



## Comparação de Classificadores de Complexos QRS Desenvolvidos Utilizando Técnicas de Aprendizado de Máquina

### Comparison of QRS Complexes' Classifiers Developed Using Machine Learning

### Comparación de clasificadores de complejos QRS desarrollados mediante técnicas de aprendizaje automático

Guilherme Bachega Gomes<sup>1</sup>, Rômulo César Silva<sup>2</sup>, Adriana Tokuhashi Kauati<sup>3</sup>,  
Lucas Guilherme Hübner<sup>4</sup>

1 Discente de Ciência da Computação na Universidade Estadual do Oeste do Paraná - UNIOESTE, Foz do Iguaçu, Paraná (PR), Brasil.

2 Professor adjunto de Ciência da Computação na Universidade Estadual do Oeste do Paraná - UNIOESTE, Foz do Iguaçu, Paraná (PR), Brasil.

3 Professora associada de Engenharia Elétrica na Universidade Estadual do Oeste do Paraná - UNIOESTE, Foz do Iguaçu, Paraná (PR), Brasil.

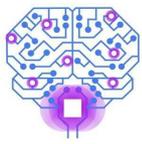
4 Professor assistente de Engenharia de Software e Análise e Desenvolvimento de Sistemas na Uniamérica Descomplica, Foz do Iguaçu, Paraná (PR), Brasil.

Autor correspondente: Guilherme Bachega Gomes

E-mail: guizobachegagomes@gmail.com

### Resumo

Doenças cardiovasculares são a maior causa de morte no mundo e sua prevenção é feita através de seu diagnóstico precoce<sup>(1)</sup>. Em 2019, cerca de 17,9 milhões de mortes por doenças cardiovasculares ocorreram globalmente<sup>(1)</sup>. Em especial as arritmias cardíacas que podem ser diagnosticadas por eletrocardiografia<sup>(2)</sup>. Estudos propuseram modelos com algoritmos de aprendizado de máquina para classificação de batimento cardíaco, utilizando a base de dados do *MIT-BIH Arrhythmia Database*<sup>(3-7)</sup>. Este trabalho apresenta três modelos preditivos, baseados na derivação D2 da base MIT-BIH, utilizando Árvore de Decisão, rede neural *Multilayer Perceptron* e *Deep Neural Network* com dois tipos de balanceamento da base de dados para a classificação de 10 arritmias. Os algoritmos foram treinados utilizando *5-fold stratified cross-validation* e sua performance, em *F1-Score*, foram submetidos à análise



estatística, tendo a *Deep Neural Network*, em ambas as bases, obtido o melhor desempenho.

**Descritores:** Aprendizagem de Máquina; Arritmia cardíaca; Eletrocardiograma.

## Abstract

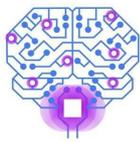
Cardiovascular diseases are the most common cause of death and its early diagnosis is key to prevention<sup>(1)</sup>. In 2019, there were 17,9 million deaths caused by cardiovascular diseases globally<sup>(1)</sup>. In the case of arrhythmias, they can be detected through an electrocardiography<sup>(2)</sup>. Some papers have proposed and built models to heartbeat classification<sup>(3-7)</sup>. In this work, three machine learning models were built, based in the MLII lead of the MIT-BIH Database, using a decision tree, multilayer perceptron and deep neural network using two different balanced versions of the same data to classify between 10 arrhythmias. The models were trained with 5-fold stratified cross-validation and their performances, compared using de *F1-Score* metric, were statistically analyzed with the deep neural network having a better performance within both databases.

