



UNIVERSIDADE ESTADUAL DO CEARÁ
CENTRO DE CIÊNCIAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
MESTRADO ACADÊMICO EM CIÊNCIA DA COMPUTAÇÃO

DIEGO ALYSSON BRAGA MOREIRA

UM SISTEMA E-HEALTH PARA ANÁLISE E DETECÇÃO DE ANOMALIAS
CARDÍACAS

FORTALEZA – CEARÁ

2018

DIEGO ALYSSON BRAGA MOREIRA

UM SISTEMA E-HEALTH PARA ANÁLISE E DETECÇÃO DE ANOMALIAS CARDÍACAS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências e Tecnologia da Universidade Estadual do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Ciência da Computação

Orientador: Prof. Emérito Dr. Joaquim Celestino Júnior

Co-Orientador: Prof. Dr. Rafael Lopes Gomes

FORTALEZA – CEARÁ

2018

Dados Internacionais de Catalogação na Publicação

Universidade Estadual do Ceará

Sistema de Bibliotecas

Moreira, Diego Alysson Braga.

Um Sistema e-Health para Análise e Detecção de Anomalias Cardíacas [recurso eletrônico] / Diego Alysson Braga Moreira. - 2018.

1 CD-ROM: il.; 4 ¼ pol.

CD-ROM contendo o arquivo no formato PDF do trabalho acadêmico com 87 folhas, acondicionado em caixa de DVD Slim (19 x 14 cm x 7 mm).

Dissertação (mestrado acadêmico) - Universidade Estadual do Ceará, Centro de Ciências e Tecnologia, Mestrado Acadêmico em Ciência da Computação, Fortaleza, 2018.

Área de concentração: Redes de Computadores.

Orientação: Prof. Ph.D. Joaquim Celestino Júnior.

Coorientação: Prof. Ph.D. Rafael Lopes Gomes.

1. Aprendizagem de Máquina. 2. Internet das Coisas. 3. e-Health. 4. Arritmias Cardíacas. 5. ECG. I. Título.

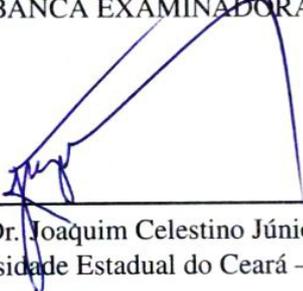
DIEGO ALYSSON BRAGA MOREIRA

UM SISTEMA E-HEALTH PARA ANÁLISE E DETECÇÃO DE ANOMALIAS CARDÍACAS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências e Tecnologia da Universidade Estadual do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Ciência da Computação

Aprovada em: 20 de Novembro de 2018

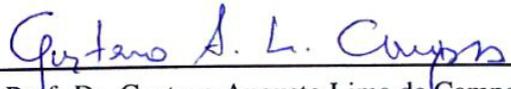
BANCA EXAMINADORA



Prof. Emérito Dr. Joaquim Celestino Júnior (Orientador)
Universidade Estadual do Ceará – UECE



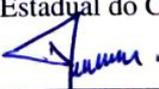
Prof. Dr. Rafael Lopes Gomes (Co-Orientador)
Universidade Estadual do Ceará – UECE



Prof. Dr. Gustavo Augusto Lima de Campos
Universidade Estadual do Ceará – UECE



Prof. Dra. Thereza Maria Magalhães Moreira
Universidade Estadual do Ceará – UECE



Prof. Dr. José Neuman de Souza
Universidade Federal do Ceará - UFC

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, em alguns momentos, a esperança para seguir. Minha esposa, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

AGRADECIMENTOS

Primeiramente a Deus que permitiu que tudo isso acontecesse, ao longo de minha vida, e não somente nestes anos como universitário, mas em todos os momentos.

Aos meus pais, pelo amor, incentivo e apoio incondicional. A minha esposa, por todas as dificuldades encontradas e enfrentadas juntos.

A esta universidade, seu corpo docente, direção e administração que oportunizaram a janela que hoje vislumbro um horizonte superior, eivado pela acendrada confiança no mérito e ética aqui presentes.

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos. Principalmente ao professores Celestino e Rafael, pela orientação, apoio e confiança.

Meus agradecimentos ao amigo Bruno Lima, companheiro de trabalhos e que fez parte da minha formação.

“Construí amigos, enfrentei derrotas, venci obstáculos, bati na porta da vida e disse-lhe: Não tenho medo de vivê-la. ”

(Augusto Cury)

RESUMO

O potencial promissor das tecnologias emergentes de Internet das Coisas (IoT) para dispositivos e sensores médicos interconectados tem desempenhado um papel importante na indústria de cuidados de saúde da próxima geração para a qualidade do atendimento ao paciente. Devido ao número crescente de idosos e portadores de deficiência, existe a necessidade urgente de uma infraestrutura de monitorização da saúde para análise dos dados destes pacientes, com intuito de evitar mortes. Healthcare IoT tem um potencial significativo pelo constante monitoramento por meio de sensores, capazes de recolher e enviar dados em tempo real. Apesar dos grandes avanços em Healthcare, uma pessoa é capaz de gerar 1 TB de informações em um dia, milhões de dados potencialmente utilizáveis em aplicações, como marketing, redes sociais e e-Health. A quantidade de dados gerados torna obrigatório o uso de técnicas que sejam capazes de analisar e gerar resultados em tempo hábil. Para diagnosticar possíveis arritmias cardíacas em pacientes pela análise do ECG (Eletrocardiograma), propõe-se um sistema de suporte à decisão, capaz de gerir os dados dos pacientes, analisando e retornando possíveis complicações cardíacas. Em uma primeira fase, para extração dos pontos de interesse da onda ECG utilizaram-se técnicas para tratamento da onda. Como uma segunda fase, técnicas como médias móveis foram utilizadas para extração das informações desejadas que serão então analisadas e classificadas por algoritmos de aprendizagem de máquina em uma terceira fase. Este sistema é capaz de classificar com alta acurácia as anomalias de um ECG humano em um curto período de tempo.

Palavras-chave: Internet das Coisas. e-Health. Arritmias Cardíacas. ECG. Aprendizagem de Máquina.

ABSTRACT

The promising potential of emerging internet of things technologies for interconnected medical devices and sensors has played an important role in the next-generation healthcare industry for the quality of patient care. Due to the growing number of elderly and disabled, there is an urgent need for a health monitoring infrastructure to analyze patient data avoiding deaths. Healthcare IoT has significant potential through constant monitoring through sensors, capable of collecting and sending data in real time. Despite major breakthroughs in Healthcare, a person is able to generate 1TB of information in one day, millions of potentially usable data in applications such as marketing, social networking, and eHealth. The amount of data generated makes it mandatory to use techniques that are capable of analyzing and generating results in a timely manner. In order to diagnose possible cardiac arrhythmias in patients through ECG (Electrocardiogram) analysis, a decision support system is proposed, capable of managing patient's data, analyzing and returning possible cardiac complications in these patients. To extract the points of interest of the ECG wave, techniques were used to treat the wave. As a second phase, techniques such as moving averages were used to extract the desired information that will then be analyzed and classified by machine learning algorithms in a third phase. This system is able to classify with high accuracy the anomalies of a human ECG in a short period of time.

Keywords: Internet of Things. e-Health. Heart Arrhythmias. ECG. Machine Learning.

LISTA DE FIGURAS

Figura 1 – Soma Vetorial das Atividades Cardíacas	22
Figura 2 – Eventos elétricos de um ECG	23
Figura 3 – Onda periódica formado por cossenos de 5, 10, 20 e 50 Hz	28
Figura 4 – FT da onda periódica formado por cossenos de 5, 10, 20 e 50 Hz	28
Figura 5 – Onda não periódica formada por cossenos de 5, 10, 20 e 50 Hz	29
Figura 6 – FT da onda não periódica formado por cossenos de 5, 10, 20 e 50 Hz	29
Figura 7 – STFT de onda formada por 4 diferentes frequências, 250 Hz, 200 Hz, 100 Hz e 50 Hz a cada 250 ms	31
Figura 8 – STFT de onda formada por 4 diferentes frequências e maior intervalo de suporte	32
Figura 9 – Algoritmo de decomposição DWT	36
Figura 10 – Resposta de Frequência em filtros Butterworth passa-baixa	38
Figura 11 – Representação da distância entre os Hiperplanos H1 e H2.	44
Figura 12 – Diferentes configurações de Kernel para Classificação de pétalas.	49
Figura 13 – Modelo de neurônio artificial não-linear.	51
Figura 14 – Diagrama de aprendizagem supervisionada.	53
Figura 15 – Gráfico da função Logística e sua derivada.	60
Figura 16 – Gráfico da função TanH e sua derivada.	61
Figura 17 – Gráfico da função ReLU e sua derivada.	62
Figura 18 – Exemplo de plataforma de Sensores IoT eHealth	71
Figura 19 – Pontos de Importância no sinal ECG e Distâncias entre ondas.	72
Figura 20 – Fluxo de Detecção dos Pontos de Importância.	72
Figura 21 – a) Sinal ECG Original. b) Sinal após a remoção da variação da linha de base. c) Sinal após a remoção da variação da linha de base e aplicação dos filtros Butterworth e Wiener	73
Figura 22 – Sinal após remoção do complexo QRS.	75
Figura 23 – Onda com pontos fiduciais detectados.	76
Figura 24 – Blocos formados pelos períodos e segmentos compreendidos entre uma onda P à T	77
Figura 25 – Classificação dos blocos ECG pelos algoritmos de aprendizagem de máquina.	78

Figura 26 – Detecção dos pontos Q, R e S no registro 103 do MIT-BIH. 82

LISTA DE TABELAS

Tabela 1 – Comparação entre os Trabalhos Relacionados.	68
Tabela 2 – Resultados da detecção do complexo QRS.	82
Tabela 3 – Resultados da detecção das ondas P e T	83
Tabela 4 – Parâmetros dos Algoritmos de Aprendizagem.	84
Tabela 5 – Resultado dos Algoritmos de Aprendizagem para as 19 diferentes anomalias por paciente	85
Tabela 6 – Resultado dos Algoritmos de Aprendizagem para as 19 diferentes anomalias por blocos	86
Tabela 7 – Melhor conjunto de parâmetros para os Algoritmos de Aprendizagem.	86
Tabela 8 – Resultado dos Algoritmos de Aprendizagem para as 4 diferentes anomalias por bloco	87

