batimentos por minuto, é a análise do eletrocardiograma (ECG).

Sendo assim, o ECG é essencial para prever, detectar e diagnosticar diversos problemas cardíacos, como a fibrilação atrial, por ser uma das técnicas não invasivas mais utilizadas para auxiliar nesse diagnóstico (Queiroz *et al.*, 2017, p. 1). Queiroz *et al.*, (2017) investigam a variação na tensão que ocorre em um intervalo t de batimento cardíaco utilizando curtose. Já Kachuee *et al.*, (2018) propõem um método baseado em redes neurais convolucionais profundas para a classificação dos batimentos cardíacos, capaz de classificar com precisão cinco diferentes arritmias.

Este artigo se propõe a extrair todo o batimento cardíaco de um ECG e agrupar dois grupos, indivíduos com sinais de Fibrilação Atrial e Fibrilação Atrial Intra Cardíaca, usando Estatísticas de alta ordem (EOS), e posteriormente realizando a classificação em três algoritmos de Aprendizado de Máquina (AM).

2. Fundamentação Teórica

2.1 Fibrilação Atrial

Segundo o Hospital Israelita Albert Einstein (2020), a Fibrilação atrial (FA) é a arritmia cardíaca sustentada mais frequente e é responsável por 33% de todas as internações por arritmia. Ela ocorre entre 1% e 2% na população geral, aumentando significativamente com o envelhecimento e com a presença de doenças cardíacas.

O hospital define ainda que a FA é uma arritmia cardíaca caracterizada pela completa desorganização da atividade elétrica dos átrios (câmaras superiores do coração) e consequente perda da contração atrial. Já Neto *et al.*, (2018) definem ainda a FA como sendo uma arritmia supraventricular caracterizada por atividade elétrica atrial desorganizada, secundária a múltiplos focos de despolarização atrial. Dentre manifestações clínicas da FA, destacam-se pela gravidade os quadros de tromboembolismo e de instabilidade hemodinâmica.

2.1 Fibrilação Atrial Intra Cardíaca

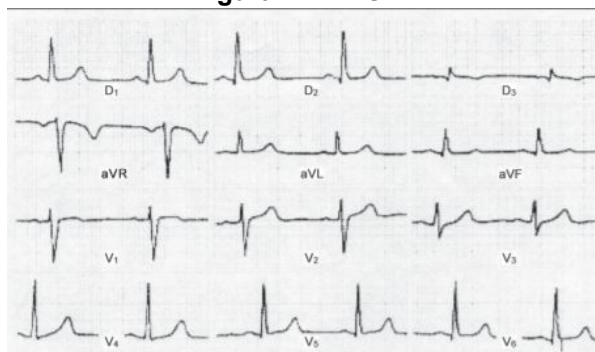
Para Richter *et al.*, (2010), os registros de potenciais cardíacos de eletrodos em contato direto com o coração é chamado um eletrograma intracardíaco (EGM). EGMs intracardíacos registram, portanto, a atividade elétrica local do coração, ou seja, o tecido cardíaco ao redor do eletrodo em contato. Isto pode auxiliar para o propósito de guiar o cateter de ablação para os locais atriais nos quais a arritmia se origina ou que representam substratos de arritmia.

Richter *et al.*, (2010) apontam que a abordagem intra cardíaca vem sendo amplamente utilizada, devido às observações a seguir: o complexo padrão elétrico observado durante a FA ser explicado com várias ondas que se propagam ao longo de várias rotas ao longo dos átrios, além dos dados disponíveis também apoiarem um mecanismo focal, de acordo com quais condutores, localizados principalmente nas veias pulmonares, desencadeando e sustentando a propagação da atividade elétrica nos átrios.

2.2 Eletrocardiograma

Strapazzon *et al.*, (2016) definem o eletrocardiograma (ECG) como um procedimento simples e rotineiro de grande importância para diagnósticos de patologias cardíacas. Este procedimento corresponde a uma representação gráfica da atividade elétrica do coração, que resulta na sua interpretação por apresentar ondas, segmentos e intervalos possíveis de medir e identificar alterações no coração.