**Keywords:** Machine Learning; Heart arrhythmia; Electrocardiogram.

## Resumen

Las enfermedades cardiovasculares son la mayor causa de muerte en el mundo y su prevención se realiza a través de su diagnóstico precoz<sup>(1)</sup>. En 2019, se produjeron alrededor de 17,9 millones de muertes por enfermedades cardiovasculares en todo el mundo<sup>(1)</sup>. En particular, las arritmias pueden ser diagnosticadas a través de un examen electrocardiográfico<sup>(2)</sup>. Varios trabajos propusieron y construyeron modelos con algoritmos de aprendizaje automático para la clasificación de ritmos cardíacos<sup>(3-7)</sup>. Este trabajo construyó tres modelos predictivos, basados en la derivación D2 de la base de datos MIT-BIH, utilizando un árbol de decisión, una red neuronal *multilayer perceptron* y una red neuronal profunda con dos tipos de balanceo de bases de datos para la clasificación de 10 arritmias. Los algoritmos fueron entrenados mediante *5-fold stratified cross-validation* y sus desempeños, medidos en *F1-Score*, fueron sometidos a análisis estadístico, con la red neuronal profunda, en ambas bases, obteniendo mejor desempeño.

**Descriptores:** Aprendizaje Automático; Arritmia cardiaca; Electrocardiograma.



## Introdução

As doenças cardiovasculares são, atualmente, a maior causa de morte no mundo. Em 2019, foram estimadas cerca de 17,9 milhões de mortes por doenças cardiovasculares globalmente, sendo 32% do total de mortes que ocorrem no planeta<sup>(1)</sup>.

A prevenção de morte por distúrbios no sistema cardiovascular é indissociável do seu diagnóstico e acompanhamento precoce. Nesse contexto, tratando das arritmias cardíacas, cuja forma de diagnosticá-las é por exame do eletrocardiograma (ECG), o qual permite o registro das atividades elétricas do coração<sup>(2)</sup>. Uma arritmia cardíaca significa que há a presença de anormalidade no mecanismo controlador do ritmo cardíaco e definida como alterações na frequência, formação e/ou condução do impulso elétrico pelo miocárdio<sup>(2)</sup>.

Comumente, as arritmias cardíacas não são percebidas pelo paciente, que desconhece possuir um problema no coração, mas que pode sofrer sintomas decorrentes dessa condição<sup>(8)</sup>. Assim, é relevante aprimorar os meios de realizar diagnósticos, buscando torná-los informatizados, automatizados e mais precisos.

Um complexo QRS é um conjunto formado pelas ondas Q, R e S, e representam a despolarização elétrica dos ventrículos. Este trabalho apresenta a construção de três classificadores de complexos QRS desenvolvidos a partir da análise de sinais de ECG. Os classificadores foram construídos utilizando os algoritmos de árvore de decisão, *Multilayer Perceptron* e *Deep Learning* e seus desempenhos foram comparados estatisticamente.

Foram escolhidas 11 classes de batimentos cardíacos, adotado como critério de escolha a presença de pelo menos 83 instâncias na base de dados do MIT-BIH, permitindo assim a comparação com outro trabalho<sup>(6)</sup>, e devido às demais classes possuírem poucas instâncias. As classes escolhidas são:

- Batimento Normal (NORMAL);
- Bloqueio de Ramo Esquerdo (*Left Bundle Branch Block Beat* - LBBB);
- Bloqueio de Ramo Direito (*Right Bundle Branch Block Beat* - RBBB);
- Contração Atrial Prematura Aberrante (*Aberrated Atrial Premature Beat* - ABERR);
- Contração Ventricular Prematura (*Premature Ventricular Contraction* - PVC);



- Fusão de Batimento Normal com Ventricular (*Fusion of Ventricular and Normal Beat* - FUSION);
- Contração Juncional Prematura (*Junctional Premature Beat* - NPC);
- Contração Atrial Prematura (*Atrial Premature Contraction* - APC);
- Batimento de Escape Ventricular (*Ventricular Escape Beat* - VESC);
- Batimento de Escape Juncional (*Junctional Escape Beat* - NESC);
- *Flutter* Ventricular (*Ventricular Flutter Wave* - FLWAV).

