

Classificação automática de sinais ECG utilizando Discrete Wavelet Transform

Rafael Francisco de Oliveira

Departamento de Computação

Universidade Federal de Ouro Preto - UFOP

Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

Email: rafael.fo@aluno.ufop.edu.br

Abstract—O processo de identificação e classificação de doenças cardiovasculares (DCVs), especialmente as arritmias em eletrocardiogramas (ECGs) pode ser muito exaustivo para um ser humano e, sujeito a erros, uma vez que seria necessário uma análise de batimento a batimento em horas ou até mesmo dias de registros de ECG capturados. A automatização desse processo torna-se desejável e atualmente, diversas técnicas de inteligência artificial têm sido aplicadas. Neste estudo, com o objetivo de obter um diagnóstico preciso dos tipos de DCVs, de acordo com a arritmia em batimentos cardíacos de ECG, um modelo de classificação automática de arritmias em ECG é proposto baseado em *Discrete Wavelet Transform* - (DWT) com função geradora db4. O modelo é avaliado em cinco classificadores: *XGradientBoost*, *AdaBoost*, *RandomForest*, *GradientBoost* e *SVM*. Testando no conjunto de dados criado pela parceria entre a Universidade de Chapman e o Hospital Shaoxing Escola de Medicina da Universidade de Zhejiang, com uso de paradigma *inter-patient*, o modelo proposto alcançou sensibilidade para as classes AFIB, GSVT, SB, SR de 82%, 83%, 99%, 91%, respectivamente, para o classificador *XGradientBoost*. Ainda, a predição positiva alcançou 85%, 89%, 98%, 81%, respectivamente. Comparado com os classificadores testados, F1-score obteve um aumento de 25,35%. Os resultados obtidos mostram potencial de implementação em cenários reais com o uso de técnicas *ensembles*.

Keywords: Arritmia; SVM; Boosting; Eletrocardiograma; Batimentos; ECG; inteligência artificial.

I. INTRODUÇÃO

De acordo com a Organização Mundial de Saúde (OMS), doenças cardíacas são a principal causa de morte no planeta. Embora boa parte das doenças cardíacas possam ser diagnosticadas com antecedência, o número de casos continua a crescer. Estima-se que mais de 75% dos casos acontecem em países pobres e em desenvolvimento e 17,9 milhões de pessoas morreram de doenças cardíacas em 2019, representando 32% das mortes globais [1]. Isso se deve ao fato de que, medidas de prevenção, aparentemente, não estão sendo adequadas, por isso o diagnóstico precoce é de fundamental importância.

O principal exame para diagnóstico de doenças cardíacas é chamado Eletrocardiograma (ECG). Segundo Cohen [2], o ECG foi um dos primeiros sinais estudados pelas técnicas de processamento de sinais biomédicos e até os dias de hoje é a técnica mais utilizada no diagnóstico de doenças do coração. Isto é possível por causa da simplicidade do exame, o qual é um processo não-invasivo.

Um dos problemas cardíacos detectável pelo exame de ECG são as arritmias. Arritmias são batimentos cardíacos anormais

ou irregulares. Há duas categorias de arritmias: as arritmias formadas por um único batimento cardíaco irregular, denominadas de arritmias morfológicas, e as arritmias formadas por um conjunto irregular de batimentos, denominadas arritmias rítmicas.

O processo de identificação e classificação de arritmias pode ser muito exaustivo para um ser humano e, ainda sujeito a erros, visto que seria necessário uma análise de batimento a batimento em horas ou até mesmo dias de registros de ECG capturados. Portanto, existe a necessidade de estudos por soluções que mitigue esses erros. Uma alternativa consiste no uso de inteligência artificial para classificação automática dos padrões de arritmia. Atualmente, métodos baseados em aprendizagem de máquina, do inglês *machine learning*, tornaram-se bastante populares, obtendo resultados expressivos [3]. Porém, ainda há muito espaço para melhorias, especialmente quando se considera uma avaliação mais rigorosa [4].

Uma grande dificuldade para o uso em campo prático de técnicas estado-da-arte, como as baseadas em aprendizagem de máquina, é a sua integração com os equipamentos médicos. Para que as soluções propostas possam ser utilizadas em monitores de sinais vitais ou desfibriladores, por exemplo, a *Association for the Advancement of Medical Instrumentation* (AAMI) [4] propõe normas para desenvolvimento de projetos na área médica.

Além disso, deve-se seguir outras normas e boas práticas de desenvolvimento para construção de software para equipamentos médicos-hospitalares. Manter a conformidade com todas as normas e boas práticas é ainda um grande obstáculo para a introdução de novas técnicas na indústria. As abordagens publicadas na literatura que não seguem esses requisitos tendem a ser superestimados e assim, sem grande valor prático para fabricantes de equipamentos médicos [5].

O processo de classificação de arritmias divide-se em três partes principais: pré-processamento, extração de características e classificação. Na etapa de pré-processamento, são aplicados métodos e técnicas com o objetivo de preparar o sinal de modo que possa ser extraído as melhores informações para caracterização do mesmo, tais como filtros para eliminar a interferência de ruídos no sinal ECG, normalização para especificar intervalo de valores para todos os atributos, amostragem para conjunto de dados desbalanceados. Para etapa de extração de características, há diversos métodos na lit-

eratura, dentre eles a *Discrete Wavelet Transform (DWT)*, que permite obter uma representação multirresolução de alguns padrões da estrutura do sinal no domínio tempo-frequência. Em seguida, aplica-se algum modelo classificador, como por exemplo, Support Vector Machine, redes neurais artificiais, Random Forest, os quais são treinados com as características extraídas do sinal.

Nesse projeto, propõe-se a classificação automática de arritmia a partir de sinais de eletrocardiograma. Para isso, a *Discrete Wavelet Transform (WT)* é utilizada nas etapas de extração de características e a classificação é realizada por meio de cinco classificadores, sendo eles: *XGradientBoost*, *AdaBoost*, *RandomForest*, *GradientBoost* e *SVM*. A proposta é a realização de um teste comparativo entre os modelos em relação à performance. Os experimentos foram validados utilizando o conjunto de dados produzido pela parceria entre a Universidade de Chapman e o Hospital Shaoxing Escola de Medicina da Universidade de Zhejiang, disponível em [6] diante do paradigma *inter-patient*.

O trabalho é organizado como segue. Seção II mostra uma breve revisão da literatura referente às etapas necessárias para um sistema de classificação de ECG, bem como os métodos existentes. Seção III define o conjunto de dados utilizado para avaliação do modelo. A metodologia proposta para a classificação de ECG é descrita na Seção IV. Além disso, na Seção V, a descrição dos experimentos e resultados são mostrados. Na Seção VI, as discussões dos resultados obtidos e comparações com outros trabalhos da literatura são apresentados. Finalmente, as observações finais são reportadas na Seção VII.