Figura 1 – ECG



Fonte: Feldman e Goldwasser (2004)

2.3 Aprendizado de Máquina

Choy *et al.*, (2016) relata em seu artigo que dentro das inovações da ciência de dados, AM é uma classe de técnicas e área de pesquisa isso permite que os computadores aprendam como humanos e extraiam ou classifiquem padrões. As máquinas podem ainda ser capazes de analisar mais conjuntos de dados e extrair recursos de dados que humanos podem não ser capazes de fazer.

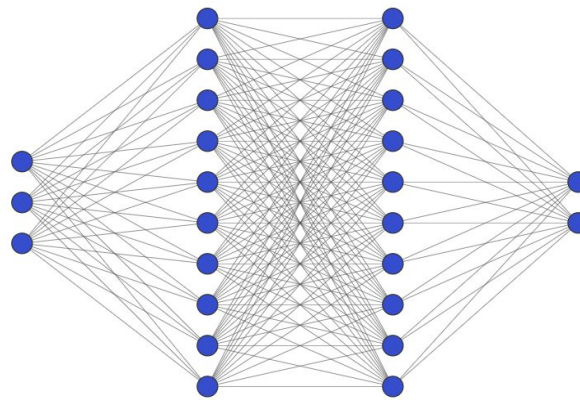
Além disso, o autor define ainda o AM é um método de ciência de dados que fornece aos computadores a capacidade de aprender sem ser programado com regras explícitas. Essa técnica permite a criação de algoritmos que podem aprender e fazer previsões. Em contraste com algoritmos baseados em regras, o AM aproveita a maior exposição a grandes e novos conjuntos de dados e tem a capacidade de melhorar e aprender com a experiência, como Redes Neurais, K-nearest Neighbors.

2.4 Redes Neurais Artificiais

Haykin (2001) define as RNA, mais conhecidas como redes neurais, como estruturas complexas interligadas por elementos de processamento simples (neurônios), que possuem a capacidade de realizar operações, como cálculos em paralelo, para o processamento de dados e representação de conhecimento (HAYKIN, 2001, p.27).

O autor ressalta ainda as propriedades e capacidades que tornam as RNAs potencialmente úteis são: não-linearidade: um neurônio artificial pode utilizar funções lineares ou não-lineares; mapeamento de Entrada-Saída: com base em exemplos de entrada e saída, a RNA é capaz de adaptar-se para minimizar o erro de mapeamento. Dentre as estruturas conhecidas desses modelos, temos a MLP (Multilayer Perceptron), que, de maneira geral, possui uma camada de entrada (sem função computacional), uma ou mais camadas ocultas e uma camada de saída. Veja na Figura 2 o modelo da MLP.

Figura 2 – RNA MLP



Fonte: Autor (2020)

Neste artigo, foi utilizado a RNA MLP com as configurações de 3 neurônios na camada de entrada, 2 camadas com 100 neurônios, e 2 neurônios na camada de saída.

2.5 K-Nearest Neighbors (KNN)

Segundo Phan e Martins (2018), o KNN é um dos algoritmos de classificação estatística prospectiva usados para classificar objetos com base em exemplos de treinamento mais próximos no espaço de recursos. É um algoritmo de aprendizado lento onde a função KNN é aproximada localmente e todos os cálculos são adiados até a classificação. Nenhum modelo ou aprendizado real é realizado durante a fase de treinamento, embora um conjunto de dados de treinamento seja necessário, ele é usado apenas para preencher uma amostra do espaço de pesquisa com instâncias cuja classe é conhecida, por este motivo, este algoritmo também é conhecido como aprendizado preguiçoso algoritmo. Isso significa que os pontos de dados de treinamento não são usados para fazer nenhuma generalização e todos os dados de treinamento são necessários durante a fase de teste.