## Trabalhos Relacionados

Alfaras *et al.*<sup>(3)</sup> desenvolveram um classificador baseado em uma técnica de aprendizado de máquina, inspirada no cérebro, denominada de *Echo State Networks*, utilizando a base de dados MIT-BIH e a derivação D2 para um classificador binário.

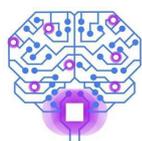
Llamedo e Martinez<sup>(4)</sup> apresentaram um classificador ternário de batimentos cardíacos, utilizando a base MIT-BIH e algoritmos de *floating feature*, com o foco em maior capacidade de generalização. Foram utilizados atributos extraídos das séries RR e outros que foram computados por diferentes escalas e amostras do eletrocardiograma em ambas as derivações.

Shaker *et al.*<sup>(5)</sup> construíram um modelo para classificar entre 15 classes de batimentos cardíacos. A base do MIT-BIH é balanceada com *Generative Adversarial Networks* e duas redes neurais convolucionais para extrair e reduzir atributos e também para a classificação propriamente dita.

Hübner<sup>(6)</sup> testou três classificadores para classificar entre 12 tipos de batimentos cardíacos presentes na base MIT-BIH. Utilizou uma árvore de decisão, uma rede MLP e um baseado em *Random Forest* combinados com diversos dados de treino: características extraídas da base, séries temporais brutas e uma mistura entre os dois. O algoritmo *Random Forest* obteve o melhor desempenho.

No trabalho proposto por Zhang *et al.*<sup>(7)</sup>, é realizada seleção de atributos por ranqueamento *one-versus-one* e um estágio de busca embutidos utilizando *Support Vector Machines* para classificar entre quatro classes representando agrupamentos de arritmias utilizando a base MIT-BIH. Método foca na seleção de subconjuntos de atributos que melhor distinguem entre as classes utilizando a comparação *one-versus-one*.

Resultados do material bibliográfico desta seção são descritos na Tabela 7.



## Métodos

Foi utilizada a base de dados de domínio público *MIT-BIH Arrhythmia Database*, oriunda de laboratórios do *Boston's Beth Israel Hospital* e do *Massachusetts Institute of Technology*. Trata-se de uma coleção de 48 trechos, com meia hora de duração cada, de exames eletrocardiográficos ambulatoriais coletados de 47 pacientes entre os anos de 1975 e 1980. Desse grupo, 23 exames foram selecionados de forma randômica de um registro de 24 horas de duração de 4.000 pacientes, dos quais 60% são intra-hospitalares e 40% são extra-hospitalares. Os outros 25 exames também selecionados da mesma amostra e escolhidos especificamente por conterem arritmias mais raras, porém, ainda clinicamente significativas. Possui as derivações D2 e V1<sup>(9)</sup>. Neste trabalho, é utilizada apenas a derivação D2.

Além dos registros de batimentos cardíacos, cardiologistas anotaram para cada um deles a classificação e o ponto que marca o pico da onda R (e conseqüentemente o pico do complexo QRS).

Para extrair cada registro como uma série temporal, a partir de cada pico do complexo QRS, foram considerados 100 pontos anteriores e 200 posteriores ao seu pico (com ele incluso) como um único batimento cardíaco, totalizando uma série temporal de 301 pontos que representa uma instância de complexo QRS. Essa decisão justifica-se por assim ser possível extrair a maior parte da informação de um ciclo cardíaco em particular<sup>(10)</sup>. Após esse processo, tem-se como resultado a base de dados, com suas respectivas proporções, registrada na Tabela 1.

