

CLASSIFICAÇÃO DE PADRÕES DE ECG UTILIZANDO AS ABORDAGENS DE REGRESSÃO LINEAR MÚLTIPLA E ALGORITMO K-MEANS

ALEX V. ECHEVERRIA U. SILVA

¹UFG – Universidade Federal de Goiás
INF – Instituto de Informática
Goiânia - GO

Resumo: Este artigo propõe um classificador para detecção de extra-sístoles ventriculares utilizando a regressão linear múltipla e o algoritmo de clusterização K-Means². Os dados de ECG MIT-BIH foram utilizados como estudo de caso. Os resultados mostram 87% de acerto do classificador, sendo que os erros ocorridos se devem ao fato da análise dos sinais em questão serem correspondentes a outro canal de ECG.

Palavras Chaves: Classificação de sinais ECG, Regressão Linear Múltipla, K-Means.

1 INTRODUÇÃO

Extra-sístole ventricular (ESV), ou contração ventricular prematura, é o disparo prematuro do estímulo de contração do ventrículo. Este comportamento é comum em indivíduos saudáveis, porém em diversas cardiopatias, como infarto, estenose aórtica, febre reumática ou insuficiência cardíaca de qualquer origem, podem causar estas disfunções. Nestes casos há possibilidade de piora do quadro clínico do paciente levando a fibrilação ventricular ou até óbito súbito [1].

As ESV podem ser percebidas como falha do batimento; a ESV não é percebida, mas sim o batimento sinusal seguinte exacerbado. Quando ESVs são muito frequentes, particularmente quando ocorrem a cada 2º batimento cardíaco, são possíveis os sintomas hemodinâmicos leves, pois a frequência sinusal foi efetivamente reduzida à metade. Os sopros de ejeção existentes podem ser exacerbados em virtude do aumento do enchimento e da contratilidade cardíaca após a pausa compensatória.

As extra-sístoles ventriculares possuem características ou morfologias que podem ser observadas no sinal de Eletrocardiograma (ECG). Elas podem aparecer, dentre várias outras formas, com o seguimento ST e onda T com direções opostas ao QRS, ou com o QRS prematuro e alargado (forma bizarra). A identificação desses padrões é de suma importância para uma avaliação sintomática do

quadro clínico do paciente. Em alguns casos, os tratamentos podem ser corrigidos por medicamentos, mas existem casos em que é necessário neutralizar a região de ativação do estímulo. Para isto é fundamental a detecção desta região.

Existem casos em que a detecção destes batimentos ectópicos só é possível através da verificação do sinal de ECG por um longo período de tempo, onde o uso do holter é indicado. A verificação destes “sinais alterados” no ECG é feita manualmente por um médico, o que gera um grande esforço e riscos de possíveis erros.

Este trabalho propõe uma forma automatizada de classificar os batimentos normais dos extra-sistólicos através da Regressão Linear Múltipla (RLM) e do algoritmo K-Means; podendo oferecer auxílio de uma análise adicional ao médico.

2 A REGRESSÃO LINEAR MÚLTIPLA

A RLM é empregada quando é necessário estabelecer uma relação envolvendo mais de uma variável independente (x) e a variável de interesse (y).

$$y = x_1b_1 + x_2b_2 + K + x_nb_n$$

Dispondo-se de um conjunto de dados de calibração X_{cal} e Y_{cal} onde, a melhor estimativa no sentido de mínimos quadrados para o vetor de coeficientes b é dado pela pseudo- inversa de Moore-Penrose [2]:

$$\hat{b} = (X_{cal}^T X_{cal})^{-1} X_{cal}^T y_{cal}$$

Em posse do coeficiente de regressão (\hat{b}) é possível fazer a classificação em um novo conjunto de testes no intuito de obter a estimativa de y .

$$\hat{y} = X_{teste} * \hat{b}$$

3 K-MEANS

O K-Means é um algoritmo de aprendizado não-supervisionado (não necessita de rótulos que informam a classe a qual os dados pertencem) simples e popular, o qual aprende por meio da clusterização (agrupamento por similaridade de características) dos dados.

Seu funcionamento se dá através da seleção, de antemão, do K número de clusters (agrupamentos) a serem realizados pelo algoritmo. Em seguida, o algoritmo define K centroides aleatórios e calcula a distância de cada ponto (elemento do conjunto de dados) aos centroides, modificando o posicionamento desses centros de modo que fique o mais perto possível de todos os pontos de cada grupo (média da posição de todos eles, o que dá o nome ao algoritmo “means” – média, do inglês).