II. TRABALHOS RELACIONADOS

Um sistema completo de classificação e reconhecimento de arritmias em ECG pode ser dividido em três partes principais: 1) pré-processamento, 2) extração de características e 3) classificação.

Para melhor extrair informações úteis do sinal, algumas técnicas são aplicadas. A etapa de pré-processamento é onde ocorre a remoção de ruído do sinal, como por exemplo, interferência na linha de energia e *baseline wandering*. Vários métodos tem sido propostos, como filtros-bandas [7], [8], métodos baseados em *wavelet transform* [9], [10], filtros adaptativo [11], [12], decomposição empírica [13], e análise de componente independente [14]. Deve-se observar que a aplicação excessiva de filtros pode ocasionar perda de informações úteis para a classificação. T. Wang et al. [15] e Dias et al. [16] utilizaram dois filtros de média móvel, um deles com largura de 200 ms e outro com largura de 600 ms para remoção do *baseline wandering* do sinal, ocasionado pela respiração ou pelo movimento do paciente. L. Wang et al. [10] propuseram um método de seleção *wavelet* com limiar adaptativo para redução de ruído em ECG baseado em funções *wavelet*. O limiar foi ajustado dinamicamente utilizando heurística e o número de camadas da decomposição da função *wavelet*. Blanco-Velasco et al. [17] utilizaram modo de decomposição empírica para remoção de altas frequências

de ruídos e *baseline wandering*. Em [18], a decomposição adaptativa de Fourier foi escolhida para diminuição o ruído dos sinais de ECG.

A detecção de arritmias com base na classificação de ECG tornou-se um tópico de pesquisa muito popular na literatura nos últimos anos [5]. Os métodos tradicionais utilizam características extraídas do sinal para a classificação dos batimentos cardíacos, como, por exemplo, o intervalo RR, coeficientes *wavelets*, amplitude, complexo QRS, onda P, duração, etc. Os principais pontos e intervalos, denominados de pontos fiduciais, de um sinal ECG são mostrados na Fig. 1. Dentro os pontos apresentados, temos o segmento QRS, o qual normalmente é utilizado durante o processo de segmentação de um sinal de ECG em batimentos menores.

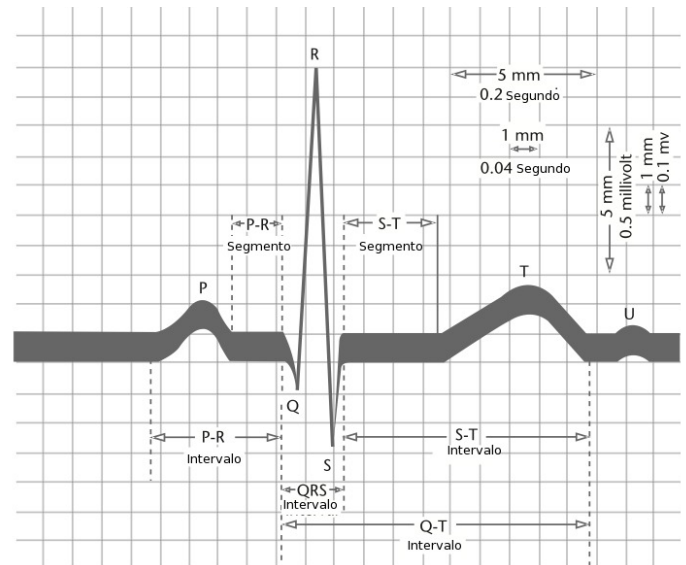


Fig. 1: Pontos fiduciais de um batimento normal de ECG correspondente à derivação II [19].

Após a aplicação dos filtros, no pré-processamento pode ser ainda aplicado outras técnicas a fim de adequar o sinal de forma que as melhores informações possam ser extraídas, tais como normalização e amostragem. Alguns trabalhos utilizam uma etapa de segmentação antes da extração de características, onde os sinais são divididos em batimentos individuais de acordo com um tamanho fixo de amostras. Essa etapa de segmentação requer detecção das ondas QRS e pontos fiduciais dos batimentos [15]. No entanto, outros trabalhos da literatura normalmente utilizam a posição da onda R rotulada no conjunto de dados [5]. Uma abordagem diferente é proposta por Ye et al [20], onde propuseram a adição sintética de um tremor (*jitter*) na posição da onda R. O tremor é modelado com média zero e desvio padrão de 5 amostras, isto é, 13,9 ms. Por ser um conjunto de dados relativamente recente, diferente do tradicional MIT-BIH por exemplo, nesse trabalho não foi realizado a etapa de segmentação, sendo o sinal utilizado por completo.

A identificação dos sinais ocorre a partir de características extraídas dos sinais de ECG. Com isso, várias características

e métodos têm sido utilizados na literatura em diversas propostas. De Chazal *et al* [21] combinaram características rítmicas, duração de onda e características morfológicas. Zheng *et al* [22] utilizaram mais de 200 características no classificador utilizado, tais como idade, gênero, características do canal II do ECG além de características extraídas dos 12 canais presente no banco de dados utilizado. L. Wang *et al.* [10] propuseram um método de seleção de características baseado em transformação *wavelet* com limiar adaptativo, visando reduzir o ruído do sinal de ECG. Rakshit e Das [13] propuseram uma metodologia utilizando o modo de decomposição empírica (*empirical mode decomposition (EMD)*) e filtro de comutação média adaptativa (*adaptive switching mean filter (ASMF)*) para decompor os sinais ECG, gerando um conjunto de funções intrínsecas (*intrinsic mode functions (IMFs)*) com base na complexidade do sinal para remover a interferência de ruído. A *Wavelet Transform* pode ser utilizada para extrair características, uma vez que a mesma realiza uma decomposição adaptativa em tempo-frequência do sinal de ECG. Pela representação multirresolução é possível descrever a estrutura do sinal por apenas alguns coeficientes do domínio *Wavelet* [23]. A Fig. 2 exemplifica a decomposição de um sinal ECG em dez sub-bandas de frequências.