LISTA DE ABREVIATURAS E SIGLAS

Acc	Acurácia
ADC	Conversor Analógico para Digital
AV	Atrioventricular
CNN	Redes Neurais Convolucionais
CPU	Unidade Central de Processamento
CWT	Transformada de Wavelet Contínua
DAC	Conversor digital para analógico
DCV	Doenças Cardiovasculares
DWT	Transformada Discreta de Wavelet
ECG	Eletrocardiograma
FC	Frequência Cardíaca
FFT	Transformada Rápida de Fourier
FIR	Filtro de Resposta ao Impulso Finita
FN	Falso Negativo
FNN	Rede Neural Feedforward
FP	Falso Positivo
FT	Transformada de Fourier
IoT	Internet das Coisas
MIT-BIH	Base de dados do Hospital Beth Israel em conjunto com o Instituto de tecnologia de Massachusetts
MLP	Perceptron Multicamadas
P+	Valor Preditivo
QMF	Filtros de Espelhos Quadrados
ReLU	Função de Ativação Retificada
RNA	Redes Neurais Artificiais
RNN	Redes Neuras Recorrentes
Se	Sensibilidade
Sp	Especificidade
STFT	Transformada de Fourier de Tempo Curto
SV	Vetores de Suporte
SVD	Decomposição em Valores Singulares

SVM	Máquinas de Vetores de Suporte
TB	Terabyte
TDIDT	Top-Down Induction of Decision Tree
TIC	Tecnologias de Informação e Comunicação
UTI	Unidade de Tratamento Intensivo
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
WT	Transformada de Wavelet
Y	Índice Youden

SUMÁRIO

1	INTRODUÇÃO	16
1.1	MOTIVAÇÃO	17
1.2	OBJETIVOS	18
1.2.1	Objetivo Geral	18
1.2.2	Objetivos Específicos	18
1.3	CONTRIBUIÇÕES	18
1.4	ORGANIZAÇÃO DOS CAPÍTULOS	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	E-HEALTH	20
2.2	ECG	21
2.2.1	Ondas P	21
2.2.2	Complexo QRS	22
2.2.3	Ondas T	24
2.2.4	Segmentos e Intervalos	24
2.3	TRANSFORMADAS WAVELET	25
2.3.1	Transformada de Fourier	26
2.3.2	Transformada de Fourier de tempo curto	30
2.3.3	Transformada Contínua de Wavelet	32
2.4	TRANSFORMADA DISCRETA DE WAVELET	34
2.5	FILTROS DIGITAIS	37
2.5.1	Filtro Butterworth	37
2.5.2	Filtro Wiener	39
2.5.3	Remoção da Variação da Linha de Base	39
2.6	ÁRVORE DE DECISÃO	40
2.7	FLORESTA ALEATÓRIA	41
2.8	MÁQUINAS DE VETORES DE SUPORTE	42
2.9	REDES NEURAS ARTIFICIAIS	50
2.9.1	Função de Ativação	51
2.9.2	Regras de Aprendizagem	52
2.9.3	Aprendizagem Profunda	56
2.9.3.1	Perceptrons de Múltiplas Camadas (MLP)	57

2.9.3.2	Rede Neural Convolutacional	62
2.9.3.3	Rede Neural Recorrente (RNN)	64
3	TRABALHOS RELACIONADOS	65
4	PROPOSTA	69
4.1	FILTRAGEM DA ONDA E DETECÇÃO DOS PONTOS DE IMPORTÂNCIA.	71
4.2	CLASSIFICAÇÃO DOS PACIENTES.	75
4.2.1	Blocos de interesse:	76
4.2.2	Classificação de Anomalias:	77
5	RESULTADOS	79
5.1	DETECÇÃO DOS PONTOS DE IMPORTÂNCIA.	79
5.2	CLASSIFICAÇÃO DAS ANOMALIAS DOS PACIENTES.	83
6	CONCLUSÕES E TRABALHOS FUTUROS	88
	REFERÊNCIAS	89
	GLOSSÁRIO	90

1 INTRODUÇÃO

Doenças cardiovasculares (DCV) representam a principal causa de mortes no Brasil. No período de 2004 a 2014 tais doenças foram responsáveis por 3.493.459 óbitos, 29% do total, portanto calcula-se uma morte a cada 40 segundos. As DCV causam o dobro de mortes em comparação com todos os tipos de câncer juntos, 2,3 vezes mais que todas as causas externas, como acidente e violência, três vezes mais que as doenças respiratórias e 6,5 vezes mais que todas as infecções incluindo a AIDS. O alerta, a prevenção e o tratamento adequados podem reverter essa grave situação (Sociedade Brasileira de Cardiologia, 2017). No tratamento das cardiopatias é necessário o constante monitoramento cardíaco do paciente. Esse monitoramento é tipicamente realizado por um eletrocardiógrafo.

Várias pesquisas têm sido desenvolvidas e, geralmente, têm abordado vários problemas pertinentes aos processos que podem ser automatizados em ambientes hospitalares. Com isso, outros requisitos têm surgido com a aplicação da computação à área médica, desta forma, colaborando para melhorar e definir um conjunto de novos processos para este ambiente.

Ressalta-se, ainda, que a automação na área de saúde necessita de muitos cuidados, efetivamente por tratar-se de vidas humanas e, principalmente, porque alguns processos demandam aplicações críticas – por exemplo, monitoramento de pacientes crônicos submetidos a cirurgias cardíacas e que estão internados em Unidades de Terapia Intensiva (UTI) e precisam de um monitoramento crítico (Begg et al., 2007) . Assim, a necessidade de inovação tecnológica e, conseqüentemente o surgimento de novos produtos para a área é imprescindível, visto que a maioria dos hospitais ainda realiza de forma manual muitos procedimentos que já poderiam ser realizados automaticamente, estes aspectos dificultam também o controle de dados e o gerenciamento das informações, com possibilidade de erros de diagnósticos, que poderiam ser evitados com tal tipo de monitoramento.

A maioria dos sistemas utilizados no monitoramento destas situações depende de um sistema de comunicação para transmitir os dados adquiridos pelo eletrocardiógrafo para o servidor central. Muitos métodos computacionais foram propostos para automatizar o processo de análise de ondas de eletrocardiograma (ECG), ajudando médicos a diagnosticar anomalias em exames de ECG. Mesmo que tenham atingido desempenhos muito bons, vários deles dependem de algoritmos computacionalmente complexos e podem ser muito custosos em termos de uso da Unidade de Processamento (CPU) e memória. O principal objetivo deste trabalho é propor um método capaz de classificar as ondas elétricas do ECG, com alta acurácia e rápida resposta, em

diversos tipos de anomalias.

A classificação automática do ECG utiliza técnicas de diversas áreas da computação, tais como estatísticas computacionais, reconhecimento de padrões, redes neurais, máquinas de vetor suporte, dentre outras. A análise de ECG inicia-se, normalmente, com a detecção das ondas, sendo que esse estágio deve ser iniciado por algum algoritmo de detecção de QRS, um complexo formado por ondas Q, R e S que representa a despolarização do ventrículo e em uma segunda fase as ondas P e S. A extração de Características é a fase que segue a detecção de cada uma das ondas, e é responsável por transformar o sinal original do ECG para um vetor de características significativas e que auxiliem no processo da classificação. Por fim, é efetuada a classificação das ondas do ECG. Nessa fase, diferentes tipos de classificadores e técnicas de reconhecimentos de padrões podem ser empregados, permitindo identificar as anomalias presentes em cada ondas recebida.

Este cenário pode ainda estender-se para sistemas que recebem e analisam uma grande quantidade de dados, sendo imposto assim todos os pré-requisitos pontuados, em um universo cada vez maior de informação, exigindo maior velocidade de processamento e análise. Esta é uma das características atuais dos sistemas de informação, incluindo, pela internet das coisas, cada vez mais dados, como de dispositivos de saúde.

Este trabalho propõe um sistema de suporte à decisão, capaz de tratar desde a aquisição dos pontos relevantes em um ECG, interpretação destes sinais e sua classificação em um cenário e-Health de monitoramento de pacientes.

1.1 MOTIVAÇÃO

Devido ao aumento no número de casos de mortes relacionadas a DVC, faz-se necessário um mecanismo de monitoramento de pacientes. É necessário o constante monitoramento no tratamento das cardiopatias, pois a identificação precoce, ou suspeita, reduz a taxa de mortalidade significativamente.

Logo, espera-se que tecnologias como IoT (Internet das coisas) possam minimizar o número de óbitos advindos de problemas cardíacos. Tal tecnologia tem capacidade para monitorar as atividades biológicas de um paciente por meio de sensores, capazes de se comunicar e transferir os dados recolhidos, possibilitando, assim, a análise constante de pacientes.

O sistema proposto neste trabalho visa receber os sinais dos pacientes, tais como ECG, Oxigenação sanguínea e Temperatura corporal. Analisar os dados recebidos e retornar o

estado de saúde em relação a problemas cardíacos. A arquitetura utilizada também se preocupa com a grande quantidade de dados gerados por tais pacientes, provendo um sistema com grande velocidade de resposta para a análise dos dados.

1.2 OBJETIVOS

Com o aumento da população idosa no Brasil, entre o período de 2005 e 2018, a proporção de pessoas com mais de 60 anos de idade superou a quantidade da média mundial, partindo de 9,8% para 17,4%, chegando até 24,5% segundo projeções para 2042 (DEMOGRÁFICO, 2018). Visto que a idade é um dos fatores de risco de problemas cardíacos, faz-se necessário um sistema computacional capaz de diagnosticar tais doenças afim de um diagnóstico mais rápido e precoce.

1.2.1 Objetivo Geral

Desenvolver um Sistema de suporte à decisão e-Health capaz de receber, preparar e analisar os dados dos pacientes, detectando possíveis Arritmias Cardíacas, com o objetivo de diagnosticar precocemente tais doenças nos pacientes monitorados, de forma rápida e precisa, reduzindo a taxa de mortalidade.

1.2.2 Objetivos Específicos

Para a construção de tal sistema, deve-se partir desde o recebimento dos dados, detecção e classificação das ondas do ECG até a detecção de anomalias cardíacas.

- a) Realizar leitura dos dados disponíveis na base de dados do Hospital Beth Israel em conjunto com o Instituto de tecnologia de Massachusetts (MIT-BIH).
- b) Descrever as principais ondas presentes no exame ECG, assim como seus segmentos e intervalos.
- c) Detectar anomalias cardíacas através de múltiplas técnicas de aprendizado de máquinas, gerando saídas com as possíveis anomalias detectadas.

1.3 CONTRIBUIÇÕES

A elaboração deste trabalho traz as seguintes contribuições:

- a) Elaboração de uma técnica capaz de extrair os pontos fiduciais de um ECG.

- b) Aplicação de técnicas de aprendizagem de máquina para classificação de anomalias cardíacas a partir dos pontos fiduciais de um ECG.
- c) Utilização de otimizadores de aprendizagem profunda com o objetivo de melhorar os resultados da detecção de anomalias.

1.4 ORGANIZAÇÃO DOS CAPÍTULOS

Esta dissertação está organizada em seis capítulos. O capítulo 2 apresenta uma visão geral das principais técnicas utilizadas no trabalho, partindo das especificações médicas de um ECG humano e uma revisão das técnicas necessárias para extração dos pontos fiduciais de um ECG, como transformadas de wavelet e filtros digitais. Por fim, é realizado um estudo sobre técnicas de aprendizagem de máquina.

No capítulo 3 realiza-se um levantamento dos principais trabalhos relacionados com o proposto, em cada uma das diferentes fases, detecção dos pontos fiduciais e classificação de anomalias cardíacas.

