



Classificação de Episódios de Fibrilação Atrial por Análise do ECG com Redes Neurais Artificiais MLP e LSTM

PEDRO HENRIQUE BORGHI DE MELO

Dissertação para obtenção do grau de Mestre em:

Engenharia Industrial – Ramo Engenharia Eletrotécnica

Este trabalho foi efetuado sob orientação de:

Prof. Dr. João Paulo Ramos Teixeira

Profa. Dra. Renata Coelho Borges

Bragança

Novembro de 2020

Classificação de Episódios de Fibrilação Atrial por Análise do ECG com Redes Neurais Artificiais MLP e LSTM

PEDRO HENRIQUE BORGHI DE MELO

Dissertação para obtenção do grau de Mestre em:

Engenharia Industrial – Ramo Engenharia Eletrotécnica

Este trabalho foi efetuado sob orientação de:

Prof. Dr. João Paulo Ramos Teixeira

Profa. Dra. Renata Coelho Borges

Bragança

Novembro de 2020

Este trabalho é dedicado a *Yahweh*, à minha família e aos meus amigos e amigas.

Agradecimentos

Agradeço primeiramente a Deus por me dar a vida, me sustentar e me capacitar física e emocionalmente. Por nunca desistir de mim. Por me manter a vida e por direcioná-la. Agradeço por suas misericórdias e por seu amor incondicional. Agradeço com toda a minha vida. Toda honra e glória à Ele.

Agradeço a minha família por me apoiar, por acreditar, investir e por cuidar de mim com tanto zelo e carinho. Agradeço por estarem dispostos o tempo todo para me acolher e por sempre me desejarem por perto. A eles expresse meu compromisso de sempre viver por eles.

Agradeço aos meus amigos e amigas que em todos os momentos estiveram dispostos a me ajudar, que acreditaram em mim compartilhando suas vidas comigo e me permitindo fazer parte da deles. A eles expresse que daria minha própria vida por eles.

Agradeço aos meus orientadores Professor Doutor João Paulo Ramos Teixeira e Professora Doutora Renata Coelho Borges, pelo apoio, investimento, credibilidade e motivação durante e além das atividades deste trabalho. Mais do que orientadores, os considero como amigos.

Agradeço as instituições que me formaram, nomeadamente a Universidade Tecnológica Federal do Paraná – campus Cornélio Procopio e o Instituto Politécnico de Bragança, me proporcionando condições de infraestrutura, aulas, alimentação, bolsas e conhecimento. Agradeço a todos os professores, servidores e funcionários.

Agradeço ainda, a todas as pessoas que de alguma forma contribuíram para a realização deste mestrado, independente da intensidade, duração ou visibilidade. Todos vocês fazem parte desta conquista.

Pedro Henrique Borghi

“[...] Porquanto a vida é mais preciosa do que o alimento, e o corpo mais importante do que as roupas. [...] Considerando que vós não podeis fazer nada em relação às pequenas coisas da vida, por que vos preocupais com todas as outras? [...] Buscai, pois, em primeiro lugar, o Reino de Deus, e todas as demais coisas vos serão providenciadas. [...] Por isso, onde estiverem os vossos bens mais preciosos, certamente aí também estará o vosso coração.”

Jesus de Nazaré

Resumo

A fibrilação atrial (AF) é uma doença cardíaca que afeta aproximadamente 1% da população mundial, sendo a anomalia cardíaca mais comum. Apesar de não ser uma causa direta de morte, frequentemente está associada ou gera outros problemas que ameaçam a vida humana, como o derrame e a doença da artéria coronária. As principais características da AF são: a alta variação do ritmo cardíaco, o enfraquecimento ou desaparecimento da contração atrial e a ocorrência de irregularidades nas atividades dos ventrículos. O diagnóstico da AF é realizado por um médico especialista, principalmente através da inspeção visual de gravações de eletrocardiograma (ECG) de longo termo. Tais gravações podem chegar a várias horas, e são necessárias pois a AF pode ocorrer a qualquer momento do dia. Dessa forma surgem os problemas quanto ao grande volume de dados e as dependências de longo termo. Além disso, as particularidades e as variabilidades dos padrões de deformação de cada sujeito fazem com que o problema esteja também relacionado com a experiência do cardiologista. Assim, a proposta de um sistema computacional de auxílio ao diagnóstico médico baseado em inteligência artificial se torna muito interessante, uma vez que não sofre com a fadiga e é fortemente indicado para lidar com dados em grande quantidade e com alta variabilidade. Portanto, neste trabalho foi proposta a exploração de modelos de aprendizagem de máquina para análise e classificação de sinais ECG de longo termo, para auxiliar no diagnóstico da AF. Os modelos foram baseados em redes neuronais artificiais do tipo Multi-Layer Perceptron (MLP) e Long Short-Term Memory (LSTM). Utilizam-se os sinais da base de dados MIT-BIH Atrial Fibrillation, sem remoção de ruído, tendências ou artefatos, numa etapa de extração de características temporais, morfológicas, estatísticas e em tempo-frequência sobre segmentos de contexto variável (duração em segundos ou contagem de intervalos entre picos R). As características do sinal ECG utilizadas, foram: duração dos intervalos R-R (RRi) consecutivos, perturbação Jitter, perturbação Shimmer, entropias de Shannon e energia logarítmica, frequências instantâneas, entropia espectral e transformada Scattering. Sobre estes atributos foram aplicadas diferentes estratégias de normalização por Z-score e valor máximo absoluto, de forma a normalizar os indicadores de acordo com o contexto do sujeito ou local do segmento. Após a exploração de várias combinações destas características e dos parâmetros das redes MLP, obteve-se uma acurácia de classificação para a metodologia 10-fold cross-validation de 80,67%. Entretanto, notou-se que as marcações do pico das ondas R advindas da base de dados

eram imprecisas. Dessa forma, desenvolveu-se um algoritmo de detecção do pico das ondas R baseado na combinação entre a derivada do sinal, a energia de Shannon e a transformada de Hilbert, resultado em uma acurácia de marcação dos picos R de 98,95%. A partir das novas marcações, determinou-se todas as características e em seguida foram exploradas diversas estruturas de redes neuronais MLP e LSTM, sendo que os melhores resultados em acurácia/exatidão para estas arquiteturas foram, respectivamente, 91,96% e 98,17%. Em todos os testes, a MLP demonstrou melhora de desempenho à medida que mais características foram sendo agregadas nos conjuntos de dados. A LSTM por outro lado, obteve os melhores resultados quando foram combinados 60 RRI e as respectivas entropias das ondas P, T e U.

Palavras Chave: Fibrilação Atrial, ECG, MLP, LSTM, Jitter, Shimmer, Transformada Scattering, Entropia.

Abstract

Atrial fibrillation (AF) is a heart disease that affects approximately 1% of the world population, being the most common cardiac anomaly. Although it is not a direct cause of death, it is often associated with or generates other problems that threaten human life, such as stroke and coronary artery disease. The main characteristics of AF are the high variation in heart rate, the weakening or disappearance of atrial contraction and the occurrence of irregularities in the activities of the ventricles. The diagnosis of AF is performed by a specialist doctor, mainly through visual inspection of long-term electrocardiogram (ECG) recordings. Such recordings can take several hours and are necessary because AF can occur at any time of the day. Thus, problems arise regarding the large amount of data and long-term dependencies. In addition, the particularities and variability of the deformation patterns of each subject make the problem also related to the cardiologist's experience. Thus, the proposal for a computational system to aid medical diagnosis based on artificial intelligence becomes very interesting, since it does not suffer from fatigue and is strongly indicated to deal with data in large quantities and with high variability. Therefore, in this work it was proposed to explore machine learning models for the analysis and classification of long-term ECG signals, to assist in the diagnosis of AF. The models were based on artificial neural networks Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM). The signals from the MIT-BIH Atrial Fibrillation database are used, without removing noise, trends or artifacts, in a stage of extracting temporal, morphological, statistical and time-frequency features over segments of variable context (duration in seconds or counting intervals between peaks R). The features of the ECG signal used were: duration of consecutive R-R (RRi) intervals, Jitter disturbance, Shimmer disturbance, Shannon entropies and logarithmic energy, instantaneous frequencies, spectral entropy and Scattering transform. On these attributes, different normalization strategies were applied by Z-score and absolute maximum value, to normalize the indicators according to the context of the subject or location of the segment. After exploring various combinations of these features and the parameters of the MLP networks, the accuracy of classification for the 10-fold cross-validation methodology was 80.67%. However, it was noted that the annotations of the peak of R waves from the database were inaccurate. Thus, an algorithm for detecting the peak of R waves was developed based on the combination of the derivative of the signal, the Shannon energy, and the Hilbert transform, resulting in an accuracy of marking the R peaks of 98.95%.

From the new markings, all features were determined and then several structures of neural networks MLP and LSTM were explored, and the best results in accuracy for these architectures were, respectively, 91.96% and 98.17%. In all tests, MLP showed improvement in performance as more features were added to the data sets. LSTM, on the other hand, obtained the best result when 60 RRi and the respective entropies of the P, T and U waves were combined.

Keywords: Atrial Fibrillation, ECG, MLP, LSTM, Jitter, Shimmer, Scattering Transform, Entropy.

Índice de Figuras

Figura 1 – Comparação do comportamento cardíaco	4
Figura 2 - Sistema de derivações proposto por Wilson	6
Figura 3 - Composição básica do eletrocardiograma	7
Figura 4 - Comparação entre normalização Z-score (2ª linha/vetor) e pelo máximo absoluto (3ª linha/vetor)	25
Figura 5 - Exemplo da decomposição pela transformada Scattering	31
Figura 6 - Modelo fundamental do Perceptron.....	33
Figura 7 - Funções de ativação não lineares.....	34
Figura 8 - Estrutura básica RNN	35
Figura 9 - Unidade neuronal LSTM e organização básica de uma camada LSTM.....	36
Figura 10 - Composição genérica de um modelo LSTM bidirecional	37
Figura 11 – Representação normalizada dos sinais relativos à etapa de remarcação do pico das ondas R para o sujeito 14	61

Índice de Tabelas

Tabela 1 - Organização da metodologia dos modelos primário e secundário	40
Tabela 2 – Localização percentual dos pontos de início e final das ondas T, U e P dentro de um intervalo R-R	49
Tabela 3 - Distribuição do número de segmentos gerados para cada sujeito num contexto de 30 R Ri.....	56
Tabela 4 - Balanço de classes para diferentes valores do fator μ	57
Tabela 5 - Parâmetros estruturais e acurácia para os dois melhores modelos na metodologia A	57
Tabela 6 - Resultados de acurácia [%] sobre o conjunto de teste para diferentes combinações de características, pela metodologia B	58
Tabela 7 - Acurácia do algoritmo proposto para detecção do pico das ondas R.....	60
Tabela 8 – Quantidade de períodos de silêncio, por sujeito, identificados pela metodologia proposta no modelo secundário	62
Tabela 9 - Distribuição do número de segmentos gerados para cada sujeito num contexto de 60 R Ri.....	63
Tabela 10 - Proporção da distribuição de classes implementada no modelo secundário, para o contexto 60 R Ri e com $\mu = 10\%$	63
Tabela 11 - Proporção da distribuição de classes implementada no modelo secundário, para o contexto 60 segundos e com $\mu = 10\%$	64
Tabela 12 - Desempenho em acurácia no conjunto de teste dos melhores modelos MLP entre os explorados no modelo secundário. Nota para a apresentação, em vermelho, do modelo de pior desempenho.....	66
Tabela 13 - Desempenho em acurácia no conjunto de teste das variações de parâmetros estruturais da LSTM e diferentes combinações de características para os conjuntos de entrada	68
Tabela 14 – Comparação entre alguns sistemas desenvolvidos em outros trabalhos e os melhores modelos MLP e LSTM deste estudo, a nível de acurácia para detecção de pelo menos AF	70

Acrónimos

AF	Fibrilação Atrial
CNN	Rede Neuronal Convolutacional
CWT	Transformada Wavelet Contínua
DWT	Transformada Wavelet Discreta
ECG	Eletrocardiograma
Hz	<i>Hertz</i>
IA	Inteligência Artificial
KNN	<i>K-nearest neighbor</i>
Log-Sig	Função Logística Sigmoides
LSTM	<i>Long Short-Term Memory</i>
MLP	Perceptron Multicamadas
RNN	Rede Neuronal Recorrente
RRi	Intervalo entre picos R
SVM	<i>Support Vector Machine</i>
Tan-Sig	Função Tangente Sigmoides

Índice

Capítulo 1: Introdução	1
1.1 Fibrilação Atrial.....	3
1.2 Eletrocardiograma	5
1.3 Estado da Arte.....	7
1.4 Justificativa.....	16
1.5 Objetivos	17
1.5.1 Objetivos Primários	17
1.5.2 Objetivos Específicos	17
Capítulo 2: Materiais e Métodos.....	19
2.1 Base de Dados.....	19
2.2 Recursos Computacionais	19
2.3 Transformada De Hilbert.....	20
2.4 Entropias.....	21
2.4.1 Entropia de Shannon	21
2.4.2 Entropia Logaritmo da Energia	22
2.4.3 Energia de Shannon	22
2.5 Jitter	22
2.6 Shimmer.....	23
2.7 Normalização Z-Score	24
2.8 Normalização Pelo Valor Máximo Absoluto	25
2.9 Média Móvel.....	25
2.10 Frequências Instantâneas.....	26
2.11 Entropia Espectral.....	27
2.12 Transformada Wavelet.....	28
2.12.1 Wavelet Scattering	29

2.13	Algoritmos de Inteligência Artificial.....	31
2.13.1	Feed Forward Network.....	32
2.13.2	Long Short-Term Memory	34
Capítulo 3:	METODOLOGIA	39
3.1	Aquisição e Organização dos Sinais	40
3.2	Modelo Primário	41
3.2.1	Segmentação	41
3.2.2	Extração de características.....	42
3.2.2.1	Normalização das características.....	43
3.2.3	Conjuntos de dados para a aprendizagem de máquina.....	43
3.2.3.1	Metodologia A.....	44
3.2.3.2	Metodologia B.....	44
3.3	Modelo Secundário	45
3.3.1	Algoritmo de detecção de pico das ondas R.....	45
3.3.2	Segmentação	47
3.3.3	Extração de características.....	48
3.3.3.1	Normalização de características	50
3.3.4	Conjuntos de dados e experimentação - redes MLP.....	51
3.3.5	Conjuntos de dados e experimentação - redes LSTM	52
Capítulo 4:	Resultados e Discussão	55
4.1	Modelo Primário	55
4.2	Modelo Secundário	59
4.3	Comparação com o Estado da Arte.....	69
Capítulo 5:	Conclusões	72
5.1	Trabalhos Futuros	74
	REFERÊNCIAS.....	76

Capítulo 1: Introdução

Neste capítulo são iniciadas as principais ideias relativas ao desenvolvimento deste trabalho. Primeiro é apresentada uma discussão sobre sinais biomédicos quanto sua natureza e aplicação, bem como a utilização destes em ambientes computacionais. Alinhado a isto, são apresentados a fibrilação atrial, a anormalidade cardíaca mais comum, e o eletrocardiograma, uma ferramenta de monitoração do comportamento cardíaco. Ao mesmo tempo, são apresentadas as motivações para utilização de técnicas de inteligência artificial no diagnóstico da fibrilação atrial e como esta pode ser utilizada neste trabalho através do estudo do estado da arte. Baseando-se nesta introdução, são apresentados as justificativas e os objetivos para o desenvolvimento da proposta.

Sinais biomédicos são uma fonte importante de informações acerca da qualidade de vida dos seres vivos. Especialmente nos seres humanos, são utilizados para elaboração de diagnósticos, aplicações de reabilitação e tratamento dos mais diversos. Sua aquisição é feita por meio de sensores específicos para cada aplicação, sendo que estes podem ir desde câmeras de reconhecimento facial e sensores de temperatura até eletrodos invasivos e sensores embarcados em próteses. Para a sua utilização, é necessário um sistema de condicionamento que executa funções de ajustes e adequações dos sinais, como atenuações de artefatos, ajuste de escala e transformação do domínio analógico ao digital. Em algumas aplicações podem ainda conter uma etapa de armazenamento. Grande parte destes sistemas retorna as observações na forma de sinais elétricos digitalizados que podem ser interpretados por circuitos lógicos locais ou enviados a um computador por protocolos de comunicação (Andersen *et al.*, 2019; Kaplan Berkaya *et al.*, 2018; Miotto *et al.*, 2017; Webster, 2009).

Os tipos de sensores do sistema condicionador e o processamento relacionado constituem a estrutura básica para visualização e a extração de informações acerca da atividade do objeto de interesse, seja interna ou externamente. A partir da visualização, ou de um conjunto de visualizações, desta e outras atividades ao longo do tempo, no caso de diagnósticos médicos, é possível ao especialista inferir sobre a condição do objeto ou do organismo como um todo, dependendo do nível de abrangência do sistema. Dessa forma, os sinais biomédicos contribuem como uma fonte de informação fundamental para um diagnóstico mais confiável. Ainda, em alguns casos é imprescindível que se faça uso destes, como na detecção de tumores por imagens de ressonância magnética e na determinação de problemas cardiovasculares. Neste último, a

representação gráfica da atividade elétrica do coração proporciona recursos fundamentais para avaliação morfológica e operacional do sistema cardíaco, como a determinação do ritmo cardíaco, a forma e duração dos eventos do ciclo cardíaco e a aparição de padrões anormais (Kaplan Berkaya *et al.*, 2018; Webster, 2009).

Entre as anomalias do comportamento cardíaco mais comuns, a fibrilação atrial é a mais frequente. Estima-se que a AF acometa pelo menos 1% da população mundial, podendo esta proporção ser maior devido ao caráter muitas vezes assintomático das ocorrências. A sua identificação se faz basicamente pela observação de variações no ritmo cardíaco e de irregularidades nas atividades do ciclo cardíaco. Estes padrões precisam ser reconhecidos pelo cardiologista que, para isso, deve analisar gravações extensas da atividade elétrica do coração em busca destes padrões. Muitas vezes, esta análise não se resume a um único sinal, já que vários “pontos de vista” do coração são produzidos de acordo com as referências e posicionamentos dos sensores no paciente. Além disso, para uma maior contextualização e a identificação de anormalidades associadas, essas outras observações se tornam fundamentais (Andersen *et al.*, 2019; Dang *et al.*, 2019; Fuster *et al.*, 2006; Kaplan Berkaya *et al.*, 2018; Sharma *et al.*, 2019c; Yildirim *et al.*, 2019).

Devido à grande quantidade de dados e a variabilidade que carregam, sistemas automáticos de detecção de anomalias e classificação de eventos têm sido propostos com o objetivo de facilitar a inspeção dos sinais e aumentar a confiabilidade do diagnóstico médico (Andersen *et al.*, 2019; Dang *et al.*, 2019; Faust *et al.*, 2018b; Kaplan Berkaya *et al.*, 2018; Manikandan and Soman, 2012; Sharma *et al.*, 2019c; Yildirim, 2018; Yildirim *et al.*, 2019). Com o uso da inteligência artificial (IA) em sistemas computacionais, por exemplo, evita-se o efeito da fadiga humana e pode-se reduzir o tempo de análise ainda fazendo uso de vários recursos extraídos de todas as observações simultaneamente. Especificamente ao uso da inteligência artificial, estruturas de aprendizagem de máquina baseadas em redes neurais artificiais são de ampla aplicação para sinais biomédicos, uma vez que são fortemente indicadas para tarefas que envolvam reconhecimentos de padrões, um elevado volume de dados, uma alta variabilidade e complexidade dos sinais e nas quais há um processo de decisão que envolve múltiplas variáveis.

Entretanto, estes modelos de inteligência necessitam de exemplos suficientemente representativos para a sua aprendizagem. Além disso, deve-se levar em conta a disponibilidade dos recursos computacionais para executar a IA, o tempo de processamento das amostras, no caso de aplicações em tempo real, e as limitações tanto de diversidade de exemplos quanto da

qualidade dos sinais. É tarefa do projetista definir, extrair e organizar os recursos que irão compor a estrutura e o treinamento, de forma que o modelo seja capaz de generalizar a informação contido nos exemplos satisfatoriamente para quaisquer outras amostras. É razoável afirmar que a experiência deste profissional é altamente relevante neste processo, uma vez que este deve ter conhecimento suficiente da natureza dos sinais e de quais recursos são descritivos no seu comportamento. No caso da fibrilação atrial, destacar as características rítmicas dos sinais e a morfologia da atividade dos átrios, por exemplo, pode ser um fator contribuinte para a IA realizar efetivamente a sua detecção (Andersen *et al.*, 2019; Dang *et al.*, 2019; Kaplan Berkaya *et al.*, 2018; Miotto *et al.*, 2017; Sharma *et al.*, 2019c; Yildirim, 2018).

1.1 Fibrilação Atrial

Fibrilação atrial (AF – *Atrial Fibrillation*) é a mais comum e sustentada perturbação rítmica cardíaca. Apesar de ser frequentemente associada a doenças cardíacas estruturadas, em muitos pacientes não se nota a presença de qualquer destas doenças. Por definição, a fibrilação atrial é uma taquiarritmia supraventricular caracterizada principalmente por uma prolongada e não uniforme variação do ritmo cardíaco causada por uma ativação descoordenada dos átrios e por uma irregularidade na atividade dos ventrículos. A partir da inspeção da atividade elétrica cardíaca, nota-se a atenuação, inexistência ou substituição da contração atrial por um comportamento oscilatório variável em amplitude, forma e duração. Associado a este comportamento, nota-se uma resposta ventricular rápida e irregular, sendo esta dependente da condição eletrofisiológica do nó atrioventricular e de outros tecidos condutores associados (Andersen *et al.*, 2019; Dang *et al.*, 2019; Faust *et al.*, 2018b; Fuster *et al.*, 2006; Markides and Schilling, 2003). A Figura 1 apresenta, a partir do eletrocardiograma, a comparação genérica entre um comportamento cardíaco normal (a) e a fibrilação atrial (b).

Apesar de poder ocorrer isoladamente, é possível que outras anomalias cardíacas sejam observadas simultaneamente, como o *flutter* atrial e a taquicardia atrial, principalmente devido a más-formações nas estruturas de condução do potencial de ação no coração. Ainda, a ação de medicamentos para prevenir a recorrência da fibrilação atrial pode colaborar para o surgimento do *flutter* atrial, que é caracterizado pela presença de ondas do tipo dente de serra (notadas no ECG) na ativação atrial (Fuster *et al.*, 2006; Markides and Schilling, 2003).

Acerca dos sintomas, a fibrilação atrial pode causar cansaço, indisposição e dores no peito, mesmo quando o sujeito está em repouso. Portanto, leva a uma considerável perda na

qualidade de vida e da capacidade de executar tarefas que envolvam algum tipo de esforço físico. Apesar de não ser uma causa direta de ameaça à vida, pode levar ou estar associada a outras complicações, como por exemplo disfunção ventricular, doença da artéria coronária, infartos, entre outras. Com o aumento da idade, principalmente após os 60 anos, verifica-se que as ocorrências de períodos de fibrilação atrial levam a complicações mais severas e frequentes, sendo necessário um acompanhamento também mais frequente da situação. Mesmo que seja fortemente aleatória, pode se sustentar por diversas horas e necessitar de intervenções medicamentosas (Andersen *et al.*, 2019; Faust *et al.*, 2018b; Fuster *et al.*, 2006; Markides and Schilling, 2003).

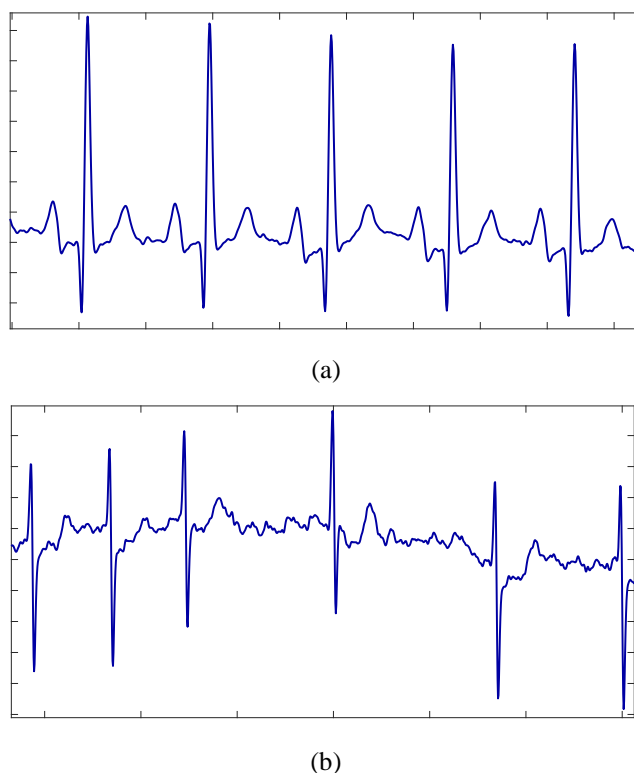


Figura 1 – Comparação do comportamento cardíaco: (a) Normal. (b) Fibrilação Atrial. Fonte: Autoria própria.

Considerando que recorrente é a presença de dois ou mais períodos de AF, quanto a duração e a persistência da AF, esta pode basicamente ser dividida em 3 tipos: paroxismal, quando a AF recorrente é encerrada espontaneamente; persistente, quando a AF recorrente se mantém por mais de uma semana e é necessária a aplicação de tratamento específico para encerrá-la; e permanente, quando a ocorrência de AF não consegue ser interrompida mesmo após administração de medicamentos. Deve-se ressaltar que os tipos não precisam ser exclusivos, uma vez que, por exemplo, um paciente pode apresentar diversos casos de AF

paroxismal e só então um caso persistente. Por outro lado, eventos de AF podem ser recorrentes por mais de sete dias e não necessitam de intervenção de tratamentos para serem interrompidos, sendo assim a AF considerada paroxismal. Com relação a duração, as irregularidades podem ocorrer em períodos curtos de poucos segundos até várias horas, ou então serem incessantes. Pode ainda estar associada a outros problemas de ordem pulmonar ou cardíaca, por exemplo (Fuster *et al.*, 2006; Markides and Schilling, 2003).