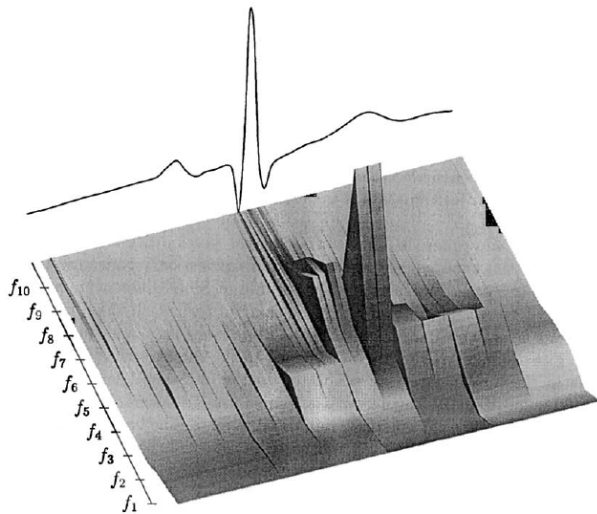


Fig. 2: Representação de multirresolução de um sinal ECG pelos coeficientes *wavelet* dividido em 10 sub-bandas [23].

Os métodos de classificação que utilizam técnicas de aprendizagem de máquina apresentam limitações de performance quando lidam com conjuntos de dados desbalanceados. Em particular, esses conjuntos de dados atingem um valor preditivo positivo baixo. Além disso, baixa sensibilidade para as classes com menor quantidade de amostras, ocasionando em resultados enviesados. Adicional ao problema de conjuntos de dados desbalanceados, métodos que não levam em consideração normas de separação de dados em conjuntos de treinamento e teste tendem a obter resultados altos e enviesados.

Com o objetivo de tornar o modelo em desenvolvimento mais robusto e adaptável a um cenário no mundo real, al-

guns trabalhos na literatura exploraram dois paradigmas de validação de dados: *intra-patient* e *inter-patient* [24]. No paradigma *intra-patient*, o conjunto de treino e teste é formado por dados de um mesmo paciente, enquanto que no paradigma *inter-patient*, proposto por De Chazal *et al.* [21], o conjunto de treino e teste são obtidos de pacientes diferentes. Esse último permite uma generalização mais próxima de um cenário do mundo real.

Mousavi e Afghah [24] propuseram um modelo baseado em redes neurais de aprendizado profundo (*deep learning*) aplicado com metodologia *sequence to sequence* e algoritmo SMOTE para balanceamento das classes durante a etapa de treinamento. Os autores avaliaram o modelo com trabalhos estado-da-arte da literatura de acordo com os paradigmas apresentados anteriormente. O modelo proposto obteve performance superior em relação a todos os trabalhos comparados, em ambos paradigmas, obtendo acurácia de 99,92% no paradigma *intra-patient* e 99,53% no paradigma *inter-patient*.

Para melhor avaliação do desempenho real dos classificadores e garantir uma competição justa entre modelos propostos na literatura, alguns trabalhos [8], [16], [21], [25] fizeram o uso do paradigma *inter-patient* com o mesmo objetivo. Logo, este trabalho propõe o uso do mesmo paradigma no desenvolvimento metodológico.

Após a extração das características do sinal de ECG, um classificador de aprendizado de máquina é treinado utilizando as características do sinal. A escolha do classificador ocorre de acordo com o problema a ser tratado. Na literatura há diversos modelos, dos quais SSVM [26], *Random Forest* [25], *AdaBoost* [27], *XGBoost* [28] e *GradientBoost* [28] são utilizados e comparados neste trabalho.

III. CONJUNTO DE DADOS

Esse trabalho utiliza o conjunto de dados produzido pela parceria entre a Universidade de Chapman e o Hospital Shaoxing Escolade Medicina da Universidade de Zhejiang [6].

O conjunto de dados é composto por 10.646 sinais de ECGs de pacientes, sendo 5,956 homens e 4,690 mulheres. Dentre esses pacientes, 17% apresentam arritmia normal e 83% apresentam pelo menos algum tipo de arritmia. A idade dos pacientes variam de 4 a 98 anos. Cada sinal apresenta informação de 12 derivações (configurações das posições dos elétrodos no paciente) com duração de 10 segundos e amostrados a uma taxa 500 Hz. Nesse trabalho, a derivação II foi a utilizada.

O conjunto de dados apresenta 11 tipos de arritmias (classes) com diferentes números de amostras. Devido a presença de arritmia com menos de 10 amostras, denominadas de classes minoritárias, cardiologistas propuseram um agrupamento em quatro grupos durante a composição do conjunto de dados [22]. A Tabela I mostra o detalhamento de todas as arritmias presentes no conjunto de dados, bem como o agrupamento e amostragem de cada uma.

O conjunto de dados ainda contém 11 características dos sinais, sendo eles: taxa ventricular, taxa atrial, duração QRS, intervalo QT, QT corrigido, eixo R, eixo T, contagem QRS,

Tabela I: Descrição das arritmias (classes) do conjunto de dados e o agrupamento propostos por cardiologistas.

Classes	Frequência	Agrupamento	Total
SR (<i>Sinus Rhythm</i>)	1826	SR	1826
AFIB (<i>Atrial Fibrillation</i>) AF (<i>Atrial Flutter</i>)	1780 445	AFIB	2225
ST (<i>Sinus Tachycardia</i>) SVT (<i>Supraventricular Tachycardia</i>) AT (<i>Atrial Tachycardia</i>)	1568 587 121	GSVT	2706
AVNRT (<i>Atrioventricular Node Reentrant Tachycardia</i>)	16		
AVRT (<i>Atrioventricular Reentrant Tachycardia</i>)	8		
SAWR (<i>Sinus Atrium to Atrial Wandering Rhythm</i>)	7		
SA (<i>Sinus Atrium</i>)	399		
SB (<i>Sinus Bradycardia</i>)	3889	SB	3889
Total			10646

QOnset, QOffset e TOffset, além da idade e do gênero de cada paciente.

Como cada registro no conjunto de dados corresponde a um paciente diferente, o paradigma *inter-patient* foi aplicado, formado um conjunto de dados para treinamento diferente do conjunto de dados para teste, permitindo assim, uma generalização do modelo classificador para um cenário real. Durante a classificação, o conjunto de dados foi dividido em conjunto de treinamento e teste de forma randômica, a taxa de 80%-20%, respectivamente. O conjunto de treinamento foi composto por 8517 amostras e o conjunto de teste por 2129 amostras.

IV. METODOLOGIA

A seguir, é apresentado a metodologia proposta para classificação automática de arritmias em ECG. O diagrama da Fig. 3 mostra as etapas do sistema de classificação proposto: pré-processamento, extração de características e classificação.

