



## PROCESSAMENTO DIGITAL DE SINAIS APLICADO A REDUÇÃO DE DIMENSÃO DE SINAIS CARDÍACOS PARA ENTRADA DE ALGORITMOS INTELIGENTES

Ericksson Lucas Silva Roque<sup>1</sup> – Unifesspa

*erickssonlucas@hotmail.com*

Diego de Azevedo Gomes<sup>2</sup> - Unifesspa

*diagomes.pa@gmail.com*

**Agência Financiadora:** FAPESPA

**Eixo Temático/Área de Conhecimento:** Engenharia Biomédica

### 1. INTRODUÇÃO

A associação de computação em nuvem e os algoritmos inteligentes podem fazer com que identificação de características importantes de sinais vitais possam ser executadas de maneira compartilhada. Nesse tipo de cenário o paciente dispõe de um dispositivo de aquisição de sinais médicos, como o eletrocardiograma (ECG), o qual é enviado para um servidor que aglomera sinais de vários pacientes, os quais podem ser processados em conjunto. Tal configuração possui a vantagem de reduzir os custos com o dispositivo do paciente em relação a processamento de dados, pois este processamento é realizado remotamente. Além disso, com os dados reunidos em um só local, pode-se aplicar algoritmos inteligentes modernos como os de aprendizagem profunda, os quais necessitam de um grande volume de amostras para seu funcionamento pleno.

As vantagens nesta arquitetura são muitas, porém há de se considerar as limitações relacionadas à transmissão, armazenamento e processamento. Nos ambientes futuros em que se espera que uma grande quantidade de dispositivos de vários tipos estarão conectados em rede, é necessário que se tomem precauções em relação a quantidade de dados transmitida por cada dispositivo a fim de permitir o acesso simultâneo e a manutenção de níveis de qualidade de serviço. Uma forma de se tomar esta medida é aplicar métodos de compressão de dados e de representação alternativa de dados, fazendo assim com que a quantidade de dados enviada seja reduzida. Além disso, esta medida reduz quantidade total de dados armazenada por aplicação de nuvem, permitindo que mais aplicações possam utilizar este processamento remoto.

Finalmente, ao mantermos a quantidade de registros, porém com menos dados representando-os, podemos lançar mão de algoritmos inteligentes do tipo aprendizado profundo que terão grandes chances aprender o comportamento daqueles sinais, mas que provavelmente realizaram tal tarefa em um tempo menor e com um menor uso de recursos, devido a redução na dimensão de cada registro avaliado [1].

#### 1.1 Eletrocardiograma

As doenças cardiovasculares (DCV) são as principais causas de morte e hospitalização na Europa em ambos os sexos, pessoas maiores de 40 têm uma maior propensão para mortalidade, no Brasil estima-se, em média, cerca de 7,4 milhões de óbitos por DCV [2], as doenças mais frequentes são Acidente Vascular Cerebral (AVC), Doença Cardíaca Isquêmica (DCI) e a insuficiência Cardíaca [3], as DCV são divididas em dois grupos, doenças cardíacas e dos vasos sanguíneos, neste trabalho é abordado apenas as doenças cardíacas .

---

<sup>1</sup>Graduanda em Engenharia Elétrica - Universidade Federal do Sul e Sudeste do Pará

<sup>2</sup>Doutor em Engenharia Elétrica - Professor Titular Adjunto da Universidade Federal do Sul e Sudeste do Pará (FACEEL/IGE/Unifesspa). Coordenadora do Programa de Extensão Relações Étnico-Raciais e Cidadania.

# VI Seminário de Iniciação Científica

Pesquisa na Amazônia: Novos cenários

27 a 29 de Outubro de 2020

On-line pela plataforma Google Meet

UNIFESSPA | PROPIT

A técnica mais utilizada para diagnosticar tais doenças é o ECG, que é um método não invasivo, simples e indolor, onde os pulsos elétricos são computados graficamente para o diagnóstico médico, geralmente o ECG é feito quando o paciente tem suspeita de alguma doença cardíaca, também o diagnóstico é frequentemente feito em exames de rotina de pessoa com meia idade [4].

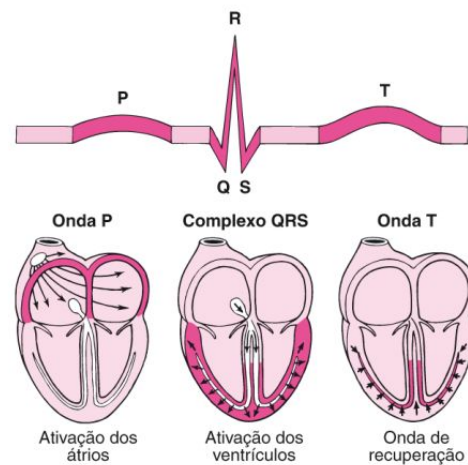


Figura 1: Estrutura padrão de um sinal de ECG.