O diagnóstico da AF é constituído principalmente a partir do histórico da condição cardíaca do paciente, relato de sintomas e pela observação da atividade elétrica do coração. Por não se conseguir determinar o momento em que ocorrerá, é necessário armazenar o comportamento elétrico do coração por diversas horas e em seguida inspecioná-lo visualmente. O reconhecimento dos padrões da AF é, portanto, diretamente afetado pelo conhecimento prévio e experiência do especialista, seja no projeto do modelo de aprendizagem de máquina ou do cardiologista ao inspecionar as gravações manualmente (Dang *et al.*, 2019; Faust *et al.*, 2018b; Fuster *et al.*, 2006; Markides and Schilling, 2003).

1.2 Eletrocardiograma

O método mais comum de se observar a atividade elétrica do coração é através do eletrocardiograma (ECG). Neste, elementos sensíveis à diferença de potencial elétrico (eletrodos) captam a propagação do potencial de ação nos tecidos condutores de determinadas regiões do coração. A região a ser observada depende da disposição destes elementos e o referencial, sendo que para isso, foram definidos padrões de posicionamentos sobre os quais as referências de medida do potencial elétrico geram as chamadas derivações (Faust *et al.*, 2018b; Kaplan Berkaya *et al.*, 2018; Webster, 2009). A Figura 2 ilustra a padronização do sistema de Wilson.

Quanto aos tipos de eletrodos superficiais, estes variam de acordo com a idade do paciente, o mecanismo de fixação e a reutilização. Quanto a idade, para neonatais, por exemplo, são utilizados eletrodos com diâmetro inferior aos utilizados para adultos. Quanto ao mecanismo de fixação, podem ser do tipo ventosa com anéis metálicos, terminação plana, pinça ou clips. Quando aplicados a cenários onde há movimentação constante, como manobras de reanimação, monitorização prolongada ou provas de esforço, devem ter alto poder de adesão. Ainda, no caso dos eletrodos com terminação plana, podem estar associados a um gel condutor, que facilita a troca de cargas elétricas entre a pele e o eletrodo. A reutilização está basicamente

associada ao descarte dos elementos, onde os do tipo ventosa e clips são os mais reutilizáveis e os do tipo terminação plana, os menos duráveis (Webster, 2009).

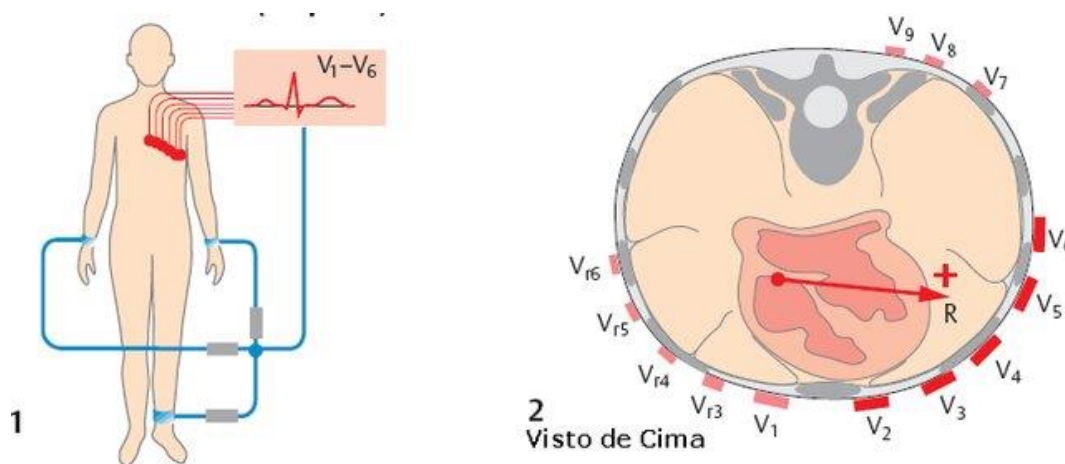


Figura 2 - Sistema de derivações proposto por Wilson.

Por ser um método não invasivo, a recolha do ECG é simples e depende apenas de um sistema de aquisição adequado, o eletrocardiógrafo. A aplicação basicamente é orientada pelo cardiologista ou por profissionais especializados e pode ser relativa a casos de prestação de socorro imediatos, monitorização em ambiente controlado ou em regime prolongado, este último sendo o cenário de estudo deste trabalho. Neste, são armazenadas as gravações de ECG em um sistema portátil que permanece conectado continuamente ao paciente durante algumas horas ou dias. Podem ser encontradas nestes sistemas, etapas de detecção local de anormalidades na forma de pré-processamento e até comunicação sem fio por rádio frequência (Andersen *et al.*, 2019; Faust *et al.*, 2018b; Kaplan Berkaya *et al.*, 2018).

Como resultado da monitoração prolongada, são obtidos grandes volumes de dados sob a presença dos mais diversos artefatos. Além disso, estas gravações podem conter períodos de silêncio, que são períodos onde há desconexão dos eletrodos com o sistema de aquisição ou saturação nos níveis de leitura, ocasionando ausência de informação. Ainda, nota-se a variação da amplitude dos sinais e a introdução de tendências nas medidas devido a características intrínsecas do sistema de aquisição. Para uma atenuação destes e outros efeitos, é interessante a utilização de uma etapa de processamento que procura a redução dos artefatos, a equalização dos níveis de energia do sinal e o realce dos eventos do ciclo cardíaco, sem a perda de informação útil em diferentes escalas (quaisquer transições reais, sejam rápidas ou lentas), que

no caso são as próprias especificidades da atividade elétrica do coração de cada sujeito (Andersen *et al.*, 2019; Faust *et al.*, 2018b; Kaplan Berkaya *et al.*, 2018; Miotto *et al.*, 2017).

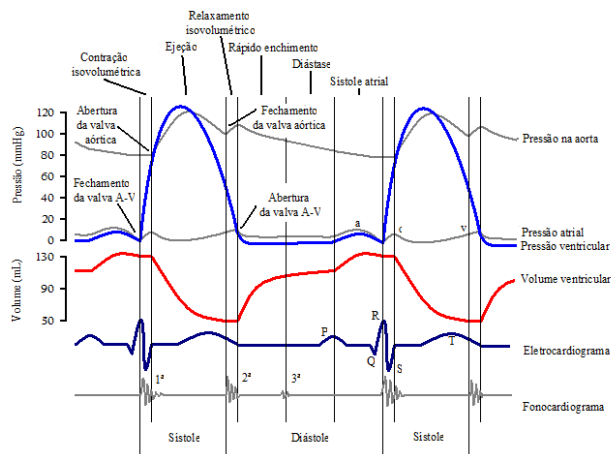


Figura 3 - Composição básica do eletrocardiograma; Fonte: DestinyQx, CC BY-SA 2.5 <<https://creativecommons.org/licenses/by-sa/2.5>>, via Wikimedia Commons.

Um ciclo cardíaco normal (Figura 3) é constituído pela seguinte sequência de eventos, relacionados com as suas respectivas representações no ECG. Primeiramente é dada a ativação elétrica no nodo sinoatrial, que é transmitida ao nodo atrioventricular através dos tratos internodais e interatrial (contração dos átrios). O nodo atrioventricular retarda a propagação do potencial de ação antes deste chegar ao feixe de His e depois à rede de Purkinje. As fibras da rede de Purkinje são responsáveis pela contração dos ventrículos. Em outras palavras, o estímulo elétrico do potencial de ação é iniciado no nodo sinoatrial e transmitido ao nodo atrioventricular (intervalo PR no ECG) onde no caminho ocorre a despolarização (contração) dos átrios (direito e em seguida, esquerdo (surgimento da onda P no ECG)). A repolarização (relaxamento) dos átrios ocorre basicamente ao mesmo tempo que a despolarização dos ventrículos (complexo QRS no ECG). A repolarização ventricular ocorre em seguida e é representada pela onda T no ECG (Andersen *et al.*, 2019; Fuster *et al.*, 2006; Kaplan Berkaya *et al.*, 2018; Webster, 2009).

1.3 Estado da Arte

Nos recentes últimos anos, foram propostos diversos algoritmos baseados em gravações de ECG para classificação de patologias e batimentos cardíacos (Kaplan Berkaya *et al.*, 2018; Lecun *et al.*, 2015; Miotto *et al.*, 2017). A utilização de métodos automáticos para

interpretação de ECG auxilia na questão das variações de ontologias médicas adotadas para realização de diagnósticos (Faust *et al.*, 2018a). Além disso, são capazes de processar horas de gravações em alguns segundos (Saadatnejad *et al.*, 2019; Yildirim *et al.*, 2019). Ainda, podem ser projetados para trabalhar com diversas representações simultaneamente, como é o caso da combinação de várias derivações do ECG e características extraídas, possibilitando a construção de conjuntos de dados com características que são potencialmente relevantes para diferenciação de comportamentos (Borghi *et al.*, 2020; Kaplan Berkaya *et al.*, 2018). Porém, aqui se encontra o grande desafio na utilização destes modelos, a extração e escolha das características que melhor representem os eventos.

Muito se tem proposto nesta área, em abordagens que exploram características temporais, estatísticas, morfológicas e em diferentes domínios, como é o caso do domínio das frequências (Kaplan Berkaya *et al.*, 2018). Entretanto, a escolha e utilização de uma característica ou um conjunto delas, depende principalmente da experiência do projetista do sistema. Este por sua vez, deve considerar parâmetros que se estendem desde a qualidade do sinal até o tipo de modelo de aprendizagem e a capacidade computacional disponível na aplicação (Andersen *et al.*, 2019; Dang *et al.*, 2019; Kaplan Berkaya *et al.*, 2018; Miotto *et al.*, 2017; Sharma *et al.*, 2019c; Yildirim, 2018).

Entre as etapas que constituem a maioria das propostas estão a aquisição e condicionamento dos sinais de ECG, a extração de certas características, a escolha das características mais relevantes, o treinamento do modelo de aprendizado e o pós-processamento para avaliação e apresentação dos resultados ao utilizador (Faust *et al.*, 2018a; Kaplan Berkaya *et al.*, 2018; Miotto *et al.*, 2017). Vale ressaltar que as etapas são ajustadas para garantir robustez e invariabilidade de entrada ao sistema, de forma que se obtenha uma generalização de todos os comportamentos esperados (Kaplan Berkaya *et al.*, 2018; Miotto *et al.*, 2017; Yildirim, 2018; Yildirim *et al.*, 2019).

Primeiro, com relação aos eventos cardiovasculares observados, grande parte dos trabalhos referem-se às arritmias cardíacas, sendo que a base de dados predominante é a MIT-BIH (Kaplan Berkaya *et al.*, 2018; Yildirim *et al.*, 2019). Os batimentos podem ser tipificados em 5 grandes classes, segundo a Associação para Avanços em Instrumentação Médica (AAMI – *Association for the Advancement of Medical Instrumentation*): Normal, Batimento Ectópico Supraventricular, Batimento Ectópico Ventricular, Batimento Fundido e Batimento Desconhecido. Dentro destas, 15 subclasses de batimentos estão contidas, como por exemplo o batimento atrial prematuro e o batimento ventricular prematuro. Eventos rítmicos, como a

fibrilação atrial, a bradicardia sinódica e a taquicardia ventricular, em grande parte se referem a comportamentos de um conjunto de batimentos e também podem ser observados a partir dos sinais contidos na base MIT-BIH (Andersen *et al.*, 2019; Kaplan Berkaya *et al.*, 2018). Os sistemas podem então ser especializados a um determinado comportamento, como nos casos dos estudos sobre fibrilação atrial (Andersen *et al.*, 2019; Borghi *et al.*, 2020; Dang *et al.*, 2019; Faust *et al.*, 2018b), ou abordar vários deles simultaneamente (Saadatnejad *et al.*, 2019; Yildirim, 2018; Yildirim *et al.*, 2019). Evidentemente, quanto maior o número de classes utilizadas, maior é a complexidade do sistema assim como a importância da representatividade para o modelo de aprendizagem.

Na construção dos exemplos para os modelos de aprendizagem supervisionados (nos quais existe a presença de uma referência), a extração de características morfológicas e temporais começa com uma adequada marcação do pico das ondas R dos batimentos (Kaplan Berkaya *et al.*, 2018; Manikandan and Soman, 2012). Quanto a isso, Manikandan e Soman (Manikandan and Soman, 2012) propuseram um método de detecção dos picos R baseado na derivada do sinal. A partir desta, é determinada a energia de Shannon que submetida à transformada de Hilbert, destaca as regiões aproximadas dos picos no sinal original nos pontos de cruzamento de zero ascendente. Por fim, uma busca do máximo real é realizada nos entornos dos pontos obtidos. O método mostra-se robusto nos casos onde o sinal contém variações de amplitude e na presença de anormalidades ou períodos de silêncio (momentos em que não há informação da atividade cardíaca, como por exemplo quando há desconexão dos eletrodos) no ECG.

Acerca do pré-processamento dos dados de entrada, observa-se o uso de métodos diversos para: ajuste da frequência de amostragem entre conjuntos de dados, seleção das derivações de interesse ou mais relevantes, atenuação de ruídos e artefatos, redução dos efeitos de linha de base e *offset*, detecção de eventos do ciclo cardíaco, segmentação da entrada em conjuntos de batimentos e/ou períodos de gravação ou entre eventos, referenciação, balanceamento quantitativo dos conjuntos de dados e normalização (Dang *et al.*, 2019; Kaplan Berkaya *et al.*, 2018; Oh *et al.*, 2019, 2018; Pławiak, 2018; Tan *et al.*, 2018). Como métodos para atenuação de ruídos e artefatos, encontram-se: Análise das Componentes Principais (PCA - *Principal Component Analysis*) (Alickovic and Subasi, 2015), transformação Wavelet (Lin *et al.*, 2014; Sahoo *et al.*, 2017; Singh and Tiwari, 2006) e filtragem convencional (Kaplan Berkaya *et al.*, 2018). Em contrapartida, alguns trabalhos não implementam uma etapa de pré-processamento complexa, utilizando sinais de ECG brutos (*raw ECG*) como entrada para as

etapas seguintes, onde no máximo esses são segmentados e rearranjados em estruturas matriciais (Acharya *et al.*, 2017b; Oh *et al.*, 2018; Yildirim, 2018; Yildirim *et al.*, 2018).

A etapa de extração de características pode, basicamente, ser dividida entre manual ou abstrata, sendo a segunda realizada intrinsecamente por estruturas de aprendizagem profunda (*deep learning*), em camadas neuronais diversas. A extração manual de características depende principalmente da experiência e do conhecimento prévio de especialistas e médicos (Li *et al.*, 2019). São propostas comuns para compor o conjunto de características, como recursos e extratores de recursos: características morfológicas (Lin and Yang, 2014; Sahoo *et al.*, 2017; Yeh *et al.*, 2009; Zhang *et al.*, 2014), características temporais (Andersen *et al.*, 2019; Huang *et al.*, 2014; Lin and Yang, 2014; Sahoo *et al.*, 2017; Yeh *et al.*, 2009; Zhang *et al.*, 2014), hermite basis function (HBF) (Park *et al.*, 2008), *higher order statistics* (HOS) (Elhaj *et al.*, 2016; Kutlu and Kuntalp, 2012; De Lannoy *et al.*, 2011; Martis *et al.*, 2013a, 2013c; Park *et al.*, 2008; Prasad *et al.*, 2013), Transformada Discreta do Cosseno (DCT – *Discrete Cosine Transform*) (Martis *et al.*, 2013b), Transformada de Fourier Discreta (DFT – *Discrete Fourier Transform*) (Pławiak, 2018), características personalizadas (Cheng and Dong, 2017), auto codificadores (*auto-encoders*) (Rahhal *et al.*, 2016), transformação Wavelet (Elhaj *et al.*, 2016; Kumar *et al.*, 2017; Kutlu and Kuntalp, 2012; Oh *et al.*, 2017; Sahoo *et al.*, 2017; Sharma *et al.*, 2019b, 2019a, 2019c; Thomas *et al.*, 2015; Yildirim, 2018), *Common Spatial Pattern* (CSP) (Oh *et al.*, 2017), entropias diversas (Udyavara Rajendra Acharya *et al.*, 2018; Sharma *et al.*, 2019b, 2019a, 2019c; Tripathy *et al.*, 2016), decomposição de modo variacional (VMD – *Variational Mode Decomposition*) (Tripathy *et al.*, 2016), modelos auto regressivos (Alickovic and Subasi, 2015), auto codificadores esparsos empilhados (SSAE – *Stacked Sparse Autoencoders*) (Jianli Yang *et al.*, 2018) e blocos convolucionais (Zihlmann *et al.*, 2017).

Com relação à aprendizagem e classificadores, é possível separá-los em duas grandes abordagens com diversos algoritmos específicos. A primeira, aprendizagem de máquina tradicional (*conventional machine learning*), tem como algumas das ramificações: k-vizinhos próximos (KNN - *K-Nearest Neighbor*) (Alickovic and Subasi, 2015; Kutlu and Kuntalp, 2012; Oh *et al.*, 2017; Prasad *et al.*, 2013), *support vector machine* (SVM) (Alickovic and Subasi, 2015; Cheng and Dong, 2017; Elhaj *et al.*, 2016; Kumar *et al.*, 2017; De Lannoy *et al.*, 2011; Park *et al.*, 2008; Pławiak, 2018; Sahoo *et al.*, 2017; Sharma *et al.*, 2019b, 2019a), campos randômicos condicionais (*Conditional Random Fields*) (De Lannoy *et al.*, 2012), sistema neuronal evolucionário (Pławiak, 2018), *Random Forest* (RF) (Tripathy *et al.*, 2016), análise discriminante linear (LDA – *Linear Discriminant Analysis*) (Yeh *et al.*, 2009) e Perceptron

multicamada (MLP – *Multi-Layer Perceptron*) (Thomas *et al.*, 2015). A segunda, baseada em aprendizagem profunda (*deep learning*), substituiu, na maioria das aplicações, a necessidade de extração manual de características, abstraindo em uma única estrutura a expressão e a generalização de características não lineares da entrada (Li *et al.*, 2019; Lynn *et al.*, 2019; Yildirim, 2018). São estruturas usuais dessa abordagem: redes neurais convolucionais (CNN – *Convolutional Neural Network*) (Acharya *et al.*, 2017b, 2017a; U. Rajendra Acharya *et al.*, 2018; Andersen *et al.*, 2019; Dang *et al.*, 2019; Fan *et al.*, 2017; Wu *et al.*, 2018; Yildirim *et al.*, 2018; Zubair *et al.*, 2016), redes neurais recorrentes (RNN – *Recurrent Neural Network*) (Andersen *et al.*, 2019; Latif *et al.*, 2018; Wang *et al.*, 2019; Zhang *et al.*, 2017), LSTM (*Long Short-Term Memory*) (Chauhan and Vig, 2015; Dang *et al.*, 2019; Oh *et al.*, 2018; Tan *et al.*, 2018; Yildirim, 2018; Yildirim *et al.*, 2019; Zihlmann *et al.*, 2017), *Gated Recurrent Unit* (GRU) (Lynn *et al.*, 2019), auto codificadores convolucionais (CAE – *Convolutional Autoencoder*) (Oh *et al.*, 2019; Yildirim *et al.*, 2018).

J. Zhang (Zhang *et al.*, 2019) utilizou duas camadas LSTM empilhadas com entrada composta por sinais ECG brutos e mesclou a sua saída com um conjunto de características morfológicas e temporais extraídas manualmente, denominado pelo autor como recursos de domínio, para determinar a usabilidade de sinais ECG obtidos por pessoas inexperientes em plataformas mobile, para diferenciação entre duas classes. Obteve acurácia de 93,5% no conjunto de entrada público (PhysioNet Challenge 2011) e 87,94% no conjunto privado construído pelos autores.

R. He (He *et al.*, 2019) utilizou a combinação entre redes residuais convolucionais (abstração de características locais) e LSTM bidirecional (abstração de características globais) para diferenciar 9 classes de comportamentos cardíacos, obtendo F1-score de 0,806 para a base de dados *China Physiological Signal Challenge 2018* (CPSC). Neste trabalho fez-se uso dos sinais brutos, aplicando segmentação temporal e não os submetendo a processos de extração manual de características.

M. Moridani (Moridani *et al.*, 2019) desenvolveu um sistema automático de diferenciação entre sinais de ECG normais e anormais baseado em support vector machine (SVM). Como pré-processamento utilizou o método Pam and Tompkins para detecção dos picos R e a partir deste, determinação dos complexos QRS, segmentação em janelas de batimentos, intervalos entre picos R e a determinação da *heart rate variability* (HRV). Como características extraídas, pode-se citar a latência e a amplitude dos sinais obtidos anteriormente e características do domínio tempo-frequência com dimensões fractais e transformadas

Wavelet. Fez-se uso do método *principal component analysis* (PCA) para determinação das características mais significativas. Obteve uma acurácia de 98,3% sobre os dados da base *PhysioNet Challenge 2016*.

S. Chauhan (Chauhan and Vig, 2015) propôs um sistema de classificação de anormalidades em sinais de ECG baseado no erro obtido entre a diferença da predição de redes LSTM profundas previamente treinadas apenas com dados sem anormalidades e o sinal original. O método para avaliar o erro foi constituído pela distribuição Gaussiana multivariada e o resultado de acurácia média adquirido para diferenciação entre 5 classes da base *PhysioBank MIT-BIH Arrhythmia* foi de 96,00%.

R. Li (Li *et al.*, 2019) apresentou um novo método conjunto para redes LSTM bidirecionais, *Attention*, que pondera as saídas destas de modo a direcionar mais “atenção” para amostras mais relevantes. O modelo é aplicado no desenvolvimento de um sistema de classificação automática de 5 classes de comportamentos em sinais ECG e utiliza como entrada, gravações da base de dados *PhysioNet MIT-BIH arrhythmia*. Como artifício para remoção de ruídos, utilizou a transformada Wavelet contínua (CWT). As gravações são segmentadas e alguns conjuntos de dados são construídos valendo-se de sequências de intervalos R-R e características morfológicas dos eventos P-QRS-T. A acurácia obtida para a proposta foi de 99,49%.

H. Lynn (Lynn *et al.*, 2019) desenvolveu parte de um sistema de reconhecimento de indivíduos a partir de sinais de ECG, pelo que para isso utilizou *deep gated recurrent units* (*deep GRUs*) bidirecionais. A partir dos dados obtidos nas bases *PhysioNet ECG-ID* e *PhysioNet MIT-BIH Arrhythmia*, realizou a remoção de ruídos e tendências, normalização, detecção de picos R (Pam and Tompkins) e segmentação das gravações. Propôs comparações e associações com outras estruturas de redes como LSTM e CNN. A proposta alcançou a acurácia de 98,50%.

J. Tan (Tan *et al.*, 2018) utilizou a combinação sequencial entre CNN e LSTM para criar um sistema automatizado de detecção de doenças relacionadas a artéria coronária (CAD). Para isso, utilizou as bases de dados *PhysioNet Fantasia* e *PhysioNet St Petesburg Institute of Cardiology Technics*. Como artifício para atenuação de ruídos e linha de base das gravações fez uso da transformada Wavelet discreta (DWT). Utilizou como entrada uma janela de gravações de 5s e a geração de novos dados a partir da combinação de três sinais num modelo inovador para balanceamento dos conjuntos. Não houve extração manual de características. De acordo com as abordagens para teste do modelo, foram obtidas acurácia de 99,85% e 95,76%.

M. Sharma (Sharma *et al.*, 2019c) desenvolveu um modelo de sistema automatizado para diferenciação de arritmias chocáveis (*shockable*) e não chocáveis (*non-shockable*) a ser aplicado em desfibriladores de uso interno e externo ao ambiente hospitalar, auxiliando em manobras de ressuscitação e prevenção de morte cardíaca súbita. A proposta tinha como objetivo demonstrar uma alternativa aos métodos de extração de características baseadas na morfologia dos batimentos. Para isso, faz uso de uma metodologia de decomposição Wavelet otimizada de dois canais em forma de banco de filtros ortogonais de três bandas. Os dados de entrada, providos das bases *MIT-BIH Arrhythmia*, *MIT-BIH Fatal Ventricular Arrhythmia* e *Ventricular Tachyarrhythmia from Creighton University*, são segmentados em janelas de 2s e, apesar de não detalhar, são filtrados para atenuação de ruídos. Dos coeficientes obtidos na decomposição, são extraídas as características de entropia Fuzzy, Renyi e amostral. Pela utilização do classificador SVM cúbico, obteve acurácia de 98,90%.

B. Murugesan (Murugesan *et al.*, 2018) propôs o desenvolvimento de um sistema automático de classificação de comportamentos de três classes de sinais ECG, baseado em redes neurais profundas e utilizando segmentos de gravações de 2s. Aplicação da técnica de transferência de aprendizagem para adaptabilidade da proposta a diferentes bases de dados, *PhysioNet MITDB*, *LTDB* e *LTAADB*. Implementação de redes CNN e LSTM em paralelo e empilhadas, sem extração manual de características ou processos de filtragem. O melhor resultado de acurácia obtida para o modelo CLSTM, foi de 97,60%.

H. Kutlu (Kutlu and Avcı, 2019), diferentemente dos trabalhos citados anteriormente, utilizou um conjunto de sistemas composto por CNN+DWT+LSTM para classificação de tumores hepáticos e cerebrais. Para isso, valeu-se de imagens de ressonância e tomografias extraídas das bases *Radiology Laboratory of Firat University Research Hospital*. Para essas, não houve pré-processamento e as características extraídas foram obtidas por meio da rede CNN pré-treinada AlexNet em conjunto com decomposição por DWT. A acurácia resultante do desempenho da proposta foi de 99,10%.

J. Yang (Jun Yang *et al.*, 2018), assim como (Kutlu and Avcı, 2019), utilizaram a combinação CNN+DWT+LSTM para classificação de sinais, entretanto aqui para *motor imagery EEG* (MI-EEG), como uma ferramenta de auxílio no reconhecimento de conteúdo útil para dispositivos de interface cérebro-computador (BCI). Como pré-processamento, destaca-se a atenuação de artefatos com ICA (*Independent Component Analysis*). No processo de extração de características, é gerado um conjunto destas pela combinação entre CNN e DWT. Ainda, utilizou o método *batch normalization* (BN) para aprimorar o treinamento. A acurácia da

proposta para o conjunto de dados público (BCI *Competition datasets 3 and 4*) foi de 87,40% enquanto para o conjunto privado (gerado pelos autores) foi de 86,70%.

Ö. Yildirim (Yildirim *et al.*, 2018) apresenta um modelo profundo de redes neurais convolucionais unidimensionais para diferenciação de 17 classes de comportamentos cardíacos em gravações de ECG de 10s para aplicações em tempo real. Baseando-se em *PhysioNet MIT-BIH Arrhythmia*, os dados foram normalizados e filtrados para atenuação da componente constante, entretanto, não houve extração manual de características. A acurácia média obtida para a proposta foi de 91,53%.