Os autores ressaltam ainda as etapas de treinamento e classificação presentes neste algoritmo. Na fase de treinamento, os exemplos de treinamento são vetores (cada um com um rótulo de classe) em um espaço de recurso multidimensional. Nesta fase, os vetores de recursos e rótulos de classe das amostras de treinamento são armazenados. Já na fase de classificação, K é uma constante definida pelo usuário, uma consulta ou ponto de teste (vetor não rotulado) é classificado pela atribuição de um rótulo, que é o mais recorrente entre as K amostras de treinamento mais próximas desse ponto de consulta.

Neste artigo, foi utilizado o KNN com as configurações K igual a 10, para a classificação da cardiopatia.

2.6 Support Vector Machine (SVM)

Huang *et al.*, (2017) define o SVM como um método poderoso para construir um classificador. Esse método objetiva criar uma fronteira de decisão entre duas classes que possibilite a previsão de rótulos de um ou mais vetores de características.

Os autores ressaltam ainda que essa fronteira de decisão, conhecido como hiperplano, é orientado de forma que esteja o mais longe possível dos pontos de dados mais próximos de cada uma das classes presentes. Tais pontos mais próximos são chamados de vetores de suporte, dando origem ao nome do método. Sendo assim, o hiperplano ideal pode

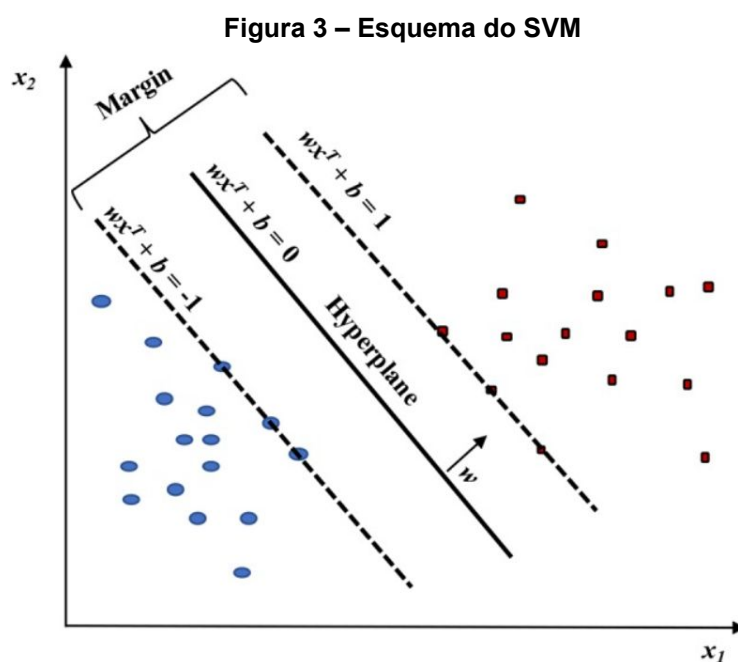
então ser definido como o que separa os dados e maximize a margem, respeitando as seguintes equações (HUANG *et al.*, 2017, p. 1). :

$$wx^T + b = 0 \quad (1)$$

$$wx_i^T + b \geq +1 \quad (2)$$

$$wx_i^T + b \leq -1 \quad (3)$$

Nas Equações acima, x representa os valores dos pesos, x o vetor de entrada e b o valor do bias. Sendo a Equação (1) representando o hiperplano ótimo, a Equação (2) parametrizando os dados que representam uma classe 1, e a Equação (3) representa os dados da classe 2. Na Figura 3 é ilustrado o esquema do modelo do SVM na classificação binária.



Fonte: Huang *et al.*, (2017)

Neste artigo, foi utilizado o SVM com as configurações do Kernel linear, para a classificação da cardiopatia.