A proposta das técnicas para esta detecção e aplicação dos algoritmos de aprendizagem de máquina são apresentadas no capítulo 4, com o detalhamento total destas fases.

O capítulo 5 exhibe os resultados obtidos pela aplicação da proposta do capítulo anterior, além de definir quais dados serão utilizados e quais parâmetros foram atribuídos em cada método.

Por fim, no capítulo 6, tem-se as conclusões deste trabalho e sugestões de melhorias para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Devido ao grande crescimento da quantidade de objetos conectados à internet, temos instintivamente a ideia de internet das coisas. Ambiente este que tem como princípio a interconectividade dos objetos a ele conectado. Atualmente, temos diversos tipos de objetos que se integram a esta rede. Como exemplo de objetos integrados à rede temos desde sensores que monitoram determinada área a objetos do cotidiano, como torradeiras e geladeiras. Por este motivo, podemos perceber que tal rede está totalmente incorporada em nosso cotidiano, tendo grande impacto em nossa vida.

Esta rede já está em operação em vários meios, no transporte, automação de indústria, saúde e detecção de desastre são alguns exemplos. Pela internet das coisas podemos então, conectar e operar múltiplos objetos. Desta forma, pela sua cooperação, formar um sistema complexo, onde os objetos inseridos no meio podem sentir e atuar. Podemos por exemplo, monitorar e controlar a temperatura de um determinado ambiente por meio da rede.

IoT (Internet das coisas) oferece uma grande oportunidade para o mercado, pois tal rede estabelece ligação entre múltiplas plataformas e áreas de atuação. Espera-se para 2020 uma quantidade de 212 bilhões de entidades ligadas a tal rede (GANTZ; REINSEL, 2012). Um dos mercados com maior participação neste processo de desenvolvimento é o de cuidados médicos.

2.1 E-HEALTH

Devido a maior preocupação da saúde humana e o avanço da tecnologia, hoje temos a disposição uma ampla rede de suporte a sistemas de saúde. Uma destas áreas é a e-health. Esta área tem como intuito dar suporte à medicina por meio do uso de tecnologias de informação e comunicação (TIC). Desta forma, as informações coletadas via dispositivos podem ser estudadas e enviadas para instituições interessadas, sendo elas hospitais, clínicas ou planos de saúde.

Dentre as aplicações existentes para IoT, uma das mais promissoras e com maior impacto econômico é a de cuidados médicos, a qual espera-se um crescimento entre \$1.1 a \$2.5 trilhões anualmente pela economia global de 2025. Espera-se que tal área tenha um crescimento superior às demais aplicações em 2025.

Pela interação das Redes IoT com a tecnologia e-health, esperam-se sistemas capazes de monitorar, analisar e enviar informações sobre determinado paciente em tempo real. O potencial promissor das tecnologias emergentes de Internet de Coisas (IoT), para dispositivos e sensores médicos interconectados tem desempenhado papel importante na indústria de cuidados

de saúde, visando a qualidade e o pronto atendimento ao paciente.

Devido ao número crescente de idosos e deficientes, há necessidade urgente de infraestrutura de monitorização de saúde em tempo real, para analisar dados de saúde dos pacientes visando à diminuição do número de mortes.

HealthCare IoT é um novo conceito, ligado aos aspectos de monitoramento e diagnóstico de pacientes por sensores e atuadores. Os sinais vitais monitorados, por sensores, têm sido utilizados isolados ou em conjunto como forma de prever possíveis problemas à saúde dos pacientes. Dentre exames e sinais vitais importantes podemos citar o ECG (Eletrocardiograma), a pressão sanguínea, oxigenação, dentre outros.

2.2 ECG

Historicamente, o eletrocardiograma (ECG) tem como origem o trabalho de Galvani, com as primeiras demonstrações de atividades elétricas cardíacas e pelos registros de potenciais elétricos por Waller. A partir do galvanômetro de corda, inventado por volta de 1901 pelo médico holandês Willem Einthoven, primeiro eletrocardiógrafo prático, forneceu-se um método imediato e confiável para registrar a atividade elétrica gerada pelo coração. Tal pesquisa gerou o que conhecemos atualmente como ferramenta clínica útil e de ampla utilização para diagnósticos de doenças associadas às ondas elétricas cardíacas e, por conseguinte, aos músculos cardíacos, o exame de ECG.

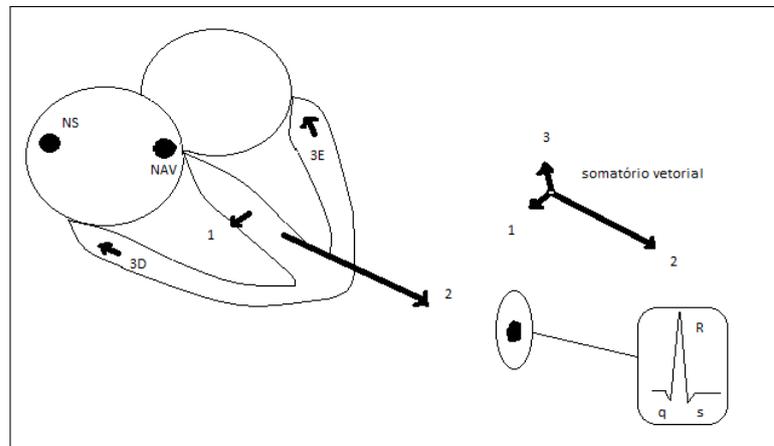
O ECG é o processo de gravação das atividades elétricas do coração por um período no qual se utilizam eletrodos na pele. Um registro gráfico das flutuações desta atividade causada por um fluxo constante de moléculas de sódio, cloro, cálcio e potássio nas células do corpo humano.

De acordo com as atividades elétricas e a soma vetorial destas atividades, como pode ser observado na Figura 1, temos como resultante, as ondas do ECG, registradas em forma de gráfico. As principais ondas de um ECG são as ondas P, Q, R, S e T. A imagem Figura 2 apresenta o ciclo cardíaco humano e as fases em que os pontos de importância são apresentados.

2.2.1 Ondas P

A onda P corresponde à despolarização atrial. Estendendo-se ao longo do tempo necessário para a excitação do nódulo sinoatrial. A duração das ondas P tendem a estar entre 0,08 a 0,11 segundos em indivíduos adultos, podendo variar de acordo com a idade.

Figura 1 – Soma Vetorial das Atividades Cardíacas



Fonte: SOMA (2013)

Quanto à amplitude de tais ondas, a tensão máxima em indivíduos sadios está entre 0,25 a 0,30 mV. Em casos de taquicardia, a tensão da onda aumenta, porém não ultrapassa seu limite máximo.

A ausência da onda P constitui arritmias, bloqueios atriais, fibrilação atrial ou flutter e ritmo juncional. A fibrilação atrial pode ser prenunciada por uma onda P alargada, entalhada ou bimodal. Em relação a sua amplitude, ondas com medida entre 2,5 mm e 3,0 mm e duração de 100 ms a 120 ms podem corresponder a hipertrofia atrial. Amplitudes maiores que 3 mm e apiculada podem corresponder ao crescimento do átrio direito. Durações maiores que 120 ms com entalhes indicam hipertrofia do átrio esquerdo. Após a onda P, tem-se um conjunto de ondas chamado complexo QRS.

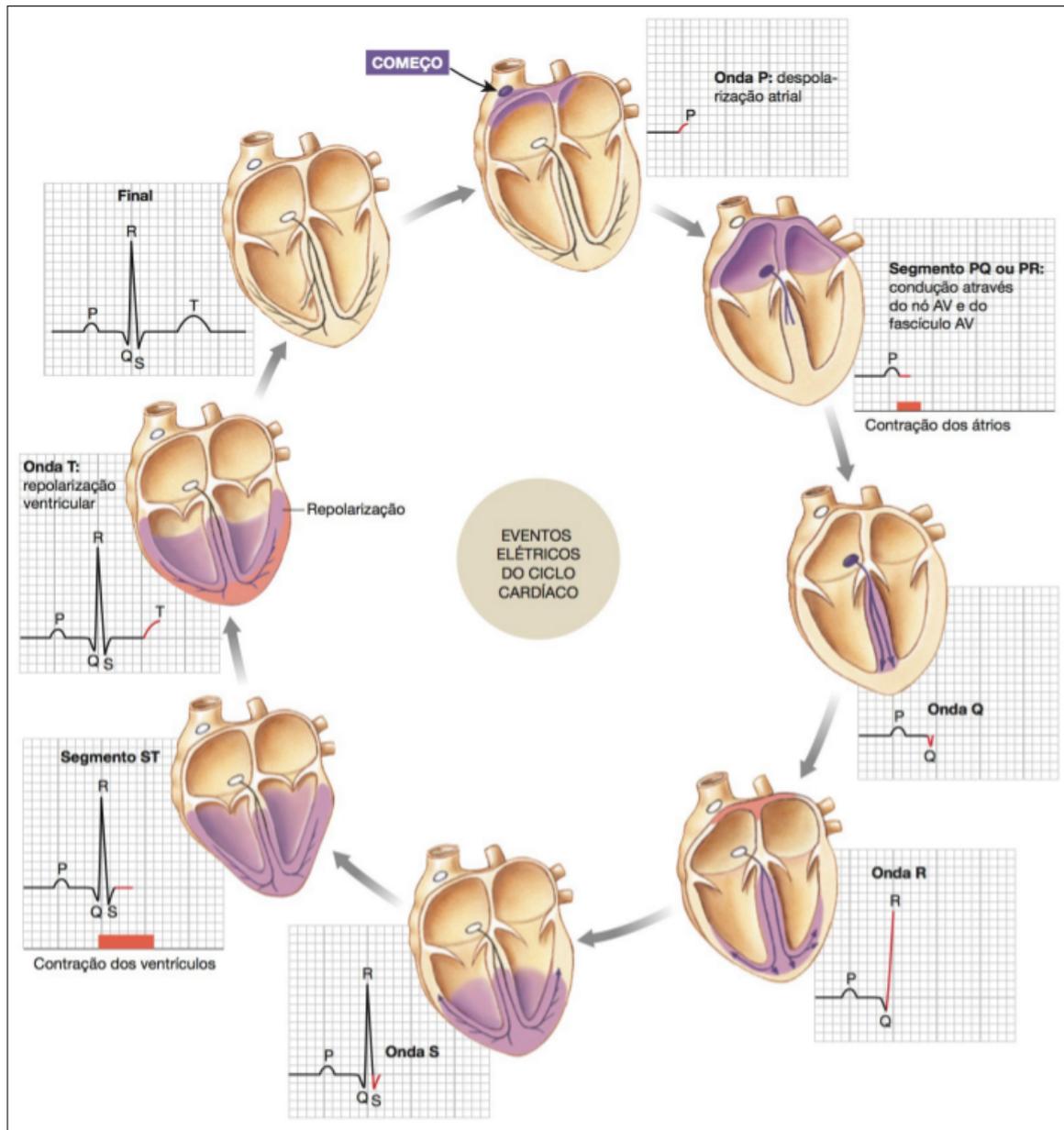
2.2.2 Complexo QRS

Normalmente com intervalo entre 0,05 a 0,10 segundos e desvio padrão de 0,016 segundos, com diferentes padrões em relação à idade, o QRS é a principal onda do ECG e corresponde à despolarização ventricular.