**Tabela 1** - Proporção de 11 tipos de batimentos cardíacos da base de dados

Classe	Quantidade	Proporção
NORMAL	74.934	73,83%
LBBB	8.069	8,05%
RBBB	7.252	7,14%
ABERR	150	0,14%
PVC	7.129	7,02%
FUSION	802	0,79%
NPC	83	0,0078%



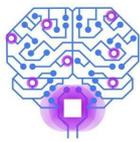
APC	2.543	2,50%
VESC	106	0,10%
NESC	229	0,22%
FWAV	192	0,18%
<b>Total</b>	<b>101.489</b>	<b>100%</b>

Os algoritmos selecionados para este trabalho foram: árvore de decisão, rede neural *Multilayer Perceptron* e *Deep Neural Network* (DNN), sendo os dois primeiros implementados na biblioteca *Scikit-learn*<sup>(11)</sup> e o último nas bibliotecas *Keras*<sup>(12)</sup> e *Tensor Flow 2*<sup>(13)</sup> para a linguagem de programação *Python 3*. Estes algoritmos foram selecionados para comparar seus desempenhos com os resultados de outro trabalho<sup>(6)</sup>.

Para a árvore de decisão, utilizou-se o algoritmo *DecisionTreeClassifier*<sup>(11)</sup> com os seguintes parâmetros: entropia como critério de ganho de informação, *best split* como estratégia de *split*, e sem limite para o número máximo de profundidade. O algoritmo não utiliza nenhuma técnica de poda.

Para a rede neural *Multilayer Perceptron*, foi aplicado o algoritmo *MLPClassifier*<sup>(11)</sup> com os seguintes parâmetros<sup>(6)</sup>: uma camada de entrada com o número de neurônios igual ao número de atributos da base de dados mais um, uma camada oculta com 165 neurônios com função de ativação ReLU, uma camada de saída com o número de neurônios igual ao número de classes com função de ativação *Softmax*, *Backpropagation* e otimização ADAM para o ajuste de pesos, taxa de aprendizado inicial de 0.3, taxa de momentum de 0.2, número de épocas máxima igual a 500 (ou caso não haja melhora na convergência por até 10 épocas), lote de treino de 200 amostras, e grupo de validação de 10% dos dados de treino.

Para a DNN, foi construído um modelo utilizando o *Google Colab*. Para a construção da rede, utilizou-se a classe *Sequential* com camada de entrada com um neurônio da classe *Flatten* (que apenas converte os dados para uma matriz unidimensional, sem realizar nenhum cálculo), 5 camadas ocultas com 50 neurônios e função de ativação ReLU e uma camada com o número de neurônios igual ao número de classes utilizando a função *Softmax*. Em relação aos seus parâmetros:



otimização SGD, função de perda *sparse categorical cross entropy*, lote de treino de 32 amostras, 500 épocas e grupo de validação de 20% dos dados de treino.

## Pré-processamento

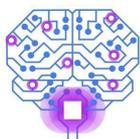
O principal foco foi equilibrar melhor as proporções entre as categorias de batimentos e também extrair características da base de dados para que o algoritmo de árvore de decisão não fosse treinado com os dados brutos das séries temporais.

Inicialmente, como característica para a árvore de decisão, foi calculada (para cada batimento) as distâncias entre o pico da onda R em relação ao pico da onda R do batimento anterior e do posterior. Importante frisar que, para cada eletrocardiograma, o primeiro e o último batimento cardíaco foram descartados, já que não é possível calcular a distância entre a onda R anterior (no caso do primeiro) ou da onda R posterior (no caso do último). Dessa forma, os atributos da base de dados foram compostos pelos pontos da série temporal, a distância da onda R anterior, distância de onda R posterior e sua respectiva classe.

Para tratar da discrepância entre o número de instâncias de cada classe, foram definidas duas bases de dados balanceadas através do processo de *undersampling*, em que parte das instâncias de classes predominantes são retiradas da base para que seus números de exemplos sejam mais próximos das classes com menos instâncias.