4 ESTUDO DE CASO

Como estudo de caso, foi utilizado o Arrhythmia Database do MIT-BIH, banco de dados que está acessível no site Physionet. Nesta base de dados constam registros ECG de diversos pacientes, e todos os registros tiveram os batimentos normais e extra-sistólicos identificados por uma equipe médica. Esses dados consistem de valores contínuos, que representam a amplitude do batimento cardíaco (em mV), em 82 momentos de captação (amostras obtidas no tempo). Para este trabalho somente o paciente “2” do MIT-BIH será utilizado; ele contém 393 registros de batimentos cardíacos normais e 106 extra-sístoles ventriculares como mostra a Figura 1.

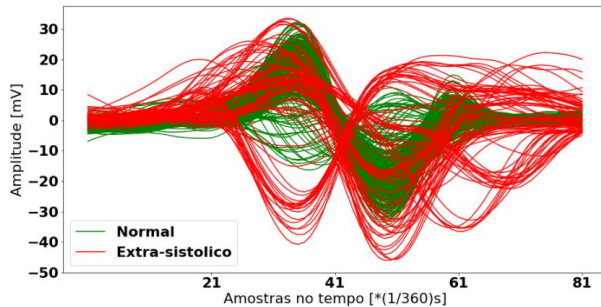


Figura 1: Registro ECG(MLII) referente ao paciente 2.

Cada batimento foi ajustado com 50 amostras, a direita do ponto de marcação do MIT e 50 a esquerda, resultando em 100 unidades. Depois dessa ordenação os sinais foram suavizados com um filtro Savitzky-Golay, com polinômio de segunda ordem e janela de 21 pontos, no intuito de reduzir variações simétricas da linha de base e possíveis ruídos. Resultando em uma redução para 81 unidades, como nota-se na Figura 1. O canal de ECG utilizado é o MLII com frequência de amostragem de 360Hz.

Após este pré-processamento, os dados foram divididos aleatoriamente em dois conjuntos, como mostra a Tabela 1:

	Batimentos normais	Extra-sístoles Ventriculares
Calibração	294	80
Teste	99	26

Tabela 1: Divisão dos dados

É possível observar na Figura 1 que existem diversos registros extra-sistólicos com características semelhantes aos batimentos normais.

5 RESULTADOS

5.1 ABORDAGEM POR REGRESSÃO LINEAR

Ao realizar a classificação utilizando a MLR, houveram somente 13 erros respectivos aos sinais extra-sistólicos, como mostra a Figura 2.

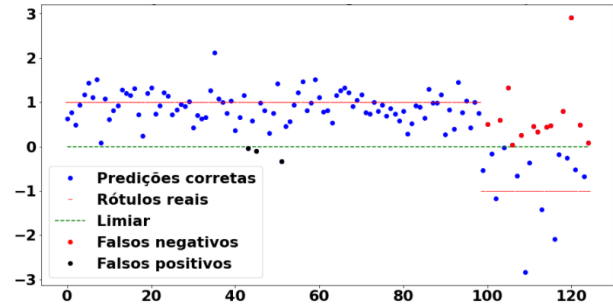


Figura 2: – Classificação dos batimentos normais (valores positivos), extra-sistólicos (valores negativos) e os erros (o).

Os sinais respectivos aos erros na classificação são apresentados em azul na Figura 3, onde é possível verificar que não há um padrão bem definido. Alguns batimento classificados incorretamente apresentam demasiados ruídos que podem ser provenientes de um mal-posicionamento dos eletrodos na realização do exame.

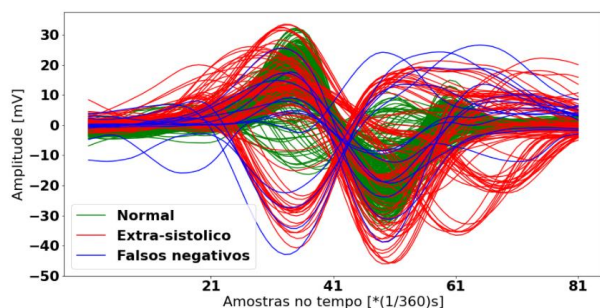


Figura 3: – Registro total com os erros (MLR).