2.7 Análise dos Componentes Principais (PCA)

Castells *et al.*, (2006) introduz o PCA como sendo uma técnica estatística que objetiva o condensamento das informações de um grande conjunto de variáveis correlacionadas em algumas variáveis ("componentes principais"), enquanto não desperdiça a variabilidade presente no conjunto de dados. Os autores salientam que os componentes principais são derivados como uma combinação linear das variáveis do conjunto de dados, com pesos escolhidos de maneira que esses componentes se tornem necessariamente não correlacionados. Cada componente contém novas informações sobre o conjunto de dados e é ordenado de forma que os primeiros componentes respondam pela maior parte da variabilidade.

Demonstrando a importância dessa técnica, inclusive em sinais de ECG, Castells *et al.*, (2006) relata que o PCA é empregado para lidar com diversos problemas na análise de ECG, como compressão de dados, detecção e classificação de batimento, redução de ruído, separação de sinal e extração de recursos. A análise de componentes principais

tornou-se uma ferramenta importante para lidar com muitos desses problemas com sucesso e foi considerada pela primeira vez para o propósito de recuperação eficiente de armazenamento de ECG.

2.8 Estatísticas de Alta Ordem

No início da década de 90, especialmente, foi notado um aumento no interesse nas EOS e suas aplicações. Conforme destaca Boreli (2018), foi verificada a aplicação de cumulantes em diversos campos do conhecimento, como em sonares, biomedicina, processamento de dados, reconstrução de imagens e etc.

Estas estatísticas fornecem mais informações do que as disponíveis simplesmente providas através da média e da variância de um processo. Dessa forma, pode-se afirmar que elas permitem uma melhor forma de discriminar processos. Então, para melhor entender e iniciar uma abordagem além da variância e média dos conjuntos, neste artigo é utilizado a curtose, assimetria, que são definidos abaixo

$$\sigma_x^2 = E(X^2) - ((E(X))^2) \quad (4)$$

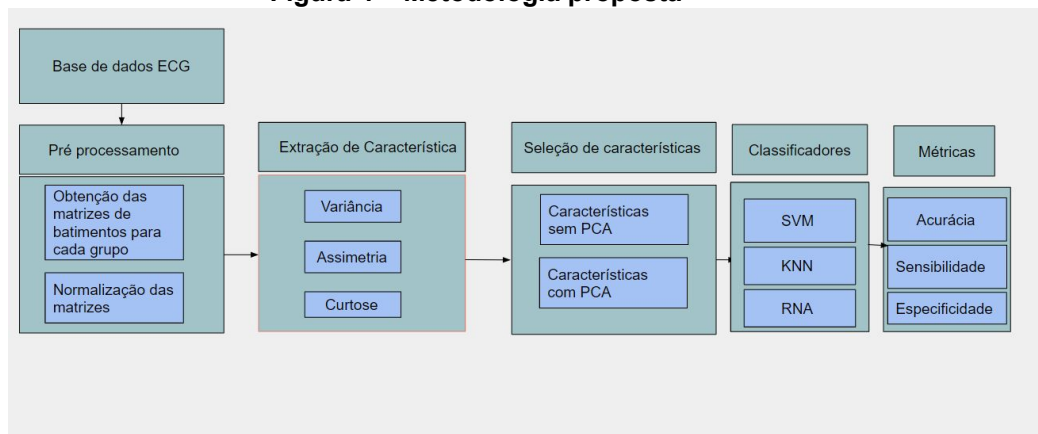
$$\lambda_x = E [((X - E(X))\sigma^{-1})^3] \quad (5)$$

$$\kappa_x = E [((X - E(X))\sigma^{-1})^4] \quad (6)$$

3. Materiais e Métodos

Na Figura 4, a metodologia utilizada neste artigo é ilustrada. As bases de dados a serem utilizadas foram definidas, separando-as em dois grupos: sinais de indivíduos com FA e indivíduos com FA intra cardíaca. Foi realizado o pré-processamento dos sinais do banco de dados, organizando-os para a extração das características. Nesta etapa, são calculados os valores de variância, assimetria e curtose do conjunto de dados de cada base.