O sinal ECG é adquirido da seguinte forma, quando o pulso cardíaco passa pelo coração, um pulso flui para os tecidos adjacentes, e uma parte deste sinal atinge a superfície da pele, resultando em uma diferença de potencial elétrica [5].

A associação de computação em nuvem e os algoritmos inteligentes podem fazer com que identificação de características importantes de sinais vitais possam ser executadas de maneira compartilhada. Nesse tipo de cenário o paciente dispõe de um dispositivo de aquisição de sinais médicos, como o eletrocardiograma (ECG), o qual é enviado para um servidor que aglomera sinais de vários pacientes, os quais podem ser processados em conjunto. Tal configuração possui a vantagem de reduzir os custos com o dispositivo do paciente em relação a processamento de dados, pois este processamento é realizado remotamente. Além disso, com os dados reunidos em um só local, pode-se aplicar algoritmos inteligentes modernos como os de aprendizagem profunda, os quais necessitam de um grande volume de amostras para seu funcionamento pleno.

## 1.2 Base de Dados

O banco de dados foi composto por, MIT-BIH Arrhythmia Database, que pode ser acessado no repositório público PhysioNet [12]. Este banco de dados é composto por 48 registros de ECG de dois canais, 30 min de duração, obtido de 47 pessoas no Laboratório de Arritmia BIH (Hospital Beth Israel), entre os anos de 1975 e 1979. A aquisição foi realizada com frequência de amostragem de 360 sps por canal, com 11 bits por amostras, o que rendeu uma resolução de 10 mV. Neste trabalho, avaliamos 17 classes de patologias cardíacas diferentes.

## 2. MATERIAS E MÉTODOS

Para realizar o trabalho, toda a modelagem foi feita dentro da linguagem de programação Python, As vantagens de escolher o Python dentre várias outras linguagens, são as amplas ferramentas que a linguagem dispõe para Ciência de Dados e Aprendizado Computacional, é gratuita e extremamente versátil.

### 2.1 Extração de Características

Os processos foram divididos em duas partes, primeiramente na extração de características utilizando a Densidade Espectral de Potência (PSD, do inglês *Power Spectral Density*) através do método de Welch mostra como a potência do sinal se acumula em torno de cada componente espectral [6]. A vantagem de se aplicar esta modelagem, é que a partir da tomada das médias dos vários periodogramas, o sinal obtido ao final do processo é representado por um conjunto de valores menor que o original. E a segunda parte com as Reduções de Dimensões, utilizando a Análise da Componente Principal (PCA, do inglês *Principal Component Analysis*) cuja seu principal objetivo é reduzir a dimensionalidade de um conjunto de dados, mantendo a máxima variação entre os sinais.

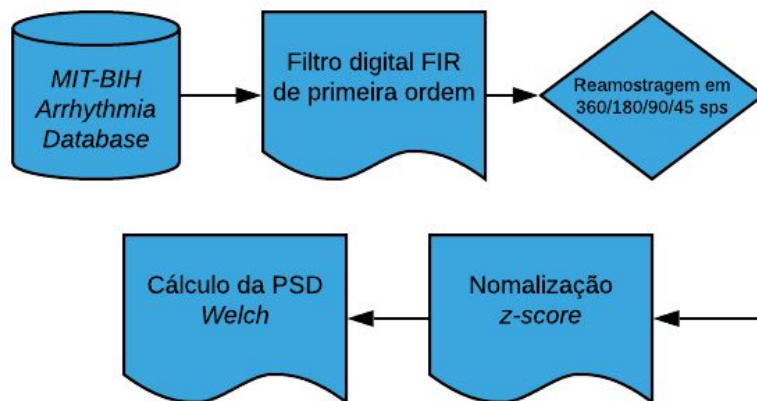


Figura 2: Extração de características.

Na Figura 2, está discriminado os passos para a extração. Após extrair os sinais do banco de dados, o sinal é filtrado de forma a poder ser reamostrado sem que ocorra o *aliasing* [8], o mesmo é feito aplicando um filtro digital Finite Impulse Response (FIR), em seguida é feita a reamostragem do sinal que está originalmente com uma frequência de amostragem de 360 sps, como o janelamento usado no tempo é de 10 segundos, cada sinal possui 3600 amostras. Como um dos objetivos do trabalho é verificar qual o impacto na escolha da frequência de amostragem, é feita a reamostragem do sinal em três novas frequências de amostragem 180, 90 e 45 sps. Algumas vezes os atributos dos vetores dos dados possuem magnitudes muito diferentes, o que pode tornar essas assinaturas menos discriminativas para a etapa de classificação. Para resolver este problema é preciso normalizar os dados, isto é, transformá-los para que se enquadrem em determinado padrão, a técnica utilizada para a normalização é a *z-score*, que transformam os dados com média 0 e variância igual a 1 [8], partindo para o cálculo da PSD.