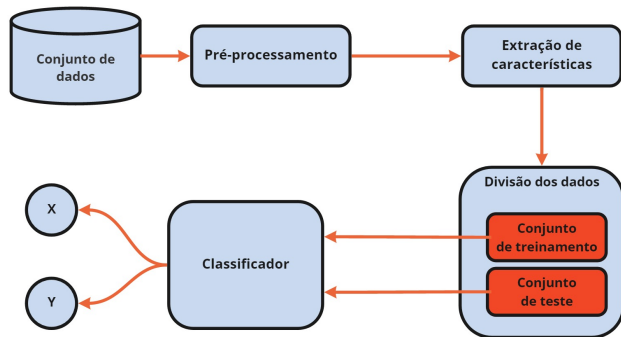


Fig. 3: Diagrama do sistema de classificação de arritmia.

A. Pré-processamento

Nessa etapa, uma análise descritiva permite verificar como os dados estão presentes no conjunto de dados, bem como sua distribuição, presença de valores nulos, *outliers* e duplicatas, correlações de atributos, dentre outros. Nesse trabalho,

esta etapa consistiu no agrupamento das classes de arritmias proposto pelos cardiologistas conforme a Tabela I, remoção de duplicatas, remoção de dois atributos desnecessários (*File-Name* e *Beat*) para a classificação e preenchimento de valores nulos por zero devido a baixa presença no conjunto de dados.

O processo de normalização foi realizado apenas para o classificador *Support Vector Machine* (SVM), sendo a normalização utilizada pela Fórmula 1 no intervalo de 0-1.

$$X' = \frac{X - X_{\text{mínimo}}}{X_{\text{máximo}} - X_{\text{mínimo}}} \quad (1)$$

onde X é o valor de cada amostra, $X_{\text{mínimo}}$ é a amostra mínima, $X_{\text{máximo}}$ é a amostra máxima e X' a amostra normalizada.

Quanto ao processo de remoção de ruído, o conjunto de dados disponibiliza os sinais brutos e os sinais com remoção de ruídos, não sendo necessário essa etapa no pré-processamento. No artigo responsável pela divulgação do conjunto de dados [22], Zheng *et al* utilizaram uma sequência de três filtros para tal processo, sendo um filtro passa-baixa *Butterworth* com passa-banda de 50 Hz e faixa de 60 Hz, suavizador *LOESS* e *Non Local Means* (NLM).

B. Extração de características

Os sinais são compostos por diferentes frequências. Por isso, nessa etapa de extração de características, os sinais são convertidos para o domínio tempo-frequência utilizando a DWT. Como os diferentes tipos de sinais exibem características de frequência diferentes, essa diferença de comportamento é capaz de gerar características para cada sinal, as quais posteriormente tornam-se um vetor de características para entrada em um classificador.

Wavelet Transform (WT) é uma série composta por operações de escalonamento e deslocamento em base de uma função geradora chamada de *mother wavelet*, podendo ser contínuas ou discretas. WT pode representar uma informação local do sinal de ECG no domínio do tempo e frequência para ajustar as sub-bandas de alta e baixa frequência separadamente através das janelas de tempo e de frequência. WT

tem sido utilizada em diferentes metodologias, como detecção de batimentos [29] e extração de características [30].

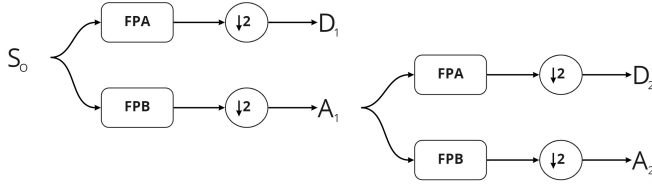


Fig. 4: Processo de decomposição do sinal em duas camadas [10].

A *Continuous Wavelet Transformation (CWT)* realiza o escalonamentos e deslocamentos de forma contínua enquanto que a *Discrete Wavelet Transform (CWT)* realiza as operações de forma discreta. As fórmulas (2) e (3) [31] representam as funções contínuas e discretas, respectivamente:

$$X_w(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (2)$$

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k2^j}{2^j}\right) \quad (3)$$

onde $x(t)$ é o sinal original, $\bar{\psi}(t)$ é a função *wavelet* contínua, $[a > 0, a \in \mathbb{R}^{+*}, j]$ são fatores de escalonamento e $[b, j]$ são fatores de deslocamento.

A DWT decompõe o sinal de ECG iterativamente, e em cada iteração, o sinal é decomposto a partir do valor médio da frequência do sinal de entrada. Para isso, utiliza-se de filtros de passagem alta (FPA) e filtros de passagem baixa (FPB). A Fig. 4 mostra o processo de decomposição de um sinal em duas camadas da WT, por exemplo.

O sinal de entrada S_0 é filtrado na sub-banda de alta frequência D_1 e na sub-banda de baixa frequência A_1 , e então a A_1 é filtrado novamente em sub-bandas de alta e baixa frequência, D_2 e A_2 , respectivamente. O processo de decomposição do sinal S_0 pode ser resumido pela fórmula (4):

$$S_0 = D_1 + D_2 + A_2 \quad (4)$$

Após a decomposição, a depender do tipo da função geradora, é possível compor o sinal original novamente utilizando os coeficientes gerados com pouca perda de informação.

A função *db4* foi utilizada no processo de extração de características por ser semelhante a onda R presente em cada sinal. Essa função gera 10 sub-bandas diferentes para cada sinal. Em cada sub-banda é extraída 12 características, conforme a Tabela II, totalizando 120 características para cada sinal. *Zero crossing* e *mean crossing* significa quantas vezes o sinal cruzou ($y = 0$) e $y = \text{media}(y)$, respectivamente.

C. Classificação

Após aplicar as etapas de pré-processamento e extração de características, as características obtidas são enviadas para um

Tabela II: Conjunto de características extraídas de cada sinal.

CARACTERÍSTICAS	
• 5-percentil	• 25-percentil
• 75-percentil	• 95-percentil
• mediana	• média
• desvio padrão	• variância
• média quadrática	• entropia
• <i>zero crossing</i>	• <i>mean crossing</i>

classificador. Nesse trabalho, é abordado cinco classificadores: *Support Vector Machine (SVM)*, *XGradientBoost*, *AdaBoost*, *RandomForest* e *GradientBoost*.

SVM foi proposto por Vapnik [32], *SVM* são métodos que constroem classificadores a partir da criação de hiperplanos em um espaço n -dimensional, isto é, desenha “linhas” em um espaço n -dimensional que são capazes de separar exemplos de diferentes classes [33], [34]. Estes hiperplanos são criados com o auxílio de pontos localizados nas bordas dos conjuntos, chamados *vetores de suporte*. Normalmente, os dados em análise não são facilmente separáveis em hiperplanos, uma vez que os dados, em sua maioria, são não-lineares. Como solução, propõe-se que os dados em espaço original sejam mapeados em um hiper espaço n -dimensional ou espaço em dimensão superior para facilitar a separação dos dados.