Figura 4 – Metodologia proposta



Fonte: Autor (2020)

Em seguida, foram selecionadas as características representadas por estatísticas de alta ordem, e aplicado o PCA, sendo colocados como entrada para os classificadores usados neste artigo. Ao final do processo, os valores da métrica de classificação são retornados para avaliação do algoritmo, para posterior comparação dos mesmos nas duas etapas.

3.1 Base de dados

Foram utilizados os conjuntos de dados Intracardiac Atrial Fibrillation Database e o MIT-BIH Atrial Fibrillation Database, ambos disponíveis em Goldberger *et al.*, (2000). O banco de dados dos sinais de pacientes com FA contém 23 registros, sendo todos utilizados nesta análise. O banco de dados de sinais de pacientes com FA intra cardíaca contém 8 pacientes, todos os quais são usados.

3.2 Pré-processamento

Foram adquiridos sinais de ECG característicos da derivação DII, a mais utilizada no mundo. Foi utilizada toda a duração do sinal, amostrada na frequência de 256 Hz, para extrair os batimentos de cada paciente para análise e posterior extração das características.

Após isso, cada sinal selecionado foi segmentado para obtenção do respectivo batimento, conforme proposto por Queiroz *et al.*, (2017). Dessa forma, os batimentos de cada grupo foram agrupados, gerando uma matriz A por concatenação dos batimentos do grupo de FA, e uma matriz B dos batimentos do grupo com FA intra cardíaca, conforme descrito nas equações abaixo.

$$M_{n,m} = [B_{n,a} \ B_{n,b} \ \dots \ B_{n,z}] \quad (7)$$

onde n representa o número de batimentos, e m representa o total de todas as colunas de todos os batimentos.

Concomitante a isto, a média de seu conjunto foi subtraído do sinal, dividindo o resultado pela entropia de Shannon, dada pela Equação (8).

$$M_{n,m} = \frac{M_{n,m} - \frac{1}{n} \sum_1^n M_{n,m}}{-\sum_1^n P_{n,m}(\log_2 \frac{1}{n})} \quad (8)$$

3.3 Extração de características

A metodologia de extração foi adaptada usando estatística de alta ordem, proposta por Queiroz *et al.*, (2017). Um vetor foi obtido para cada uma das estatísticas associadas: variância, curtose e assimetria, que serão as entradas dos classificadores, representadas por σ_X^2 , κ_X e λ_X , respectivamente.

3.4 Métricas de avaliação dos algoritmos

Neste artigo foram utilizados os valores de acurácia, sensibilidade e especificidade, descritos pela Equação 9, Equação 10 e Equação 11 a seguir, para verificar o desempenho dos classificadores.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \times 100 \quad (9)$$

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \times 100 \quad (10)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \times 100 \quad (11)$$

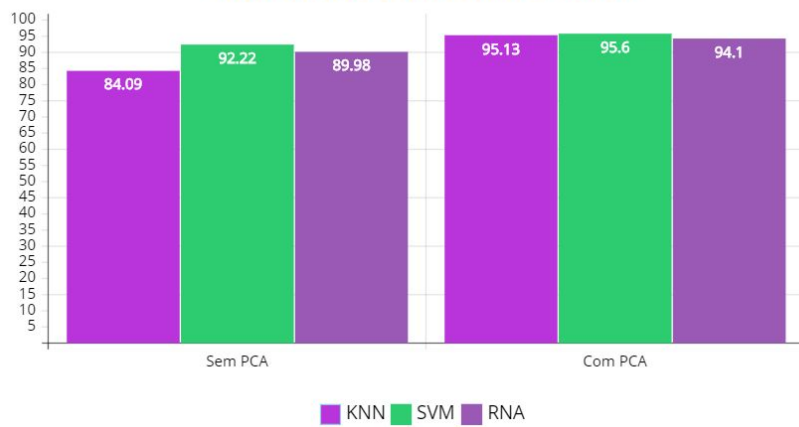
Nas equações, VP corresponde ao número de verdadeiros positivos, VN aos verdadeiros negativos, FP para registros de falsos positivos e FN para classificações de falsos negativos.