S. Chauhan (Chauhan *et al.*, 2019) dá continuidade no trabalho (Chauhan and Vig, 2015) associando à saída do modelo anterior um estágio de classificação que se baseia no erro de predição da rede e o sinal original para diferenciar 7 classes de anomalias no ECG. A partir dos dados da base *PhysioNet MIT-BIH Arrhythmia* foram realizadas segmentações em períodos de um minuto e a referência associada a estes segmentos foi alterada de modo a determiná-los como sendo patológicos ou não. A extração de características é feita com base no vetor de erro obtido entre a saída preditora da etapa anterior e a entrada, sendo realizadas análises estatísticas sobre os dados e arranjando os resultados matricialmente para a etapa de classificação. As estruturas de classificadores propostas, *multi layer perceptron* (MLP), *logistic regression* (LR) e SVM, apresentaram respectivamente, acurácia de 42,86%, 51,43% e 50,00%. Além disso, foi desenvolvido um método de background para avaliar o desempenho e confiabilidade das propostas.

O. Faust (Faust *et al.*, 2018b) utilizou conjuntos de segmentos de 100 intervalos entre picos R consecutivos (RRi) em redes neuronais recorrentes do tipo LSTM para a detecção de fibrilação atrial. A acurácia obtida pelos autores foi de 98,51% na utilização da técnica *10-fold cross-validation* e 99,77% no teste para sujeitos não apresentados durante o treino (*blindfold validation*).

H. Dang (Dang *et al.*, 2019) utilizou as gravações de ECG da base *MIT-BIH Atrial Fibrillation* sem filtragem para a criação de dois conjuntos de dados baseados em: 100 RRi (conjunto A) e 100 sequências de 5 amplitudes referentes a cada onda (P, Q, R, S e T) dos 100 batimentos (conjunto B). Os conjuntos foram submetidos a um processo de normalização Z-score e inseridos numa combinação entre redes neuronais convolucionais (CNN) e LSTM bidirecionais (BLSTM) para detecção de fibrilação atrial, tendo obtido uma acurácia de 96,59%.

R. S. Andersen (Andersen *et al.*, 2019) utilizou segmentos de 30 RRI também sobre a combinação entre CNN e BLSTM para a mesma classificação de fibrilação atrial. O conjunto de treinamento foi construído sobre a base de dados *MIT-BIH Atrial Fibrillation* sendo a acurácia média obtida em *5-fold cross-validation* de 97,80%. Os autores ainda propuseram a aplicação do modelo para as gravações da base *MIT-BIH Arrhythmia*, onde a acurácia obtida foi 89,30%.

Ö. Yildirim (Yildirim, 2018) desenvolveu uma adaptação a entrada de redes LSTM pela implementação de um camada baseada na transformada wavelet. A partir dela, os sinais de ECG, retirados da base *PhysioBank MIT-BIH Arrhythmia*, são decompostos em várias bandas de frequência a diferentes escalas, gerando um conjunto de novas sequências que expressam as dependências em frequência da entrada. Utilizando esta proposta, o autor obteve acurácia de 99,39% na classificação de 5 tipos de batimentos cardíacos.

Ö. Yildirim (Yildirim *et al.*, 2019) propôs um sistema de classificação de arritmias baseado na codificação dos batimentos cardíacos utilizando Auto Codificadores Convolucionais (CAE). O objetivo foi reduzir a quantidade de amostras na entrada do modelo de aprendizagem (LSTM) e ao mesmo tempo favorecer a diferenciação entre as classes, no caso, cinco classes foram utilizadas. A medida de perda de informação na reconstrução pelo CAE foi de 0,70%. A acurácia de classificação obtida no estudo foi de 99,11% no modelo com uso do CAE e de 99,23% no modelo sem. Entretanto, houve redução do tempo de processamento (treino e teste) em cerca de 7 vezes quando aplicada a codificação (já considerando o tempo de sua execução), o que viabiliza um estudo aprofundado de sua implementação em tempo real.

S. Saadatnejad (Saadatnejad *et al.*, 2019) desenvolveu um sistema de classificação de batimentos cardíacos direcionado para aplicações em tempo real em dispositivos *wearable*. Na proposta foram construídos três conjuntos de dados para a entrada do modelo de aprendizagem: segmentos de tamanho fixo aproximadamente centrados no pico R (detecção pelo algoritmo *Pam and Tompkins*) de cada batimento cardíaco, quatro parâmetros R-R (referentes ao contexto de vizinhança de cada batimento e do paciente) e decomposição pela transformada wavelet discreta. Estes conjuntos são apresentados, a partir de uma combinação, como exemplos para dois modelos de aprendizagem baseados em redes LSTM. Em um deles, é utilizada uma etapa de seleção entre os conjuntos baseado no método PCA (*Principal Component Analysis*). Os dois modelos são executados simultaneamente e seus resultados são combinados para

determinação da classificação final. A acurácia da proposta esteve sempre acima de 98,3% para todas as bases de dados utilizadas, sendo que houve diferenciação entre 7 e depois 5 classes.

Utilizando sinais polissonográficos (PSG) (agregação de sinais biomédicos de diversas fontes, como por exemplo algumas derivações de eletroencefalograma (EEG), eletro-oculograma (EOG), ECG, SaO₂ (saturação), etc), Warrick (Warrick *et al.*, 2019) propôs um sistema de detecção da perda de sono de pacientes utilizando a extração de características a partir da transformada Scattering (ST) de segunda ordem. Baseando-se nesta técnica, são determinados mapas de características relativos a etapas de decomposição em diferentes níveis, aplicação de não linearidade e função escala. A representação obtida é introduzida em uma camada chamada convolucional profundamente separável que mantém o aspecto temporal enquanto adequa a dimensionalidade para a rede BLSTM. As grandes contribuições deste trabalho, pode-se dizer que foram a demonstração da melhora de representação dos dados pelo uso da ST relativamente ao trabalho anterior dos autores e a confirmação da suspeita de relevância das bandas de frequência “infra lentas” do EEG, sugerida por trabalhos adjacentes.

1.4 Justificativa

De acordo com o estado da arte e pela observação das características da fibrilação atrial e das gravações de ECG de longo termo, sistemas de auxílio ao diagnóstico médico se mostram de grande valor, especialmente pelo uso de inteligência artificial. Portanto, propõe-se o estudo da classificação de segmentos de sinais ECG para detecção de fibrilação atrial pelo uso de técnicas de análise de séries temporais baseadas na extração de recursos morfológicos, temporais e estatísticos dos sinais a serem utilizados em algoritmos de aprendizagem de máquina.

Em uma das ferramentas sugeridas para o desenvolvimento da proposta, propõe-se a utilização da transformada Scattering baseada na função de Gabor (Morlet wavelet) (Warrick *et al.*, 2019) para o mapeamento do ECG em diferentes níveis de escala tempo-frequência. A utilização da transformada Scattering em sinais biomédicos para sistemas de aprendizagem de máquina é quase inexistente, mas sinaliza ter alto potencial pelos resultados encontrados em outras aplicações (Warrick *et al.*, 2019). A utilização da Scattering pode ser interessante por conta dos seguintes aspectos: redução de dimensionalidade do sinal (levando a uma redução do tempo de processamento do modelo de aprendizagem); pelo conjunto de operações aplicadas (transformadas wavelet, não linearidade e média) conduzir a informação de uma forma muito

parecida com as redes neuronais; e pelo destaque de comportamentos observáveis apenas no domínio das frequências por decomposições em múltiplas escalas. Esta decomposição em múltiplas escalas é interessante no sentido da própria codificação do sinal, onde como resultado da transformação Scattering, se obtém um conjunto de dados no formato matricial, à prova de translações e deslocamentos na entrada. Além disso, mantêm-se os recursos temporais do sinal original, que podem ser interpretados por uma inteligência artificial dedicada à análise de comportamentos com dependência temporal.

Os modelos de inteligência artificial utilizados podem ser do tipo rede neuronal artificial devido à sua habilidade já consolidada em abstrair recursos de sinais de entrada complexos, com variabilidade entre sujeitos e qualidade prejudicada pela presença de ruído, artefatos e períodos de silêncio. Além disso, são recomendadas para lidar com conjuntos de dados muito extensos, como é o caso de gravações de ECG de longo termo.

1.5 Objetivos

Nesta seção são apresentados os objetivos que orientam a proposta deste trabalho. A criação dos objetivos foi dada a partir do estudo do estado da arte e das justificativas apresentados nas seções anteriores.

1.5.1 Objetivos Primários

O trabalho tem como objetivo principal a experimentação da utilidade das redes neuronais recorrentes do tipo *Long Short-Term Memory* (LSTM) para a identificação de patologias cardíacas. Para isso, propõe-se a utilização de gravações de eletrocardiogramas (ECG) de longo termo e o desenvolvimento dos sistemas em linguagem Matlab.

1.5.2 Objetivos Específicos

Como objetivos específicos são apresentadas a utilização das transformadas Wavelet, a extração dos parâmetros Jitter e Shimmer de segmentos dos ECG e a utilização de redes neuronais do tipo Perceptron Multi Camadas (MLP).

Entre as patologias cardíacas encontradas na literatura, neste trabalho especificamente são abordados os comportamentos relativos à fibrilação atrial. O sistema proposto, portanto,

tem como objetivo específico a classificação de segmentos de ECG de longo termo entre fibrilação atrial e não fibrilação atrial.

Capítulo 2: Materiais e Métodos

Nesta seção são apresentados os materiais e métodos utilizados neste trabalho. Os materiais se referem a base de dados de onde foram extraídas as gravações de ECG utilizadas e os recursos computacionais que envolveram as implementações das propostas. Os métodos apresentados são referentes às técnicas de processamento de sinal e inteligência artificial que foram aplicadas para compor a proposta.

2.1 Base de Dados

A base de dados utilizada foi a *MIT-BIH Atrial Fibrillation* (Goldberger *et al.*, 2003; Moody and Mark, 1992) disponível pública e gratuitamente em (MIT-BIH Atrial Fibrillation Database v1.0.0). Nesta base de dados, estão disponíveis 25 gravações de ECG de longo termo (duração de 10 horas para cada gravação) entre as quais 23 são representadas tanto pelo sinal ECG de duas derivações quanto por suas respectivas anotações de detecção do complexo QRS e ritmo associado. As outras duas gravações apenas contém as anotações de batimentos e ritmo e, portanto, são descartadas. Os pacientes cujas gravações foram obtidas apresentaram comportamento ritmo de fibrilação atrial na sua maioria do tipo paroxismal.

As anotações rítmicas foram construídas a partir do diagnóstico de médicos especialistas e possuem as seguintes classes: fibrilação atrial, *flutter* atrial, ritmo juncional AV (atrioventricular) e normal (compreendendo quaisquer outros ritmos). A detecção dos complexos QRS, foi realizada via algoritmo computacional automático.

Das características de gravação, os sinais foram obtidos a partir de um Holter portátil de monitorização contínua e digitalizados a uma frequência de amostragem de 250 Hz com 12-bits de resolução sobre um intervalo de ± 10 mV. As gravações originais foram realizadas no Hospital Beth Israel de Boston utilizando sistema de aquisição com largura de banda típica entre 0,1 Hz e 40 Hz.

2.2 Recursos Computacionais

Para conduzir os experimentos foi utilizado um computador portátil com processador Intel® Core™ i7-6500U 2.5 GHz, 8 GB de memória RAM e placa de vídeo GeForce 920MX

2 GB. O sistema proposta foi implementado e executado no software MatLab® 2020a, utilizando as *Toolboxes* de *Deep Learning*, *Wavelet*, *Global Optimization*, *Signal Processing* e WFDB (Goldberger *et al.*, 2000; Silva and Moody, 2014).

2.3 Transformada De Hilbert

A transformada de Hilbert é uma técnica de processamento de sinais utilizada principalmente nas áreas de telecomunicações, biomédica e de sistemas elétricos de potência. Consiste na decomposição de sinais reais para sua forma complexa (analítica) pela utilização dos conceitos de envelope e pré-envelope. Apesar de envolver um processo de transformação, o sinal resultante permanece no domínio do tempo, sendo apenas sua representação alterada (Benitez *et al.*, 2001; Manikandan and Soman, 2012).

Uma das principais aplicações desta técnica sobre sinais de ECG, inclusive utilizada neste trabalho, é a determinação de regiões com alta probabilidade de detecção do pico da onda R. Como resultado da transformação do ECG, obtém-se na região do complexo QRS, uma função de simetria ímpar, onde seu ponto de cruzamento de zero indica o instante de alta probabilidade de ocorrência do pico R (Benitez *et al.*, 2001; Manikandan and Soman, 2012).

Matematicamente, a transformada de Hilbert, $\hat{x}(t)$, de um sinal real, $x(t)$, pode ser expressa como apresenta a Equação 1. Nota-se que a transformada em questão é a convolução do sinal de entrada, que no caso é o ECG, com o sinal $\frac{1}{\pi t}$.

$$\hat{x}(t) = \mathcal{H}[x(t)] = \frac{1}{\pi t} * x(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \quad (1)$$

No domínio da frequência, pela utilização da transformada de Fourier, tem-se que, $\hat{X}(f)$ é a multiplicação ponto a ponto entre a função $sgn(f)$ com a transformada do próprio sinal de entrada, $X(f)$, ponderada pela constante $(-j)$. Esta relação é descrita matematicamente pela Equação 2, onde a definição da função $sgn(\cdot)$, na frequência, é dada em seguida. Note-se que o efeito da transformada sobre o sinal é a aplicação de um atraso de fase igual a -90° para frequências positivas e um avanço de 90° para negativas.

$$\hat{X}(f) = \mathcal{F} \left[\frac{1}{\pi t} * x(t) \right] = \mathcal{F} \left[\frac{1}{\pi t} \right] \cdot \mathcal{F}[x(t)] = -j \cdot sgn(f) X(f) \quad (2)$$

$$\text{sgn}(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases}$$

No domínio do tempo, após aplicação da transformada de Fourier inversa sobre $\hat{X}(f)$, o resultado da transformada de Hilbert aparece como o fator complexo do sinal analítico, $y(t)$, representado pela Equação 3. Na forma polar, as Equações 4 e 5 apresentam, respectivamente, o envelope e o pré-envelope complexo do sinal $y(t)$.

$$y(t) = x(t) + j\hat{x}(t) \quad (3)$$

$$B(t) = \sqrt{x(t) + \hat{x}(t)} \quad (4)$$

$$\phi(t) = \tan^{-1} \left(\frac{\hat{x}(t)}{x(t)} \right) \quad (5)$$

2.4 Entropias

As entropias, como utilizadas neste trabalho, são representações do nível de organização dos sinais de ECG. Este caráter de organização pode ser descrito matematicamente de diversas maneiras de acordo com o critério e o objetivo da observação. Por exemplo, neste trabalho, com as entropias, pretende-se ter uma noção da quantidade de informação contida nos segmentos dos ECG e a intensidade de complexidade dos sinais. No caso do objeto de estudo ser o comportamento de fibrilação atrial, a utilização de janelas longas como entradas para determinação das entropias selecionadas é fortemente indicada, uma vez que estas fornecem maior contextualização sobre as dependências temporais do segmento e podem ser mais descritivas que janelas curtas (Udyavara Rajendra Acharya *et al.*, 2018).

2.4.1 Entropia de Shannon

Fundamentada na teoria da informação de Shannon, esta medida de entropia fornece uma observação acerca da distribuição de probabilidade da informação ao longo do segmento de ECG, $x(n)$ (Udyavara Rajendra Acharya *et al.*, 2018; Manikandan and Soman, 2012). Pode ser interessante devido as variações de morfologias das ondas dentro de um período de observação de vários batimentos cardíacos ou sobre as ondas isoladamente. A Equação 6

descreve esta relação, onde se pode notar que o resultado é um escalar ponderado pela informação ao longo da amplitude do sinal.

$$E_{sh} = -\sum_{n=1}^N (|x(n)| \cdot \log_2(|x(n)|)) \quad (6)$$

2.4.2 Entropia Logaritmo da Energia

Segundo (Udyavara Rajendra Acharya *et al.*, 2018; AydIn *et al.*, 2009), este tipo de entropia estima a intensidade de complexidade do sinal, e pode ser descrita pela Equação 7. Esta complexidade pode ser notada como a distribuição da energia de um sinal aleatório, em escala logarítmica, ao longo do tempo (AydIn *et al.*, 2009).

$$E_{logen} = \sum_{n=1}^N \log_2(x(n)^2) \quad (7)$$

2.4.3 Energia de Shannon

A energia de Shannon, Equação 8, é uma transformação não linear que pondera a energia de um sinal, $x(n)$, pela própria distribuição de energia (Manikandan and Soman, 2012). Essa representação faz com que se tenha uma grande expressividade dos locais do sinal com muita informação (caso do complexo QRS), ao mesmo tempo que reduz os efeitos de componentes de ruído de baixa amplitude (Manikandan and Soman, 2012).

$$E_{shen}(n) = -x(n)^2 \cdot \log_2(x(n)^2) \quad (8)$$

2.5 Jitter

Jitter pode ser definido como uma medida de perturbação temporal em sinais periódicos (Teixeira *et al.*, 2013). A perturbação observada é relacionada pela variação do período fundamental do sinal ciclo a ciclo ao longo de uma janela (Teixeira *et al.*, 2013). A princípio, quatro parâmetros Jitter são considerados, os quais se diferem pela quantidade de ciclos utilizados como referência de comparação.

As Equações de 9 a 12 descrevem os quatro parâmetros Jitter utilizados neste trabalho, sendo T o tempo de duração de cada um dos períodos dentro de uma janela de N períodos consecutivos. O Jitter absoluto ($J1$) é a média, em segundos, da diferença absoluta entre pares de períodos consecutivos dentro da janela N . O Jitter relativo ($J2$) é a medida do Jitter absoluto normalizado pela média dos períodos que constituem a janela. O Jitter RAP (*Relative Average Perturbation*) ($J3$) é a medida de perturbação relativa entre cada período e seu contexto até dois períodos vizinhos adjacentes. O Jitter PPQ5 ($J4$) é a medida de perturbação relativa de cinco ponto que leva em consideração a diferença absoluta de cada período e seu contexto até quatro períodos vizinhos adjacentes (Teixeira and Gonçalves, 2016).

$$J1 = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_{i+1} - T_i| \text{ [s]} \quad (9)$$

$$J2 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_{i+1} - T_i|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \text{ [%]} \quad (10)$$

$$J3 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-1} \left| T_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \text{ [%]} \quad (11)$$

$$J4 = \frac{\frac{1}{N-1} \sum_{i=3}^{N-2} \left| T_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \text{ [%]} \quad (12)$$

Apesar de sinais ECG não serem completamente periódicos, nota-se uma clara periodicidade do ciclo cardíaco ao longo do tempo. A fibrilação atrial por sua vez, insere uma alta variação nesta periodicidade, alterando o ritmo cardíaco em uma curta janela de tempo. Portanto, a determinação dos parâmetros de Jitter pode colaborar como um recurso que expõe a quantidade desta variação, contribuindo para a diferenciação entre um segmento que contém AF ou não.

2.6 Shimmer

Shimmer pode ser definido como a medida de variação de amplitude ciclo a ciclo de um sinal periódico dentro de uma janela de N ciclos. A partir de uma referência de amplitude de cada ciclo, quatro parâmetros Shimmer podem ser determinados como medidas absolutas e relativas de perturbação dentro da janela (Teixeira *et al.*, 2013; Teixeira and Gonçalves, 2016). As Equações de 13 a 16 descrevem os parâmetros Shimmer utilizados neste trabalho, onde N é o número de ciclos contidos na janela e A são as medidas de amplitude de cada ciclo.

O Shimmer absoluto ($S1$) é a média absoluta da razão logarítmica entre a amplitude de dois períodos consecutivos. O Shimmer local ($S2$) é a diferença absoluta média entre cada par de períodos consecutivos na janela N , normalizada pela média de amplitude dos períodos que constituem a janela. O Shimmer APQ3 ($S3$) é a medida de perturbação em amplitude de três pontos que representa a média da diferença absoluta entre cada período e seus dois vizinhos adjacentes. Já o Shimmer APQ5 ($S4$) representa a mesma perturbação anterior, mas agora com um contexto de vizinhança de quatro vizinhos adjacentes a cada período, ambas medidas normalizadas pela amplitude média da janela.

$$S1 = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 * \log \left(\frac{A_{i+1}}{A_i} \right) \right| [dB] \quad (13)$$

$$S2 = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 [\%] \quad (14)$$

$$S3 = \frac{\frac{1}{N-1} \sum_{i=2}^{N-1} \left| A_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 [\%] \quad (15)$$

$$S4 = \frac{\frac{1}{N-1} \sum_{i=3}^{N-2} \left| A_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} A_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 [\%] \quad (16)$$

Uma vez que é característico da AF a deformação do complexo QRS, a determinação das medidas de Shimmer para este trabalho pode ser interessante como um recurso que representa esta perturbação e colabora na diferenciação entre a presença ou não da AF no segmento.

2.7 Normalização Z-Score

A normalização dos dados de entrada é uma etapa essencial para modelos de aprendizagem de máquina baseados em redes neurais artificiais. Tanto por conta da maioria dos algoritmos de treinamento destes modelos estarem relacionados com a noção de descida do gradiente quanto pela redução de redundâncias e inconsistência na presença de anomalias nos exemplos de entrada (Oh *et al.*, 2018; Sharma *et al.*, 2018; Silva *et al.*, 2019). Ainda, a normalização ajuda a reduzir os efeitos de variabilidade entre pacientes, mantendo os comportamentos alinhados a uma mesma escala (U. Rajendra Acharya *et al.*, 2018).

O método de normalização utilizado neste trabalho foi a normalização padrão Z-score (Standardized z-scores - MATLAB). A Equação 17 descreve matematicamente este método,

onde o sinal $x(n)$ é descontado de seu valor médio, \bar{x} , e em seguida dividido por seu desvio padrão, S . O resultado, $z(n)$, é a versão de média aproximadamente zero e desvio padrão unitário de $x(n)$.

$$z(n) = \frac{x(n) - \bar{x}}{S} \quad (17)$$

2.8 Normalização Pelo Valor Máximo Absoluto

A normalização pelo valor máximo de um vetor qualquer, $x(n)$, é definida pela relação apresentada pela Equação 18, onde $\tilde{x}(n)$ é a representação de $x(n)$ numa escala de intervalo $[-1; 1]$ (Manikandan and Soman, 2012). A Figura 4 apresenta um exemplo comparativo entre a normalização do tipo Z-score e pelo máximo absoluto. É de se notar que a normalização por Z-score padroniza o vetor exemplo no sentido estatístico de média e desvio padrão, não se importando com o máximo ou mínimo, enquanto a normalização por valor máximo absoluto transforma a escala de amplitude num intervalo de $[-1, 1]$, padronizando as medidas estatísticas de máximo e mínimo e negligenciando a média e desvio padrão.

$$\tilde{x}(n) = \frac{x(n)}{\max_{n=1}^N (|x(n)|)} \quad (18)$$

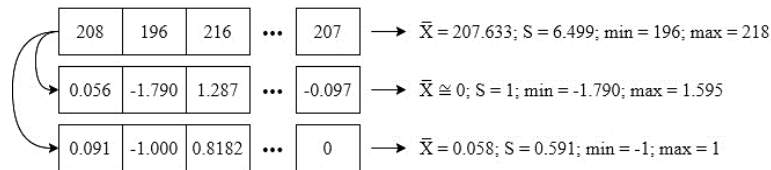


Figura 4 - Comparação entre normalização Z-score (2ª linha/vetor) e pelo máximo absoluto (3ª linha/vetor).
Fonte: Autoria própria.

2.9 Média Móvel

A média móvel, ou também, média deslizante, é a ponderação das amostras de um sinal, x , pela aplicação de uma função janela, w , de tamanho ímpar, $L + 1$. A função janela é aplicada sobre um segmento do sinal, centrado em uma determinada amostra, gerando uma amostra de saída. O processo é repetido ordenadamente para todos os segmentos possíveis a

partir de um deslizamento equivalente a um passo amostral (Proakis and Manolakis, 2007). A Equação 19 apresenta a relação matemática da média móvel utilizada, sendo negligenciados os casos especiais desta quando nas bordas do sinal.

$$M(n) = \frac{1}{L+1} \sum_{m=n-\frac{L}{2}}^{n+\frac{L}{2}} x(m).w\left(m - \left(n - \frac{L}{2}\right)\right) \quad (19)$$

Como função janela, neste trabalho foi utilizado um trem de impulsos unitários, ou seja, uma janela do tipo retangular. Com isso, a ponderação aplicada sobre o contexto dentro dos segmentos foi do tipo simples.

2.10 Frequências Instantâneas

As frequências instantâneas de um sinal não estacionário, como os de ECG por natureza, permitem a observação da variação do seu conteúdo espectral ao longo do tempo (Boashash, 1992). Basicamente, por conta destes sinais não serem limitados no tempo (são limitados em banda, no domínio da frequência) é necessário que um janelamento seja aplicado para delimitar uma região de observação e ponderar o comportamento do sinal por partes. Este processo, assim como na média móvel, envolve uma função janela. Neste caso, e particularmente neste trabalho com a janela de Hamming, esta função é utilizada com a finalidade de reduzir os efeitos de borda por conta do truncamento (atenuar a ocorrência de ecos no espectro de frequência). O contexto dos segmentos passa a considerar as amostras laterais (próximas a borda) com muito menos importância do que as amostras centrais (Boashash, 1992; Harris, 1978). Entretanto, em segmentos adjacentes há uma sobreposição de amostras, que de certa forma compensa esse efeito de importância.