Em uma base foi efetuado um balanceamento estrito, em que a quantidade de instâncias em todas as classes é igual à quantidade de instâncias da classe minoritária. Já na outra base, foi realizado um balanceamento *ad hoc* (arbitrário) por parte do autor, em que o objetivo foi o de não realizar um *undersampling* tão acentuado e apenas evitar valores muito discrepantes entre classes com mais instâncias e menos instâncias. O resultado de cada balanceamento está descrito na Tabela 2.

Em seguida ao balanceamento, fez-se a normalização dos pontos de todas as séries temporais de modo que a média entre elas seja igual a 0 e o desvio padrão seja igual a 1. Esse processo é importante tanto para obter melhor treinamento das redes neurais quanto para a extração das componentes principais para o algoritmo de árvore de decisão<sup>(14)</sup>.

**Tabela 2** - Proporção de 11 tipos de batimentos cardíacos da base de dados com balanceamento estrito e balanceamento arbitrário

<b>Classe</b>	<b>Quantidade (estrito)</b>	<b>Proporção (estrito)</b>	<b>Quantidade (arbitrário)</b>	<b>Proporção (arbitrário)</b>
NORMAL	83	9,09%	250	11,20%
LBBB	83	9,09%	250	11,20%
RBBB	83	9,09%	250	11,20%
ABERR	83	9,09%	150	6,73%
PVC	83	9,09%	250	11,20%
FUSION	83	9,09%	250	11,20%
NPC	83	9,09%	83	3,73%
APC	83	9,09%	250	11,20%
VESC	83	9,09%	106	4,76%
NESC	83	9,09%	200	8,96%
FWAV	83	9,09%	192	8,62%
<b>Total</b>	<b>913</b>	<b>100%</b>	<b>2.231</b>	<b>100%</b>

As componentes principais tanto da base de dados balanceada quanto da normalizada são extraídas para a árvore de decisão. O número de componentes principais extraídas é o mínimo que representa uma variância de 90% entre os atributos, sendo 9 para base estritamente normalizada e 10 para arbitrariamente normalizada, respectivamente.

A proporção entre o número de instâncias de todas as bases de dados de cada algoritmo foram mantidas. Dessa forma, cada algoritmo foi treinado com duas bases de dados: uma estritamente balanceada e outra arbitrariamente balanceada. A árvore de decisão foi treinada, tendo como atributos, as componentes principais e a distância da onda R anterior e posterior em relação à onda R daquele batimento. As redes neurais foram treinadas, tendo como atributos, os pontos normalizados de cada série temporal.

Para facilitar a identificação da combinação modelo e base, foi adotada uma sigla que abrevia o nome do algoritmo utilizado, seguidos de hífen e a base de dados utilizada (exemplo: AD-B1). Assim, define-se:



- AD: Árvore de decisão;
- MLP: *Multilayer Perceptron*;
- DNN: *Deep Neural Network* (Rede Neural Profunda);
- B1: Base 1 (estritamente balanceada);
- B2: Base 2 (balanceamento arbitrário).

### Métricas de análise de desempenho e método de treino

Foram escolhidas três métricas para a avaliação do desempenho dos classificadores: precisão, sensibilidade e *F1-Score*<sup>(14)</sup>.

A precisão pode ser definida como a proporção de instâncias corretamente classificadas, dada pela equação (1):

$$P = \frac{TP}{TP + FP} \quad (1)$$

Onde *TP* equivale a verdadeiros positivos e *FP* equivale a falsos positivos.

A sensibilidade pode ser definida como a proporção de acerto de verdadeiros positivos, dada pela equação (2):

$$S = \frac{TP}{TP + FN} \quad (2)$$

Onde *TP* equivale a verdadeiros positivos e *FN* equivale a falsos negativos.

O *F1-Score* é definido como a média harmônica entre precisão e sensibilidade, dado pela equação (3):

$$F1 = 2 \times \frac{P \times S}{P + S} \quad (3)$$

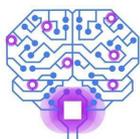
Como método de treino dos algoritmos, foi selecionado o *stratified k-fold cross-validation* com 5 *folds* (sem embaralhamento), mantendo a proporção das classes em cada um deles. Este número de *folds* permite que o conjunto de teste não fique tão reduzido (como seria no caso de 10 *folds*, por exemplo).