4. Resultados

Este artigo analisou os batimentos extraídos do ECG para pacientes com FA e de sinais de indivíduos com FA intra cardíaca, a fim de classificá-los. Para a etapa de classificação, foram geradas matrizes, onde cada coluna é representada por variância, assimetria e curtose, respectivamente.

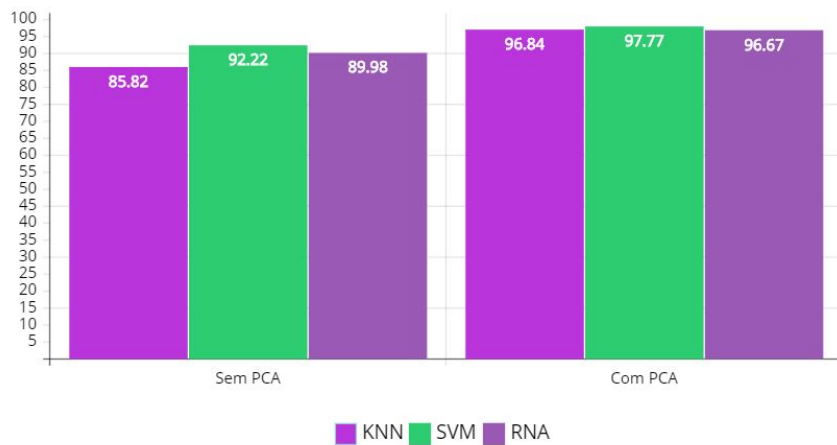
Tais matrizes foram as entradas dos classificadores KNN, SVM e RNA para verificar qual algoritmo de classificação apresenta maior acurácia, sensibilidade e especificidade. Além disso, foi realizada a comparação dos resultados com PCA nos conjuntos de dados e sem a aplicação do mesmo. Os resultados estão expostos nas Figura 5, Figura 6 e Figura 7.

Figura 5 – Exemplo de figura
Acurácia dos classificadores



Fonte: Autor (2020)

Figura 6 – Exemplo de figura
Sensibilidade dos classificadores



Fonte: Autor(2020)

Figura 7 – Exemplo de figura
Especificidade dos classificadores

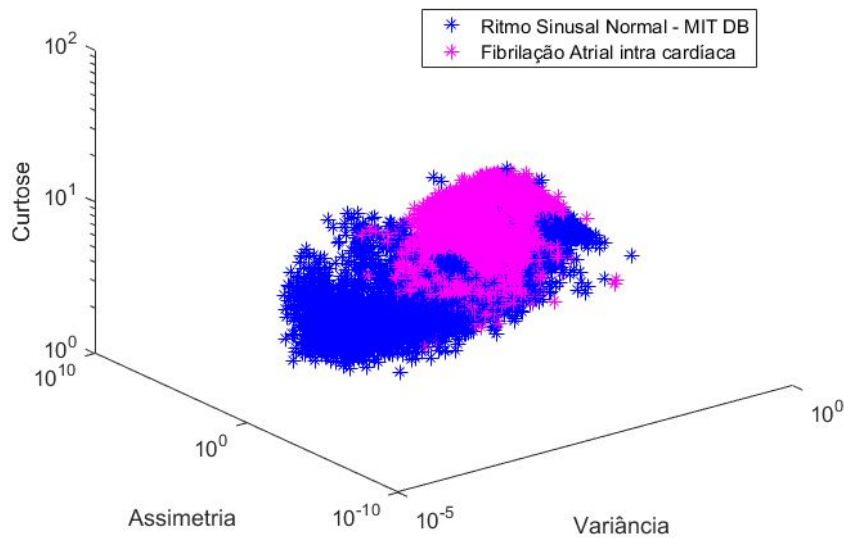


Fonte: Autor (2020)

5. Discussões

Este artigo apresentou uma análise das características três dimensões, com e sem aplicação de PCA, para a classificação dos batimentos com FA e FA intra cardíaca. Na Figura 8, a análise sem PCA mostra que os dados estão bastante agrupados, por terem quase as mesmas características, e tratarem da mesma cardiopatia, porém com abordagens de análise diferentes.