Cada segmento por sua vez é submetido à transformada de Fourier de curto-termo que converte a representação do sinal no tempo para tempo-frequência. Ou seja, cada instante do sinal original, de acordo com uma determinada resolução, passa a ser representado pelo perfil espectral do segmento relacionado, daí a ideia de frequências instantâneas. Como resultado, obtém-se uma matriz de três eixos, onde: o eixo das abscissas (eixo 'x') permanece sendo o tempo, o eixo das ordenadas (eixo 'y') exibe as escalas de frequência analisadas e o eixo de cota (eixo 'z') expressa a intensidade da frequência (Boashash, 1992). As frequências instantâneas são determinadas através da relação apresentada pela Equação 20, onde $P(t, f)$ é

o espectrograma do sinal de entrada calculado pela transformada de Fourier de curto-termo. As variáveis t e f da Equação 20 são respectivamente o tempo e a frequência.

$$f_{inst}(t) = \frac{\int_0^{\infty} f \cdot P(t,f) df}{\int_0^{\infty} P(t,f) df} \quad (20)$$

Portanto, de acordo com a variação do perfil espectral dos sinais de ECG ao longo do tempo, pode-se encontrar padrões significativos para descrever determinados comportamentos de anomalia ou não. Assim, mesmo que esta indicação possa parecer complexa no sentido de quantidade de amostras, acredita-se que os modelos de aprendizagem de máquina sejam capazes de compreender as informações de dependência tempo-frequência assim como nos sinais representados apenas pelo tempo. Não só isso, mas acredita-se que alguns padrões são mais bem reconhecidos quando representados no domínio das frequências.

2.11 Entropia Espectral

A entropia espectral de um sinal é a medida da distribuição de informação acerca do seu espectro de potência (Vakkuri *et al.*, 2004). A partir do espectrograma de potência do sinal de entrada, $S(t, f)$, em que t e f são respectivamente as variáveis de tempo e frequência, a distribuição de probabilidade para cada tempo t é apresentada pela Equação 21, onde m é o indicador de frequências (assim como f) para critério de normalização (Bein, 2006; Shen *et al.*, 1998). Ou seja, dado um instante t , o valor em cada frequência m , é normalizado de acordo com todos os valores no mesmo instante t .

$$P(t, m) = \frac{S(t,m)}{\sum_f S(t,f)} \quad (21)$$

O cálculo da entropia espectral instantânea (ou seja, para cada instante t) é realizado através da Equação 22, que utiliza a noção de informação de Shannon (Bein, 2006; Shen *et al.*, 1998). Resumindo, para cada instante t , é determinada uma noção de informação espectral acerca do espectro de potência relativo à contribuição da amostra $x(t)$ do sinal de entrada.

$$H(t) = - \sum_{m=1}^N P(t, m) \cdot \log_2 P(t, m) \quad (22)$$

2.12 Transformada Wavelet

Como dito, sinais de ECG podem ser vistos como sendo não estacionários por natureza (Kaplan Berkaya *et al.*, 2018; Li, 2019; Sharma *et al.*, 2019c). Uma conversão dos sinais no tempo para o domínio das frequências baseada na transformada de Fourier, não é a mais adequada neste caso já que considera o sinal como um todo, negligenciando a noção de localidade e, portanto, a característica de variação espectral ao longo do tempo. Mesmo com a utilização da transformada de Fourier de curto-termo, que analisa o sinal por pequenos trechos, ainda aparecem dificuldades relacionadas com a inflexibilidade entre as resoluções de tempo e frequência da ferramenta. A transformada Wavelet, por outro lado, se mostra superior neste tipo de análise tempo-frequência, especialmente de sinais não estacionários, pois permite controlar a noção de localidade tanto no tempo como na frequência (Bruna and Mallat, 2013; Li, 2019; Nair and Paul Joseph, 2014; Sharma *et al.*, 2019c).

A transformada Wavelet pode ser definida como a convolução de um sinal, $x(t)$, com um núcleo de formato específico, $\psi_{a,\tau}(t)$, denominado Wavelet, limitado no tempo (Bruns, 2004). O núcleo, é uma representação baseada numa função modelo chamada Wavelet mãe $\psi(t)$, que por sua vez tem um comportamento característico de acordo com sua família. O núcleo pode variar a função modelo com relação a sua escala e localização. A Equação 23 descreve a relação de dilatação/compressão (a) e deslocamento (τ) entre a Wavelet mãe e o núcleo, sendo que o núcleo pode ser a própria Wavelet mãe (Li, 2019; Nair and Paul Joseph, 2014).

$$\psi_{a,\tau}(t) = \frac{1}{|a|^{\frac{1}{2}}} \cdot \psi\left(\frac{t-\tau}{a}\right) \quad (23)$$

No caso deste trabalho utilizou-se a função Wavelet mãe da família Morlet, descrita pela Equação 24, onde α é a amplitude, $\beta \ll 1$ é ajustado para que $\int \psi(t) dt = 0$ (ortogonalidade), $\sigma = 0,85$ é o desvio padrão do envelope Gaussiano e f é a frequência associada (Bruna and Mallat, 2013; Bruns, 2004).

$$\psi_f(t) = \alpha(e^{itf} - \beta) \cdot e^{(-|t|^2/2\sigma^2)} \quad (24)$$

A transformada Wavelet contínua (CWT) de um sinal $x(t)$, é então definida pela Equação 25. A transformada Wavelet discreta (DWT) por sua vez pode ser obtida a partir da discretização da função $\psi_{a,\tau}(t)$, Equação 26, onde $a = b^j$ para $j \in \mathbb{Z}$ e $b > 0$, $\tau = b^j kT$ para $k \in \mathbb{Z}$ e $T > 0$. Sendo j a escala de frequência, k a escala de tempo e T uma constante relacionada à Wavelet mãe (Li, 2019; Nair and Paul Joseph, 2014).

$$W(a, \tau) = x(t) * \psi_{a,\tau}(t) = \frac{1}{|a|^{\frac{1}{2}}} \int_{-\infty}^{+\infty} x(t) \overline{\psi\left(\frac{t-\tau}{a}\right)} dt \quad (25)$$

Uma análise em múltiplas escalas, pela DWT, permite que o sinal seja decomposto em várias bandas de frequência diferentes. O efeito da decomposição é equivalente à aplicação do sinal sobre um banco de filtros passa faixa com diferentes frequências centrais. Comumente, esta representação é na forma cascata, onde o sinal a cada nível, é submetido a uma decomposição por filtragem passa baixo e passa alto, e em seguida subamostrado. Os coeficientes, após as filtragens passa alto, são chamados de coeficientes de detalhe, enquanto os das filtragens passa baixo são chamados de coeficientes de aproximação, que por sua vez são submetidos a uma nova etapa de decomposição caso não se tenha alcançado o nível desejado (Kaplan Berkaya *et al.*, 2018; Li, 2019; Yildirim, 2018).

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (26)$$

2.12.1 Wavelet Scattering

Uma das maiores dificuldades encontradas nas tarefas de diferenciação dos tipos de comportamento do ECG é a variabilidade dos sinais, seja entre sujeitos ou entre os padrões das anomalias. Essas variabilidades podem, sem grande rigor, ser comparadas a deformações morfológicas e temporais vistas no ECG. Além disso, a interferência causada pela ação de ruídos, artefatos e desvios também caracteriza um tipo de variabilidade.

Portanto, ao mesmo tempo que se deseja transformar as representações de modo a tornar os sinais invariantes a pequenas mudanças (como o ruído), as caracterizações dos padrões que distinguem as classes devem ser mantidas. Em suma, o que se deseja é que pequenas variações, quando transformadas, sejam mantidas pequenas em comparação com grandes variações transformadas. Entretanto, operadores de transformação invariantes a translação (por

exemplo), como a autocorrelação e a transformada de Fourier, são instáveis para alguns tipos de deformação. No caso da transformada de Fourier, quanto mais rápida é a variação do sinal de entrada, mais instável a representação se torna nas altas frequências.

Uma representação tempo-frequência desejada para os sinais que são suscetíveis a tais variações, é aquela que, tendo uma referência livre de deformações, $x(t)$, sua versão com algum tipo de deformação, $x_c(t)$, pode ser relacionada linearmente através de um operador, H , estável e Lipschitz contínuo às deformações.

Como alternativa às ondas senosoidais da transformada de Fourier, pode-se utilizar a transformada Wavelet, cuja análise é realizada pelas versões escaladas da função Wavelet mãe. Por conta da localidade das ondas Wavelet, a transformação é estável a deformações. Contudo, é covariante a translações por causa das operações de convolução. A transformada Scattering resolve este problema associando aos coeficientes da transformação Wavelet, uma não linearidade (função de módulo) e uma função de média. Neste caso, a escolha do operador de módulo é para preservar a energia do sinal, quando este é complexo. Além disso, a operação de média é baseada numa Gaussiana escalável, uma vez que a Wavelet mãe é da família Morlet.

O módulo da convolução entre a função Wavelet e o sinal (transformada Wavelet) é apresentado pela Equação 27, onde λ representa a localidade do filtro passa faixa, consequente da escala da função Wavelet. A partir de uma série de decomposições sucessivas, organizadas por um caminho $p = (\lambda_1, \lambda_2, \dots, \lambda_m)$, obtém-se a relação de cascata da Equação 28.

$$U[\lambda]x = |x * \psi_\lambda| \quad (27)$$

$$U[p]x = \left| \left| |x * \psi_{\lambda_1}| * \psi_{\lambda_2} \right| \dots * \psi_{\lambda_m} \right| \quad (28)$$

A transformada Scattering é então definida pela Equação 29, onde a função de média Gaussiana é aplicada sobre o resultado da Equação 28. A variável u representa a posição da janela de análise da Scattering, $S[p]x(u)$, sobre o sinal, a qual pode ser subamostrada até o tamanho 2^J . Supondo uma deformação $x_c(u) = x(u - c)$, a aplicação da média à escala 2^J , faz com que, deslocamentos $|c| \ll 2^J$, resultem na aproximação $S[p]x \approx S[p]x_c$. A Figura 5 exemplifica o processo de decomposição Scattering para até três níveis de profundidade.

$$S[p]x(u) = \left| \left| |x * \psi_{\lambda_1}| * \psi_{\lambda_2} \right| \dots * \psi_{\lambda_m} \right| * \phi_{2^J}(u) \quad (29)$$

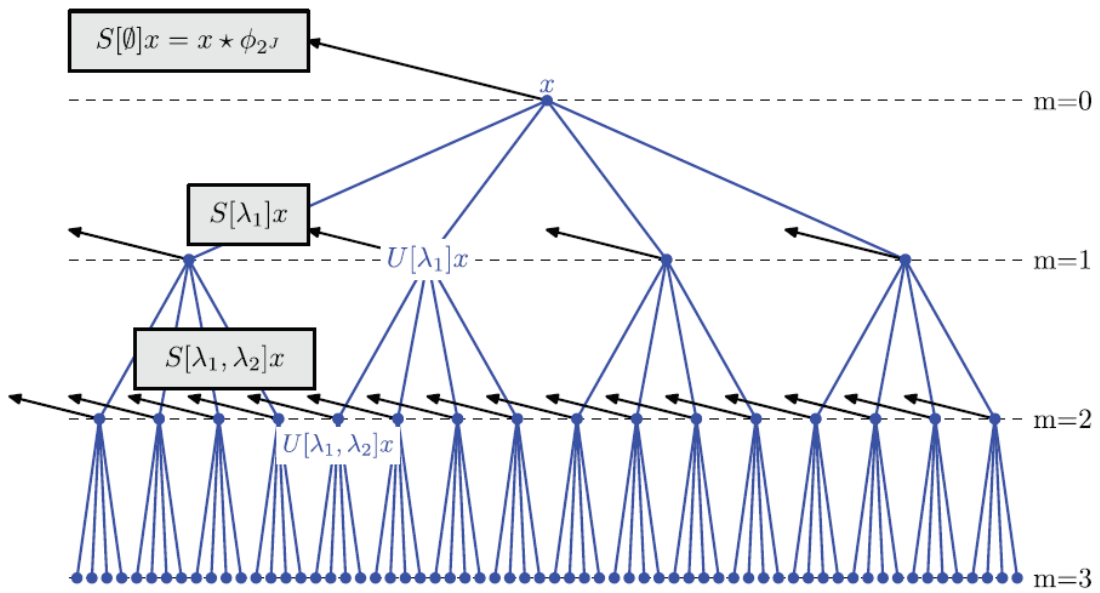


Figura 5 - Exemplo da decomposição pela transformada Scattering. Fonte: Bruna and Mallat, 2013

2.13 Algoritmos de Inteligência Artificial

Os algoritmos de inteligência artificial utilizados neste trabalho foram baseados nas redes neurais artificiais. Estas redes, são modelos de aprendizagem que a partir de treinamento apropriado, são capazes de executar funções de controle, previsão, classificação, entre outras.

Inspiradas na estrutura do neurónio humano, as redes possuem como unidade fundamental, o neurónio artificial. Neste, as informações presentes na entrada, chamadas de estímulos, são ponderadas de acordo com os chamados pesos sinápticos e somadas em conjunto com o limiar de ativação. O resultado, um escalar chamado de potencial de ativação, é aplicado a uma função de ativação, na maioria das vezes não linear, para produzir a saída. A Equação 30 apresenta o processo descrito para um único neurónio, onde \mathbf{x} é o vetor de entrada, \mathbf{w} é o vetor de pesos sinápticos, b é o limiar de ativação, φ é a função de ativação e y a saída.

$$y = \varphi(\mathbf{x} \cdot \mathbf{w} + b) \quad (30)$$

Este conjunto de operações transforma a informação dos dados da entrada para um domínio de representação de mais alta dimensão, chegando a um nível de abstração da informação. À medida que mais neurónios são associados, a capacidade de abstração da

informação da entrada é elevada e ao mesmo tempo a representação da saída passa a estar em função de múltiplas variáveis, dependentes dos parâmetros da rede. Esta nova representação é utilizada para fins de previsão, controle e classificação. No caso deste trabalho, a saída do modelo é utilizada para classificar a informação da entrada em um comportamento de AF ou não AF.

Para diferentes aplicações e de acordo com a complexidade do problema, tem-se diferentes modelos de neurónios e arranjos de neurónios. Neste trabalho, foram abordadas as redes do tipo *Feed Forward* e *Long Short-Term Memory*. A primeira estrutura é caracterizada pelo sentido de propagação da informação ser da entrada para a saída, não havendo desvios ou realimentações. Além disso, é utilizado como unidade o Perceptron. A segunda é uma estrutura baseada na análise de contexto temporal da entrada e possui não só o recurso de realimentação como também os de memória e estado de célula. A unidade de processamento é formada por estruturas chamadas portas (*gates*) que analisam as amostras da entrada para determinadas funções de tomadas de decisão, como esquecimento e atualização de informação e a geração da saída. Com isso, este modelo consegue reter as informações relevantes e descartar aquelas com baixa importância de acordo com o propósito do problema.

2.13.1 *Feed Forward Network*

A principal característica de redes neuronais do tipo *Feed Forward* é o sentido de propagação da informação, sendo este diretamente da entrada para a saída (Chauhan *et al.*, 2019). Basicamente, os valores apresentados na entrada da rede são ponderados pelos pesos sinápticos de cada neurónio da primeira camada, somados e aplicados a uma determinada função de ativação. O processo se repete para a próxima camada que utiliza como entrada, as saídas dos neurónios da camada anterior. Assim, todo o conteúdo da entrada percorre as camadas da rede diretamente e ao mesmo tempo, sem retornos ou desvios de fluxo (Alickovic and Subasi, 2015).

A unidade básica mais comum para as redes deste tipo é o Perceptron (Figura 6). Já descrito pela Equação 30, este tipo de neurónio foi proposto por (McCulloch and Pitts, 1943) sendo aplicado mais tarde por (Rosenblatt, 1958). Ao longo dos anos, foram sendo experimentados diversos arranjos baseados nas variações do número de camadas, números de neurónios, modos de interconexão e funções de ativação dos neurónios para as redes. A rede Perceptron Multi Camada (MLP) é um tipo destes arranjos onde todos os neurónios de uma

camada são conectados com a camada seguinte e assim sucessivamente até a camada de saída (Sahoo *et al.*, 2017).

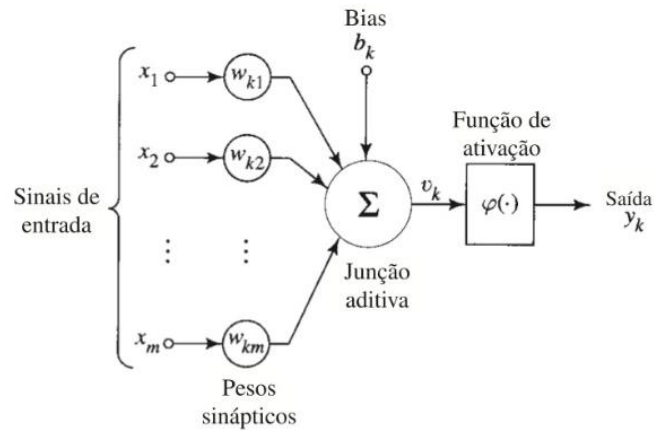


Figura 6 - Modelo fundamental do Perceptron.

Porém, independentemente do tipo de estrutura, as redes neurais executam determinada função após um treinamento. Este treinamento pode ser do tipo supervisionado ou não supervisionado, este quando a rede busca a diferenciar os comportamentos de entrada apenas com os exemplos apresentados. O treino supervisionado ocorre quando para cada exemplo apresentado na entrada, existe uma referência esperada na saída. A aprendizagem neste caso, acontece pela atualização dos pesos sinápticos de cada neurônio pela propagação, da saída para a entrada (*backpropagation*), da diferença entre o comportamento esperado e o obtido na saída da rede. Esta diferença, denominada erro, nos casos dos algoritmos de treino Levenberg-Marquardt e *Resilient Backpropagation*, é aplicada aos pesos sinápticos no sentido derivativo parcial a cada neurônio e a cada camada. Assim, durante o treinamento, espera-se que a rede busque reduzir o erro na saída de acordo com alguma função custo (Haykin, 2007; Sahoo *et al.*, 2017). Tipicamente, esta função é o erro quadrático médio (MSE), o qual foi utilizado neste trabalho e é descrito matematicamente pela Equação 31, onde y é o valor esperado, \hat{y} é a saída da rede e n é o número de observações.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (31)$$

Por conta de os algoritmos de treino estarem relacionados com a noção de derivada da Equação 30, é necessário que a função de ativação dos neurônios seja diferenciável para todo

domínio real. Dessa forma, neste trabalho foram utilizadas as funções de ativação tangente sigmoide (Tan-sig) e logística sigmoide (Log-sig), as quais são apresentadas pela Figura 7.

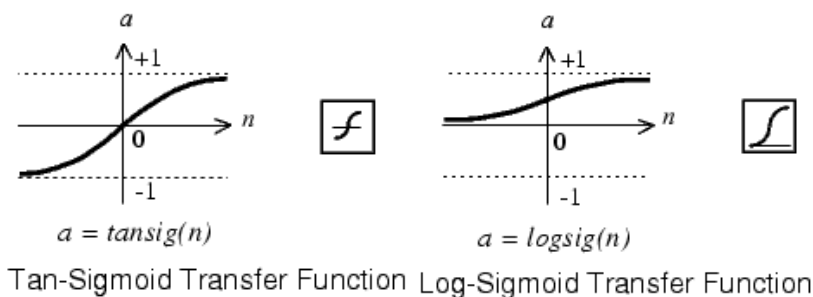


Figura 7 - Funções de ativação não lineares do tipo: (a) Tangente Sigmoide; (b) Logística Sigmoide.

Para se determinar a melhor combinação de parâmetros para o modelo baseado na MLP, a abordagem mais comum é a exploração extensiva das combinações por comparação do nível de acurácia de cada uma.

2.13.2 Long Short-Term Memory

Quando o recurso da realimentação está presente no modelo da rede neuronal, essa passa a ser chamada rede neuronal recorrente (RNN). A realimentação faz com que as entradas dos neurónios não são dependentes apenas das amostras de entrada instantânea, mas também das saídas passadas. Além disso, redes do tipo *feed forward* assumem que as amostras de entrada são independentes entre si, sendo que para séries temporais e sequências, essa suposição não é totalmente válida (Andersen *et al.*, 2019). Dessa forma, é razoável notar que essa estrutura é mais adequada do que a rede neuronal *feed forward* para analisar séries temporais, sequências ou amostras que tenham dependência temporal, uma vez que possui o artifício de levar em consideração essas tais dependências ao longo do tempo além da entrada atual (Dang *et al.*, 2019; Faust *et al.*, 2018b; Yildirim, 2018). Sinais como os de ECG, possuem essa característica de dependência temporal e portanto, sua classificação pelo uso de RNNs pode ser favorecida (Andersen *et al.*, 2019).

Os modelos mais simples de RNNs, chamados aqui de convencionais (Figura 8), apresentam sérios problemas de instabilidade do gradiente durante o treinamento, podendo chegar ao ponto de seu desaparecimento ou explosão (Faust *et al.*, 2018b; Saadatnejad *et al.*, 2019; Yildirim, 2018).

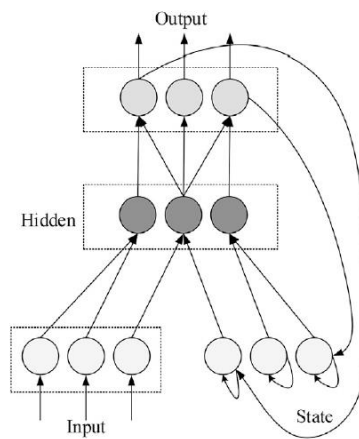


Figura 8 - Estrutura básica RNN; Fonte: Yildirim, 2018

Uma arquitetura que substitui a unidade básica do Perceptron por outra mais complexa é a *Long Short-Term Memory* (LSTM) (Yildirim, 2018). A LSTM utiliza dois artifícios principais que a torna fortemente superior na análise de séries temporais comparativamente ao modelo convencional. Primeiro, os neurónios possuem um conjunto de portas que analisam as amostras instantâneas e utilizam o contexto interno de aprendizagem da rede para atualizar o próprio contexto e gerar a saída, levando em consideração o que reter, inserir ou esquecer das novas amostras (Andersen *et al.*, 2019; Dang *et al.*, 2019; Faust *et al.*, 2018b; Yildirim *et al.*, 2019). Segundo, é inserido o conceito de estado de célula, onde a informação flui entre os neurónios sem ser submetida a funções de ativação, apenas sendo atualizada pelas portas (Andersen *et al.*, 2019; Dang *et al.*, 2019). Com isso, evita-se o problema do desaparecimento do gradiente das estruturas convencionais ao mesmo tempo que é potencializada a capacidade de extrair informações ao longo do tempo (Yildirim, 2018). A composição de um neurônio LSTM é apresentada pela Figura 9.

Cada uma das portas é composta por um neurônio do tipo Perceptron com função de ativação sigmoide. A quantidade de entradas destes neurónios é igual para todos e referente à soma do número de blocos LSTM e do número de amostras do sinal (Yildirim, 2018). As portas de um bloco LSTM são: esquecimento, entrada e saída, descritos pelas Equações 32 a 34, onde os vetores \mathbf{u} , \mathbf{w} são os pesos sinápticos associados respectivamente as entradas do sinal \mathbf{x} e do estado da célula \mathbf{h} .

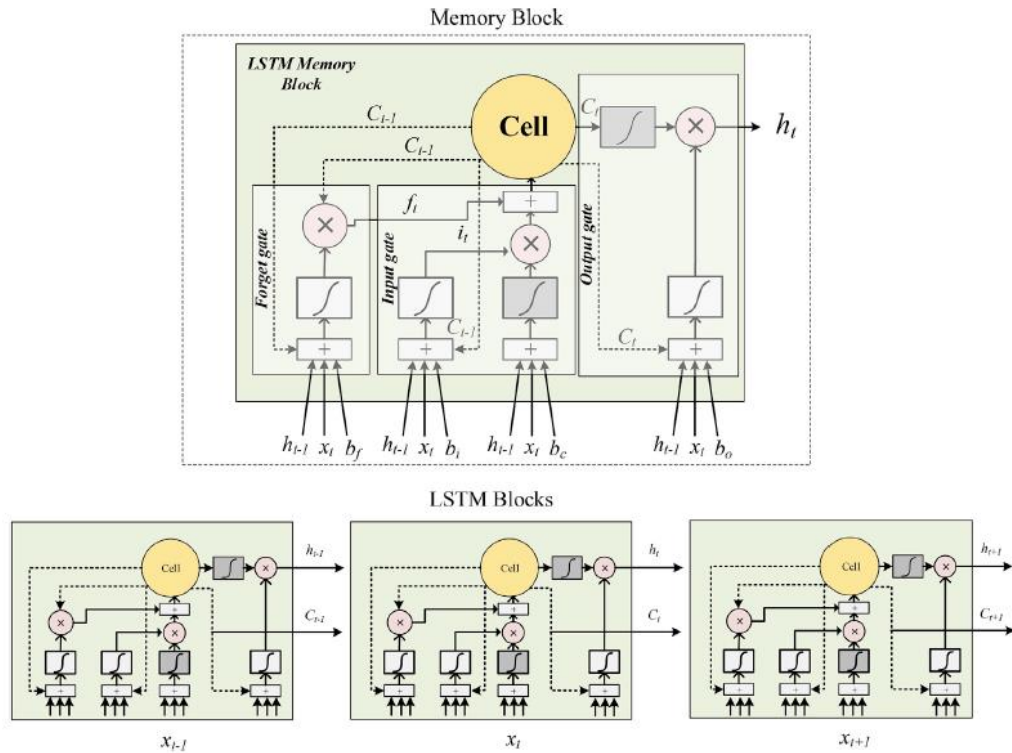


Figura 9 - Unidade neuronal LSTM e organização básica de uma camada LSTM. Fonte: Yildirim, 2018

Inicialmente, a porta de esquecimento verifica qual informação deve ser esquecida e com que intensidade, aplicando à saída de sua porta, f_t , o estado da célula. Em seguida, a porta de entrada, i_t , determina as informações mais relevantes na entrada atual e atualiza o estado da célula após a combinação entre a saída da porta e o contexto da informação atual. A porta de saída, o_t , determina a informação que vai compor a saída do bloco a partir da combinação entre o contexto de entrada e o estado da célula totalmente atualizado (Andersen *et al.*, 2019; Dang *et al.*, 2019; Yildirim, 2018).

$$f_t = \varphi(b_f + \mathbf{u}_f^T \mathbf{x}_t + \mathbf{w}_f^T \mathbf{h}_{t-1}) \quad (32)$$

$$i_t = \varphi(b_i + \mathbf{u}_i^T \mathbf{x}_t + \mathbf{w}_i^T \mathbf{h}_{t-1}) \quad (33)$$

$$o_t = \varphi(b_o + \mathbf{u}_o^T \mathbf{x}_t + \mathbf{w}_o^T \mathbf{h}_{t-1}) \quad (34)$$

A formação do estado da célula, h_t , é realizada pela combinação entre a saída da porta de saída, o_t , e a ativação do contexto interno (memória), c_t , pela função tangente hiperbólica. Este estado da célula no bloco atual é introduzido no bloco seguinte para compor sua entrada. Assim, o contexto interno é atualizado a cada amostra e utilizado juntamente com a entrada

para formar o estado da célula, que representa a saída do modelo (Andersen *et al.*, 2019; Yildirim, 2018). As Equações 35 a 37 apresentam as relações para o estado da célula e o contexto interno. Em aplicações de classificação, as saídas dos blocos LSTM podem tanto serem utilizadas como um todo (todas as saídas são utilizadas) ou apenas a saída do último instante de tempo (saída do último bloco).