Algoritmos de aprendizagem de máquina combinados (*Ensembles*) são modelos compostos que combinam uma série de classificadores de baixo desempenho com o objetivo de criar um classificador aprimorado, ou seja, são meta-algoritmos que combinam vários métodos de aprendizado de máquina em um único modelo preditivo para aumentar o desempenho. Os algoritmos combinados podem diminuir a variância utilizando a abordagem de ensacamento (*bagging*), reforço (*boosting*) e empilhamento (*stacking*).

Bagging significa agregação de *bootstrap*. Ele combina vários aprendizados de forma a reduzir a variação das estimativas. Por exemplo, a floresta aleatória (*Random Forest*) treina a árvore de decisão M , é possível treinar M diferentes árvores em diferentes subconjuntos aleatórios dos dados e realizar a votação para a previsão final.

Os algoritmos de *boosting* são um conjunto do classificador de baixa precisão para criar um classificador de alta precisão. O classificador de baixa precisão oferece uma precisão melhor do que jogar uma moeda. O classificador altamente preciso (ou classificador forte) oferece uma taxa de erro próxima de 0. O algoritmo de *boosting* pode rastrear o modelo que falhou na previsão precisa. Esses algoritmos são menos afetados pelo problema de sobreajuste do modelo (*overfitting*). *XGradientBoost*, *AdaBoost* e *GradientBoost* fazem parte desse grupo de algoritmos.

Por último, *stacking* é uma técnica de aprendizado de conjunto que combina várias previsões de modelos de classificação com base em um novo conjunto de dados. Esses novos dados são tratados como dados de entrada para outro classificador.

Os modelos são ajustados usando qualquer função de

perda diferenciável e arbitrária (*loss function*) e algoritmo de otimização de gradiente descendente (*gradient descent*). Isso dá à técnica o nome de “reforço de gradiente” (*gradient boosting*), já que o gradiente de perda é minimizado conforme o modelo é ajustado, de forma semelhante ao utilizado em uma rede neural.

V. RESULTADOS

A. Configurações

Os experimentos foram conduzidos em um ambiente Linux Mint 20 e Python 3.8 com Intel®i9-10900 com 10 cores físicos (20 threads) de 2.80GHz, 128GB RAM DDR4 e GPU RTX 3090 de 24GB de RAM GDDR6X.

Como medida de avaliação dos modelos propostos, três métricas foram utilizadas: predição positiva ($+P^k$), sensibilidade (Se^k) e F1-score (F_s^k).

As métricas são baseadas nas medidas de verdadeiro positivo (TP^k), falsos negativos (FN^k), verdadeiros negativos (TN^k) e falsos positivos (FP^k), de acordo com as Fórmulas 6, 5 e 7.

$$+P^k = \frac{TP^k}{TP^k + FP^k} \times 100 \quad (5)$$

$$Se^k = \frac{TP^k}{TP^k + FN^k} \times 100 \quad (6)$$

$$F_s^k = 2 \frac{Se^k(+P^k)}{Se^k + (+P^k)} \quad (7)$$

onde $k \in \{SR, AFIB, GSVT, SB\}$ indica as classes correspondente de cada métrica.

B. Resultados

O conjunto de dados foi dividido em conjunto de treinamento e conjunto de teste a uma taxa de 80%-20%, respectivamente. Para cada classificador, o conjunto de treinamento foi dividido utilizando *10-fold cross-validation* pelo método *StratifiedKFold* da biblioteca *Scikit-learn* do python. Este método permite que a divisão de *folds* ocorra preservando o percentual de amostras de cada classe. A validação cruzada (*cross-validation*) com *10-fold* utiliza *9-folds* como treinamento e *1-fold* como teste, realizando a variação dos *folds* a cada iteração. Posteriormente, o classificador é treinado novamente no conjunto de treinamento com todas as amostras e validado no conjunto de teste. A Figura 5 mostra como o conjunto de dados foi dividido em treinamento e teste. Ainda, a Tabela III mostra o número de amostra de cada classe para o conjunto de treinamento e o conjunto de teste.

VI. DISCUSSÕES

Ambos classificadores apresentam um desempenho considerável no conjunto de dados utilizado, porém o *XGradient-Boost* apresentou o melhor desempenho comparado aos outros classificadores. A métrica curva ROC (*Receiver operating characteristic*) foi utilizada como métrica de comparação. A curva ROC é um gráfico que mostra a performance de modelos

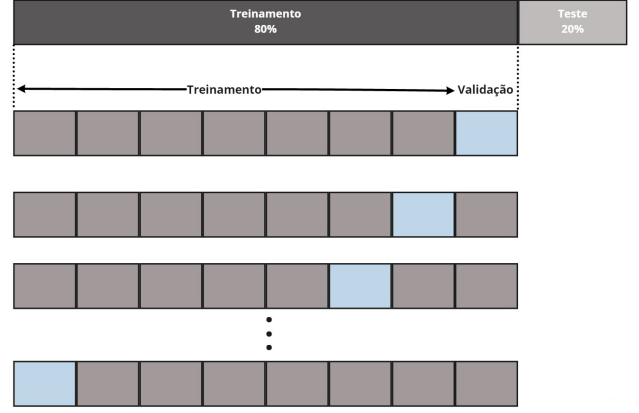


Fig. 5: Divisão do conjunto de dados em treinamento e teste.

Tabela III: Número de classes para cada classe de acordo com os conjuntos de treinamento e teste.

Classes	Treinamento	Teste	Total
AFIB	1786	439	2225
SR	1454	372	1826
SB	3121	768	3889
GSVT	2155	551	2706
Total	8516	2130	10646

de classificação considerando todos os limites de classificação baseado em dois parâmetros: a taxa de verdadeiros positivos (*True Positive Rate* - *TPR*) ou sensibilidade e a taxa de falsos positivos (*False Positive Rate* - *FPR*) ou especificidade definidas pelas fórmulas 8 e 9, respectivamente. Idealmente, um modelo é considerado ótimo quando ele apresenta *TPR* igual a 1 e *FPR* igual a 0, ou seja, o modelo consegue prever corretamente todas as classes. A área abaixo da curva ROC (*AUC*) é a medida que diz o quão bem o modelo previu corretamente as classes, portanto quanto mais próximo de 1 melhor o modelo de classificação.

$$TPR = \frac{TP}{TP + FN} \quad (8)$$

$$FPR = \frac{FP}{FP + TN} \quad (9)$$

A. Support Vector Machine

O modelo *SVM* da biblioteca *Scikit-learn* do python apresenta diversos hiperparâmetros, parâmetros utilizados para controlar o processo de aprendizado dos modelos e não são aprendidos durante o processo de treinamento. Dentre esses hiperparâmetros, dois apresentam grande influência no processo de classificação do *SVM*: o parâmetro de regularização (C) e o coeficiente de kernel (γ). A técnica utilizada para estimar os melhores valores para esses parâmetros foi a *RandomizedSearchCV* da biblioteca *Scikit-learn*. Essa biblioteca realiza uma busca randômica entre um intervalo de valores para cada hiperparâmetro, retornando o melhor resultado de

acordo com alguma métrica específica. Há ainda, a biblioteca *GridSearchCV* que realiza o mesmo processo de otimização, porém a busca é feita no conjunto formado pela combinação de todos os valores do intervalo de busca, portanto quanto maior o intervalo e o conjunto de dados, maior o tempo de processamento.