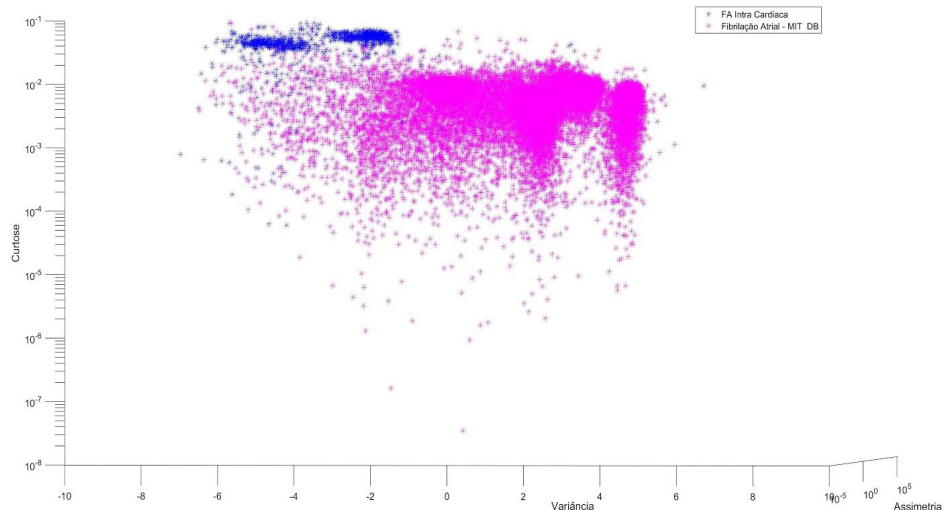
Figura 8 – Dados sem PCA



Fonte: Autor(2020)

Com isso, foi realizado nos conjunto de dados uma técnica que pudesse discriminar melhor tais valores. Para tal, a fim de melhorar a representação dessas características, utilizou-se, então, o PCA para rotação desses dados, resultando na Figura 9.

Figura 9 – Dados com PCA



Fonte: Autor (2020)

Como visto nas Figuras 5, 6, e 7, o resultados das classificações com a utilização do PCA foram melhores e comprovadas pelas métricas de avaliação dos próprios algoritmos, descritos nas Equações 9, 10, e 11.

O SVM teve melhor desempenho devido a sua facilidade de definição de parâmetros. Já para a RNA, necessita-se de estimar e definir muito bem esses valores de forma empírica para garantir a convergência e capacidade de generalização.

Ou seja, para se chegar ao melhor resultado é necessário testar diversas arquiteturas diferentes, aumentando ou diminuindo o número de camadas ocultas, realizar variações nas taxas de aprendizado, momentum e número de épocas de treinamento. Já o KNN, por ter um treinamento lento e também ser necessário a estimação do número de K, esse algoritmo teve um desempenho abaixo dos demais utilizados.

6. Conclusões

Neste artigo, a eficácia do uso de estatísticas de alta ordem para extrair características e classificar doenças cardíacas, como fibrilação atrial, foi reforçada. Além disso, mostrou-se a utilização de modificação dos dados, evidenciando uma diferença do desempenho dos dados originais e os dados rotacionados.