Um dos motivos de sua grande importância deve-se ao complexo QRS servir como referência para a contagem da frequência cardíaca (FC), período definido a partir do intervalo entre uma onda R e a seguinte ou intervalo R-R.

Existe maior facilidade para se calcular a FC a partir do intervalo R-R, visto que este é o máximo de amplitude, sendo a onda mais relevante e facilmente isolada. O intervalo normal entre dois complexos QRS sucessivos é aproximadamente 0,83s. Em um intervalo de

Figura 2 – Eventos elétricos de um ECG



Fonte: Ferraz (2018)

ondas determinado por linhas de calibração de 1s, a frequência cardíaca será 60 batimentos/min. Logo, a frequência cardíaca de $60/0,83$ vezes por minuto, ou 72 batimentos/min.

A maior parte da energia do QRS está concentrada entre 5 Hz e 15 Hz, porém nestas frequências existem muitas interferências de outras ondas, principalmente ondas de ruídos musculares, necessitando, assim, de algoritmos para remoção de ruídos em sistemas de detecção desta onda.

Quanto a anormalidades, ondas Q com duração maior que 30ms e amplitudes maiores que 3 mm ou tamanho maior que 25% do QRS correspondem a áreas ventriculares eletricamente

inativas. Complexos QRS de baixas amplitudes, menores que 8 mm de amplitude podem ser observados em obesos, enfisema pulmonar, mixedema, miocardite e derrame pericárdico. Altas amplitudes podem relatar hipertrofia no ventrículo esquerdo. Algumas alterações morfológicas podem indicar hipertrofias e bloqueio de ramo.

2.2.3 Ondas T

A onda T corresponde à repolarização ventricular e tem sua importância na amplitude, que está abaixo de 0,6 mV. Geralmente, o registro é uma deflexão arredondada e lenta. Habitualmente com polaridade igual à do complexo QRS.

Ondas T de altas amplitudes sugere hiperpotassemia. Podem apresentar pequenas amplitudes em tabagistas, após ingestão de álcool ou hiperventilação.

2.2.4 Segmentos e Intervalos

Relacionado às ondas citadas, tem-se, ainda, alguns segmentos e intervalos importantes, que podem prover informações quanto à saúde dos pacientes.

Um destes é o intervalo RR, já citado anteriormente por sua grande importância no cálculo da frequência cardíaca. A onda R tem a maior amplitude do ECG, facilitando, assim, sua detecção e sendo utilizada como referência. A duração normal deste intervalo varia entre 0,12 s a 0,22 s.

O intervalo PR representa a despolarização atrial e o retardo fisiológico do estímulo ao passar pelo nó atrioventricular (AV). Medido desde o início da onda P até a onda R, tem duração padrão de 0,12 s a 0,20 s.

Representando a sístole elétrica ventricular, o intervalo QT, tem duração variante de acordo com a frequência cardíaca, assim sendo necessário ajustar seu valor à frequência cardíaca para permitir a comparação deste intervalo em diferentes frequências cardíacas. Para definir o intervalo QT corrigido, utiliza-se a fórmula de Bazett, dada na Equação 2.1

$$QTc = QT / \sqrt{R-R} \quad (2.1)$$

O segmento ST corresponde ao período de inatividade entre a despolarização e o início da repolarização ventricular. Este segmento normalmente é isoeletrico e medido desde o final do complexo QRS até o início da onda T. Diferente dos intervalos, o segmento ST não

contém onda.

Uma vez que a onda R, por ter maior amplitude é utilizada como gatilho para a detecção das demais ondas, a melhor forma de detectar o término da sístole ventricular é no momento em que o movimento cardíaco é mínimo, a partir do intervalo RT e não pelo intervalo QT, uma vez que se utiliza a onda R. Existem diferenças entre tais intervalos, especialmente em casos de baixas frequências cardíacas.

Para a análise das ondas de um ECG, aqui apresentados, faz-se necessário um mecanismo de anulação de ruídos, visto que os sistemas de coleta podem introduzir ruídos de diferentes fontes, como movimentação do paciente ou rede elétrica. Para extração destes sinais indesejados, utilizaram-se técnicas como filtros digitais e Transformadas Discretas.

2.3 TRANSFORMADAS WAVELET

Os métodos de análise, sinais de ondas, tempo e frequência, têm uma grande utilização na análise e decomposição de tais sinais, uma vez que oferecem uma interpretação simultânea do sinal, tanto no âmbito de tempo quanto de frequência, permitindo, desta forma que, sinais locais, intermitentes e transientes sejam analisados, uma vez que alguns destes sinais podem ser suplantados devido à média inerente aos métodos apenas espectrais, como na Transformada Rápida de Fourier (FFT) (SARGOLZAEI; FAEZ; SARGOLZAEI, 2009).

As transformações matemáticas são utilizadas em sinais para a obtenção de informações adicionais. Existem inúmeras transformações matemáticas, a Transformada de Fourier é uma das mais populares.

A grande maioria dos sinais, na prática, são representados no domínio do tempo, ou seja, mensurado em função do tempo e normalmente no outro eixo está a amplitude, porém esta nem sempre é a melhor forma de representação de uma onda. Vários casos necessitam de uma informação que está oculta pela representação, a frequência.

O espectro de frequência é basicamente a frequência em que os componentes destes sinais apresentam-se. Em altas frequências, o número de ocorrências de uma variável matemática ou física são consideradas aceleradas, no contrário, em baixas frequências estas mudanças são mais suaves. Esta grandeza física é medida em ciclos por segundo ou mais comumente em Hertz (Hz), onde 1 Hz significa que um determinado evento é repetido uma vez por segundo.

Uma forma de determinar a frequência de um sinal que está no domínio do tempo é pela Transformada de Fourier (FT, do inglês Fourier Transform). Ao utilizar a FT em um sinal

no domínio do tempo, temos como resultado uma representação em frequência e amplitude. Esta informação é útil em muitos casos, uma vez que uma determinada informação não pode ser lida no sinal original, domínio do tempo, porém é vista no domínio da frequência.

Como exemplo, o sinal do Eletrocardiograma (ECG), que representa a atividade elétrica do coração, é reconhecido por cardiologistas por sua forma, qualquer variância desta forma é normalmente considerada um sintoma de condições patológicas. Apesar dos especialistas utilizarem usualmente estes sinais no domínio do tempo, com auxílio de gráficos com grades, as doenças podem não ser facilmente observadas em seu domínio original de tempo, utilizando-se do sinal no domínio da frequência pode-se facilitar a detecção de certos diagnósticos.

Uma característica da FT é a reversibilidade, ou seja, uma vez aplicada a um sinal é possível reverter esta onda gerada para a onda original. Porém tem-se acesso somente a um dos domínios, assim ao utilizar a FT, obtêm-se informações da frequência, porém as informações de quando estas frequências ocorrem são ocultadas. Estas informações não são requeridas quando se trata de sinais estacionários.

Sinais estacionários são sinais que suas frequências não se alteram durante o tempo decorrido. Desta forma, a informação de tempo não se torna necessária, pois todos componentes de frequência existem em todos os momentos.

2.3.1 Transformada de Fourier

No século 19, em 1822 o matemático francês Jean Baptiste Joseph Fourier provou que qualquer função periódica pode ser expressa por somas infinitas de funções exponenciais complexas periódicas. Esta propriedade das funções periódicas foi generalizada após um período para funções não periódicas. Após a generalização, esta propriedade tornou-se uma importante ferramenta para cálculos computacionais.

A transformada de Fourier decompõe o sinal em funções exponenciais complexas de diferentes frequências, pela Equação 2.2 e Equação 2.3.

$$X(f) = \int_{-\infty}^{\infty} x(t) * e^{-2j\pi ft} dt \quad (2.2)$$

$$x(t) = \int_{-\infty}^{\infty} X(f) * e^{-2j\pi ft} df \quad (2.3)$$

Onde t é o tempo, f a frequência e x o sinal atual. Neste caso, x representa o sinal no

domínio do tempo, enquanto X , no domínio da frequência. Sendo $X(f)$, a equação Equação 2.3 a transformada de Fourier de $x(t)$, enquanto o inverso da transformada de Fourier de $X(f)$ é calculada por $x(t)$, Equação 2.2.

O sinal $x(t)$ é multiplicado com o termo exponencial em uma frequência f e então integrado. O termo exponencial pode ser escrito como:

$$\text{Cos}(2\pi ft) + j.\text{Sin}(2\pi ft) \quad (2.4)$$

Esta equação é formada pela integral de uma parte real do cosseno da frequência f e uma parte imaginária do seno da frequência f , adicionando todos os pontos da onda na análise por meio deste produto. Com altos valores de resultante, diz-se que o sinal $x(t)$ tem um componente espectral dominante na frequência f , pois o produto das altas amplitudes, componentes de f , e o termo sinusoidal resultam em um termo de alto valor. Caso contrário, com baixos valores resultantes, o sinal não tem grandes quantidades de f em sua composição. O resultado zero, representa que a onda não contém componentes da frequência f . A integral da Equação 2.2 é realizada para cada valor de f , de mais infinito ao menos infinito.

Visto que a integral corresponde a todos os tempos possíveis, como descrito anteriormente, não há importância em que tempo um determinado componente de frequência f apresenta-se, seja qual for o tempo, têm o mesmo efeito em relação à integração. Assim, em um sinal resultante de uma FT, vemos se uma determinada frequência existe ou não no sinal original, porém não temos informações sobre o tempo que este sinal ocorreu.

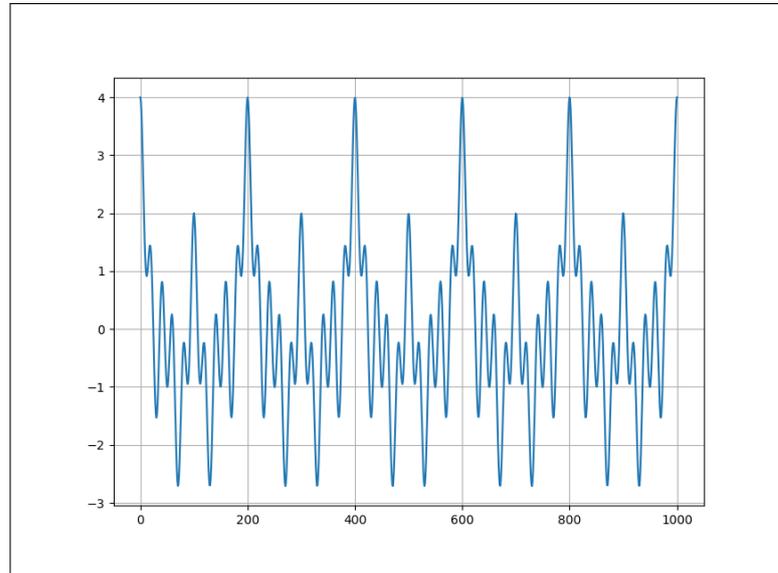
Por este motivo, FT não é ideal nos casos em que o sinal tem variação de frequência em relação ao tempo, como temos em sinais não estacionários, assim como o ECG é classificado. Pode-se observar este exemplo ao analisar duas ondas que originalmente são diferentes, porém os resultados da FT indicam ondas muito parecidas, visto que não se tem a indicação de quando um dado evento ocorreu.