Utilizando 200 iterações, os melhores parâmetros retornados pela *RandomizedSearchCV* foram $C = 13,4$, $\gamma = 0,2$ e $\text{kernel} = \text{rbf}$, sendo os mesmos utilizados na classificação do conjunto de dados. As Tabelas IV e V mostram as matrizes de confusão do *SVM* para o conjunto de treinamento e teste, respectivamente.

Tabela IV: Matriz de confusão do conjunto de treinamento do *SVM*.

		Predição			
		AFIB	GSVT	SB	SR
Verdade	AFIB	1565	117	45	59
	GSVT	131	1853	26	145
	SB	9	0	3108	4
	SR	67	61	62	1264

Tabela V: Matriz de confusão do conjunto de teste do *SVM*.

		Predição			
		AFIB	GSVT	SB	SR
Verdade	AFIB	359	47	10	23
	GSVT	65	429	11	46
	SB	6	3	759	0
	SR	25	35	11	301

O modelo *SVM* apresentou uma performance considerável na predição das quatro classes do conjunto de dados, tanto no conjunto de treinamento quanto no conjunto de teste. Considerando o conjunto de teste, onde o modelo deve apresentar a melhor performance possível, o número de amostras de cada classe foi relativamente baixo, com o modelo *SVM* conseguindo desempenho acima de 78% para todas as classes.

B. Gradient Boost

Como mencionado anteriormente sobre a dificuldade de estimação dos hiperparâmetros, alguns foram configurados manualmente, sendo eles: $n_estimator=500$, $criterion='entropy'$ e $learning_rate=0,1$ para o restantes dos modelos comparados.

As matrizes de confusão do conjunto de treinamento e do conjunto de teste do *Gradient Boost* são mostradas na Tabela VI e VII, respectivamente. Nesse modelo, o desempenho ficou acima de 82% para o conjunto de teste. A classe SB apresentou a melhor predição com 99%, errando apenas 5 amostras e a classe GSVT apresentou a pior predição de 82%, errando 95 amostras.

C. AdaBoost

Esse modelo não apresentou uma boa performance, conforme as matrizes de confusões da Tabela VIII e IX mostram para o conjunto de teste. A melhor performance foi para a

Tabela VI: Matriz de confusão do conjunto de treinamento do *Gradient Boost*.

		Predição			
		AFIB	GSVT	SB	SR
Verdade	AFIB	1785	0	0	1
	GSVT	2	2149	0	4
	SB	0	0	3121	0
	SR	0	0	0	1454

Tabela VII: Matriz de confusão do conjunto de teste do *Gradient Boost*.

		Predição			
		AFIB	GSVT	SB	SR
Verdade	AFIB	374	39	8	18
	GSVT	41	456	5	49
	SB	3	1	763	1
	SR	21	14	0	337

classe SB com 99%. A classe GSVT apresentou 269 predições incorretas, obtendo performance de apenas de 51%. A classe SR também apresentou baixa performance de 58%, somando 153 predições incorretas, enquanto que a classe AFIB apresentou 76% de performance de acertos.

Tabela VIII: Matriz de confusão do conjunto de treinamento do *AdaBoost*.

		Predição			
		AFIB	GSVT	SB	SR
Verdade	AFIB	1355	179	42	210
	GSVT	894	1075	17	1694
	SB	6	19	3096	0
	SR	225	302	0	927

Tabela IX: Matriz de confusão do conjunto de teste do *Gradient Boost*.

		Predição			
		AFIB	GSVT	SB	SR
Verdade	AFIB	335	44	11	49
	GSVT	233	282	5	31
	SB	4	1	762	1
	SR	73	80	0	219

D. Random Forest

Random Forest apresentou uma ótima performance para o conjunto de treinamento, com predição de 100% para todas as classes. Quanto ao conjunto de teste, a menor classificação foi a classe AFIB, onde o modelo errou 107 amostras, obtendo 75% de acerto. As outras classes apresentaram predição acima de 80%, tendo a classe SB apresentado o melhor em desempenho, chegando a 99% de desempenho. A classe GSVT e SR apresentaram desempenho de 80% e 91%, respectivamente. O detalhamento do desempenho é mostrado nas Tabelas X e XI.

E. XGradient Boost

Por último, as Tabelas XII e XIII mostram o desempenho do modelo XGradient Boost. A classe SB apresentou novamente

Tabela X: Matriz de confusão do conjunto de treinamento do *Random Forest*.

Verdade	Predição			
	AFIB	GSVT	SB	SR
AFIB	1786	0	0	0
GSVT	0	2155	0	0
SB	0	0	3121	0
SR	0	0	0	1454

Tabela XI: Matriz de confusão do conjunto de teste do *Random Forest*.

Verdade	Predição			
	AFIB	GSVT	SB	SR
AFIB	332	58	16	33
GSVT	40	441	6	64
SB	3	0	764	1
SR	24	6	0	342

a melhor performance de 99%. A classe SR apresentou performance de 90%, enquanto que 83% foi para a classe GSVT e 82% para a classe AFIB.

Tabela XII: Matriz de confusão do conjunto de treinamento do *XGradient Boost*.

Verdade	Predição			
	AFIB	GSVT	SB	SR
AFIB	1781	3	0	2
GSVT	4	2105	0	46
SB	0	0	3121	0
SR	0	0	0	1454

Tabela XIII: Matriz de confusão do conjunto de teste do *XGradient Boost*.

Verdade	Predição			
	AFIB	GSVT	SB	SR
AFIB	360	44	8	27
GSVT	37	459	6	49
SB	3	1	763	1
SR	23	12	0	337

Pelos resultados apresentados anteriormente, percebe-se que a classe SB foi a que apresentou o melhor desempenho com média de acertos de 99% para o conjunto de teste em todos os modelos comparados. A classe AFIB teve o melhor desempenho nos modelos *SVM* e *GradientBoost* a uma taxa de 85%. O *XGradientBoost* apresentou o melhor desempenho de 83% para a classe GSVT enquanto que para a classe SR, o modelo *Random Forest* apresentou 91% de desempenho.