### Resultados e Discussão

Os resultados da execução de cada algoritmo, média da precisão, sensibilidade e *F1-Score* para cada base de dados, são apresentados na Tabela 3.

**Tabela 3** - Média da precisão, sensibilidade e *F1-score* para cada algoritmo/base de dados com balanceamento estrito

Algoritmo - Base	Precisão	Sensibilidade	<i>F1-Score</i>
AD - B1	78% (13)	77% (16)	76% (14)



<b>AD - B2</b>	79% (11)	79% (11)	79% (11)
<b>MLP - B1</b>	79% (8)	78% (10)	77% (9)
<b>MLP - B2</b>	79% (11)	79% (14)	78% (12)
<b>DNN - B1</b>	92% (4)	91% (6)	91% (4)
<b>DNN - B2</b>	94% (5)	94% (5)	94% (5)

Em seguida, foi realizado um teste estatístico a fim de verificar se há diferença estatística significativa entre os modelos com ambas as bases de dados.

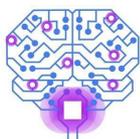
Os valores utilizados no teste são o desempenho, de cada *fold*, da média da medida *F1-Score* em cada uma das 11 classes para as duas bases de dados. O desempenho de cada algoritmo, em cada *fold* e para as duas bases de dados, é mostrado na Tabela 4.

**Tabela 4** - Média do *F1-score* para cada algoritmo/base de dados em cada *fold*

<b>Fold</b>	<b>AD - B1</b>	<b>AD - B2</b>	<b>MLP - B1</b>	<b>MLP - B2</b>	<b>DNN - B1</b>	<b>DNN - B2</b>
1	82%	83%	78%	74%	91%	88%
2	75%	77%	78%	75%	94%	88%
3	80%	77%	81%	77%	97%	93%
4	82%	79%	76%	78%	98%	93%
5	78%	89%	82%	82%	90%	91%

Fez-se um teste de normalidade para verificar se os dados são paramétricos (seguem uma distribuição normal). Para tal, utilizou-se o teste de *Shapiro-Wilk*, onde a hipótese nula é definida como os dados serem paramétricos e o valor  $p$  para se rejeitar essa hipótese é 0,05. Os resultados estão apresentados na Tabela 5. Como não foi possível rejeitar a hipótese nula, assumiu-se que os dados seguem a distribuição normal, portanto, um teste estatístico paramétrico é adequado.

Devido ao uso da mesma base de dados para os algoritmos de árvore de decisão e as redes neurais MLP e DNN, sendo que para a árvore de decisão houve a extração das componentes principais, mas ainda assim dos mesmos dados, pode-se assumir que os grupos são pareados.

**Tabela 5** - Valor-p resultante do teste de normalidade *Shapiro-Wilk*

Base	Algoritmo		
	AD	MLP	DNN
B1	0,39	0,56	0,50
B2	0,38	0,67	0,10

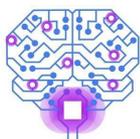
Como a comparação é feita entre 6 grupos, sendo a base de dados estritamente e arbitrariamente balanceada para cada algoritmo, os dados seguem distribuição normal e os grupos são pareados (por se tratar dos mesmos dados), foi escolhido o teste ANOVA *Repeated Measures* para verificar a existência de diferença estatística significativa entre o desempenho das abordagens.

Antes de aplicar o ANOVA *Repeated Measures*, o teste de *Brown-Forsythe* foi realizado para garantir que há homogeneidade nas variâncias entre os 6 grupos. Define-se a hipótese nula como não haver diferença estatisticamente significativa entre as variâncias e a condição para rejeitar essa hipótese um valor-*p* menor que 0,05. Após sua aplicação, o teste resultou em um valor-*p* de 0,93.