A onda formada pelo sinal apresentado na Equação 2.5 é formado por componentes de frequências de 5 HZ, 10 HZ, 20 HZ e 50 HZ, porém estas frequências ocorrem a todo o momento neste sinal, ou seja, uma onda periódica, como pode ser observado na Figura 3.

$$x(t) = \text{cos}(2\pi 5t) + \text{cos}(2\pi 10t) + \text{cos}(2\pi 20t) + \text{cos}(2\pi 50t) \quad (2.5)$$

Observa-se na Figura 4 a FT da onda representada na Equação 2.5, onde percebe-se

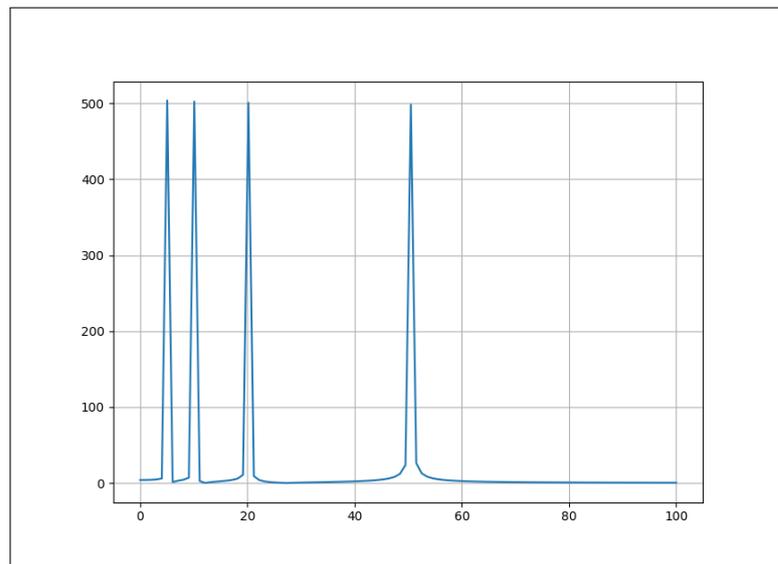
Figura 3 – Onda periódica formado por cossenos de 5, 10, 20 e 50 Hz



Fonte: Robi Polikar (1996)

que os componentes de ondas com maior significância são os de 5, 10, 20 e 50 Hz, uma vez que o sinal original é formado por tais componentes.

Figura 4 – FT da onda periódica formado por cossenos de 5, 10, 20 e 50 Hz

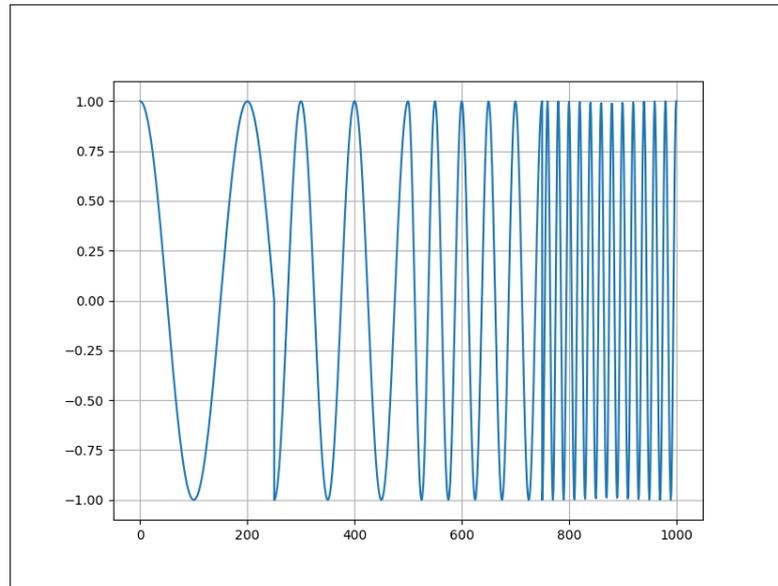


Fonte: Adaptado de Robi Polikar (1996)

Outro sinal é apresentado na Figura 5, nesta um sinal cosseno não periódico, onde seus componentes de onda não ocorrem em tempos iguais, formado pelos mesmos componentes da onda, Equação 2.5. Nesta, o sinal cosseno tem sua frequência acrescida no decorrer do tempo. A cada 360 ms o sinal realiza uma alteração, percorrendo os componentes de onda de 5, 10, 20 e

50.

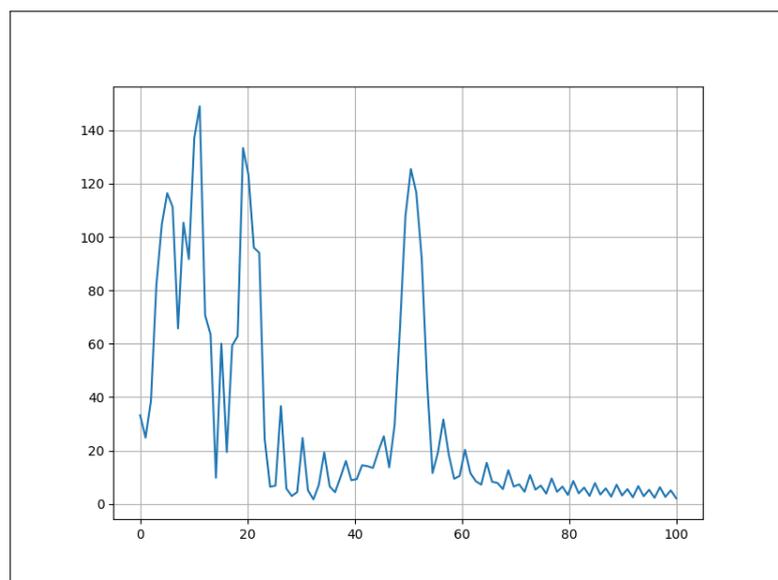
Figura 5 – Onda não periódica formada por cossenos de 5, 10, 20 e 50 Hz



Fonte: Adaptado de Robi Polikar (1996)

A onda apresentada anteriormente, na Figura 5, é formada por quatro componentes de onda, são estes, 5, 10, 20 e 50 Hz, que podem ser observadas em sua FT, Figura 6, que são os componentes de maior representatividade.

Figura 6 – FT da onda não periódica formado por cossenos de 5, 10, 20 e 50 Hz



Fonte: Adaptado de Robi Polikar (1996)

Observa-se por deste exemplo que sinais totalmente diferentes, um periódico e outro

não periódico, apresentam uma FT que indica uma informação muito similar, com exceção dos pequenos sinais apresentados na FT da Figura 6 que representa a mudança entre as frequências da onda não periódica.

Este exemplo ilustra que a FT não distingue de forma eficiente os sinais apresentados, mesmo estes tendo diferentes características. Esta característica está relacionada à falta de informações do tempo de ocorrência destes componentes, dificultando a utilização desta técnica em ondas não estacionárias.

Para solucionar a deficiência da FT com sinais não estacionários, propôs-se uma nova técnica chamada de Transformada de Fourier de tempo curto (STFT, do inglês short-term Fourier transform) que propõe dividir o sinal em porções estacionárias.

2.3.2 Transformada de Fourier de tempo curto

A partir da observação das ondas e suas frequências, é possível constatar que existem espaços de tempo, mesmo que curtos, que uma determinada onda é periódica. Pode-se, então, criar uma janela estreita o bastante para comportar esta parte estacionária do sinal. Diferente da FT o sinal é dividido em pequenos segmentos que são considerados estacionários.

A partir do início do sinal, uma função de janela é multiplicada pela onda em análise. O produto é considerado um novo sinal e aplicado a FT. Se esta porção é estacionária, como assumido, então, tem-se a representação dos primeiros períodos do sinal. Por um deslocamento da janela para a próxima posição, onde será repetido o processo até que chegue ao fim do sinal. Este processo é realizado via Equação 2.6

$$STFT_X^{(\omega)}(t', f) = \int_t [x(t)\omega(t-t')]e^{-j2\pi ft} dt \quad (2.6)$$

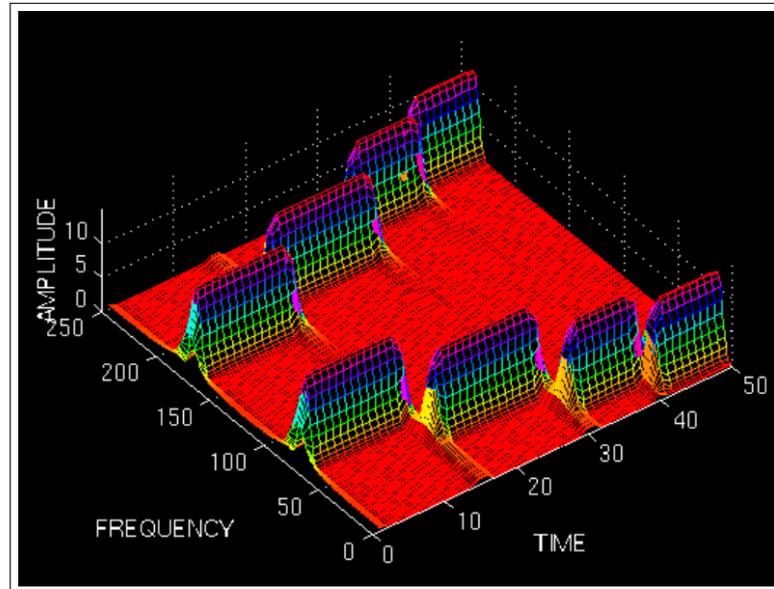
onde $x(t)$ é o sinal analisado e $\omega(t)$ a função janela.

Ao analisar um sinal com 4 diferentes frequências, não periódicas, em um intervalo de 1000 ms variando sua frequência de 250 Hz, 200 Hz, 100 Hz e 50 Hz a cada 250 ms, obtêm-se uma STFT como na Figura 7.

Observa-se no gráfico ilustrado na Figura 7 quatro diferentes picos, relacionados às frequências presentes na onda, em diferentes tempos. A STFT provê uma representação composta por componentes de tempo e frequência composição dos FT de intervalo de onda.