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (35)$$

$$\tilde{c}_t = \tanh(b_c + \mathbf{u}_c^T \mathbf{x}_t + \mathbf{w}_c^T \mathbf{h}_{t-1}) \quad (36)$$

$$h_t = o_t \tanh(c_t) \quad (37)$$

O modelo de rede LSTM também pode ser utilizado de modo a analisar as dependências temporais da entrada no sentido reverso. Ou seja, uma segunda camada de blocos LSTM é utilizada em paralelo à primeira, na qual o sinal da entrada é introduzido da última amostra para a primeira (Figura 10). Esta aplicação, chamada bidirecionalidade, permite que a rede leve em consideração o contexto da informação tanto no sentido direto quanto no reverso, aumentando a sua capacidade de observar, por exemplo, comportamentos transitórios (Andersen *et al.*, 2019; Dang *et al.*, 2019; Faust *et al.*, 2018b; Yildirim, 2018).

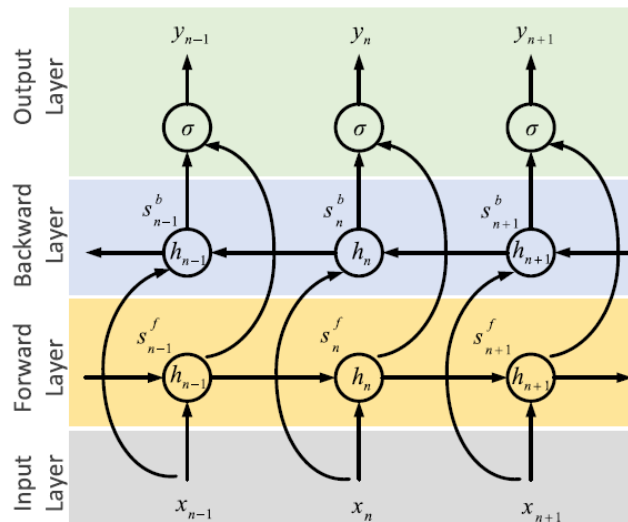


Figura 10 - Composição genérica de um modelo LSTM bidirecional. Fonte: Andersen *et al.*, 2019

A Equação 38 descreve a combinação entre as saídas das duas redes, onde o índice sobrescrito f representa as saídas da rede no sentido direto e o b no sentido reverso. É de se notar que a estrutura neuronal de combinação é do tipo Perceptron com função de ativação σ .

$$y_t = \sigma(w_f s_t^f + w_b s_t^b + b_h) \quad (38)$$

Para treinar as redes LSTM, uma modificação no algoritmo *backpropagation* tradicional é implementada. A modificação, baseada no gradiente descendente estocástico, utiliza a função custo de entropia cruzada binária entre as distribuições de probabilidade do valor esperado e da previsão (Andersen *et al.*, 2019; Faust *et al.*, 2018b).

Capítulo 3: Metodologia

Nesta seção serão apresentadas as metodologias utilizadas neste trabalho para implementação, experimentação e avaliação das propostas. A seção é organizada da seguinte forma.

Primeiro são apresentadas as metodologias de aquisição e organização dos sinais da base de dados *MIT-BIH Atrial Fibrillation* no ambiente Matlab®. Em seguida, a seção é direcionada para metodologia de exploração do modelo, aqui chamado primário, de testes com a rede MLP. Nesta subseção do modelo primário são apresentadas as etapas de segmentação, extração de características e condicionamento dos sinais. Ainda, são apresentadas as etapas de construção dos conjuntos de treino e teste do modelo MLP bem como os critérios de avaliação.

A subseção seguinte explora o desenvolvimento dos modelos secundários. São apresentadas as etapas de detecção de picos R (aplicada devido a dificuldades encontradas durante a exploração do modelo primário), segmentação (agora por dois critérios), extração de características, determinação da transformada Scattering, normalização e organização dos conjuntos de dados e, finalmente, a exploração dos modelos MLP e LSTM. A Tabela 1 apresenta a organização da metodologia usada neste trabalho.

Fundamentalmente, o processo de construção da proposta se iniciou com uma busca extensiva sobre aplicações de sistemas de aprendizagem de máquina em sinais biomédicos, especificamente sinais de ECG. Nesta busca, pretendeu-se perceber a base de dados utilizada, o objetivo, se classificação, previsão ou outras, as patologias envolvidas e os métodos e técnicas utilizadas. Baseado neste estudo (apresentado na seção de Estado da Arte), escolheu-se um conjunto de sistemas para direcionar a construção do modelo deste trabalho, servindo de inspiração e fundamento para adaptação aos objetivos propostos. A partir de uma observação mais específica dos trabalhos selecionados (por exemplo, um olhar sobre as motivações e justificativas das utilizações de certas metodologias e modelos), definiu-se as possíveis associações entre técnicas de diferentes trabalhos e a potencial exploração de outras que não haviam sido testadas. Exemplo desta etapa é a combinação das informações acerca do ritmo cardíaco e das entropias dentro de um segmento (técnicas já exploradas) com a determinação dos parâmetros Jitter e Shimmer para medida de perturbação (proposta inovadora). Além disso, essa combinação de características é inovadora também no sentido de serem extraídas de sinais ECG na forma bruta (sem pré-processamento) para serem aplicadas sobre redes MLP.

Tabela 1 - Organização da metodologia dos modelos primário e secundário.

	<i>Contexto</i>	<i>Características</i>	<i>Normalização</i>	<i>Redes Neurais</i>
<i>Modelo Primário</i>	30 RRI $\mu = 10\%$, 30% e 50%	RRI, Jitter, Shimmer, Entropias Shannon e Energia Logarítmica	Z-score	MLP
<i>Modelo Secundário</i>	60 RRI $\mu = 10\%$	RRI, Jitter, Shimmer, Entropias Shannon e Energia Logarítmica para o segmento e Shannon para ondas P, T e U	Máximo Absoluto	MLP
	60 RRI $\mu = 10\%$	RRI, Jitter, Shimmer, Entropias Shannon e Energia Logarítmica para o segmento e Shannon para ondas P, T e U	Máximo Absoluto	LSTM
	60s $\mu = 10\%$	Intervalo 60s, Frequências Instantâneas, Entropia Espectral e Transformada Scattering	Z-score	LSTM

Portanto, a etapa de análise do estado da arte e a associação de suas perspectivas com técnicas alternativas, são de importâncias extremamente elevadas especialmente para a relevância deste trabalho. Após esta etapa, iniciou-se a fase de implementação da proposta no software Matlab®, primeiramente pela aquisição dos sinais e anotações da base de dados, apresentada na subseção a seguir.

3.1 Aquisição e Organização dos Sinais

Para aquisição dos sinais e respectivas anotações contidas na base de dados *MIT-BIH Atrial Fibrillation*, utilizou-se a *toolbox* WFDB para Matlab® (Silva and Moody, 2014; WFDB Toolbox for MATLAB and Octave). A partir dela, funções de requisição web transferem o conteúdo de cada paciente da base de dados para variáveis no ambiente Matlab®. Para cada paciente, são capturadas as gravações das duas derivações do ECG, a frequência de amostragem da digitalização e as anotações rítmica e de localização dos complexos QRS, sendo estas representadas pelos pares tipo e instante. Das anotações de ritmo, o instante se refere à amostra

onde se iniciou e o tipo se refere à classe do comportamento diagnosticado. Das anotações dos complexos QRS, o instante é uma localização aproximada do pico R e o tipo a classe do batimento.

Os dados foram armazenados numa variável do tipo estrutura que diferencia os dados por paciente e pelo nome da base de dados, possibilitando uma posterior agregação de vários dados de quaisquer bases na mesma estrutura. Para uma manipulação mais ágil e organizada das informações, a estrutura foi salva no formato de arquivo, sendo carregado somente quando necessário.

3.2 Modelo Primário

O modelo primário diz respeito à experimentação da tarefa de classificação de segmentos de ECG pelo uso da rede neuronal MLP. Para compor a estrutura das entradas da rede, utilizou-se a extração de características temporais, morfológicas e estatísticas. Como caracterização temporal, utilizou-se a determinação dos intervalos entre picos R consecutivos e os parâmetros de perturbação Jitter. Como caracterização morfológica, utilizou-se a determinação dos parâmetros Shimmer. Para caracterização estatística, utilizou-se as medidas de informação baseadas nas entropias de Shannon e energia logarítmica. Todos os recursos foram determinados para janelas de sinal bruto entre 30 intervalos R-R consecutivos, sendo este tamanho de intervalo um valor típico em aplicações similares. Procurou-se não aumentar este tamanho uma vez que isto poderia gerar dificuldades excessivas para o modelo generalizar o comportamento do segmento. A limitação principal neste caso é a própria simplicidade do Perceptron em comparação com a capacidade do neurônio LSTM, por exemplo.

3.2.1 Segmentação

A segmentação dos sinais foi realizada baseando-se na diferença em amostras entre picos R consecutivos, possível pelas marcações advindas da base de dados. Ou seja, utilizando as marcações dos picos R, determinou-se um vetor com as diferenças consecutivas dos seus valores. Esse vetor pôde ser convertido para escala temporal pela divisão pela frequência de amostragem.

Em seguida, agrupou-se conjuntos de 30 intervalos R-R consecutivos deste vetor de intervalos para formação dos segmentos, sendo que entre segmentos adjacentes o deslocamento

foi de 10 intervalos R-R. Com isso, entre dois segmentos consecutivos existe a repetição de 20 valores. Entretanto, esta abordagem permite o aumento da quantidade de exemplos comparativamente com o deslocamento do tamanho da janela (zero sobreposições entre segmentos). Este aumento no número de exemplos é valioso para modelos de aprendizagem de máquina já que a insuficiência de dados é um problema recorrente na etapa de treinamento destes modelos.

Além dos segmentos baseados no intervalo entre picos R, usando as mesmas referências, são gerados segmentos com as amplitudes dos respectivos picos R das duas derivações, utilizadas posteriormente no cálculo dos parâmetros Shimmer.

Para cada segmento construído, foi atribuída uma referência de classe de acordo com a proporção de comportamento fibrilação atrial que este contém. A proporção é representada pelo fator μ de intervalo $[0,1]$, sendo que neste trabalho, foram explorados os valores 0,1, 0,3 e 0,5. Um fator $\mu = 0,5$ significa que metade ou mais dos intervalos R-R no segmento de 30, devem estar contidos em períodos de fibrilação atrial para que o segmento receba a classe fibrilação atrial. Isto é verificado através das anotações de ritmo dos sinais, comparando as amostras de início de cada comportamento com os instantes de cada pico R e em seguida calculando a respectiva proporção. No modelo primário, apenas duas classes foram atribuídas aos segmentos: “AF” ou “não AF”, estando em “não AF” compreendidos quaisquer outros comportamentos que não são AF. Numericamente, a referência das classes é feita por representação binária $[0 \ 1]$ para AF e $[1 \ 0]$ para não AF. Assim, a rede MLP, que, portanto, possui dois neurónios de saída, pode levar cada neurônio de saída à especialização de uma única classe.

Para avaliar o balanceamento entre as duas classes, após a segmentação, é determinada a razão entre a quantidade de segmentos de cada classe e o total de segmentos. A prevalência de exemplos de uma das classes pode levar a rede a ter dificuldades em compreender o comportamento da outra classe.

3.2.2 Extração de características

Para cada segmento foram determinados: os quatro parâmetros de Jitter (absoluto, relativo, RAP e PPQ5) utilizando os intervalos R-R como entrada; os quatro parâmetros Shimmer (absoluto, local, APQ3 e APQ5) utilizando as amplitudes dos picos R das duas derivações, separadamente, como entrada (total de 8 parâmetros); as entropias de Shannon e

energia logarítmica utilizando o sinal original, para cada derivação, contido nos limites do segmento (total de 4 medidas).

Com isso, além do próprio vetor com os valores dos 30 intervalos R-R consecutivos, cada segmento passar a conter os recursos de Jitter (4), Shimmer (8) e entropias (4), respectivos ao comportamento dos sinais ECG na sua duração, num total de 46 características.

3.2.2.1 Normalização das características

Cada característica é submetida à normalização z-score da seguinte forma: cada segmento 30 RRi é normalizado isoladamente e cada parâmetro Jitter, Shimmer e entropias, um por vez, são normalizadas considerando todos os segmentos como um todo. Ou seja, o comportamento da variação temporal dentro de cada 30 RRi é mantido específico para si somente, enquanto os comportamentos refletidos pelas outras características levam em consideração todo o universo do sujeito, permanecendo exclusivos a especificidade do sujeito e relativos ao comportamento geral dos sujeitos.

Ao final da normalização, em noção estatística, cada 30 RRi deve possuir, separadamente, média aproximadamente 0 (zero) e desvio padrão 1 (um). Entretanto, apenas entre todos os segmentos de cada sujeito, é que cada parâmetro das outras características possui média aproximadamente zero e desvio padrão 1. Após a normalização, foi realizada a organização das características através da concatenação dos valores, resultando, portanto, em um conjunto de vetores de dimensão 46.

3.2.3 Conjuntos de dados para a aprendizagem de máquina

A preparação dos conjuntos de exemplos para as etapas de treino, validação e teste da MLP levou em consideração não apenas a proporção no número de segmentos, mas também a separação de sujeitos. Com isso, a rede pode ser avaliada, no teste, com relação a sua capacidade de generalizar não apenas dados não vistos no treino, como também sujeitos novos.

Além disso, duas metodologias distintas de formação de conjuntos foram propostas para: A, explorar a variação dos parâmetros estruturais da rede em busca da melhor combinação para a aplicação deste trabalho; e B, explorar diferentes combinações de características e proporção de referenciação dos segmentos, μ , a partir dos melhores modelos obtidos em A.

3.2.3.1 Metodologia A

Esta metodologia para geração dos conjuntos de dados de entrada da MLP é baseada na separação de 15 sujeitos para etapa de treino, 3 para validação e 5 para teste. A cada iteração da exploração dos parâmetros da MLP, os sujeitos que compõem cada um dos conjuntos são sorteados aleatoriamente, sendo que cada sujeito só pertence a um único conjunto por vez. A partir do sorteio dos sujeitos, seus respectivos segmentos são agrupados de acordo com o conjunto em que pertencem, sendo que as etapas de treino e validação utilizam os segmentos também de modo aleatório, não importando a sequencialidade nestes conjuntos.

A metodologia da variação dos parâmetros da MLP seguiu a seguinte ordem:

- Fator de proporção de referenciação fixo $\mu = 0,3$;
- Organização dos conjuntos de treino, validação e teste;
- Repetição do treino e teste de cada combinação de parâmetros por cinco vezes para reduzir o efeito da atribuição aleatório dos pesos sinápticos iniciais e os mínimos locais da função custo;
- Variação das funções de ativação das camadas escondidas e de saída entre tangente sigmoideal e logística sigmoideal;
- Variação do número de neurónios em cada camada escondida de 40 a 195; Variação do número de camadas escondidas de 1 a 4;
- Algoritmo de treinamento *Resilient Backpropagation*.

A avaliação de cada combinação foi baseada na acurácia média produzida a partir da aplicação aleatória de segmentos do conjunto de teste sobre a rede treinada. Foram selecionados os dois melhores modelos, cujas acurácias médias eram as maiores.

3.2.3.2 Metodologia B

A metodologia B, utilizada no modelo primário final, foi baseada no método de validação *10-fold cross-validation*. Este método consiste na geração de 10 diferentes combinações de sujeitos para treino e teste, sendo que cada sujeito pertence a um único conjunto por combinação e pertence ao conjunto de teste uma vez em dez. O desempenho é medido a partir da média entre a acurácia dos dez testes. O conjunto de validação durante o treino da rede

foi definido aleatoriamente a partir de todas as amostras do conjunto de treino, sendo apenas parametrizada pela proporção de 15% deste conjunto.

A partir dos dois melhores modelos obtidos durante a metodologia A, aplicou-se a exploração de diferentes valores para o fator de proporção de referência ($\mu = [0,1; 0,3; 0,5]$) e para cada μ , quatro combinações de características de formação dos conjuntos de entrada da MLP. As quatro combinações geradas são as que se seguem, onde os índices subscritos representam a derivação do ECG a que pertencem:

- Comp_01: 30 RRI;
- Comp_02: Comp_01 + J1 + J2 + J3 + J4;
- Comp_03: Comp_02 + S1₁ + S1₂ + S2₁ + S2₂ + S3₁ + S3₂ + S4₁ + S4₂;
- Comp_04: Comp_03 + E_{sh1} + E_{sh2} + E_{logen1} + E_{logen2}.

3.3 Modelo Secundário

O desenvolvimento do modelo secundário surgiu do interesse em extrair e agregar características morfológicas das ondas P, T e U às características já existentes do modelo primário. A representação destas ondas pode ser útil na descrição do comportamento cardíaco, especialmente para detecção de anomalias. Por exemplo, a partir da deformação, atenuação ou desaparecimento da onda P, pode-se ter fortes indícios da ocorrência de fibrilação atrial.

Entretanto, notou-se uma imprecisão nas marcações dos picos R advindas da base de dados utilizada. Portanto, propôs-se a implementação de um novo algoritmo para remarcação dos picos R dos sinais. A partir da nova marcação, todo o conjunto de características anteriormente utilizado precisou ser atualizado.

Com a nova composição do conjunto de características, explorou-se novamente o desempenho da MLP. Além disso, explora-se no modelo secundário a utilização da LSTM, uni e bidirecional, em conjunto com a etapa de transformação Scattering.

3.3.1 Algoritmo de detecção de pico das ondas R

O algoritmo proposto foi inspirado e fundamentado nos trabalhos de (Benitez *et al.*, 2001; Kaur *et al.*, 2019; Manikandan and Soman, 2012).

Inicialmente, aplicou-se uma operação de filtragem dos sinais de ECG da derivação I com um filtro FIR (*Finite Impulse Response*) passa banda, com frequências de passagem entre 5 Hz e 15 Hz, de ordem 240 e projetado pelo método da janela a partir da função janela de Hamming de ordem 60 autoconvoluída. A operação de autoconvolução é descrita pela Equação 39, em que w_1 pode ser considerada a janela de Hamming e w_2 a janela resultante da autoconvolução por ordem p . Esta, definida como a quantidade de convoluções somada a um. A janela w_2 possui, portanto, ordem $M = p \cdot L$, sendo L a ordem da janela w_1 .

$$w_2 = (w_1 * w_1) * w_1 * \dots \quad (39)$$

O benefício da utilização desta técnica para projetos de filtros FIR é que se obtém bordas de transição mais rápidas, atenuação nas bandas de rejeição maior e com menores oscilações quando comparadas ao projeto do mesmo filtro com a janela original, mas com ordem aumentada até a ordem da janela autocovoluída. Isso significa que a filtragem passa banda é mais seletiva e tem menor interferência das bandas laterais, potencializando um resultado mais promissor. No caso, o objetivo da filtragem é somente de reduzir fortemente os artefatos dos sinais e ressaltar os complexos QRS.

O custo computacional das operações de convolução pode ser negligenciado devido ao projeto do filtro ser feito apenas uma vez, ou seja, numa aplicação em tempo real por exemplo, os coeficientes do filtro não seriam gerados a cada operação de filtragem e sim previamente armazenados. Inclusive, neste trabalho são utilizados os coeficientes da resposta em frequência do filtro, devido a agilidade do processo de filtragem quando realizado no domínio das frequências.

A partir dos sinais filtrados, são determinadas as suas derivadas, $d(n)$, de acordo com a diferença entre amostras consecutivas (Equação 40). O resultado é então submetido a uma operação de normalização pelo valor máximo absoluto de cada sinal separadamente (Equação 18). Em seguida aplica-se a transformação não linear da Equação 8, energia de Shannon, sobre os sinais normalizados para se obter um sinal com os complexos QRS destacados e com baixa variância entre si, mesmo se possuem grande variação em amplitude.

$$d(n) = x(n + 1) - x(n) \quad (40)$$

Para aplicar um alisamento sobre o resultado da energia de Shannon e atenuar as oscilações de alta frequência, utiliza-se um filtro de média móvel de janela $L_{m1} = 152,7$ ms. O sinal, agora suavizado, é aplicado na transformada de Hilbert, Equação 1, para obtenção da respectiva função de simetria ímpar, que por sua vez indica uma alta probabilidade de ocorrência do pico R na região de cruzamento de zero positivo. Entretanto, para reduzir os deslocamentos de baixa frequência da função ímpar, sobre ela subtrai-se a sua versão filtrada com média móvel retangular de duração $L_{m2} = 2,5$ s. Agora, supostamente os pontos de cruzamento de zero positivos estarão mais evidentes e uma etapa de busca pelo máximo local no sinal original a partir destes pontos é implementada para detecção do pico R real.

A avaliação da efetividade do modelo foi realizada pela verificação da existência ou não de uma marcação original (da base de dados) numa região de contexto de 150 ms centrada no novo pico R.

Além disso, desenvolveu-se um mecanismo para se evitar períodos de silêncio dos sinais. O método é baseado no intervalo entre um pico R e o seguinte. Se a duração for maior do que um limite empírico de 3s, os picos são considerados desconectados. Como resultado desta etapa tem-se a numeração crescente dos picos de acordo com a sua conectividade: números iguais em picos adjacentes significa que os mesmos estão “conectados”, ou seja, o intervalo entre eles é menor do que 3s; caso a numeração seja diferente, significa que estão distantes a mais de 3s, portanto, considerados desconectados. Assim, picos desconectados são evitados de estarem no mesmo segmento, reduzindo a interferência de períodos de silêncio na determinação dos Jitter, por exemplo.

3.3.2 Segmentação

Assim como a etapa de segmentação do modelo primário, a segmentação do modelo secundário utilizou os intervalos entre picos R consecutivos como referência. Entretanto, agora utiliza-se como amostras de referência as remarcações geradas na subseção anterior. Além disso, o contexto de cada segmento passa a ser 60 RRi uma vez que no modelo secundário utilizou-se as redes do tipo LSTM, que possuem grande capacidade de análise temporal de longo termo. Ainda, pretendeu-se experimentar o desempenho da MLP para um contexto de conjunto mais extenso comparativamente ao modelo primário.

Para posterior cálculo dos parâmetros Shimmer, sobre as mesmas marcações de picos R utilizadas para determinar os intervalos R-R do segmento, determinou-se um vetor de amplitudes para cada uma das derivações.

Para que fosse posteriormente determinado o mapa de características da transformada Scattering, segmentaram-se os sinais das duas derivações em intervalos de tempo fixo de 60s. Essa janela foi determinada como uma aproximação grosseira da duração de 60 RRi. Portanto, o modelo secundário passa a ter dois conjuntos de segmentos paralelos, um baseado nos 60 intervalos R-R e outro baseado na duração fixa de 60s. Quanto ao deslocamento entre segmentos adjacentes, este foi mantido como sendo 10 RRi para o conjunto baseado nos RRi e 10s para o conjunto baseado nos 60s.

A atribuição das referências de classe para cada segmento, passa aqui a ser dividida em três, podendo um segmento ser considerado um comportamento de fibrilação atrial, ritmo normal ou qualquer outro. Em noção de sintaxe, a representação da referência para o conjunto da MLP foi feita em notação binária onde apenas um dígito pode estar ativo ([0 0 1]; [0 1 0]; [1 0 0]), semelhante à abordagem do modelo primário. Para o conjunto alvo (comportamento esperado) da LSTM, a representação é em forma de categoria ('1', '2', '3', por exemplo).

Com relação ao fator de proporção para atribuição das classes, utilizou-se $\mu = 0,1$ devido ao melhor desempenho observado na metodologia B do modelo primário. Aqui, o fator μ significa que:

- Se mais de 10% dos 60 RRi (ou 60s) pertencerem a um comportamento de AF (baseado na anotação de ritmo) o segmento é considerado da classe AF;
- Se não, se mais de 50% do segmento pertence a qualquer outro ritmo que não seja normal, o segmento é considerado NDA (nenhuma das anteriores);
- Se as condições anteriores não foram satisfeitas, estima-se que o comportamento seja do tipo Normal.

3.3.3 Extração de características

As características extraídas dos segmentos dependem do conjunto de análise. Aqui, para simplicidade de nomenclatura, nomeia-se “cRR” o conjunto de segmentos gerado a partir dos 60 intervalos R-R, e “cSeg” o conjunto gerado a partir da janela fixa de 60 segundos.

Do conjunto cRR, para cada segmento são extraídas as características: 4 parâmetros Jitter, baseando-se em cada vetor de 60 intervalos R-R consecutivos; 4 parâmetros Shimmer, baseando-se no vetor de 61 picos R, da primeira derivação, de cada segmento; 4 medidas de entropias, baseadas nos sinais das duas derivações, contidos nos respectivos limites de cada segmento; 3 vetores com 60 medidas cada, da entropia de Shannon das ondas P, T e U para cada um dos 60 intervalos do segmento. Antes do cálculo das entropias das ondas, as regiões em análise são interpoladas por um fator $r = 2$ para aumentar a resolução temporal nas regiões das ondas e aumentar a precisão na identificação dos limites destas ondas.

A determinação dos locais das ondas P, T e U de cada intervalo, foi realizada por inspeção visual de muitas janelas R-R sobrepostas (aproximadamente 1000 por vez) numa aproximação percentual das amostras de início e fim dos eventos. A aproximação percentual é aplicada para tornar os parâmetros independentes da duração dos intervalos. Os pontos de início e término das ondas foram definidos, percentualmente, para qualquer janela R-R, como apresentado pela Tabela 2.

Tabela 2 – Localização percentual dos pontos de início e final das ondas T, U e P dentro de um intervalo R-R.

	<i>Início [%]</i>	<i>Final [%]</i>
<i>Onda T</i>	15,00	50,00
<i>Onda U</i>	45,00	65,00
<i>Onda P</i>	70,00	92,00

Do conjunto cSeg, foram extraídas sobre as janelas das duas derivações: as frequências instantâneas, a entropia espectral e o mapa de características pela transformada Scattering. Nota-se que todas pertencem a um contexto tempo-frequência e não podem ser diretamente associadas com as características do conjunto cRR pois sua duração no tempo, para cada segmento, é variável.