Dessa forma, aplicou-se o teste ANOVA *Repeated Measures* onde a hipótese nula foi definida como não haver diferença estatisticamente significativa entre os grupos e a condição para rejeitar essa hipótese é um valor-*p* menor que 0,05. O teste resultou em um valor-*p* menor que 0,001 sendo possível afirmar que há diferença estatisticamente significativa entre os desempenhos dos algoritmos pelos critérios adotados. Por isso, fez-se um pós-teste para verificar quais algoritmos/bases de dados possuem diferença estatisticamente significativa entre si. Para tal, adotou-se o teste de *Tukey*, onde a hipótese nula é não haver diferença estatística significativa entre dois modelos, e o grau de confiança para rejeitar esta hipótese é um valor-*p* menor que 0,05. Os resultados são mostrados na Tabela 6.

**Tabela 6** - Valor-*p* do teste de *Tukey* com comparações em pares

	AD - B1	AD - B2	MLP - B1	MLP - B2	DNN - B1	DNN - B2
AD - B1	-	1,000	1,000	0,809	0,001	0,001



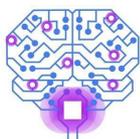
<b>AD - B2</b>	1,000	-	1,000	0,862	0,001	0,001
<b>MLP - B1</b>	1,000	1,000	-	0,906	0,001	0,001
<b>MLP - B2</b>	0,809	0,862	0,906	-	0,001	0,001
<b>DNN - B1</b>	0,001	0,001	0,001	0,001	-	0,416
<b>DNN - B2</b>	0,001	0,001	0,001	0,001	0,416	-

Analisando a Tabela 6, é possível afirmar que a DNN treinada, tanto com a base estritamente balanceada quanto com a base com balanceamento arbitrário, possuem um desempenho superior às outras abordagens, mas não possuem diferença estatisticamente significativa entre si. Em relação à árvore de decisão e à rede MLP, ambos algoritmos em ambas as bases de dados, não é possível afirmar que há diferença de desempenho entre eles.

Como não há diferença estatisticamente significativa entre o algoritmo treinado com as duas bases, foi escolhido o algoritmo DNN treinado com a base estritamente balanceada (Base 1) para comparar com o desempenho de outros trabalhos pela sua precisão, sensibilidade e *F1-Score*. É importante frisar que trabalhos que não possuem *F1-Score*, mas possuem precisão e sensibilidade tiveram seu resultado calculado por este autor. Essa comparação só pode ser feita de forma parcial e visualmente, já que diversos trabalhos utilizam classes diferentes, balanceamentos diferentes, métricas diferentes e escopos diferentes (classificar apenas arritmias consideradas letais, por exemplo). A comparação entre este estudo e outros trabalhos está descrita na Tabela 7.

**Tabela 7** - Comparação com outros trabalhos que utilizam a mesma base de dados

<b>Trabalho</b>	<b>Precisão</b>	<b>Sensibilidade</b>	<b><i>F1-Score</i></b>
<b>DNN - B1 (Proposto)</b>	92%	91%	91%
<i>Random Forests</i> com as séries temporais <sup>(6)</sup>	98,30%	98,30%	98,30%
Rede Neural Convolutiva com base de dados balanceada <sup>(5)</sup>	90,00%	97,40%	93,55%
<i>Support-vector Machine</i> <sup>(7)</sup>	86,82%	60,36%	71,21%
Modelo obtido pelo algoritmo	84%	85%	84,49%



<i>Sequential Floating Feature Selection</i> <sup>(4)</sup>			
<i>Echo State Network</i> <sup>(3)</sup>	92,70%	95,70%	94,17%

## Conclusão

Para o problema de classificação de 11 batimentos escolhidos entre os diversos presentes nos registros do *MIT-BIH Arrhythmia Database*, foram construídos três modelos para serem treinados com duas bases de dados com balanceamentos diferentes.