Como cada classe teve um determinado desempenho em diferentes modelos, o melhor modelo é considerado aquele que apresentou a melhor média em relação à todas as classes. Para isso, a curva ROC média, levando em consideração todas as classes, determinou que o modelo *XGradientBoost* foi o melhor modelo com desempenho médio de 99% para o conjunto de teste, conforme mostra a Figura 6. O pior desempenho ficou com *AdaBoost* em 90%.

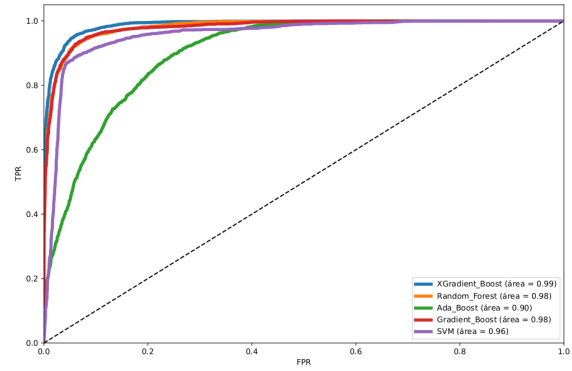


Fig. 6: Curva ROC para os modelos de classificação.

Em relação ao desempenho dos modelos comparados de acordo com as métricas de sensibilidade (Se), predição positiva ($+P$) e F1-score (F_s), as Tabelas XVI e XVII resume os resultados obtidos. Os desempenhos ficaram altos em relação ao conjunto de treinamento, como era previsto, com bastante valores próximos de 100%. Na validação dos modelos pelo conjunto de teste, os melhores resultados são apresentados em negrito na Tabela XVII de acordo com cada métrica em cada classe. A melhor acurácia foi para o modelo *GradientBoost* com valor de 91%.

Tabela XIV: Matriz de confusão do conjunto de treinamento do *XGradient Boost* no conjunto original sem adição de características pela DWT.

Verdade	Predição			
	AFIB	GSVT	SB	SR
AFIB	1632	83	2	69
GSVT	60	1932	4	159
SB	2	0	3119	0
SR	16	10	1	1427

Tabela XV: Matriz de confusão do conjunto de teste do *XGradient Boost* no conjunto original sem adição de características pela DWT.

Verdade	Predição			
	AFIB	GSVT	SB	SR
AFIB	344	61	4	30
GSVT	38	459	6	48
SB	2	1	763	2
SR	23	10	0	332

Percebe-se que o modelo *GradientBoost* apresentou os melhores resultados para quase todas as classes, exceto para a Se da classe SR. Todos os modelos apresentaram 99% de Se para a classe SB, corroborando com os resultados apresentados pelas matrizes de confusão dos modelos. O *AdaBoost* não apresentou um bom desempenho, apresentando bons resultados apenas para a classe SB. O *XGradientBoost*

apresentou melhor desempenho para $+P$ para a classe AFIB de 85% e para todas as métricas para as classes GSVT e SB. Por último, *SVM* e *Random Forest* apresentaram resultados melhores apenas para Se de 99% para a classe SB e Se de 92% para a classe SR, respectivamente.

Por último, foi realizada uma verificação do comportamento do melhor modelo em desempenho, o *XGradientBoost*, com o conjunto de dados original sem adição de características geradas pela DWT. As Tabelas XIV e XV resume os resultados dessa análise. Os resultados mostram, que a predição do conjunto de treinamento sofreu uma diminuição nos acertos das classes quando comparado aos resultados da Tabela XII. As classes AFIB e GSVT sofreram uma diminuição de previsão de acerto em 8,36% e 8,21%, respectivamente. Quanto ao conjunto de teste, a diminuição ocorreu apenas na classe AFIB de 4,4%.

Verificando as métricas de cada classe das Tabelas XVI e XVII, o conjunto original performado com o *XGradientBoost* apresentou queda de desempenho em algumas métricas em relação ao conjunto de treinamento, principalmente na Se das classes AFIB e GSVT de 9% e 8%, respectivamente e $+P$ e F_s para a classe SR de 11,34% e 6,12%, respectivamente. Quanto ao conjunto de teste, não ocorreu grandes variações, sendo as maiores quedas em relação a Se e F_s da classe AFIB de 4,87% e 3,57%, respectivamente. Portanto, percebe-se que as técnicas *ensembles* de aprendizado de máquina pode melhorar significativamente o desempenho dos modelos de classificação de arritmias em ECG.

VII. CONCLUSÕES

Neste trabalho, foi proposto um modelo de classificação automática de arritmia baseado em transformação discreta Wavelet (*Discrete Wavelet Transform - DWT*) como técnica de extração de características dos sinais de ECG. Inicialmente, os sinais sem ruídos disponibilizados pelo conjunto de dados foram normalizados e decompostos em sub-bandas de frequências de acordo com a função geradora *wavelet* db4, gerando assim 120 características. As mesmas foram combinadas com as 18 características disponíveis no conjunto de dados, totalizando 138 características por sinal. Por fim, cinco classificadores (*Support Vector Machine (SVM)*, *XGradientBoost*, *AdaBoost*, *Random Forest* e *GradientBoost*) foram utilizados e comparados de acordo com suas performances na classificação de cinco classes de arritmias. Como simulação de aplicabilidade em cenários reais, os modelos foram treinados de acordo com o paradigma *inter-patient*. Adicionalmente, a performance de classificação dos modelos foram analisadas separadamente por conjunto de treinamento e teste.

Os resultados mostraram performance de classificação bastante satisfatória, principalmente para os modelos que utilizam algoritmos de *boosting* para melhoria de performance de classificação, sendo o *GradientBoost* e *XGradientBoost* apresentando os melhores resultados entre os modelos. Isso torna claro que as técnicas que envolvem a combinação de algoritmos de aprendizado de máquina podem ajudar na mel-

horia do desempenho de projetos de classificação de arritmias de cenários reais.

AGRADECIMENTOS

Agradecimentos ao incentivo e suporte da Universidade Federal de Ouro Preto e FAPEMIG.