Para uma visualização melhor dos resultados dessa abordagem, veja a matriz de confusão da Figura 4:

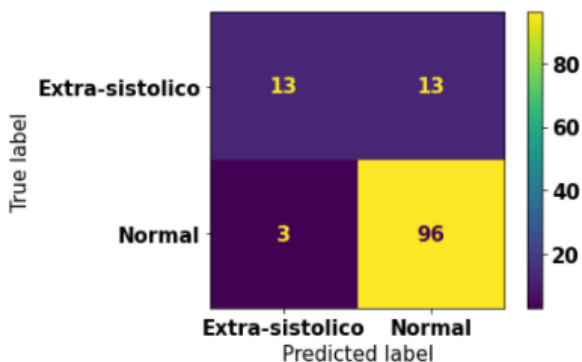


Figura 4: – Matriz de confusão dos resultados da Regressão Linear Múltipla.

5.2 ABORDAGEM K-MEANS

Ao realizar a classificação utilizando o K-Means, houveram 18 erros referentes aos sinais extra-sistólicos, como mostra a Figura 5.

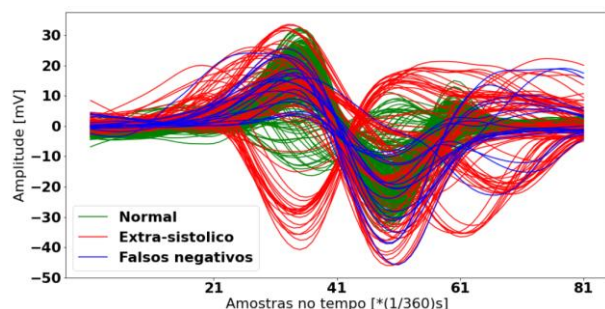


Figura 5: – Registro total com os erros (K-Means).

Os sinais respectivos aos erros na classificação são apresentados em azul na Figura 5, onde é possível verificar que eles possuem o mesmo padrão de batimentos normais e alguns extra-sistólicos. Os

resultados são ilustrados pela matriz de confusão, que revela algumas métricas dessa abordagem (Figura 6).

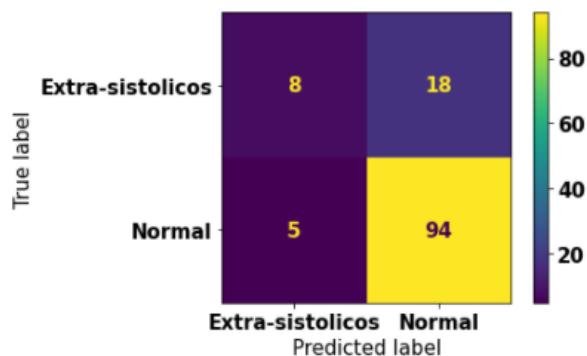


Figura 6: – Matriz de confusão dos resultados da abordagem K-Means.

6 CONCLUSÃO

Os resultados indicam que a abordagem utilizando Regressão Linear Múltipla demonstra maior eficiência em realizar previsões para esse tipo de problema, demonstrando cerca de 87% de acurácia, sendo que os 13% equivocados se dão por ruídos e outros tipos de interferência, em que um humano facilmente confundiria também.

Para a abordagem K-Means, percebe-se que sua eficiência é comprometida quando aplicada a esse tipo de problema, pois mesmo sua acurácia ser de 87% de acertos, quando se olha para as previsões envolvendo extra-sistólicos ela cai para apenas 25% de acertos de verdadeiros positivos, sendo 58% preditos como falsos negativos (batimentos extra-sistólicos classificados como normais). Fato esse que é um problema enorme para a área.

REFERÊNCIAS

What is K-Means? Disponível em: <https://pt.wikipedia.org/wiki/K-means>. Acesso em 17/09/2021.

O que é extrassístole ventricular? Disponível em: <https://www.msmanuals.com/pt-br/profissional/doencas-cardiovasculares/arritmias-e-doencas-de-conducao/extrassístole-ventricular-esv>. Acesso em 17/09/2021.

Understanding K-means Clustering in Machine Learning. Disponível em: <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>. Acesso em 18/09/2021.

Filho, A. R. G. “*Classificação de padrões de ECG utilizando regressão linear múltipla*”. Instituto Tecnológico de Aeronáutica.