Dos resultados deste trabalho, pode-se afirmar que tanto a DNN treinada com a base de dados estritamente balanceada quanto com a base com balanceamento arbitrário tiveram desempenho superior aos outros algoritmos de árvore de decisão e *Multilayer Perceptron*, considerando a métrica *F1-Score*, e que não tiveram diferença de desempenho significativa entre si. Seu desempenho foi, para a base de dados estritamente balanceada e com balanceamento arbitrário, 91% (4) e 94% (5), respectivamente.

Propõe-se, como trabalhos futuros, a comparação entre modelos ou bases de dados utilizando o desempenho de cada classificador em cada classe, especificamente. Também há de se explorar outras técnicas de balanceamento de dados, em especial as de *oversampling*, visando aumentar a quantidade de classes que o modelo consegue prever. O uso de outra derivação também é sugerida para comparar o desempenho com a derivação D2. Por fim, o uso de outra base de dados distinta da *MIT-BIH Arrhythmia Database* para verificar se há melhora no desempenho.

## Referências

1. World Health Organization. Cardiovascular diseases [Internet]. Who.int. World Health Organization: WHO; 2022. Available from: [https://www.who.int/health-topics/cardiovascular-diseases#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1)
2. Pastore C, Pinho J, Pinho C, Samesima N, Pereira-Filho H, Kruse J, et al. III Diretrizes da Sociedade Brasileira de Cardiologia sobre análise e emissão de laudos eletrocardiográficos. *Arquivos Brasileiros de Cardiologia*. 2016;106(4).
3. Alfaras M, Soriano MC, Ortín S. A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection. *Frontiers in Physics*. 2019 Jul 18;7.



4. Llamedo M, Martínez JP. Heartbeat Classification Using Feature Selection Driven by Database Generalization Criteria. *IEEE Transactions on Biomedical Engineering*. 2011 Mar;58(3):616–25.
5. Shaker AM, Tantawi M, Shedeed HA, Tolba MF. Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks. *IEEE Access*. 2020;8:35592–605.
6. Hübner LG. Classificação de ritmos cardíacos em tempo real aplicando tecnologias embarcadas [Internet] [Dissertation]. [Universidade Estadual do Oeste do Paraná]; 2020. p. 1–148. Available from: <https://tede.unioeste.br/handle/tede/5114>
7. Zhang Z, Dong J, Luo X, Choi K-S, Wu X. Heartbeat classification using disease-specific feature selection. *Computers in Biology and Medicine*. 2014 Mar;46:79–89.
8. Hall JE. *Guyton and Hall Textbook of Medical Physiology : Enhanced E-book*. London: Elsevier Health Sciences; 2012.
9. MIT-BIH Arrhythmia Database v1.0.0 [Internet]. [physionet.org](https://physionet.org). Available from: <https://physionet.org/content/mitdb/1.0.0/>
10. Can Ye, Coimbra MT, Vijaya Kumar BVK. Arrhythmia detection and classification using morphological and dynamic features of ECG signals. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. 2010 Aug;
11. User guide: contents — scikit-learn 0.22.1 documentation [Internet]. [Scikit-learn.org](https://scikit-learn.org). 2019. Available from: [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)
12. Home - Keras Documentation [Internet]. [Keras.io](https://keras.io). 2019. Available from: <https://keras.io>
13. TensorFlow. TensorFlow [Internet]. TensorFlow. 2019. Available from: <https://www.tensorflow.org>
14. Aurélien Géron. *Hands-on machine learning with Scikit-Learn and TensorFlow concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.; 2019.
15. Gifari MW, Zakaria H, Mengko R. Design of ECG Homecare:12-lead ECG acquisition using single channel ECG device developed on AD8232 analog front end. 2015 International Conference on Electrical Engineering and Informatics (ICEEI). 2015 Aug;