REFERENCES

- [1] W. H. Organization, *Cardiovascular diseases (CVDs)*, WHO, 2021. [Online]. Available: <https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- [2] A. Cohen, *Biomedical Signal Processing*. CRC Press., 1988.
- [3] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, p. 65, 2019.
- [4] ANSI/AAMI, "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," American National Standards Institute, Inc. (ANSI), Association for the Advancement of Medical Instrumentation (AAMI), 2008, ANSI/AAMI/ISO EC57, 1998-(R)2008.
- [5] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer methods and programs in biomedicine*, vol. 127, pp. 144–164, 2016.
- [6] J. Zheng, C. Rakovski, S. Danioko, J. Zhang, H. Yao, and G. Hangyuan, *A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients*, 2019. [Online]. Available: <https://doi.org/10.6084/m9.figshare.c.4560497.v2>
- [7] M. K. Das and S. Ari, "Ecg beats classification using mixture of features," *International scholarly research notices*, vol. 2014, 2014.
- [8] G. Garcia, G. Moreira, D. Menotti, and E. Luz, "Inter-patient ecg heartbeat classification with temporal vcg optimized by pso," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.
- [9] W.-H. Jung and S.-G. Lee, "Ecg identification based on non-fiducial feature extraction using window removal method," *Applied Sciences*, vol. 7, no. 11, p. 1205, 2017.
- [10] L. Wang, W. Sun, Y. Chen, P. Li, and L. Zhao, "Wavelet transform based ecg denoising using adaptive thresholding," in *Proceedings of the 2018 7th International Conference on Bioinformatics and Biomedical Science*, 2018, pp. 35–40.
- [11] Y. Wu, R. M. Rangayyan, Y. Zhou, and S.-C. Ng, "Filtering electrocardiographic signals using an unbiased and normalized adaptive noise reduction system," *Medical engineering & physics*, vol. 31, no. 1, pp. 17–26, 2009.
- [12] S. Pongponnori and X.-H. Yu, "An adaptive filtering approach for electrocardiogram (ecg) signal noise reduction using neural networks," *Neurocomputing*, vol. 117, pp. 206–213, 2013.
- [13] M. Rakshit and S. Das, "An efficient ecg denoising methodology using empirical mode decomposition and adaptive switching mean filter," *Biomedical signal processing and control*, vol. 40, pp. 140–148, 2018.
- [14] M. Sarfraz, A. A. Khan, and F. F. Li, "Using independent component analysis to obtain feature space for reliable ecg arrhythmia classification," in *2014 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2014, pp. 62–67.
- [15] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, "Automatic ecg classification using continuous wavelet transform and convolutional neural network," *Entropy*, vol. 23, no. 1, p. 119, 2021.
- [16] F. M. Dias, H. L. Monteiro, T. W. Cabral, R. Naji, M. Kuehni, and E. J. d. S. Luz, "Arrhythmia classification from single-lead ecg signals using the inter-patient paradigm," *Computer Methods and Programs in Biomedicine*, vol. 202, p. 105948, 2021.
- [17] M. Blanco-Velasco, B. Weng, and K. E. Barner, "Ecg signal denoising and baseline wander correction based on the empirical mode decomposition," *Computers in biology and medicine*, vol. 38, no. 1, pp. 1–13, 2008.
- [18] Z. Wang, F. Wan, C. M. Wong, and L. Zhang, "Adaptive fourier decomposition based ecg denoising," *Computers in Biology and Medicine*, vol. 77, pp. 195–205, 2016.
- [19] G. D. Clifford, F. Azuaje, P. McSharry *et al.*, *Advanced methods and tools for ECG data analysis*. Artech house Boston, 2006.

Tabela XVI: Performance dos modelos em relação ao conjunto de treinamento do conjunto de dados.

Modelos	Acc	AFIB			GSVT			SB			SR		
		+P	Se	F _s	+P	Se	F _s	+P	Se	F _s	+P	Se	F _s
GradientBoost	100	100	100	100	100	100	100	100	100	100	100	100	100
SVM	91	88	88	88	91	86	89	96	100	98	86	87	86
XGradientBoost	99	100	100	100	100	98	99	100	100	100	97	100	98
AdaBoost	76	55	76	64	68	50	58	98	99	99	71	64	67
Random Forest	100	100	100	100	100	100	100	100	100	100	100	100	100
XGradientBoost*	95	95	91	93	95	90	92	100	100	100	86	98	92

* XGradientBoost aplicado no conjunto original sem adição de características geradas pela DWT.

Tabela XVII: Performance dos modelos em relação ao conjunto de teste do conjunto de dados.

Modelos	Acc	AFIB			GSVT			SB			SR		
		+P	Se	F _s	+P	Se	F _s	+P	Se	F _s	+P	Se	F _s
GradientBoost	91	85	85	85	89	83	86	98	99	99	83	91	87
SVM	87	79	82	80	83	78	81	96	99	97	81	81	81
XGradientBoost	90	85	82	84	89	83	86	98	99	99	81	91	86
AdaBoost	75	52	76	62	69	51	59	98	99	99	73	59	65
Random Forest	88	83	76	79	87	80	84	97	99	98	78	92	84
XGradientBoost*	89	85	78	81	86	83	85	99	99	99	81	91	86

* XGradientBoost aplicado no conjunto original sem adição de características geradas pela DWT.

- [20] C. Ye, B. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2930–2941, 2012.
- [21] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.
- [22] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific data*, vol. 7, no. 1, pp. 1–8, 2020.
- [23] K. Sternickel, "Automatic pattern recognition in ecg time series," *Computer methods and programs in biomedicine*, vol. 68, no. 2, pp. 109–115, 2002.
- [24] S. Mousavi and F. Afghah, "Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1308–1312.
- [25] P. Yang, D. Wang, W.-B. Zhao, L.-H. Fu, J.-L. Du, and H. Su, "Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification," *Biomedical Signal Processing and Control*, vol. 63, p. 102138, 2021.
- [26] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ecgs via ensemble of classifiers," *Biomedical Signal Processing and Control*, vol. 47, pp. 41–48, 2019.
- [27] K. N. Rajesh and R. Dhuli, "Classification of imbalanced ecg beats using re-sampling techniques and adaboost ensemble classifier," *Biomedical Signal Processing and Control*, vol. 41, pp. 242–254, 2018.
- [28] H. Shi, H. Wang, Y. Huang, L. Zhao, C. Qin, and C. Liu, "A hierarchical method based on weighted extreme gradient boosting in ecg heartbeat classification," *Computer methods and programs in biomedicine*, vol. 171, pp. 1–10, 2019.
- [29] S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels, "Wavelet transform-based qrs complex detector," *IEEE Transactions on biomedical Engineering*, vol. 46, no. 7, pp. 838–848, 1999.
- [30] Q. Zhao and L. Zhang, "Ecg feature extraction and classification using wavelet transform and support vector machines," in *2005 International Conference on Neural Networks and Brain*, vol. 2. IEEE, 2005, pp. 1089–1092.
- [31] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.
- [32] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999.
- [33] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [34] B. Schölkopf, A. J. Smola, F. Bach *et al.*, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.