Os parâmetros utilizados para determinação do mapa de características pela transformada Scattering foram:

- Tamanho do sinal analisado: 15000 amostras (ou 60s a 250 Hz);
- Frequência de amostragem: 250 Hz;
- Escala de invariância: 500ms;
- Fatores de qualidade dos bancos de filtros primário e secundário: [8; 1];

O fator escala de invariância define o comprimento máximo, no tempo, que a função wavelet pode assumir, ou seja, na frequência, a largura de banda mínima associada para análise “passa-faixa” é de 2 Hz.

A determinação das frequências instantâneas e da entropia espectral foi realizada com a utilização dos sinais das duas derivações, separadamente, de cada segmento e a frequência de amostragem.

3.3.3.1 Normalização de características

A normalização dos dados do conjunto cRR, no modelo secundário, se dá de maneira diferente do que no modelo primário. Sobre o vetor de 60 RRi de cada segmento, foi aplicada a normalização, isoladamente, pelo valor absoluto máximo, mantendo apenas o contexto local intacto e negligenciando-se o contexto global dos demais segmentos.

Sobre o Jitter absoluto e o Shimmer absoluto, separadamente, aplicou-se a normalização pelo valor absoluto máximo entre todos os segmentos do mesmo sujeito, mantendo as especificidades do sujeito e tornando estes parâmetros relativos ao contexto global. Os outros três parâmetros de Jitter e de Shimmer foram apenas divididos por cem, transformando a representação percentual em decimal.

Sobre os valores de entropia do sinal todo, de cada segmento, a normalização pelo valor absoluto máximo foi aplicada apenas no contexto local do segmento, mas para os quatro valores ao mesmo tempo. Com isso, assim como os vetores de 60 RRi, estas entropias foram mantidas para refletir o comportamento de informação local do segmento.

Sobre os vetores de entropias das ondas T, U e P, separadamente, a normalização pelo valor absoluto máximo foi aplicada sobre todos os valores dentro do mesmo sujeito. Ou seja, as noções de informação acerca destas ondas se tornaram relativas ao contexto geral do comportamento do paciente, respectivamente a cada onda.

A normalização das características do conjunto cSeg foi realizada por meio de normalização Z-score sendo esta aplicada a cada segmento e característica separadamente. Assim, o contexto de cada característica de cada segmento se tornou padronizado somente ao comportamento dentro dos 60 segundos da janela.

3.3.4 Conjuntos de dados e experimentação - redes MLP

Para a investigação dos modelos baseados na MLP foram utilizados apenas os dados do conjunto cRR. A princípio, para uma avaliação abrangente dos parâmetros estruturais da rede, utilizou-se a concatenação de todas as características extraídas do conjunto cRR, ou seja, utilização total dos recursos. Nesta condição, os exemplos possuem ao todo, 252 dimensões, sendo um aumento notável relativamente aos do modelo primário (46):

- 60 intervalos R-R (RRi);
- 60 entropias das ondas P;
- 60 entropias das ondas T;
- 60 entropias das ondas U;
- Jitter absoluto;
- Jitter relativo + Jitter RAP + Jitter PPQ5;
- Shimmer absoluto;
- Shimmer local + Shimmer APQ3 + Shimmer APQ5;
- 2 entropias de Shannon;
- 2 entropias da energia logarítmica.

A geração do conjunto de treino foi baseada na seleção aleatória de 85,00% dos segmentos de cada sujeito para o conjunto cRR. Dentre os dados separados para o treinamento, 15,00% são aleatoriamente selecionados pelo algoritmo para validação durante o treino (efetivamente 12,75% do total de segmentos). Para o teste, são utilizados os 15,00% totais restantes, os quais não foram vistos durante o treinamento, mas pertencem a todos os sujeitos na mesma proporção.

Os parâmetros estruturais da MLP experimentados nesta abordagem foram os seguintes:

- Número de neurónios em cada camada escondida: 15 até 205 em passos de 10;
- Número de camadas escondidas: 1 até 3;
- Funções de ativação para as camadas escondidas e de saída: Sigmoides Tangente e Sigmoides Logísticas;
- Algoritmo de treino: *Resilient Backpropagation*;

Para avaliação de cada combinação de parâmetros, utilizou-se a acurácia/exatidão de classificação como critério. Cada combinação foi executada três vezes, sendo o melhor desempenho em acurácia sobre o conjunto de teste selecionado para representá-la. Esta abordagem procura reduzir o efeito de divergências devido a inicialização aleatória dos pesos sinápticos, aumentar a fuga de treinamentos levados a mínimos locais e proporcionar diversidade na seleção de segmentos nos conjuntos.

A partir dos 5 melhores modelos, foram experimentadas diferentes combinações de características para compor os conjuntos de entrada. Estas combinações foram geradas pela retirada uma a uma das características do fim para o começo da lista apresentada. Com isso, contando a execução com todas as características, obteve-se 10 combinações. A ordem de retirada das características foi assim determinada devido à proximidade de tal recursos com as características mais relevantes de detecção da AF. Por exemplo, as medidas de entropia de Shannon e energia logarítmica sobre o sinal do segmento todo são pouco sensíveis à localidade de variações, ou seja, caso aconteça um desaparecimento de uma das ondas P, seu impacto no valor final da entropia pode ser compensado por um possível complexo QRS com mais energia. Ainda, o Shimmer, por estar utilizando a amplitude dos picos R como parâmetro de referência, pode descrever com menor intensidade um período de AF do que a medida de perturbação temporal do Jitter. Portanto, de acordo com esta noção vinda do estado da arte, o vetor RR_i e as entropias das ondas são consideradas as mais relevantes e por isso estão presentes em quase todas as combinações.

Assim como na etapa de procura dos parâmetros estruturais da rede, nesta etapa utilizou-se a acurácia de classificação, sobre o conjunto de teste, como critério de desempenho. Além disso, cada combinação também foi executada, do início, três vezes para garantir diversidade na seleção dos segmentos e a fuga de possíveis comportamentos indesejados na etapa de treino.

3.3.5 Conjuntos de dados e experimentação - redes LSTM

A geração dos conjuntos de dados de entrada para experimentação da LSTM foi realizada tanto para o conjunto cRR quanto para o cSeg. Primeiramente em relação ao cRR, foram adotadas as seguintes combinações de características:

- Todos os recursos disponíveis (RRi + entropias das ondas P + T + U + 4 Jitter + 4 Shimmer + 2 entropias de Shannon + 2 entropias energia logarítmica);
- RRi + entropias das ondas P + T + U;
- RRi + entropia das ondas P;
- RRi.

Com relação ao conjunto de dados cSeg, foram geradas as seguintes combinações, separadamente:

- Janelas do próprio sinal original (60 segundos - 2 derivações);
- Frequências instantâneas (2 derivações);
- Entropia espectral (2 derivações);
- Mapas de características pela transformada Scattering (2 derivações).

Cada item das duas listas anteriores, representa um conjunto separado, não tendo havido concatenação entre eles. Entretanto, nota-se a aplicação da ideia de hierarquia, utilizada na seção anterior com a MLP, sobre a geração dos conjuntos a partir do conjunto cRR.

Os parâmetros estruturais da LSTM explorados neste trabalho foram, número de neurónios em uma única camada escondida e contextualização: unidirecional ou bidirecional. Para o número de neurônio, foram investigados o desempenho de 10, 50 e 100 neurónios.

O critério de avaliação utilizado foi a acurácia de classificação sobre os conjuntos de teste. Por conta do elevado tempo de treinamento das redes LSTM, apenas uma iteração de cada combinação de parâmetros e conjuntos de entrada foi realizada.

Capítulo 4: Resultados e Discussão

Nesta seção são apresentados os resultados obtidos durante as etapas de desenvolvimento e execução das propostas deste trabalho. Ainda, são promovidas algumas discussões acerca dos resultados obtidos em comparação aos comportamentos esperados e aos apresentados pelo estado da arte.

A seção é organizada da seguinte forma. Primeiramente são apresentados os resultados e discussões que envolvem os experimentos do modelo primário, ou seja, a investigação da MLP a partir dos recursos de ECG baseados em segmentos de 30 RRI. Em seguida são apresentados os resultados e discussões acerca do modelo secundário, ou seja, a exploração da MLP e LSTM a partir dos recursos de ECG extraídos de contextos de 60 RRI ou 60 segundos. Por fim, os melhores resultados de acurácia dos modelos desenvolvidos neste trabalho são comparados com os do estado da arte.

4.1 Modelo Primário

Inicialmente, com relação à segmentação dos sinais de ECG em janelas de 30 RRI, obteve-se um número total de segmentos igual a 112.797, distribuídos desigualmente pelos 23 sujeitos. Esta desigualdade deve-se ao fato de obviamente haver variações do ritmo cardíaco ao longo do tempo durante todas as gravações. Para as condições de segmentação do modelo primário, o número de segmentos para cada sujeito foi determinado pela Equação 41, onde K_{RRI} é a quantidade de segmentos gerados para um determinado sujeito, K_r é a quantidade de picos R marcados pela anotação de complexos QRS advinda da base de dados, $K_L = 30$ é a quantidade de intervalos R-R por segmento, $d = 10$ é o fator de deslocamento entre segmentos adjacentes e a função ‘*round()*’ é o truncamento do seu argumento para seu valor inteiro.

$$K_{RRI} = \text{round} \left(\frac{(K_r - 1) - K_L}{d} \right) + 1 \quad (41)$$

Ao somar o número de segmentos de todos os sujeitos, obtém-se o número total de segmentos do modelo primário. A Tabela 3 apresenta a distribuição do número de segmentos por sujeito.

A redundância total de dados nesta abordagem foi de $\frac{K_L}{d} = 3$, o que significa que, exceto nas regiões de início e final dos sinais, cada intervalo R-R é utilizado em 3 segmentos. Portanto, há aproximadamente uma triplicação dos dados utilizados. Isto provoca um aumento substancial no número de exemplos disponíveis e, apesar da redundância de dados não ser recomendada por sua tendência em conduzir o aprendizado ao sobreajuste (*overfitting*), aqui ele se justifica pela escassez de dados e por estarem associados em contextos diferentes em cada segmento. Ou seja, uma mesma janela de intervalos R-R que se repita em três segmentos está associada, nos três segmentos, a contextos diferentes de intervalos, sejam adjacentes anteriores, ou posteriores. Isto indica que a representação dos comportamentos nos segmentos em si, não é redundante, e o efeito de sobreajuste é supostamente reduzido.

Tabela 3 - Distribuição do número de segmentos gerados para cada sujeito num contexto de 30 RRI.

<i>Sujeito</i>	<i>Número de segmentos</i>	<i>Sujeito</i>	<i>Número de segmentos</i>	<i>Sujeito</i>	<i>Número de segmentos</i>
01	4398	09	4986	17	3657
02	6189	10	4551	18	4333
03	3991	11	5513	19	5927
04	4283	12	3481	20	4549
05	4785	13	5516	21	5883
06	6173	14	3927	22	3982
07	5362	15	6024	23	5953
08	3677	16	5657		

O balanceamento entre as classes AF e não AF, após a etapa de referenciação dos segmentos, para os valores de μ utilizados, é apresentado pela Tabela 4. Nota-se que há equilíbrio entre as classes para todos valores de μ , uma vez que o desvio máximo é de 5% do valor ideal. Este equilíbrio faz com que a dominância de uma classe sobre a outra, no período de treinamento, seja evitada e o perfil de generalização da rede neuronal no pós-treinamento seja mais consistente. Entretanto, o equilíbrio de classes não garante que a rede consiga distingui-las, isto é tarefa da representação dos comportamentos, elaborada pela extração de características. O que se garante, é que haja aproximadamente a mesma quantidade de exemplos disponíveis para cada classe.

Tabela 4 - Balanço de classes para diferentes valores do fator μ .

<i>Classe/μ</i>	10%	30%	50%
<i>AF</i>	55%	53%	45%
<i>Não AF</i>	45%	47%	65%

Após a determinação das características de cada segmento (de acordo com o que foi proposto no modelo primário), formadas as combinações dos conjuntos de treinamento, validação e teste, e a exploração dos parâmetros estruturais da MLP conforme a Metodologia A, obteve-se um relatório de acurácia sobre o conjunto de teste para cada modelo.

Os resultados de acurácia para os dois melhores modelos (*ML_01* e *ML_02*) durante a metodologia A, são apresentados pela Tabela 5. Vale ressaltar que a acurácia é a medida de acerto da rede referente à classificação esperada. A classificação esperada neste caso foi constituída pelo fator de proporção de referenciação $\mu = 30\%$, por apresentar o maior nível de equilíbrio entre as classes. Observa-se que ambos modelos listados, possuem 3 camadas escondidas e função de ativação tangente sigmoide para todas camadas. Dessa forma, afirma-se que a combinação destes parâmetros para a arquitetura MLP, dentro daqueles propostos para experimentação, leva ao melhor desempenho de reconhecimento das classes. Portanto, é sobre estas duas estruturas que são investigadas a variação de μ e das combinações de características.

Tabela 5 - Parâmetros estruturais e acurácia para os dois melhores modelos na metodologia A.

	<i>Nº de camadas escondidas</i>	<i>Nº de neurónios por camada escondida</i>	<i>Função de ativação</i>	<i>Acurácia média no conjunto de teste [%]</i>
<i>ML_01</i>	3	165	Tan-sig	85,45
<i>ML_02</i>	3	180	Tan-sig	86,29

Os resultados de acurácia, para a Metodologia B, sobre os dois melhores modelos (*ML_01* e *ML_02*) obtidos durante a Metodologia A, são apresentados pela Tabela 6. Nota-se uma leve melhora de desempenho em acurácia dos modelos à medida que mais recursos são concatenados nos conjuntos de entrada (direção mencionada: *Comp_01* para *Comp_04*).

Genericamente, pode-se afirmar que agregar mais informações, conforme as representações morfológicas, temporais e estatísticas propostas, acerca do comportamento das janelas de sinais fez com que os modelos reconhecessem os comportamentos das classes com mais facilidade. Essa afirmação nem sempre é verdadeira para aprendizagem de máquina, uma

vez que podem existir recursos que provocam deterioração e dificuldades para a aprendizagem das redes. Estes efeitos podem ser comparados a uma ‘confusão’ durante a aprendizagem pelo excesso de informação ou contradições. Entretanto, não é o que se observa pelos resultados apresentados.

Tabela 6 - Resultados de acurácia [%] sobre o conjunto de teste para diferentes combinações de características, pela metodologia B.

	<i>Comp_01</i>			<i>Comp_02</i>			<i>Comp_03</i>			<i>Comp_04</i>		
μ [%]	10	30	50	10	30	50	10	30	50	10	30	50
<i>ML_01</i>	76.66	73.66	76.70	80.00	72.56	77.50	79.52	74.94	78.73	80.67	75.47	80.47
<i>ML_02</i>	78.31	73.16	75.75	76.10	70.33	77.96	77.05	71.95	73.87	80.33	72.08	80.61

Basicamente, três conclusões podem ser apresentadas em relação aos resultados da Tabela 6. Em primeiro lugar, nota-se que o fator de proporção $\mu = 30\%$ sempre proporcionou resultados inferiores comparativamente com $\mu = 10\%$ e $\mu = 50\%$, e, portanto, é considerado o menos adequado para referenciação dos segmentos. Ainda, o fator $\mu = 10\%$ é considerado o mais adequado por quase sempre estar associado à melhor acurácia de cada combinação. Isto pode significar que pouca informação sobre a AF dentro do segmento é suficiente para caracterizar o comportamento da classe, e não apenas isso, mas pode significar também que, uma alta proporção (mas abaixo de 50%) da AF em segmentos classificados como ‘não AF’ pode representar uma carga comportamental muito elevada sobre segmentos considerados normais.

Em segundo lugar, notou-se que a inclusão dos parâmetros Shimmer pouco ou nada influenciaram na melhoria de acurácia dos modelos. Esta indiferença pode estar associada com o fato da referência do cálculo do Shimmer estar posta sobre a amplitude do pico das ondas R. Apesar da deformação do complexo QRS (irregularidade da atividade dos ventrículos) ser uma característica da fibrilação atrial, ela não se mostrou suficiente para colaborar na diferenciação de um comportamento de AF sobre um normal. Ou então, a extração da representação desta deformação com as ferramentas de Shimmer não foi uma escolha adequada para o modelo. Para determinar o nível de representatividade destes parâmetros sobre sinais ECG, especialmente em comportamentos de AF, pode-se buscar alterar a amplitude de referência para o pico das ondas P e/ou experimentá-las com a atenuação do ruído. Sobre o ruído e artefatos nos sinais, o Shimmer pode ser facilmente influenciado, uma vez que a presença de perturbações em amplitude é expressiva em desvios da linha de base e ruído de alta frequência. Por fim, pode-

se citar as especificidades dos pacientes quando as perturbações morfológicas consequentes da AF. Ou seja, as deformações morfológicas das ondas do ciclo cardíaco são variáveis entre os sujeitos, o que pode enfraquecer o potencial uso do Shimmer.

Em terceiro lugar, o aumento do número de características concatenadas mostrou que as estruturas MLP propostas são capazes de lidar com o aumento da complexidade de entrada (número de entradas), sinalizando um potencial para a agregação de novos recursos. Portanto, baseando-se neste indicativo, foram propostos o aumento de contexto dos segmentos e a extração de recursos relativos às morfologias das outras ondas do ECG, especialmente a onda P. Ainda, o uso das entropias de Shannon e energia logarítmica demonstrou que um parâmetro relativo à distribuição de informação sobre todo o contexto do segmento contribui para o aumento da acurácia e portanto, deve ser mantido nas próximas aplicações. Além disso, apesar das medidas de perturbação Jitter e Shimmer não demonstrarem forte contribuição no modelo primário, seus potenciais de descrição permanecem alinhados à proposta, sendo que seus pontos de referência, entretanto, devem ser melhor trabalhados, seja pela redução do ruído, melhora na localização das marcações ou critérios de seleção de recursos por sua correlação com o comportamento esperado.

Referente a isso, notou-se que uma nova perspectiva quanto à marcação do pico das ondas R era necessária, devido às imprecisões das marcações advindas da base de dados e à desconsideração dos períodos de silêncio. Ou seja, o impacto do deslocamento ligeiro das marcações dos picos e de longos intervalos entre picos (superiores a 3 segundos), pode ter provocado fortes variações nos resultados de Jitter e Shimmer, prejudicando a relevância na descrição das classes. Entretanto, uma marcação precisa de tais picos se faz necessária também para a localização precisa das demais ondas, como na metodologia proposta no modelo secundário. Pelo bom desempenho observado das entropias, propôs-se a utilização da entropia de Shannon para extrair a distribuição de informação nas regiões das ondas P, T e U, de acordo com suas localizações relativas aos picos R.

4.2 Modelo Secundário

Inicialmente, com relação ao algoritmo proposto para remarcação do pico das ondas R dos sinais, e de acordo com a metodologia para o critério de acerto, obteve-se os resultados apresentados na Tabela 7. A acurácia média total para detecção do pico R foi de 95,41%. Entretanto, nota-se uma baixa acurácia (17,51%) na detecção para o sujeito 14, indicando que

para este sujeito os picos não foram satisfatoriamente identificados. É razoável afirmar que este resultado é discrepante do desempenho geral, uma vez que se encontra longe da tendência criada pelos outros resultados.

Tabela 7 - Acurácia do algoritmo proposto para detecção do pico das ondas R.

<i>Sujeito</i>	<i>Acurácia [%]</i>	<i>Sujeito</i>	<i>Acurácia [%]</i>	<i>Sujeito</i>	<i>Acurácia [%]</i>
01	98,37	09	97,47	17	99,97
02	99,69	10	99,50	18	99,57
03	99,62	11	97,58	19	98,45
04	98,30	12	99,95	20	99,94
05	99,73	13	99,62	21	99,58
06	99,94	14	17,51	22	99,81
07	99,76	15	92,41	23	99,28
08	98,82	16	99,45		

A Figura 11 revela que na verdade, são as marcações originais dos sinais deste sujeito (14) que se encontram extremamente deslocadas em relação ao complexo QRS. Com isso, o critério de acerto proposto no algoritmo de remarcação se torna inválido nesta condição. Porém, ao se basear na acurácia obtida para todos outros sujeitos (acurácia média sem o sujeito 14: 98,95%) e pela inspeção visual da remarcação executada pelo algoritmo em alguns trechos do sinal, é suposto que na verdade houve a correta localização do pico das ondas R.

Além do mais, nota-se que o sinal Original da Figura 11 contém uma forte interferência de ruído de alta frequência, além da presença de ondas do tipo F (ondas de alta frequência características do *flutter* atrial resultantes de uma atividade vibratória anômala dos átrios). Apesar da anotação rítmica indicar que o comportamento é de fibrilação atrial ao longo de todo sinal, nota-se a agregação de outros comportamentos. Aqui supõe-se que o método de marcação automático dos sinais, aplicado pelos criadores da base de dados, pode não ter sido capaz de lidar com tais ruídos e com a presença das ondas F. A partir disso, afirma-se que o algoritmo proposto neste trabalho supera estas dificuldades, se mostrando robusto à baixa qualidade dos sinais (seja pela corrupção por ruídos ou artefatos) e na presença de comportamentos patológicos.

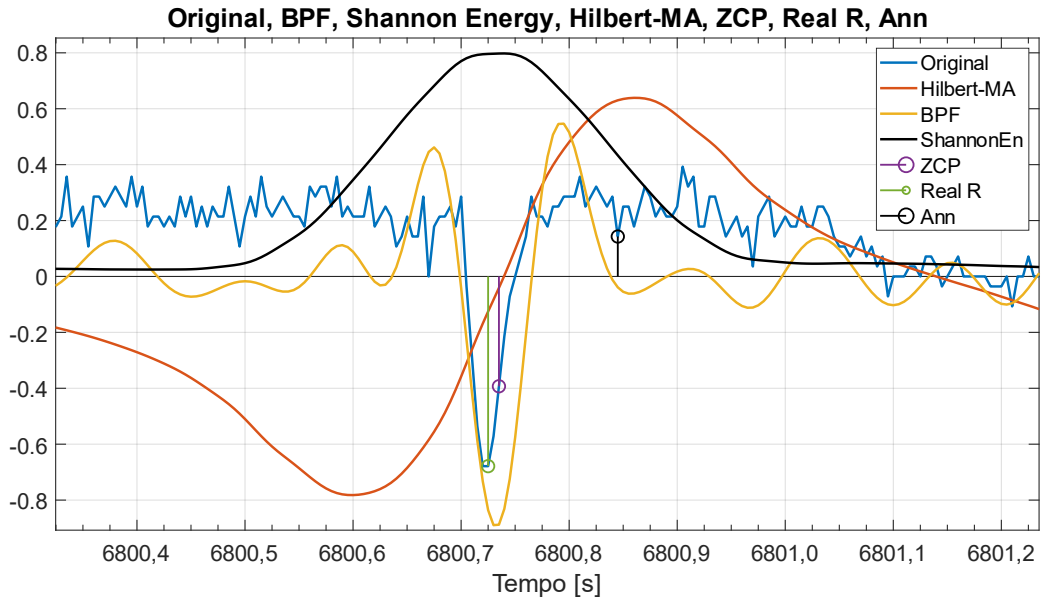


Figura 11 – Representação normalizada dos sinais relativos à etapa de remarcação do pico das ondas R para o sujeito 14. Legenda: Original: Sinal de ECG original (como obtido da base de dados) da primeira derivação; BPF: Sinal original após a aplicação da filtragem passa banda; ShannonEn: Sinal da energia de Shannon já suavizada; Hilbert-MA: Sinal da energia de Shannon após a transformação de Hilbert e subtração da média móvel; ZCP: ponto de cruzamento de zero positivo do sinal Hilbert-MA projetado sobre o sinal Original; Real R: ponto do pico R remarcado pelo algoritmo proposto e projetado sobre o sinal Original; Ann: marcação do complexo QRS advinda das anotações da base de dados.

A seguir, são apresentados os resultados acerca da segmentação dos sinais a partir das novas marcações, em conjuntos com contextos de 60 RRI e 60 segundos, separadamente.

Primeiramente para o contexto 60 RRI, o número de segmentos gerados para cada sujeito foi determinado segundo a Equação 42, onde K_{RRI} é a quantidade de segmentos gerados, K_r é a quantidade de picos R marcados, $(a - 1)$ é a contagem de períodos de silêncio do sinal (onde $a \in \mathbb{Z} \mid a \geq 1$ é a referência final de conectividade das marcações), $K_L = 60$ é a quantidade de intervalos R-R por segmento, $d = 10$ é o deslocamento entre segmentos adjacentes e $\text{round}()$ é a função de truncamento ao inteiro inferior do seu argumento.

$$K_{RRI} = \text{round} \left(\frac{(K_r - 1) - (a - 1) - (K_L - d)}{d} \right) \quad (42)$$

A quantidade de períodos de silêncio, $(a - 1)$, identificados em cada sinal, pela metodologia proposta (a partir do fator limite de 3 segundos entre picos R adjacentes), é relacionada pela Tabela 8. Uma alta quantidade de períodos de silêncio pode ser uma consequência de uma baixa qualidade no sinal (presença de artefatos do tipo desconexão dos eletrodos) ou de uma inadequada escolha do fator limite. Entretanto, percebe-se que houve uma

quantidade elevada destes períodos em apenas dois sujeitos e mesmo assim, relativamente à duração das gravações (10 horas para cada sujeito) este número é considerado aceitável. Com isso, a introdução desta noção na segmentação dos sinais, ajudou a reduzir a perturbação por longos períodos de falta de informação.

Tabela 8 – Quantidade de períodos de silêncio, por sujeito, identificados pela metodologia proposta no modelo secundário.

<i>Sujeito</i>	<i>Períodos de Silêncio</i>	<i>Sujeito</i>	<i>Períodos de Silêncio</i>	<i>Sujeito</i>	<i>Períodos de Silêncio</i>
01	21	09	22	17	0
02	2	10	0	18	0
03	6	11	1	19	0
04	0	12	1	20	0
05	0	13	0	21	2
06	0	14	0	22	4
07	0	15	0	23	3
08	4	16	0		

Após estas considerações e, ao somar a quantidade de segmentos de cada um dos 23 sujeitos (Tabela 9), obteve-se o total de 115.522 segmentos de 60 RRi. O número total de segmentos para o modelo secundário é superior ao do modelo primário por causa da remarcação do pico das ondas R. Entretanto, a mudança do tamanho de contexto de 30 para 60 RRi não causa grandes alterações no total pois o fator dominante é o deslocamento, $d = 10$, que foi mantido nos dois modelos.