## 2.2 Redução de Dimensões

Após o processo de extração de características, os dados são levados aos processos de redução de dimensão, que serão detalhados nesta seção.

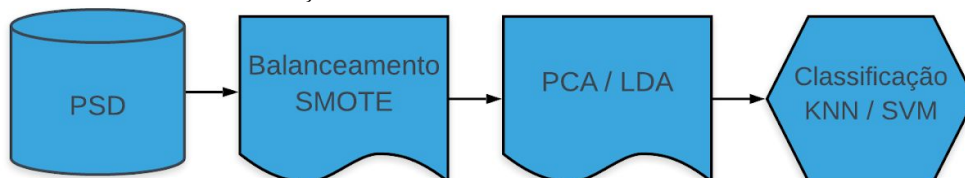


Figura 3: Redução de dimensão e classificação.

Na Figura 3, exibe os processos para a redução de dimensão. Para evitar o *overfitting*, é necessário fazer o balanceamento entre as classes [9], evitando o super treinamento das classes que tem uma quantidade

de sinais superiores às outras classes. Em seguida é feita a redução de dimensões com o PCA e seguida a classificação.

Para a classificação é utilizado dois algoritmos, primeiramente o k-Vizinho mais próximos (kNN, do inglês *k Nearest Neighbor*), é um dos classificadores mais simples de ser implementado, a ideia principal do kNN é determinar o rótulo de classificação de um conjunto de dados baseado nas amostras vizinhas advindas de um conjunto de treinamento, assim é encontrado o k vizinhos mais próximos do padrão de treinamento [9]. E também é utilizado a Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*), é um algoritmo de aprendizado supervisionado, cujo objetivo é classificar determinado conjunto de pontos de dados que são mapeados para um espaço de características multidimensional implementando um limite de decisão espacial.

### 3. RESULTADOS E DISCUSSÃO

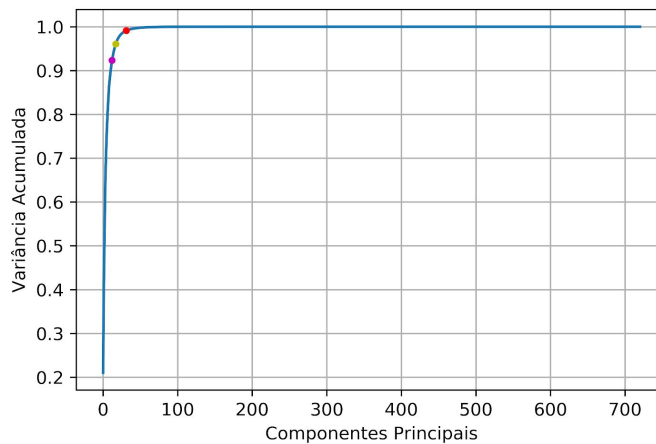


Figura 4: Variância acumulada para PCA.

Primeiramente será avaliado a compressão dos dados, comparado a cada método proposto, posteriormente é posto em avaliação os classificadores.

Na Figura 4, é exibido um gráfico onde mostra o quando variância é guardada dos sinais originais pelas PCs, como neste trabalho é abordado três tipos de variâncias, na figura é discriminado estes pontos, sendo que para uma variância de 99%, 95% e 90% os sinais podem ser representados com 31 PCs, 17 PCs e 12 PCs respectivamente, para uma frequência de amostragem 360 sps.

Freq. Amos. (sps)	360	180	90	45
<b>PCA(99%)</b>	31	31	86	91
<b>PCA (95%)</b>	17	35	41	42
<b>PCA (90%)</b>	12	23	26	26

Tabela 1: Quantidade de amostras, após a redução de dimensão.

Na Tabela 1, é exibido o comparativo de quanto cada técnica de redução é capaz de reduzir sinal original quando é feita a extração de características pela PSD, percebe-se que ocorre um fenômeno em que quanto mais se diminui a frequência de amostragem, menor é a capacidade do algoritmo, porque quanto mais o sinal é decimais eles ficam correlacionados, fazendo com que o PCA não maximize tanto suas variações [12].

Até agora foi observado apenas o quão eficaz é a técnica de redução comparando apenas a quantidade de amostras resultantes, entretanto esta análise não é suficiente quando o objetivo é a



classificação de características, para isso é feito o treinamento dos classificadores abordados e tirado a acurácia teste para todos os modelos de extração e de redução.