Contudo, percebemos que o obtido STFT representa um intervalo de tempo. Por este

Figura 7 – STFT de onda formada por 4 diferentes frequências, 250 Hz, 200 Hz, 100 Hz e 50 Hz a cada 250 ms



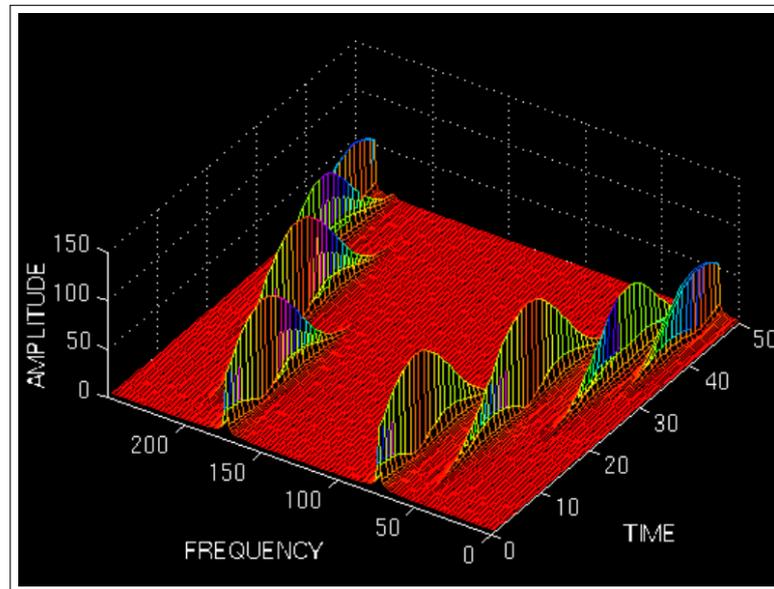
Fonte: Robi Polikar (1996)

motivo, o princípio da incerteza de Heisenberg pode ser observado nesta espécie de onda. Este princípio estabelece um limite de precisão, onde pares de propriedades de uma partícula física, ou seja, variáveis complementares, podem ser conhecidos. No contexto de análise de sinais com STFT, as propriedades físicas envolvidas são tempo e frequência. Desta forma, não é possível definir a representação exata de tempo e frequência deste sinal. Temos a capacidade de definir o intervalo de tempo que uma banda de frequências pertence.

O fator responsável por esta resolução é a largura da janela utilizada ou intervalo de suporte. No caso de uma FT, tem-se com precisão as frequências existentes em um sinal, porém não há informações de tempo. Neste contexto, FT é uma STFT com intervalo de suporte com tamanho infinito. Isto deve-se à função utilizada $\exp\{j\omega t\}$. Quanto menor o tamanho da janela, maior será a precisão temporal, porém menor a precisão da frequência. Percebe-se que os picos na Figura 7 são bem divididos em relação ao tempo, porém têm uma área de cobertura em seu componente de frequência. A STFT desta mesma onda, com um intervalo de suporte maior é representada na Figura 8.

Percebe-se que, apesar de uma boa resolução do componente frequência, tem-se uma pequena resolução do tempo. A utilização desta técnica está ligada diretamente à natureza do problema avaliado. Para solucionar a dificuldade encontrada nas STFT, com relação ao dilema da incerteza de Heisenberg, as transformadas contínuas de wavelet são utilizadas.

Figura 8 – STFT de onda formada por 4 diferentes frequências e maior intervalo de suporte



Fonte: Robi Polikar (1996)

2.3.3 Transformada Contínua de Wavelet

Como alternativa, uma vez que por STFS tem-se a dificuldade relacionada à resolução de tempo e frequência, desenvolveram-se Wavelets contínuas (CWT). Apesar de realizar uma análise parecida ao das STFT, visto que o sinal é multiplicado por uma função que é dividida em janelas e computada separadamente em seus segmentos, esta não realiza a transformada de Fourier dos sinais, além de que o tamanho da janela é adaptável, computando diferentemente para cada componente espectral. A CWT é calculada como segue na Equação 2.7:

$$CWT_x^\Psi(\tau, s) = \Psi_x^\Psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \Psi * \left(\frac{t - \tau}{s} \right) dt \quad (2.7)$$

Tem-se a transformada do sinal, formada por a translação (τ) e a escala (s). Utiliza-se também uma função de transformada, chamada de Wavelet Mãe (ψ), função com diferentes regiões são derivadas a partir desta, sendo utilizada como protótipo.

A localização da janela e sua movimentação, assim como em STFT, é realizada pela variável de translação, que corresponde às informações no domínio do tempo. A escala, uma vez que não há parâmetros de frequência, é definido como $1/freq$, que corresponde à dilatação ou compressão do sinal. Grandes escalas definem sinais mais espaçados enquanto pequenas escalas um sinal mais concentrado.

Para a análise de um dado sinal $x(t)$, uma wavelet mãe é utilizada como protótipo para a janela, que será formada pelo espaçamento, concentração ou deslocamento da onda mãe. Como exemplo de ondas que podem ser utilizadas como wavelet mãe temos a Morlet Wavelet e a função de chapéu mexicano.

No processo CWT, wavelet mãe será utilizada na janela, de forma que esta onda com o tempo será dilatada para compreender as demais frequências. Esta onda, a partir do início do sinal, é multiplicada com o mesmo, e depois integrada. Para normalização de energia do sinal, este resultado é multiplicado por $1/\sqrt{s}$. Esta operação repete-se em toda a extensão da onda em análise.

O resultado obtido nesta primeira etapa corresponde a uma linha de pontos no plano de tempo, definido pela escala utilizada. Após este processo, a escala e a translação são incrementadas por um pequeno valor continuamente no decorrer do processo, até que ele se complete.

Assim, diferente das STFT que têm resolução constante, as Transformadas de Wavelets (WT), em altas frequências, têm uma boa resolução de tempo e baixa resolução de frequência, e em baixas frequências, uma baixa resolução de tempo e boa resolução de frequência.

Como um passo para discretização das equações de FT e STFT, uma vez que os cálculos destas podem ser realizados por computadores, podem-se realizar amostragens do plano de tempo e frequência. No caso das WT, a mudança da escala pode ser utilizada para reduzir o período desta amostragem.

A partir da regra de Nyquist, em altas escalas, ou seja, baixas frequências, a taxa de amostragem pode ser diminuída. Logo, se o plano do tempo tem uma amostragem N_1 e escala s_1 , este plano pode ser representado em uma amostragem N_2 e escala s_2 , onde $s_1 < s_2$, ou seja, $f_1 > f_2$ e $N_2 < N_1$, sendo:

$$N_2 = \frac{s_1}{s_2} N_1 \quad (2.8)$$

ou

$$N_2 = \frac{f_2}{f_1} N_1 \quad (2.9)$$

Desta forma, em baixas frequências, a taxa de amostragem pode ser decrescida, diminuindo o tempo computacional. Esta operação de discretização pode ser operada sem

restrições, caso não seja necessário uma reconstrução do sinal. Porém, para reconstrução do sinal, a taxa de amostragem de Nyquist é a mínima possível.

De modo similar às FT e STFT, a discretização das transformadas de Wavelet podem ser expressas considerando a escala $s = s_0^j$, e translação $\tau = ks_0^j\tau_0$ onde $s_0 > 1$ e $\tau_0 > 0$, quando a taxa de amostragem é realizada com a amostragem diádica.

Desta forma, a função contínua de wavelet representada na Equação 2.10 pode ser representada pela Equação 2.11, inserindo-se $s = s_0^j$ e $\tau = ks_0^j\tau_0$.

$$\psi_{\tau,s} = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (2.10)$$

$$\psi_{j,k}(t) = s_0^{-j/2} \psi(s_0^{-j}t - k\tau_0) \quad (2.11)$$

Caso $\psi_{j,k}$ forme uma base ortonormal, a transformada torna-se 2.12:

$$\Psi_x^{\psi} j,k = \int x(t) \psi_{j,k}(t) dt \quad (2.12)$$

Apesar de a discretização ser realizada, tal operação tem um alto custo computacional. Para reduzir este custo, pode-se utilizar um algoritmo chamado de Transformada de Wavelet Discreta.

2.4 TRANSFORMADA DISCRETA DE WAVELET

Uma vez que a técnica de discretização das CWT possibilita a computação das mesmas, há a necessidade de um algoritmo computacionalmente eficaz para tal, além de que as discretizações das CWT não constituem as wavelets discretas, por serem muito redundantes na reconstrução dos sinais, dificultando ainda mais a computação. As Transformadas Discretas de Wavelet (DWT), por outro lado são simples de implementar, se comparadas com as CWT, além de serem constituídas com a informação necessária para reconstrução dos sinais originais.

Assim como as CWT, DWT forma uma representação tempo e escala de um sinal digital pela utilização de filtros com diferentes cortes de frequência em escalas distintas, utilizando-se de filtros passa-alta para analisar os sinais de alta frequência, assim como os de baixa frequência são analisados após os filtros de passa-baixa serem aplicados.

Nestes sinais gerados, a resolução, que define os detalhes do sinal é modificada de acordo com as operações aplicadas por meio dos filtros, enquanto que a escala é modificada pela taxa de amostragem. Assim como as CWT, os coeficientes de escala e resolução são definidos pela amostragem diádica, considerando também a escala no formato $s = s_0^j$, e translação $\tau = ks_0^j \tau_0$.

O processo de filtragem de um dado sinal $[x]$ a partir de um filtro passa-baixa de frequências médias e baixas $[h]$, corresponde a uma operação de convolução do sinal e filtro, em tempo discreto, expressa na Equação 2.13:

$$x[n] * h[n] = \sum_{-\infty}^{\infty} x[k] \cdot h[n - k] \quad (2.13)$$

Tem-se que a resolução deste sinal, após o processo de convolução é reduzida pela metade, porém sua escala é inalterada. O sinal é, então, operado com uma função de subamostragem por 2, já que metade do número de amostras é redundante, esta operação é dada pela Equação 2.14:

$$y[n] = \sum_{-\infty}^{\infty} h[k] \cdot x[n - k] \quad (2.14)$$

Pelo filtro passa faixa, a DWT analisa todo o sinal em diferentes bandas de frequência, decompondo o sinal em múltiplas frequências. Estas várias bandas são obtidas via sucessivas aplicações de filtros passa-baixa $h[n]$ e passa-alta $g[n]$. Aplica-se primeiramente os filtros no sinal original $x[h]$, e uma operação de subamostragem, resultando em dois outros sinais constituindo o primeiro nível de decomposição, que podem ser observadas pela Equação 2.15 e Equação 2.16:

$$y_{high}[k] = \sum_n x[n] \cdot g[2k - n] \quad (2.15)$$

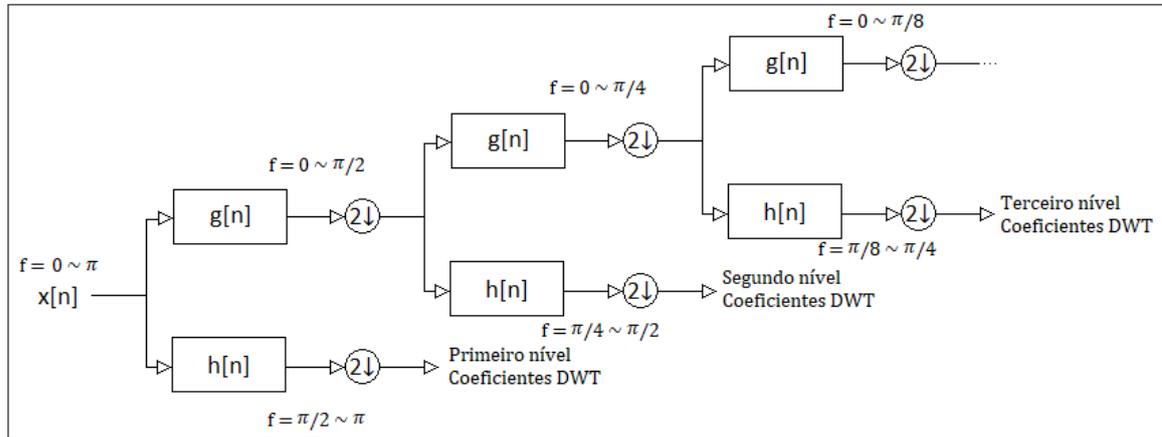
$$y_{low}[k] = \sum_n x[n] \cdot h[2k - n] \quad (2.16)$$

onde, y_{low} e y_{high} são os sinais resultantes das operações de convolução do sinal original com os filtros passa-baixa e passa-alta respectivamente, seguido de uma subamostragem.