O balanço de classes, para $\mu = 10\%$, sobre todos os segmentos gerados no contexto 60 RRi é apresentado pela Tabela 10. Nota-se que há um equilíbrio entre as classes AF e Normal e que, comparativamente a estas, a classe Outros é extremamente minoritária. Entretanto, como o objetivo da criação dessa classe foi de reduzir a presença dos demais comportamentos, que não AF ou Normal, sobre as duas classes principais, não se considera problemática esta diferença. Além disso, não foi pretendido reconhecer os padrões comportamentais contidos nesta classe (uma vez que podem ser dos mais variados) e sim, funcionar como uma certa “válvula de escape” para o modelo quando as características da entrada não se encaixarem no aprendizado majoritário das classes AF e Normal.

Tabela 9 - Distribuição do número de segmentos gerados para cada sujeito num contexto de 60 R Ri.

<i>Sujeito</i>	<i>Número de segmentos</i>	<i>Sujeito</i>	<i>Número de segmentos</i>	<i>Sujeito</i>	<i>Número de segmentos</i>
01	4510	09	5036	17	3799
02	6324	10	4659	18	4441
03	4190	11	5588	19	6030
04	4370	12	3530	20	4640
05	4908	13	5645	21	6037
06	6310	14	4062	22	4103
07	5517	15	6155	23	6045
08	3812	16	5811		

Tabela 10 - Proporção da distribuição de classes implementada no modelo secundário, para o contexto 60 R Ri e com $\mu = 10\%$.

$\mu = 10\%$	
<i>AF</i>	45,92%
<i>Normal</i>	52,95%
<i>Outros</i>	1,13%

Para o contexto de 60 segundos, o número de segmentos gerados não depende das referências de conectividade entre os picos R, uma vez que são intervalos de 60 segundos consecutivos. Entretanto, são contados a partir da primeira marcação de ritmo do sinal, já que todas as amostras devem pertencer a uma marcação rítmica para determinação da classe do segmento. Portanto, o número de segmentos gerados para cada sujeito foi determinado de acordo com a Equação 43, onde K_S é a quantidade de segmentos gerados, K_{sig} é a quantidade de amostras do sinal, j é a amostra da primeira marcação rítmica, $K_{LS} = 60 \cdot fs$ é a duração, em amostras, de cada segmento, $d_s = 10 \cdot fs$ é o deslocamento, em amostras, entre segmentos adjacentes, $fs = 250$ Hz é a frequência de amostragem da digitalização e ‘round()’ é a função de truncamento para o inteiro imediatamente anterior do seu argumento.

$$K_S = \text{round} \left(\frac{(K_{sig} - j) - (K_{LS} - d_s)}{d_s} \right) \quad (43)$$

O número de segmentos gerados para cada sujeito foi de 3.677, com exceção do sujeito 12, que proporcionou 3.324 segmentos devido a um deslocamento da sua primeira marcação de ritmo em relação às dos demais sujeitos. Portanto, ao todo foram obtidos 84.218 segmentos para o contexto de 60 segundos. A diferença deste número total para o anterior, referente ao contexto 60 RRI, se deve ao fato de o ritmo cardíaco dos pacientes ser frequentemente superior a 60bpm, resultando em mais de 60 intervalos R-R durante o período de 60 segundos. De maneira análoga, é razoável afirmar que a quantidade de informação presente num intervalo de 60 segundos é maior do que em 60 RRI, uma vez que mais batimentos estão contidos e a quantidade de amostras é proporcional à frequência de amostragem do sinal.

A partir da Tabela 11, que apresenta o nível de balanceamento entre classes do conjunto de segmentos, nota-se claramente a separação de outros comportamentos do comportamento normal, efeito não obtido no conjunto com contexto 60 RRI.

Tabela 11 - Proporção da distribuição de classes implementada no modelo secundário, para o contexto 60 segundos e com $\mu = 10\%$.

$\mu = 10\%$	
<i>AF</i>	47,98%
<i>Normal</i>	29,38%
<i>Outros</i>	22,64%

Após a formação dos conjuntos de treino, validação e teste, de acordo com a metodologia apresentada para o modelo secundário sobre os contextos de 60 RRI e 60 segundos, foram obtidos os resultados para a exploração das redes MLP e LSTM, também de acordo com a metodologia apresentada.

Primeiramente referente à experimentação da MLP, os resultados obtidos são organizados na Tabela 12. Os modelos M1-M5 representam, ordenadamente, os melhores desempenhos quando utilizadas todas as 10 características extraídas dos segmentos, ou seja, todos os recursos disponíveis sobre o conjunto de segmentos de contexto 60 RRI. O modelo M6 por sua vez, representa o pior desempenho entre os parâmetros explorados, e, portanto, não é levado adiante para experimentação das outras combinações de características. O modelo M2 é considerado o mais adequado entre os apresentados, por obter o melhor desempenho de acurácia em 60% das combinações de características propostas.

Sobre os resultados apresentados, nota-se a exclusividade do uso de uma única camada escondida entre os melhores modelos, a quantidade de neurónios sempre superior a 155 e a dominância da função de ativação tangente sigmoide. Com relação ao número de camadas escondidas, pode-se afirmar que não se fez necessária várias transformações da informação de entrada para alcançar os melhores desempenhos de reconhecimento dos padrões.

Por simplicidade de expressão, é dito que a cada camada escondida pela qual a informação passa, esta é conduzida a um nível de representação mais abstrato, num espaço multidimensional mais elevado, o qual possibilita a rede neuronal compreender os padrões da entrada. Quando os padrões contidos na entrada são muito complexos, várias transformações por muitos neurónios se fazem necessárias para que esses padrões sejam captados pela rede. Uma vez que a representação dos dados é muito bem descritiva acerca do comportamento esperado, os recursos da rede necessários para interpretá-los são proporcionalmente menores.

De acordo com esse raciocínio, conclui-se que houve uma melhora substancial na medida de acurácia devido ao aumento da significância da representação dos dados pelas novas características agregadas, pela remarcação do pico das ondas R e pelos critérios de normalização dos dados. Especialmente os dois últimos fatores mencionados, certamente levaram a uma determinação mais precisa e confiável das características (nomeadamente os parâmetros Jitter e Shimmer), bem como uma contextualização sobre os comportamentos a nível local (segmento) e global (sujeito) que favoreceram a descrição do comportamento do segmento e da especificidade de cada sujeito.

Assim como nos resultados apresentados pela Tabela 6 referentes à Metodologia B do modelo primário, na Tabela 12 também é observado que houve um aumento progressivo da acurácia de classificação sobre o conjunto de teste à medida que mais características foram sendo adicionadas. Mais uma vez nota-se que o Shimmer, neste caso o Shimmer absoluto, provoca uma leve redução de desempenho em alguns arranjos. Apesar disso, a utilização de todos os parâmetros Shimmer ao mesmo tempo sempre contribui para uma melhoria na acurácia. Vale destacar que a utilização dos intervalos R-R por si só é a que mais contribuiu, sendo assim, comprovadamente a característica mais relevante entre todas. Em conjunto, a agregação das entropias das ondas P, T e U foram as que mais contribuíram para o aumento da acurácia, representando uma melhora na classificação de 2,5%. Portanto, a suposição de que as informações acerca das outras ondas do ECG além do pico da onda R são relevantes foi confirmada. Os parâmetros Jitter de perturbação contribuíram com um aumento de

aproximadamente 1% no nível de acurácia dos modelos, sendo que especialmente o Jitter absoluto se mostrou ser o mais relevante.

Tabela 12 - Desempenho em acurácia no conjunto de teste dos melhores modelos MLP entre os explorados no modelo secundário. Nota para a apresentação, em vermelho, do modelo de pior desempenho.

(a) resultados para a utilização de 10 a 6 características.

Modelo	Nº de Nós	Nº de Camadas	Função Ativação	Acurácia no teste [%] / nº de características				
				10	9	8	7	6
M1	165	1	Tan-sig	91,96	91,46	91,20	91,17	91,24
M2	205	1	Tan-sig	91,91	91,49	91,57	91,71	91,56
M3	175	1	Tan-sig	91,78	91,38	91,42	91,16	91,24
M4	155	1	Tan-sig	91,64	91,50	91,31	90,82	91,09
M5	195	1	Tan-sig	91,56	91,61	91,49	91,31	90,95
M6	155	3	Tan-sig	87,81	-	-	-	-

(b) resultados para a utilização de 5 a 1 características.

Modelo	Nº de Nós	Nº de Camadas	Função Ativação	Acurácia no teste [%] / nº de características				
				5	4	3	2	1
M1	165	1	Tan-sig	90,66	90,33	90,07	89,10	88,01
M2	205	1	Tan-sig	90,89	90,54	90,36	89,60	87,87
M3	175	1	Tan-sig	90,76	90,53	90,12	89,48	88,04
M4	155	1	Tan-sig	90,90	90,42	89,86	89,24	88,30
M5	195	1	Tan-sig	90,35	90,44	89,95	89,22	88,16
M6	155	3	Tan-sig	-	-	-	-	-

Uma limitação potencial quanto ao modelo secundário, é referente ao critério de organização dos dados dos sujeitos nos conjuntos de treino, validação e teste. Aqui, utilizou-se os dados de todos os sujeitos em todos os conjuntos, eliminando a possibilidade de validação do modelo para sujeitos não vistos durante o treinamento. Este tipo de abordagem, reduz o impacto da representação das especificidades dos sujeitos, uma vez que exemplos de todos são apresentados no treinamento da rede. Entretanto, com a demonstração do comportamento de um maior número de sujeitos para a aprendizagem da rede, mais próxima ela estará de alcançar a invariabilidade de sujeitos, já que a superação das particularidades de cada sujeito só acontece

quando um número grande deles é fornecido. O efeito da particularidade é similar à presença de ruído branco numa imagem. É possível anular o efeito do ruído branco ao se sobrepor muitas capturas da mesma imagem em diferentes momentos. Analogamente, é possível tornar a aprendizagem invariante ao sujeito ao se sobrepor (aprender) o comportamento do maior número possível deles. Portanto, a apresentação de todos os sujeitos nesta abordagem do modelo secundário é ideal para futuras aplicações que envolvam, por exemplo, a validação das redes treinadas sobre outras bases de dados.

Dos resultados da experimentação das redes LSTM, apresentados na Tabela 13, os conjuntos de características representados pelos números ordinais de 1 a 5 são, respectivamente, onde apenas o 5º pertence ao contexto cSeg:

- 1º: Concatenação de todas as 10 características dos segmentos 60 RRI;
- 2º: Concatenação dos 60 RRI e as entropias das ondas P, T e U;
- 3º: Concatenação dos 60 RRI e a entropia das ondas P;
- 4º: Apenas os 60 RRI;
- 5º: Mapa de características da transformada Scattering.

Sobre os resultados obtidos, pode-se dizer que para o primeiro conjunto, tornar os valores escalares (como os parâmetros Jitter e Shimmer) em sequências de 60 constantes (pela repetição do escalar ao longo do tempo) para adequar à LSTM, fez com que a rede se comportasse pior do que na presença de apenas os 60 RRI. Ou seja, certamente, esta metodologia de conversão de escalares para sequências de constantes, não só foi inadequada como também deteriorou a aprendizagem.

Entretanto, a partir da observação das acurácias para os conjuntos 4º, 3º e 2º, nesta ordem, nota-se uma expressiva evolução de desempenho, especialmente após a adição da entropia das ondas P, sinalizando que esta característica, para o tipo de análise temporal realizada pela LSTM é bem relevante, confirmando a informação coletada no estado da arte. Além disso, nota-se que a agregação das entropias das ondas T e U colaboram para melhorar na tarefa de classificação da proposta, sendo que é sobre este conjunto que se obteve o melhor resultado, em acurácia, de todo o trabalho (98,17% - modelo ME).

Acerca da utilização do mapeamento pela transformada Scattering (conjunto 5º), não houve um desempenho satisfatório à medida da expectativa criada pelo potencial da ferramenta e da quantidade de recursos computacionais gastos. Estes pontos se referem principalmente à

comparação dos níveis de acurácia obtidos em relação aos demais conjuntos e aos resultados da MLP na Tabela 12. Isto pode indicar que a transformada Scattering, no caso da aplicação pretendida, precise da associação de outras etapas de transformação, como camadas convolucionais ou projeções do mapa para uma única dimensão (camada *Flatten*). Ou também, os parâmetros utilizados para decomposição dos sinais, encontrados em aplicações paralelas, podem não ser os mais adequados para análise de sinais ECG ou para utilização direta na LSTM.

Quanto à direcionalidade das arquiteturas, observa-se que não há efeito de melhoria em acurácia na utilização de um contexto bidirecional sobre um unidirecional. Uma das possíveis justificativas é a baixa quantidade de amostras que formam as sequências de entrada, ou seja, pelo contexto ser muito curto, a variação da informação ao longo do segmento tende a ser mais baixa do que em observações mais longas. Outra justificativa pode estar associada à escala das representações, já que, por exemplo, a velocidade dos eventos no sinal original está ligada a quantidade de amostras entre eles, e no caso das extrações manuais de características essa dependência amostral é prejudicada e levada a uma noção do conteúdo da entrada em si.

Entretanto, este fator de direcionalidade possivelmente deve ser mais bem observado quando os segmentos são representações do sinal original, de transformações que mantenham as localidades de cada informação ou em sequências de muitas amostras. Ou seja, supõe-se que para que a rede, na forma bidirecional, capture algum tipo de informação no sentido reverso da sequência de entrada, esta deve preservar as relações de continuidade do sinal, para que a propagação da informação crie um contexto futuro que contenha as relações necessárias de temporalidade e informação.

Tabela 13 - Desempenho em acurácia no conjunto de teste das variações de parâmetros estruturais da LSTM e diferentes combinações de características para os conjuntos de entrada.

<i>Modelo</i>	<i>Nº de Nós</i>	<i>Direcionalidade</i>	<i>Acurácia no teste [%] / conj. características</i>				
			1º	2º	3º	4º	5º
<i>MA</i>	10	uni	93,69	96,20	94,98	92,28	87,01
<i>MB</i>	10	bi	94,98	96,89	95,01	92,75	89,43
<i>MC</i>	50	uni	90,90	97,48	97,69	94,74	92,32
<i>MD</i>	50	bi	92,13	97,80	97,25	94,94	92,48
<i>ME</i>	100	uni	94,26	98,17	97,04	94,77	92,90
<i>MF</i>	100	bi	94,25	97,73	97,28	95,27	91,74

Para as características obtidas nos segmentos de 60 segundos, as frequências instantâneas, entropia espectral e segmentos do próprio sinal, não houve convergência em nenhum dos casos, o que pode indicar que a presença de ruído sobre estes recursos nas condições estruturais propostas e a não associação de mais atributos configura uma dificuldade potencial para o modelo reconhecer os padrões. Além disso, pode ser que seja necessário o empilhamento de mais camadas do tipo LSTM para elevar o nível de abstração das informações e aumentar a capacidade de análise da rede. Ainda, futuramente podem ser investigados os impactos da utilização de apenas uma derivação do sinal, observando se a utilização das duas derivações, como feito neste trabalho, provoca algum efeito de anulação ou se alguma das duas corrompe a aprendizagem da rede.

4.3 Comparação com o Estado da Arte

A partir dos resultados obtidos durante todo o trabalho e baseando-se nos métodos e nas metodologias utilizadas para o desenvolvimento e exploração das propostas, pretende-se nesta subseção, comparar brevemente o desempenho das abordagens para classificação de fibrilação atrial propostas com alguns estudos apresentados no estado da arte. A Tabela 14 apresenta a comparação direta entre os trabalhos, referentes: às bases de dados utilizadas; se houve a aplicação de pré-processamento para remoção de ruídos, tendências ou artefatos e qual o método empregado; as características extraídas e o tipo de segmentação; os métodos de análise e classificação das representações; e a acurácia obtida.

Através dos trabalhos relacionados, nota-se que naqueles onde há associação de diferentes bases de dados ocorre o aumento de exemplos e de sujeitos envolvidos, tornando o treinamento dos modelos de aprendizado mais diversificado e a validação mais consistente. Além disso, diferentes bases de dados carregam particularidades não só de sujeitos diferentes, mas também com relação aos métodos de aquisição e as condições dos sistemas e ambientes envolvidos durante as gravações. Neste sentido, para esses trabalhos que envolvem mais de uma base, pode-se afirmar que os resultados tendem a ser mais confiáveis e conseqüentemente, o sistema, mais robusto. Este não o caso das propostas desenvolvidas neste trabalho, as quais utilizaram apenas os dados contidos na base *MIT-BIH Atrial Fibrillation*. Dessa forma, afirma-se que há uma desvantagem no presente modelo frente aos demais, principalmente quanto a diversidade de sujeitos e de validação entre base de dados.

Tabela 14 – Comparação entre alguns sistemas desenvolvidos em outros trabalhos e os melhores modelos MLP e LSTM deste estudo, a nível de acurácia para detecção de pelo menos AF.

	<i>Base de Dados</i>	<i>Pré-Processamento</i>	<i>Segmentação e Extração de Características</i>	<i>Método de Classificação</i>	<i>Acurácia [%]</i>
<i>Saadatmejad et al., 2019</i>	MITDB	Não citado	Estatísticas RRi, transformada Wavelet	LSTM	98,30
<i>Acharya et al., 2017a</i>	AFDB, CUDB, MITDB	Transformada Wavelet	Segmentação em dois segundos (A) e cinco segundos (B)	CNN	Rede A: 92,50 Rede B: 94,90
<i>Faust et al., 2018b</i>	AFDB	Não citado	100 RRi	LSTM	98,51
<i>Dang et al., 2019</i>	AFDB, MITNR, MITDB	Nenhum	Segmentação em torno do pico R, detecção das ondas P, T e complexo QRS	CNN + LSTM	96,59
<i>Xia et al., 2018</i>	AFDB	Filtro Passa faixa elíptico	Segmentação em 5 segundos; transformada de Fourier de curto-termo, transformada Wavelet estacionária	CNN	98,63
<i>Andersen et al., 2019</i>	AFDB, MITDB, MITNR	Nenhum	30 RRi	CNN + LSTM	97,80
<i>Martis et al., 2013c</i>	AFDB, MITDB	Transformada Wavelet Discreta	Espectros de ordem superior	KNN	99,50
<i>Borghi et al., 2020 / Este trabalho</i>	AFDB	Nenhum	30 RRi, Jitter, Shimmer, Entropias de Shannon e energia logarítmica	MLP	80,67
<i>Este trabalho</i>	AFDB	Nenhum	60 RRi, Jitter, Shimmer, Entropias de segmento e das ondas P, T e U	MLP	91,96
<i>Este trabalho</i>	AFDB	Nenhum	60 RRi e entropias das ondas P, T e U	LSTM	98,17

Legenda: MITBD: *MIT-BIH Arrhythmia*; AFDB: *MIT-BIH Atrial Fibrillation*; MITNR: *MIT-BIH Normal Sinus Rhythm*; CUDB: *Creighton University Ventricular Tachyarrhythmia*.

Acerca do pré-processamento utilizado, o presente trabalho apresenta a vantagem de não ter utilizado técnicas de filtragem, remoção de tendências, entre outros, uma vez que estas podem comprometer informações potencialmente relevantes em bandas de frequência equivalentes às rejeitadas. Além disso, o pré-processamento demanda uma exploração exhaustiva dos parâmetros, sejam dos filtros, das decomposições Wavelet, etc, que sucede em um custo computacional elevado e em um tempo de ocupação para análise da qualidade dos resultados. Entretanto, o não tratamento do ruído nos sinais de entrada pode acarretar uma forte

limitação para aplicações em tempo real ou em cenários com muita interferência sobre as medidas, pois estas podem ser totalmente descaracterizadas ou até mesmo, invalidadas.

Com relação à segmentação e à extração de recursos, nota-se que há uma certa dominância na escolha da representação dos sinais pelos intervalos R-R, o que contextualiza bem as metodologias propostas com os demais estudos. Pelo menos relativa à fibrilação atrial, esta medida ao longo do tempo do comportamento do ritmo cardíaco produz bons resultados, mas nem sempre aparece isoladamente. A seleção dos demais recursos a serem combinados geralmente é o que define se o modelo de aprendizagem encontrará uma representação descritiva da AF ou não. Entretanto, não há um conjunto específico de métodos que possam conduzir à representação dos exemplos a um nível ótimo. O que se observa, é uma variedade de características propostas para descrever os sinais quanto a sua morfologia, temporalidade e variáveis estatísticas, atingindo um desempenho satisfatório.

Quanto aos modelos de classificação, pode-se destacar que o bom desempenho em acurácia alcançado pela proposta com a MLP é bem interessante, uma vez que sua complexidade é muito inferior aos modelos de aprendizado profundo, nomeadamente as CNN e LSTM. Tudo indica que uma nova representação da AF ainda mais descritiva e que seja invariante a sujeitos ou ruídos, associada ao conjunto utilizado, pode fazer com que o desempenho da MLP se torne equiparado aos melhores modelos do estado da arte. Entretanto, mesmo com o bom desempenho, as redes profundas ainda superam em muito as mais convencionais. O preço a se pagar por esta superioridade é o elevado custo computacional. Quanto a isso, afirma-se que para o propósito deste trabalho, na identificação da AF, os resultados obtidos pela MLP são os mais relevantes frente aos do estado da arte apresentados, uma vez que a rede é extremamente leve no sentido da complexidade da estrutura, não é necessária a aplicação de filtragem sobre os sinais de ECG e os resultados demonstram uma tendência de melhoria à medida que mais recursos são agregados. A grande desvantagem é que à medida que o número de recursos aumenta, o custo computacional envolvido para suas determinações aumenta proporcionalmente. Além disso, obteve-se uma combinação de parâmetros e características para a LSTM que se pode certamente equiparar em acurácia com os melhores modelos do estado da arte.

Capítulo 5: Conclusões

Neste trabalho foram propostas experimentações sobre modelos de aprendizagem de máquina baseados em redes neurais dos tipos MLP e LSTM para classificação de segmentos de sinais de eletrocardiograma (ECG). A tarefa de classificação principal foi para detecção de comportamentos de fibrilação atrial, onde foi obtida acurácia máxima de 98,17% entre os sistemas baseados na LSTM e 91,96% entre os sistemas baseados na MLP. Para representar os comportamentos de cada trecho do ECG extraído, foram utilizadas ferramentas matemáticas relativas à extração de características temporais, morfológicas, estatísticas e em tempo-frequência. Para características temporais, determinou-se: a medida do ritmo cardíaco através do intervalo entre picos R consecutivos e os parâmetros de perturbação temporal através do Jitter. Como características morfológicas, foram utilizados: os parâmetros de perturbação em amplitude através do Shimmer e medidas de entropia de Shannon das ondas P, T e U do ECG. Como características estatísticas, foram utilizadas medidas de entropia de Shannon e de energia logarítmica sobre o contexto de todo o segmento. Acerca das ferramentas de análise tempo-frequência, foram determinadas as frequências instantâneas, a entropia espectral e a transformada Scattering de cada segmento.

A duração de cada segmento foi explorada para valores de 30 e 60 intervalos R-R e 60 segundos, sendo que cada um destes valores teve seu conjunto específico de características associadas. De acordo com as metodologias desenvolvidas, foram realizadas experimentações com relação à combinação das características extraídas e à organização dos conjuntos de treinamento e teste dos modelos de aprendizagem. Além disso, os modelos foram observados quanto aos seus desempenhos a diferentes arranjos de sujeitos. Foi também promovida uma metodologia de normalização dos dados que visou preservar as especificidades dos sujeitos ao mesmo tempo que destacava os comportamentos locais. Para isso foram utilizadas a normalização Z-score e pelo valor máximo absoluto sobre diferentes características em diferentes maneiras. A princípio, observou-se que a introdução desta ideia acompanhou uma melhoria de acurácia do modelo MLP, entretanto, não foi proposta uma metodologia de avaliação específica para mensurar o impacto real da proposta.

Através dos resultados obtidos em acurácia, notou-se que as sequências de intervalos R-R foram as que mais contribuíram para diferenciar um comportamento de fibrilação atrial de um normal. Além disso, observou-se que após a inclusão dos recursos que trazem a ideia do

comportamento acerca das ondas P e T, principalmente, os sistemas apresentaram melhoria em acurácia significativa, indicando que para a tarefa de detecção de AF, estas informações, da maneira como foram utilizadas, são relevantes.

Os parâmetros Jitter e Shimmer, apesar de não demonstrarem uma melhoria significativa no desempenho, é notável que contribuíram. No melhor que se soube, esta foi a primeira vez que parâmetros Jitter e Shimmer foram aplicados sobre análise de ECG e, portanto, certamente devem ser melhor adaptados, principalmente com relação à referência de cálculo do Shimmer. Ficou claro, no modelo primário, que o monitoramento da contração ventricular (onda R), não contribuiu muito ou é indiferente quando se observa pelo Shimmer. Entretanto, pelo Shimmer ser uma ferramenta de análise em amplitude, a presença de ruídos e desvios moderados pode ser fatal para a medida. O Jitter por outro lado, indicou que contribui mais que o Shimmer mas ainda precisa ser melhor estudado para verificar se os contextos de vizinhança como os utilizados são os mais adequados quando analisando um comportamento de AF.

As entropias por outro lado, se mostraram uma fonte de informação um pouco mais significativas, mesmo que aplicadas ao segmento como um todo. Com relação às não convergências dos treinamentos da LSTM na utilização das frequências instantâneas, entropia espectral e o próprio sinal, isto pode ter sido causado por conta da forte presença de ruído e artefatos sobre os sinais de entrada. Pelo fato de não ter sido aplicada uma etapa de filtragem e remoção de tendências do sinal para que fossem utilizados na forma bruta, divergências como essas são recorrentes. Entretanto, notou-se que a transformada Scattering superou este tipo de dificuldade, como já era esperado. Porém o desempenho apresentado pela LSTM quando tendo como entrada a decomposição Scattering dos segmentos, não foi significativo comparado ao estado da arte ou a outras combinações de características. Por isso, afirma-se que estes recursos tempo-frequência, especialmente aqueles que não levaram a LSTM a convergir, devem ser melhor trabalhados e analisados numa etapa de seleção de características para verificar o grau de relação entre as representações e os padrões das classes. Além disso, pode-se experimentar a associação destes recursos com outros ou aplicá-los a modelos de aprendizagem diferentes. No caso da transformada Scattering, por exemplo, pode ser interessante a utilização de uma etapa de processamento de imagem, já que o mapa de decomposição resultante tem este mesmo formato, ou a utilização de redes convolucionais.