Freq. Amos.	SVM				KNN			
	360 sps	180 sps	90 sps	45 sps	360 sps	180 sps	90 sps	45 sps
PCA (99 %)	97,13%	97,25%	97,19%	97,33%	98,63%	97,18%	97,60%	97,07%
PCA (95 %)	95,07%	95,75%	95,17%	95,03%	98,44%	97,44%	97,19%	97,17%
PCA (90 %)	93,48%	91,69%	93,38%	94,62%	98,17%	97%	97,18%	97,21%

Tabela 1: Acurácia de teste para os classificadores.

A Tabela 1, exibe todas as acurácias com a classificador SVM e KNN, para o SVM Para a frequência de amostragem de 360 e 180 sps, o classificador apresentou uma acurácia de 97,13% e 97,25% respectivamente aplicando o PCA como redutores de dimensão e obteve 97,33% com uma frequência de 45 sps, para o KNN quando se avalia os Redutores de dimensões, também o PCA obteve a melhor quando os sinais estão como uma amostragem de 360 e 90 sps, em que a acurácia resultou em 98,63% e 97,6% respectivamente.

#### 4. CONSIDERAÇÕES FINAIS

Os resultados de acurácia dos classificadores revelam o alto potencial da classificação de patologias cardíacas, com aplicações de técnicas de redução dimensões, para o diagnóstico médico remoto. Para esta modelagem de classificação de anomalias cardíacas, recomenda-se o uso do KNN e SVM, que obteve uma acurácia de 98,63% e 97,33%, respectivamente, e também recomenda-se o uso da técnica de extração de características PSD e de redução de dimensão PCA, que consegue representar dados de 31 classes com 31 amostras e apresentou resultados bem expressivos.

A proposta deverá evoluir para o desenvolvimento de sistemas automáticos de classificação e validação, através da coleta de sinais e classificação remota. O trabalho apresenta parte do desenvolvimento de uma metodologia de pesquisa em busca por melhores técnicas de redução de dimensão, principalmente no que concerne aos novos paradigmas de classificação. No entanto, o trabalho também objetiva consolidar referências que permitam automação do monitoramento das patologias.

#### REFERÊNCIAS (Conforme ABNT)

- [1] L. Costa, C. Menezes, A. Santos, M. Araujo, and G. Campos, “A comparative study involving classifiers and dimensionality reduction techniques applied to facial recognition” in Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional, pp. 832–843, SBC, 2020.
- [2] E. J. Cesarino, A. L. G. Vituzzo, J. M. C. Sampaio, D. A. S. Ferreira, H. A. F. Pires, and L. d. Souza, “Assessment of cardiovascular risk of patients with arterial hypertension of a public health unit” Einstein (São Paulo), vol. 10, no. 1, pp. 33–38, 2012.
- [3] A. Macedo, A. Santos, E. Rocha, C. Perdigão, “Percepção da doença cardíaca e cerebral e dos factores de risco cardiovasculares em portugal: Estudo amalia” Rev Port Cardiol, vol. 27, no. 5, pp. 569–580, 2008.
- [4] C. Zaslavsky and I. Gus, “Idoso: doença cardíaca e comorbidades” Arquivos brasileiros de cardiologia, vol. 79, no. 6, pp. 635–639, 2002.
- [5] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, “A dynamical model for generating synthetic electrocardiogram signals” IEEE transactions on biomedical engineering, vol. 50, no. 3, pp. 289–294, 2003.
- [6] S. Haykin and B. Van Veen, “Signals and systems”. John Wiley & Sons, 2007.



The banner features a light green background with several circular icons: a lightbulb, a magnifying glass, a smartphone, a Wi-Fi symbol, a laptop, and a search icon. The text is centered and includes the event title, subtitle, dates, and platform information.

**VI Seminário de Iniciação Científica**  
*Pesquisa na Amazônia: Novos cenários*  
27 a 29 de Outubro de 2020  
On-line pela plataforma Google Meet

UNIFESSPA | PROPIT

[7] S. Carmo, L. VASCONCELOS, and M. da SA EMERENCIANO, “**Aplicação da análise de componentes independentes (ica) no controle de uma coluna de destilação de alta pureza**” in XX Congresso Brasileiro de Engenharia Química–XX COBEQ. Florianópolis–SC, 2014.

[8] P. S. Diniz, E. A. da Silva, and S. L. Netto, **Processamento Digital de Sinais: Projeto e Análise de Sistemas**. Bookman Editora, 2014.

[9] D. M. Hawkins, “**The problem of overfitting**” Journal of chemical information and computer sciences, vol. 44, no. 1, pp. 1–12, 2004.

[10] J. M. Giron-Sierra, **Digital Signal Processing with Matlab Examples**, Volume 2: Decomposition, Recovery, Data-Based Actions. Springer, 2016