O procedimento pode ser aplicado em vários níveis de decomposição. Em cada nível, o processo resultará em sinais com metade das amostras e frequência, gerando, assim, um sinal

com menos resolução de tempo, porém com o dobro de resolução na frequência. Este pode ser observado na imagem

Figura 9 – Algoritmo de decomposição DWT



Fonte: Adaptado de Robi Polikar (1996)

A partir do esquema anterior, as frequências com maior regularidade no sinal terão altas amplitudes em sua faixa de frequência no sinal DWT. As que tiverem menor quantidade, terão baixas amplitudes. As informações de localização destas ondas em relação ao tempo não serão perdidas, visto que temos todo o sinal, porém separado em frequências diferentes.

Uma importante propriedade para a reconstrução dos sinais é a relação entre as respostas ao impulso dos filtros passa-baixa e passa-alta, dado na Equação 2.17.

$$g[L - 1 - n] = (-1)^n \cdot h[n] \quad (2.17)$$

onde $h[n]$ é o filtro passa-baixa, $g[n]$ o filtro passa-alta e L o tamanho do filtro. Estes filtros são versões reversas um do outro e são chamados de Filtros de espelho quadrados (QMF), facilitando, assim, a reconstrução do sinal, dada à ortonormalidade que tais filtros formam. Para isto, a taxa de amostragem do sinal é acrescida novamente, aplicam-se os filtros e, então, incrementam-se os sinais. A equação responsável por obter o sinal reconstruído pode ser representada na Equação 2.18.

$$x[n] = \sum_{k=-\infty}^{\infty} (y_{high}[k] \cdot g[-n + 2k] + y_{low}[k] \cdot h[-n + 2k]) \quad (2.18)$$

Um dos mais famosos filtros ideais é o de Daubechies. No caso da não utilização destes tipos de filtros não é possível realizar uma reconstrução perfeita.

Observa-se a grande utilização de filtros digitais para o processo de análise de sinais. Neste trabalho também utilizaram-se outros tipos de filtros digitais para filtrar a onda ECG recebida que contém diversos ruídos de diferentes fontes.

2.5 FILTROS DIGITAIS

No processamento de sinal, o filtro tem como função a remoção ou divisão de partes do sinal, como ruídos ou faixas de frequência. Os filtros digitais usam um processador digital para realizar cálculos numéricos em valores amostrais do sinal.

O sinal de entrada analógico deve primeiro ser digitalizado por amostra, utilizando um ADC (Conversor analógico para digital). Os números resultantes, que representam valores amostrais do sinal de entrada, são transferidos para o processador, que realiza cálculos numéricos sobre eles. Se necessário, os resultados desses cálculos, que agora representam os valores amostrais do sinal filtrado, são emitidos por um DAC (Conversor digital para analógico) para converter o sinal de volta para a forma analógica. Nota-se que os sinais para filtros digitais são representados por uma sequência de números de uma tensão ou corrente.

Existem algumas vantagens dos filtros digitais sobre os filtros analógicos. O filtro digital é programável, isto significa que ele pode ser facilmente alterado, sem a necessidade de modificações em circuitos. Assim, estes filtros são facilmente planejados, implementados, testados e estáveis em tempo de execução.

Dois filtros digitais foram utilizados neste projeto, o filtro Butterworth e o Filtro Wiener. Estes filtros foram aplicados com a principal função de remoção de ruídos do ECG obtido.

2.5.1 Filtro Butterworth

O filtro Butterworth é desenvolvido de modo a ter uma resposta em frequência mais plana possível matematicamente na banda passante, ou seja, temos como resultado uma onda mais plana, o que significa uma atenuação dos ruídos.

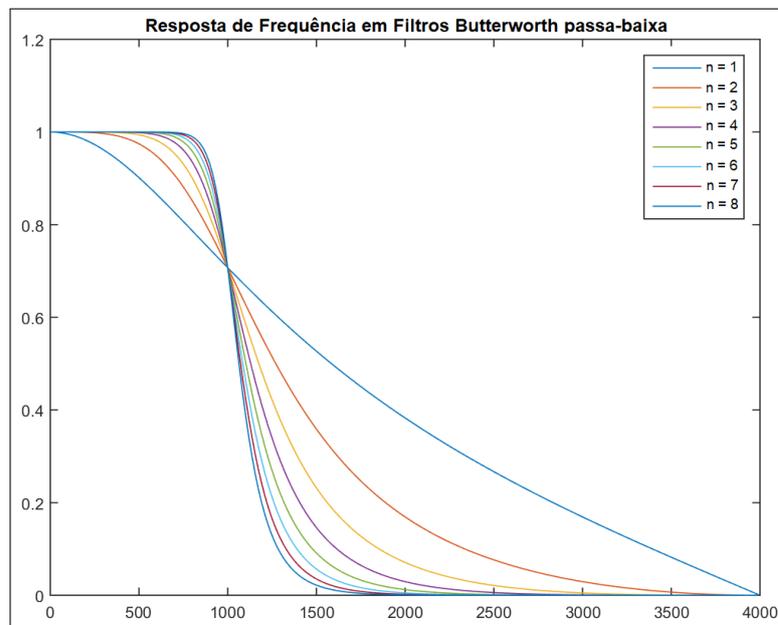
A resposta em amplitude de ordem n com o filtro passa-baixa butterworth, é dada na

Equação 2.19:

$$G_n(\omega) = |H_n(j\omega)| = \frac{1}{\sqrt{1 + (\frac{\omega}{\omega_c})^{2n}}} \quad (2.19)$$

Onde n é a ordem do filtro. Quanto maior o valor de ' n ', mais plana será a resposta do filtro, como pode ser observado na Figura 10. Na equação 2.19, ω é a frequência angular do sinal em radianos por segundo, ω_c é a frequência de corte, frequência abaixo da qual a potência na saída de um sistema é reduzida, G é o ganho total do filtro e H a função de transferência, com j sendo um número imaginário.

Figura 10 – Resposta de Frequência em filtros Butterworth passa-baixa



Fonte: Elaborada pelo Autor

O filtro passa-alta tem função oposta ao passa-baixa, ω tende a infinito enquanto o passa-baixa tende a zero. Para filtros passa-alta, tem-se a equação:

$$G_n(\omega) = |H_n(j\omega)| = \frac{|\frac{\omega}{\omega_c}|^n}{\sqrt{1 + (\frac{\omega}{\omega_c})^{2n}}} \quad (2.20)$$

A obtenção de filtros Butterworth passa-faixa pode ser realizada pela concatenação de filtros passa-baixa e passa-alta.

2.5.2 Filtro Wiener

O filtro digital Wiener foi desenvolvido por Nobert Wiener e tem como objetivo produzir uma estimativa ótima de um sinal com ruídos. Uma vez que se têm processos aleatórios estacionários, $d(n)$ e $x(n)$, estaticamente relacionados, espera-se minimizar o erro médio quadrático da estimativa de $d(n)$. Definindo o erro como $e(n) = d(n) - \hat{d}$, deseja-se calcular a seguinte minimização de ξ , Equação 2.21:

$$\min \xi = E|e(n)|^2 \quad (2.21)$$

Em Filtros Wiener para séries discretas de impulsos finitos, para um filtro de ordem p , tem-se a Equação 2.22.

$$W(z) = \sum_{n=0}^{p-1} w(n)z^{-n} \quad (2.22)$$

Onde $w(n)$ é a amostra unitária. A saída do filtro, calculada pela convolução de $w(n)$ com $x(n)$ e os coeficientes do filtro devem minimizar o erro.

Após aplicar os dois filtros apresentados anteriormente, também utilizou-se o algoritmo de remoção da variação da linha de base para preparação do sinal.

2.5.3 Remoção da Variação da Linha de Base

O sinal do ECG, obtido a partir de um eletrocardiógrafo, pode sofrer pequenas variações, chamadas de ruídos, de diferentes fontes. Uma característica que também pode estar contida neste sinal é a variação de sua linha de base, ou seja, a flutuação do sinal no decorrer do tempo, que não pode ser definida como ruído ou artefato.

Para melhor detecção dos pontos de importância a partir dos métodos computacionais, utiliza-se um método capaz de cancelar esta flutuação, utilizando as transformadas de wavelet.

Após a melhoria do sinal pela remoção da variação da linha de base e dos ruídos existentes via filtros digitais, aplicaram-se diferentes modelos de aprendizagem de máquina com o intuito de classificar as doenças e detectar possíveis anomalias em um ECG humano. Os algoritmos de aprendizagem de máquina aplicados foram Árvores de Decisão, Floresta Aleatória, Máquinas de Vetores de Suporte e Redes Neurais Artificiais.

2.6 ÁRVORE DE DECISÃO

O algoritmo Árvore de decisão é um método de aprendizagem de máquina, que visa construir um modelo de classificação ou regressão para um montante de dados, dividindo estes dados em menores porções enquanto constrói uma árvore de decisão, que é formada pelo nó raiz, nós internos, atributos da árvore, e nós folhas. O nó raiz é construído com nenhuma aresta de entrada e zero ou mais nós de saída. Os nós internos com exatamente uma aresta de entrada e dois ou mais ramos de decisão e os nós folhas representam a classificação ou decisão.

Para o processo de construção desta árvore, dá-se o nome de indução, onde constrói-se um modelo por meio do algoritmo de aprendizagem com um conjunto de dados de treino já classificados. A partir deste modelo, novas observações podem ser classificadas por um processo de indução.

Um dos mais utilizados e conhecidos algoritmos de indução é o Top-Down Induction of Decision Tree (TDIDT). Este modelo procura no conjunto de dados a melhor divisão em subconjuntos, ou seja, uma solução local. Para isto utiliza-se diferentes critérios de avaliação, como a impureza.