Sobre os modelos de aprendizagem explorados, notou-se que no caso da MLP, o refazer das características sob referências de pico das ondas R mais precisas quando as utilizadas no modelo primário, trouxe a uma melhoria expressiva de acurácia. Assim, conclui-

se que é de extrema relevância que as marcações estejam muito bem localizadas quando se realiza extração de características. Além disso, notou-se que ao melhorar a representação do conjunto de entrada, mesmo com o aumento significativo do número de amostras, a rede necessitou de menos recursos estruturais para compreender e diferenciar os padrões, nomeadamente o número de camadas escondidas foi menor. Ainda, afirma-se que, pela tendência observada de melhoria no desempenho à medida que mais informações são concatenadas à entrada, pode ser dada a continuidade da exploração do modelo agregando outros tipos de recursos aos já obtidos. Sem contar que, já neste trabalho, é fortemente recomendada uma etapa de análise da significância dos recursos extraídos. No caso do desempenho da LSTM, notou-se uma clara superioridade desta sobre a MLP, confirmando a expectativa criada sobre este tipo de aprendizado profundo. Muito se deve à sua forte capacidade em lidar com séries temporais e as dependências que carregam. Entretanto, o tempo de processamento para o treinamento é extremamente maior, o que limita a quantidade de parâmetros e de combinações que podem ser testados no desenvolvimento de um trabalho como este. Outro ponto interessante acerca da LSTM é a possibilidade de introduzir sequências de diferentes tamanhos como exemplos, o que possibilita uma futura construção de conjuntos que sejam variantes na duração.

Esta investigação faz parte do desenvolvimento de uma série de projetos mais abrangentes que têm como alvo a criação de um sistema de inteligência artificial para auxílio ao diagnóstico médico. Neste trabalho, a colaboração promovida foi acerca do desempenho das redes MLP e LSTM para análise e classificação de ECGs brutos (ou seja, sem remoção de ruídos ou tendências) e de longo termo, a partir da geração e utilização de um conjunto de características. Com isso, algumas projeções podem ser apresentadas para continuidade e/ou melhoramento deste estudo.

5.1 Trabalhos Futuros

Para trabalhos futuros quatro grandes áreas são indicadas.

A primeira, relacionada à utilização do Jitter e do Shimmer sobre sinais ECG, é sugerida a exploração de outros contextos de vizinhança mais abrangentes (para uma possível captura da perturbação de longo termo), sobre outras referências (como no caso da amplitude das ondas P e T) e uma avaliação estatística quanto à presença de ruído, artefatos e deformações.

A segunda, voltada aos modelos de aprendizagem MLP e LSTM, sugere-se que sejam agregados novos recursos extraídos dos sinais ao conjunto de características já construído. Alinhado a isto, deve ser implementada uma etapa de avaliação e seleção de características ou de transformação para redução de complexidade. Para essa etapa, pode-se utilizar por exemplo, métodos de correlação ou algoritmos genéticos, e para de transformação, auto codificadores, análise da componente principal, análise de discriminantes linear, entre outros. Além disso, é ideal a investigação de outras combinações dos parâmetros estruturais tanto da MLP quanto da LSTM, comparando com os melhores modelos obtidos neste trabalho. Por exemplo, para LSTM propõe-se aumentar o contexto dos segmentos de entrada e empilhar mais de uma camada para observação do seu desempenho sobre estas condições.

Em terceiro lugar, propõe-se a continuação da investigação das técnicas de transformação tempo-frequência, especialmente a utilização da transformada Scattering sobre os sinais de ECG. Esta investigação pode ser feita pela associação de etapas de processamento de sinais (assim como de imagens), como feito por camadas convolucionais e auto codificadores. A ideia é potencializar as descrições geradas por estas ferramentas ao mesmo tempo que se favorece das suas propriedades. Ainda, deve-se procurar corrigir as causas que levaram os métodos de aprendizagem à não convergência, além da experimentação de outras arquiteturas de redes como SVM, KNN e GRU.

Em quarto lugar, recomenda-se explorar a implementação do sistema em um protótipo físico para execução de diagnósticos em tempo real. Para isso, devem ser levados em conta parâmetros de complexidade computacional dos métodos, tempo de execução da rede neuronal para novas amostras, a dimensão da aplicação (se é voltada para dispositivos *wearable* ou para constituir um software de auxílio ao diagnóstico em clínicas e hospitais, etc) e as tecnologias disponíveis. Além disso, sugere-se que os modelos sejam avaliados, antes, quanto à variabilidade dos sinais em um novo conjunto de sujeitos, que pode ser obtido por cruzamentos de bases de dados ou por aquisição de gravações em parcerias com academias de medicina, clínicas de exames ou hospitais.

REFERÊNCIAS

Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Inf Sci (Ny)* 2017a;405:81–90. <https://doi.org/10.1016/j.ins.2017.04.012>.

Acharya U. Rajendra, Fujita H, Oh SL, Raghavendra U, Tan JH, Adam M, et al. Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network. *Futur Gener Comput Syst* 2018;79:952–9. <https://doi.org/10.1016/j.future.2017.08.039>.

Acharya Udyavara Rajendra, Hagiwara Y, Koh JEW, Oh SL, Tan JH, Adam M, et al. Entropies for automated detection of coronary artery disease using ECG signals: A review. *Biocybern Biomed Eng* 2018;38:373–84. <https://doi.org/10.1016/j.bbe.2018.03.001>.

Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, et al. A deep convolutional neural network model to classify heartbeats. *Comput Biol Med* 2017b. <https://doi.org/10.1016/j.compbiomed.2017.08.022>.

Alickovic E, Subasi A. Effect of Multiscale PCA De-noising in ECG Beat Classification for Diagnosis of Cardiovascular Diseases. *Circuits, Syst Signal Process* 2015;34:513–33. <https://doi.org/10.1007/s00034-014-9864-8>.

Andersen RS, Peimankar A, Puthusserypady S. A deep learning approach for real-time detection of atrial fibrillation. *Expert Syst Appl* 2019;115:465–73. <https://doi.org/10.1016/j.eswa.2018.08.011>.

Aydin S, Saraoğlu HM, Kara S. Log energy entropy-Based EEG classification with multilayer neural networks in seizure. *Ann Biomed Eng* 2009;37:2626–30. <https://doi.org/10.1007/s10439-009-9795-x>.

Bein B. Entropy. *Best Pract Res Clin Anaesthesiol* 2006;20:101–9. <https://doi.org/10.1016/j.bpa.2005.07.009>.

Benitez D, Gaydecki PA, Zaidi A, Fitzpatrick AP. The use of the Hilbert transform in ECG signal analysis. *Comput Biol Med* 2001;31:399–406. [https://doi.org/10.1016/S0010-4825\(01\)00009-9](https://doi.org/10.1016/S0010-4825(01)00009-9).

Boashash B. Estimating and Interpreting The Instantaneous Frequency of a Signal—Part 1: Fundamentals. *Proc IEEE* 1992;80:520–38. <https://doi.org/10.1109/5.135376>.

Borghi PH, Teixeira JP, Borges RC. Atrial fibrillation classification based on MLP networks by extracting Jitter and Shimmer parameters. *Procedia Comput. Sci., Elsevier B.V.*;

2020, p. 1–8.

Bruna J, Mallat S. Invariant scattering convolution networks. *IEEE Trans Pattern Anal Mach Intell* 2013;35:1872–86. <https://doi.org/10.1109/TPAMI.2012.230>.

Bruns A. Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *J Neurosci Methods* 2004;137:321–32. <https://doi.org/10.1016/j.jneumeth.2004.03.002>.

Chauhan S, Vig L. Anomaly detection in ECG time signals via deep long short-term memory networks. *Proc 2015 IEEE Int Conf Data Sci Adv Anal DSAA 2015* 2015:1–7. <https://doi.org/10.1109/DSAA.2015.7344872>.

Chauhan S, Vig L, Ahmad S. ECG anomaly class identification using LSTM and error profile modeling. *Comput Biol Med* 2019;109:14–21. <https://doi.org/10.1016/j.compbiomed.2019.04.009>.

Cheng P, Dong X. Life-threatening ventricular arrhythmia detection with personalized features. *IEEE Access* 2017;5:14195–203. <https://doi.org/10.1109/ACCESS.2017.2723258>.

Dang H, Sun M, Zhang G, Qi X, Zhou X, Chang Q. A Novel Deep Arrhythmia-Diagnosis Network for Atrial Fibrillation Classification Using Electrocardiogram Signals. *IEEE Access* 2019;7:75577–90. <https://doi.org/10.1109/ACCESS.2019.2918792>.

Elhaj FA, Salim N, Harris AR, Swee TT, Ahmed T. Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Comput Methods Programs Biomed* 2016;127:52–63. <https://doi.org/10.1016/j.cmpb.2015.12.024>.

Fan X, Chen R, He C, Cai Y, Wang P, Li Y. Toward Automated Analysis of Electrocardiogram Big Data by Graphics Processing Unit for Mobile Health Application. *IEEE Access* 2017;5:17136–48. <https://doi.org/10.1109/ACCESS.2017.2743525>.

Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. *Comput Methods Programs Biomed* 2018a;161:1–13. <https://doi.org/10.1016/j.cmpb.2018.04.005>.

Faust O, Shenfield A, Kareem M, San TR, Fujita H, Acharya UR. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Comput Biol Med* 2018b;102:327–35. <https://doi.org/10.1016/j.compbiomed.2018.07.001>.

Fuster V, Rydén LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, et al. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the European Society of Cardiology Committee for practice guidelines

(Writing committee to revise the 2001 guidelines for the management of patients with atrial fibrillation). *Circulation* 2006;114:257–354.

<https://doi.org/10.1161/CIRCULATIONAHA.106.177292>.

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K SH. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals 2003:5.

Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 2000;101. <https://doi.org/10.1161/01.cir.101.23.e215>.

Harris FJ. On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *Proc IEEE* 1978;66:51–83. <https://doi.org/10.1109/PROC.1978.10837>.

Haykin S. *Redes neurais: princípios e prática*. Bookman Editora; 2007.

He R, Liu Y, Wang K, Zhao N, Yuan Y, Li Q, et al. Automatic Cardiac Arrhythmia Classification Using Combination of Deep Residual Network and Bidirectional LSTM. *IEEE Access* 2019;7:102119–35. <https://doi.org/10.1109/access.2019.2931500>.

Huang H, Liu J, Zhu Q, Wang R, Hu G. A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals. *Biomed Eng Online* 2014;13:90. <https://doi.org/10.1186/1475-925X-13-90>.

Kaplan Berkaya S, Uysal AK, Sora Gunal E, Ergin S, Gunal S, Gulmezoglu MB. A survey on ECG analysis. *Biomed Signal Process Control* 2018;43:216–35. <https://doi.org/10.1016/j.bspc.2018.03.003>.

Kaur A, Agarwal A, Agarwal R, Kumar S. A Novel Approach to ECG R-Peak Detection. *Arab J Sci Eng* 2019;44:6679–91. <https://doi.org/10.1007/s13369-018-3557-8>.

Kumar M, Pachori RB, Acharya UR. Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals. *Biomed Signal Process Control* 2017;31:301–8. <https://doi.org/10.1016/j.bspc.2016.08.018>.

Kutlu H, Avcı E. A Novel Method for Classifying Liver and Brain Tumors Using Convolutional Neural Networks, Discrete Wavelet Transform and Long Short-Term Memory Networks. *Sensors (Basel)* 2019;19. <https://doi.org/10.3390/s19091992>.

Kutlu Y, Kuntalp D. Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients. *Comput Methods Programs Biomed* 2012;105:257–67. <https://doi.org/10.1016/j.cmpb.2011.10.002>.

De Lannoy G, François D, Delbeke J, Verleysen M. Weighted conditional random

fields for supervised interpatient heartbeat classification. *IEEE Trans Biomed Eng* 2012;59:241–7. <https://doi.org/10.1109/TBME.2011.2171037>.

De Lannoy G, François D, Delbeke J, Verleysen M. Weighted SVMs and feature relevance assessment in supervised heart beat classification. *Commun. Comput. Inf. Sci.*, vol. 127 CCIS, 2011, p. 212–23. https://doi.org/10.1007/978-3-642-18472-7_17.

Latif S, Usman M, Rana R, Qadir J. Phonocardiographic Sensing Using Deep Learning for Abnormal Heartbeat Detection. *IEEE Sens J* 2018;18:9393–400. <https://doi.org/10.1109/JSEN.2018.2870759>.

Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44. <https://doi.org/10.1038/nature14539>.

Li R, Zhang X, Dai H, Zhou B, Wang Z. Interpretability Analysis of Heartbeat Classification Based on Heartbeat Activity's Global Sequence Features and BiLSTM-Attention Neural Network. *IEEE Access* 2019;7:109870–83. <https://doi.org/10.1109/access.2019.2933473>.

Li W. Wavelets for electrocardiogram: Overview and taxonomy. *IEEE Access* 2019;7:25627–49. <https://doi.org/10.1109/ACCESS.2018.2877793>.

Lin CC, Yang CM. Heartbeat classification using normalized rr intervals and wavelet features. *Proc. - 2014 Int. Symp. Comput. Consum. Control. IS3C 2014*, IEEE Computer Society; 2014, p. 650–3. <https://doi.org/10.1109/IS3C.2014.175>.

Lin HY, Liang SY, Ho YL, Lin YH, Ma HP. Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals. *Irbm* 2014;35:351–61. <https://doi.org/10.1016/j.irbm.2014.10.004>.

Lynn HM, Pan SB, Kim P. A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks. *IEEE Access* 2019;7:145395–405. <https://doi.org/10.1109/ACCESS.2019.2939947>.

Manikandan MS, Soman KP. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomed Signal Process Control* 2012;7:118–28. <https://doi.org/10.1016/j.bspc.2011.03.004>.

Markides V, Schilling RJ. Atrial fibrillation: Classification, pathophysiology, mechanism and drug treatment. *Heart* 2003;89:939–43. <https://doi.org/10.1136/heart.89.8.939>.

Martis RJ, Acharya UR, Lim CM, Mandana KM, Ray AK, Chakraborty C. Application of higher order cumulant features for cardiac health diagnosis using ECG signals. *Int J Neural Syst* 2013a;23. <https://doi.org/10.1142/S0129065713500147>.

Martis RJ, Acharya UR, Lim CM, Suri JS. Characterization of ECG beats from cardiac arrhythmia using discrete cosine transform in PCA framework. *Knowledge-Based Syst* 2013b;45:76–82. <https://doi.org/10.1016/j.knosys.2013.02.007>.

Martis RJ, Acharya UR, Prasad H, Chua CK, Lim CM, Suri JS. Application of higher order statistics for atrial arrhythmia classification. *Biomed Signal Process Control* 2013c;8:888–900. <https://doi.org/10.1016/j.bspc.2013.08.008>.

McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 1943;5:115–33. <https://doi.org/10.1007/BF02478259>.

Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: Review, opportunities and challenges. *Brief Bioinform* 2017;19:1236–46. <https://doi.org/10.1093/bib/bbx044>.

MIT-BIH Atrial Fibrillation Database v1.0.0. n.d. <https://physionet.org/content/afdb/1.0.0/> (accessed April 29, 2020).

Moody GB, Mark RG. MIT-BIH Atrial Fibrillation Database 1992. <https://doi.org/10.13026/C2MW2D>.

Moridani MK, Abdi Zadeh M, Shahiazar Mazraeh Z. An Efficient Automated Algorithm for Distinguishing Normal and Abnormal ECG Signal. *Irbm* 2019;40:332–40. <https://doi.org/10.1016/j.irbm.2019.09.002>.

Murugesan B, Ravichandran V, Ram K, Preejith SP, Joseph J, Shankaranarayana SM, et al. ECGNet: Deep Network for Arrhythmia Classification. *MeMeA 2018 - 2018 IEEE Int Symp Med Meas Appl Proc* 2018:1–6. <https://doi.org/10.1109/MeMeA.2018.8438739>.

Nair SS, Paul Joseph K. Wavelet based electroretinographic signal analysis for diagnosis. *Biomed Signal Process Control* 2014;9:37–44. <https://doi.org/10.1016/j.bspc.2013.09.008>.

Oh SL, Adam M, Tan JH, Hagiwara Y, Sudarshan VK, Koh JEW, et al. Automated identification of coronary artery disease from short-term 12 lead electrocardiogram signals by using wavelet packet decomposition and common spatial pattern techniques. *J. Mech. Med. Biol.*, vol. 17, World Scientific Publishing Co. Pte Ltd; 2017. <https://doi.org/10.1142/S0219519417400073>.

Oh SL, Ng EYK, Tan RS, Acharya UR. Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types. *Comput Biol Med* 2019;105:92–101. <https://doi.org/10.1016/j.compbiomed.2018.12.012>.

Oh SL, Ng EYK, Tan RS, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput Biol Med* 2018;102:278–87. <https://doi.org/10.1016/j.compbiomed.2018.06.002>.

Park KS, Cho BH, Lee DH, Song SH, Lee JS, Chee YJ, et al. Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function. *Comput. Cardiol.*, vol. 35, 2008, p. 229–32. <https://doi.org/10.1109/CIC.2008.4749019>.

Plawiak P. Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Syst Appl* 2018;92:334–49. <https://doi.org/10.1016/j.eswa.2017.09.022>.

Prasad H, Martis RJ, Acharya UR, Min LC, Suri JS. Application of higher order spectra for accurate delineation of atrial arrhythmia. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, 2013, p. 57–60. <https://doi.org/10.1109/EMBC.2013.6609436>.

Proakis JG., Manolakis DG. *Digital Signal Processing: principles, algorithms and Applications*. 4th ed. New Jersey: Pearson Prentice Hall; 2007.

Rahhal MMA, Bazi Y, Alhichri H, Alajlan N, Melgani F, Yager RR. Deep learning approach for active classification of electrocardiogram signals. *Inf Sci (Ny)* 2016;345:340–54. <https://doi.org/10.1016/j.ins.2016.01.082>.

Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958;65:386–408. <https://doi.org/10.1037/h0042519>.

Saadatnejad S, Oveisi M, Hashemi M. LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices. *IEEE J Biomed Heal Informatics* 2019;1–1. <https://doi.org/10.1109/jbhi.2019.2911367>.

Sahoo S, Kanungo B, Behera S, Sabut S. Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities. *Meas J Int Meas Confed* 2017;108:55–66. <https://doi.org/10.1016/j.measurement.2017.05.022>.

Sharma M, Shah S, Achuth P V. A NOVEL APPROACH for EPILEPSY DETECTION USING TIME-FREQUENCY LOCALIZED BI-ORTHOGONAL WAVELET FILTER. *J Mech Med Biol* 2019a;19. <https://doi.org/10.1142/S0219519419400074>.

Sharma M, Sharma P, Pachori RB, M. Gadre V. Double density dual-tree complex wavelet transform-based features for automated screening of knee-joint vibroarthrographic signals. *Adv. Intell. Syst. Comput.*, vol. 748, Springer Verlag; 2019b, p. 279–90. https://doi.org/10.1007/978-981-13-0923-6_24.

Sharma M, Singh S, Kumar A, San Tan R, Acharya UR. Automated detection of shockable and non-shockable arrhythmia using novel wavelet-based ECG features. *Comput Biol Med* 2019c;115:103446. <https://doi.org/10.1016/j.compbimed.2019.103446>.

Sharma M, Tan RS, Acharya UR. A novel automated diagnostic system for classification of myocardial infarction ECG signals using an optimal biorthogonal filter bank. *Comput Biol Med* 2018;102:341–56. <https://doi.org/10.1016/j.compbimed.2018.07.005>.

Shen J-L, Hung J-W, Lee L-S. Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments. 1998.

Silva I, Moody GB. An Open-source Toolbox for Analysing and Processing PhysioNet Databases in MATLAB and Octave. *J Open Res Softw* 2014;2. <https://doi.org/10.5334/jors.bi>.

Silva L, Hermsdorf J, Guedes V, Teixeira F, Fernandes J, Bispo B, et al. Outliers Treatment to Improve the Recognition of Voice Pathologies. *Procedia Comput Sci* 2019;164:678–85. <https://doi.org/10.1016/j.procs.2019.12.235>.

Singh BN, Tiwari AK. Optimal selection of wavelet basis function applied to ECG signal denoising. *Digit Signal Process A Rev J* 2006;16:275–87. <https://doi.org/10.1016/j.dsp.2005.12.003>.

Standardized z-scores - MATLAB zscore. n.d. <https://www.mathworks.com/help/stats/zscore.html> (accessed April 29, 2020).

Tan JH, Hagiwara Y, Pang W, Lim I, Oh SL, Adam M, et al. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Comput Biol Med* 2018;94:19–26. <https://doi.org/10.1016/j.compbimed.2017.12.023>.

Teixeira JP, Gonçalves A. Algorithm for Jitter and Shimmer Measurement in Pathologic Voices. *Procedia Comput. Sci.*, vol. 100, Elsevier B.V.; 2016, p. 271–9. <https://doi.org/10.1016/j.procs.2016.09.155>.

Teixeira JP, Oliveira C, Lopes C. Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. *Procedia Technol* 2013;9:1112–22. <https://doi.org/10.1016/j.protcy.2013.12.124>.

Thomas M, Das MK, Ari S. Automatic ECG arrhythmia classification using dual tree complex wavelet based features. *AEU - Int J Electron Commun* 2015;69:715–21. <https://doi.org/10.1016/j.aeue.2014.12.013>.

Tripathy RK, Sharma LN, Dandapat S. Detection of Shockable Ventricular Arrhythmia using Variational Mode Decomposition. *J Med Syst* 2016;40:1–13. <https://doi.org/10.1007/s10916-016-0441-5>.

Vakkuri A, Yli-Hankala A, Talja P, Mustola S, Tolvanen-Laakso H, Sampson T, et al.

Time-frequency balanced spectral entropy as a measure of anesthetic drug effect in central nervous system during sevoflurane, propofol, and thiopental anesthesia. *Acta Anaesthesiol Scand* 2004;48:145–53. <https://doi.org/10.1111/j.0001-5172.2004.00323.x>.

Wang G, Zhang C, Liu Y, Yang H, Fu D, Wang H, et al. A global and updatable ECG beat classification system based on recurrent neural networks and active learning. *Inf Sci (Ny)* 2019;501:523–42. <https://doi.org/10.1016/j.ins.2018.06.062>.

Warrick PA, Lostanlen V, Nabhan Homs M. Hybrid scattering-LSTM networks for automated detection of sleep arousals. *Physiol Meas* 2019;40. <https://doi.org/10.1088/1361-6579/ab2664>.

Webster JG. *Medical Instrumentation: Application and Design*. 4th ed. 2009.

WFDB Toolbox for MATLAB and Octave. n.d. <https://archive.physionet.org/physiotools/matlab/wfdb-app-matlab/> (accessed April 29, 2020).

Wu Y, Yang F, Liu Y, Zha X, Yuan S. A Comparison of 1-D and 2-D Deep Convolutional Neural Networks in ECG Classification 2018.

Xia Y, Wulan N, Wang K, Zhang H. Detecting atrial fibrillation by deep convolutional neural networks. *Comput Biol Med* 2018;93:84–92. <https://doi.org/10.1016/j.combiomed.2017.12.007>.

Yang Jianli, Bai Y, Lin F, Liu M, Hou Z, Liu X. A novel electrocardiogram arrhythmia classification method based on stacked sparse auto-encoders and softmax regression. *Int J Mach Learn Cybern* 2018;9:1733–40. <https://doi.org/10.1007/s13042-017-0677-5>.

Yang Jun, Yao S, Wang J. Deep Fusion Feature Learning Network for MI-EEG Classification. *IEEE Access* 2018;6:79050–9. <https://doi.org/10.1109/ACCESS.2018.2877452>.

Yeh YC, Wang WJ, Chiou CW. Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals. *Meas J Int Meas Confed* 2009;42:778–89. <https://doi.org/10.1016/j.measurement.2009.01.004>.

Yildirim Ö. A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med* 2018;96:189–202. <https://doi.org/10.1016/j.combiomed.2018.03.016>.

Yildirim O, Baloglu UB, Tan RS, Ciaccio EJ, Acharya UR. A new approach for arrhythmia classification using deep coded features and LSTM networks. *Comput Methods Programs Biomed* 2019;176:121–33. <https://doi.org/10.1016/j.cmpb.2019.05.004>.

Yildirim O, Tan RS, Acharya UR. An efficient compression of ECG signals using deep convolutional autoencoders. *Cogn Syst Res* 2018;52:198–211.

<https://doi.org/10.1016/j.cogsys.2018.07.004>.

Yıldırım Ö, Pławiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med* 2018;102:411–20. <https://doi.org/10.1016/j.compbiomed.2018.09.009>.

Zhang C, Wang G, Zhao J, Gao P, Lin J, Yang H. Patient-specific ECG classification based on recurrent neural networks and clustering technique. *Proc. 13th IASTED Int. Conf. Biomed. Eng. BioMed 2017*, Institute of Electrical and Electronics Engineers Inc.; 2017, p. 63–7. <https://doi.org/10.2316/P.2017.852-029>.

Zhang J, Wang L, Zhang W, Yao J. A Signal Quality Assessment Method for Electrocardiography Acquired by Mobile Device. *Proc - 2018 IEEE Int Conf Bioinforma Biomed BIBM 2018 2019*:2826–8. <https://doi.org/10.1109/BIBM.2018.8621160>.

Zhang Z, Dong J, Luo X, Choi KS, Wu X. Heartbeat classification using disease-specific feature selection. *Comput Biol Med* 2014;46:79–89. <https://doi.org/10.1016/j.compbiomed.2013.11.019>.

Zihlmann M, Perekrestenko D, Tschannen M. Convolutional recurrent neural networks for electrocardiogram classification. *Comput. Cardiol.* (2010)., vol. 44, IEEE Computer Society; 2017, p. 1–4. <https://doi.org/10.22489/CinC.2017.070-060>.

Zubair M, Kim J, Yoon C. An automated ECG beat classification system using convolutional neural networks. *2016 6th Int. Conf. IT Converg. Secur. ICITCS 2016*, Institute of Electrical and Electronics Engineers Inc.; 2016. <https://doi.org/10.1109/ICITCS.2016.7740310>.