Os resultados obtidos podem servir de base para a tomada de decisões de natureza clínica, detectando arritmias de forma autônoma. Já em trabalhos futuros, diferentes doenças cardiovasculares podem ser utilizadas na metodologia e técnicas podem ser utilizadas para melhorar o pré-processamento, bem como aplicar outros classificadores para avaliar as métricas, e testagem de hiperparâmetros dos algoritmos de classificação.

Referências

BORELI, F. A. **Extração de características em sinais biológicos utilizando Estatísticas de Alta Ordem**. Minas Gerais, 175 p., 2018. Dissertação (Mestrado) – Universidade Federal de Minas Gerais.

Castells *et al.*,. Principal Component Analysis in ECG signal processing. 2006. Disponível em<https://www.researchgate.net/publication/26620236_Principal_Component_Analysis_in_ECG_Signal_Processing> Acesso em 13 set. 2020.

Choy *et al.*,. Current Applications and Future Impact of Machine Learning in Radiology. 2018. Disponível em<https://www.researchgate.net/publication/326005335_Current_Applications_and_Future_Impact_of_Machine_Learning_in_Radiology%3E>

Goldberger *et al.*,. **PhysioBank, PhysioToolkit, and PhysioNet: The MIT-BIH Atrial Fibrillation Database**. Disponível em: <<https://archive.physionet.org/physiobank/database/afdb/>> Acesso em: 14 set. 2020.

Goldberger *et al.*,. **PhysioBank, PhysioToolkit, and PhysioNet: Intracardiac Atrial Fibrillation Database**. Disponível em: <<https://archive.physionet.org/physiobank/database/iafdb/>> Acesso em: 14 set. 2020.

HAYKIN, S. **Redes neurais: princípios e prática**. Porto Alegre: Bookman, 2001.

Hospital Israelita Albert Einstein. **Fibrilação Atrial**. Disponível em: <einstein.br/especialidades/cardiologia/doencas-sintomas/fibrilacao-atrrial> Acesso em: 17 set. 2020.

Huang *et al.*,. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. 2017. Disponível em<<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5822181/>> Acesso em 13 set. 2020.

Kachuee *et al.*,. ECG Heartbeat Classification: A Deep Transferable Representation. 2018. Disponível em <<https://arxiv.org/pdf/1805.00794.pdf>> Acesso em 18 set. 2020.

Neto *et al.*,. Fibrilação atrial. **Revista QualidadeHC**, São Paulo, v.1, n.1, p. 1-4, 2018

PHAN, T.; MARTIN, K.;. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. 2018. Disponível em<https://www.researchgate.net/publication/322010741_Comparison_of_Random_Forest_k-Nearest_Neighbor_and_Support_Vector_Machine_Classifiers_for_Land_Cover_Classification_Using_Sentinel-2_Imagery>. Acesso em: 13 set. 2020.

Queiroz *et al.*,. Diagnostic decision support systems for atrial fibrillation based on a novel electrocardiogram approach. 2017. Disponível em <https://www.researchgate.net/publication/320682762_Diagnostic_decision_support_systems_for_atrial_fibrillation_based_on_a_novel_electrocardiogram_approach> Acesso em 18 set. 2020.

Richter *et al.*,. A Novel Approach to Propagation Pattern Analysis in Intracardiac Atrial Fibrillation Signals. 2010. Disponível em <https://www.researchgate.net/publication/46037472_A_Novel_Approach_to_Propagation_Pattern_Analysis_in_Intracardiac_Atrial_Fibrillation_Signals> Acesso em 18 set. 2020.

Strapazzon *et al.*,. Interpretação básica de eletrocardiograma: o conhecimento dos enfermeiros. In: SEMINÁRIO DE INOVAÇÃO E TECNOLOGIA, 16., Porto Alegre. **Anais...** Porto Alegre: UNIJUI – Centro de Tecnologia, 2016. p. 2-4.

World Health Organization (WHO). **Cardiovascular diseases**. Disponível em: <<https://www.who.int/health-topics/cardiovascular-diseases>> Acesso em: 17 set. 2020.