Uma das possíveis medidas de impureza é a entropia de um nó N , que verifica a relação de fração de elementos de uma classe no conjunto de dados de um dado ramo, mantendo o maior ganho possível para o sistema. Esta medida é dada na Equação 2.23.

$$Entropia(N) = - \sum_{C=1}^k p(C|N) \log_2 p(C|N) \quad (2.23)$$

onde, $p(C|N)$ é a fração de elementos de uma classe C no nó N , e k o número de classes. O ganho é mensurado a partir da diminuição da entropia após a divisão de um conjunto de dados para novos atributos resultantes. Esta resultante é dada pela diferença da impureza do nó pai com a soma das impurezas das partições multiplicadas pelas suas probabilidades. Uma divisão S tem um ganho associado e definido na Equação 2.24.

$$GanhoEntropia(S) = Entropia(N_p) - \sum_{j=1}^n \frac{N_j}{N_p} Entropia(N_j) \quad (2.24)$$

Tendo n como número de nós filhos resultantes na divisão, N_p o tamanho da partição associada ao nó pai e N_j a quantidade de elementos associados aos nós filhos. Os nós com o

maior ganho são divididos em novos ramos. Ramos com entropia zero são considerados nós folhas. Enquanto houver nós com entropia maior que zero, haverá divisões de novos ramos.

Outra medida baseada na impureza dos nós é o Gini, calculada via Equação 2.25.

$$Gini(N) = - \sum_{C=1}^k p(C|N)^2 \quad (2.25)$$

e o ganho desta medida é calculado por 2.26:

$$GanhoGini(S) = Gini(N_p) - \sum_{j=1}^n \frac{N_j}{N_p} Gini(N_j) \quad (2.26)$$

Uma outra medida é o erro de classificação, dada por 2.27:

$$ErroClassificao(N) = 1 - \max_C(p(C|N)) \quad (2.27)$$

sendo $\max_C(p(C|N))$ o máximo da fração de elementos das classes existentes no nó.

Outro critério que pode ser considerado é o modo de divisão das camadas da árvore (splitter). Ao criar uma nova estrutura, pode-se dividi-la por dois métodos principais. O melhor (best) seleciona o atributo de maior relevância, enquanto que no modo aleatório (random), a árvore é subdividida sem um critério exato.

Uma vez que a arvore está formada, pode-se obter resultados de classificação de acordo com suas entradas e saídas. O outro algoritmo de aprendizagem de máquinas para classificação por arvores é o de Floresta Aleatória.

2.7 FLORESTA ALEATÓRIA

Uma floresta aleatória consiste em uma coleção de árvores de decisão. Assumindo um conjunto de treino $D = (x_1, y_1), \dots, (x_n, y_n)$, constrói-se um classificador capaz de prever y a partir de x , baseando-se nos dados D . Por meio de um conjunto de classificadores $h = h(x), \dots, h_k(x)$, sendo cada $h_k(x)$ uma árvore de decisão, este conjunto chama-se de florestas aleatórias.

Os parâmetros de uma dada árvore $h_k(x)$ são dados por $\Theta_k = (\theta_{k1}, \theta_{k2}, \dots, \theta_{kp})$, que incluem a estrutura da árvore e como esta se estrutura. Estas características da h^h árvore é

definida randomicamente por parâmetros θ_k escolhidos aleatoriamente via modelos de variáveis Θ . Assim, uma árvore k recebe um classificador de acordo com a Equação 2.28.

$$h_k(x) = h(x|\Theta_k) \quad (2.28)$$

Pode-se definir uma floresta aleatória como um classificador baseado em árvores $h(x|\Theta_1), \dots, h(x|\Theta_K)$ com parâmetros Θ_k randomicamente escolhidos.

Para definir o valor final de uma classificação $f(x)$ por meio de florestas aleatórias que combinam os classificadores das árvores $h = h(x), \dots, h_k(x)$, cada uma das árvores indica uma classe, a decisão majoritária para classificação, ou a média na regressão, é definida como previsão final.

Entre as técnicas utilizadas para maior acerto na previsão dos resultados, pode-se elencar: Boosting define-se por um conjunto de técnicas menos precisas, que podem obter resultados de boa qualidade em consenso. Bagging, uso da pluralidade dos votos e randomizing, visto que para cada técnica utilizada, têm-se diferentes variáveis.

Além das técnicas baseadas em árvores aleatórias, também utilizou-se das Máquinas de vetores de suporte.

2.8 MÁQUINAS DE VETORES DE SUPORTE

Máquinas de Vetores de Suporte (SVM, do inglês, Support Vector Machines) é uma técnica de aprendizado de máquinas com resultados de aplicações comparáveis e muitas vezes superiores aos obtidos por outros algoritmos de aprendizagem. Podemos encontrar diversos trabalhos e áreas de aplicação de tal algoritmo.

Com embasamento nas teorias de aprendizado estatístico, estabelece uma série de princípios que devem ser seguidos para obtenção de classificadores com uma boa generalização. Ou seja, sua capacidade de prever corretamente a classe de novos dados do mesmo domínio.

As SVM estão no grupo de aprendizado supervisionado, uma vez que se tem a utilização de um conhecimento do ambiente por conjuntos de exemplos, entrada e saída desejada. Ou seja, a partir de um conjunto de exemplos rotulados da forma (x_i, y_i) , onde x_i representa um exemplo e y_i denota o seu rótulo, deve-se produzir um classificador capaz de prever precisamente o rótulo de novos dados a partir dos resultados anteriormente apresentados em seu treinamento. Assim, teremos uma função f , capaz de receber um dado x e fornecer uma

predição de y . Os rótulos representam o fenômeno em que se deseja fazer previsões.

Cada caso ou dado é tipicamente representado por um vetor de características. Estas características, também denotadas como atributos, expressam aspectos do caso. Um componente importante em técnicas de aprendizagem de máquina é que sejam capazes de lidar com dados imperfeitos, ou seja, ruídos. Vários tipos de dados contêm esta característica. Ou seja, as técnicas de aprendizagem de máquina devem ser robustas a ruídos presentes nos dados. Deve-se também diminuir a influência destes dados sobre o processo de indução.

O principal objetivo do algoritmo Máquinas de vetores de suporte (SVM) é encontrar um hiperplano ou plano dimensional que separe as diferentes classes do sistema. Para isto, têm-se formulações que lidam com diferentes problemas, lineares e não lineares.

No grupo das SVM lineares, pode-se separar em margens rígidas ou suaves. SVM de margens rígidas definem as fronteiras lineares a partir de dados linearmente separáveis. Para este caso, em um conjunto de treinamento T de n dados $x_i \in X$ e os rótulos respectivos $y_i \in Y$, sendo X o espaço dos dados e $Y = -1, +1$, define-se T linearmente separável se for possível separar os dados de cada classe em $+1$ e -1 a partir de um hiperplano. Estes classificadores são chamados de lineares.

Tem-se a função de um hiperplano dado pela Equação 2.29 onde $w \cdot x$ é o produto escalar de w e x , onde $w \in X$ é um vetor normal ao hiperplano e $b/\|w\|$ a distância deste hiperplano à origem, com $b \in \mathbb{R}$.

$$f(x) = w \cdot x + b = 0 \quad (2.29)$$

Esta função divide o espaço de dados em duas regiões: $w \cdot x + b > 0$ e $w \cdot x + b < 0$, pode-se então utilizar uma dada função sinal $g(x) = \text{sgn}(f(x))$ exibida na Equação 2.30, para se obter os classificadores.

$$g(x) = \text{sgn}(f(x)) = \begin{cases} -1 & \text{se } w \cdot x + b > 0 \\ +1 & \text{se } w \cdot x + b < 0 \end{cases} \quad (2.30)$$

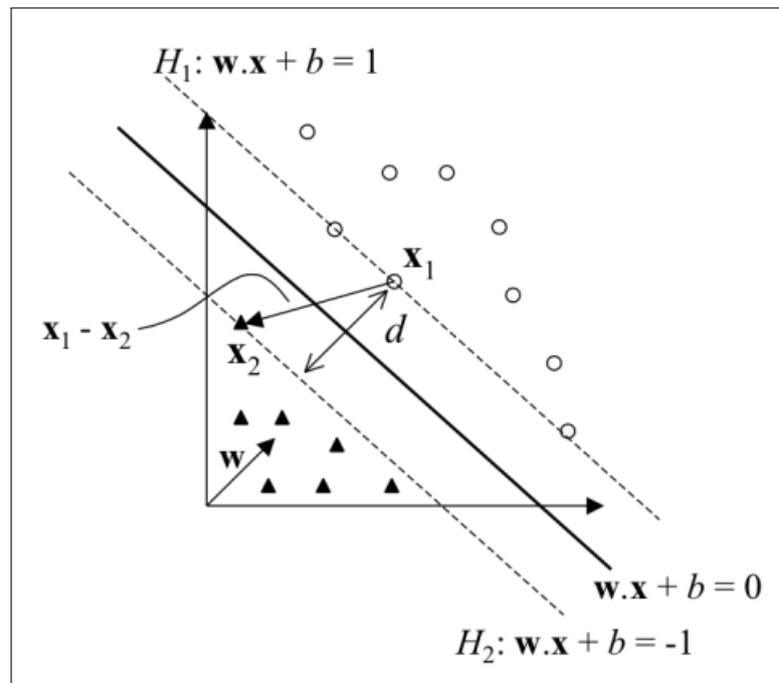
Com $f(x)$ obtêm-se infinitos hiperplanos pela multiplicação wb por uma constante. O hiperplano que mais se aproxima de $|w \cdot x_i + b|$ e satisfaça $|w \cdot x_i + b| = 1$, implica na Equação 2.31.

$$g(x) = \text{sgn}(f(x)) = \begin{cases} w \cdot x + b \geq +1 & \text{se } y_i = +1 \\ w \cdot x + b \leq -1 & \text{se } y_i = -1 \end{cases} \quad (2.31)$$

De forma resumida, esta equação pode ser representada por $y_i(w \cdot x_i + b) - 1 \geq 0, \forall (x_i, y_i) \in T$

Sendo $x_1 = w \cdot x + b = +1$ e $x_2 = w \cdot x + b = -1$, pontos dos hiperplanos H_1 e H_2 , ao projetar-se $x_1 - x_2$ em direção a w de forma perpendicular ao hiperplano definido na Equação 2.29, tem-se a distância entre os hiperplanos H_1 e H_2 . A projeção de $x_1 - x_2$ em direção a w , perpendicularmente ao hiperplano, pode ser observada na Figura 11 e esta projeção gera a Equação 2.32.

Figura 11 – Representação da distância entre os Hiperplanos H_1 e H_2 .



Fonte: Lorena Carvalho (2007)

$$(x_1 - x_2) \left(\frac{w}{\|w\|} \cdot \frac{(x_1 - x_2)}{\|x_1 - x_2\|} \right) \quad (2.32)$$

A diferença entre x_1 e x_2 resultam em $w \cdot (x_1 - x_2) = 2$, em relação à Equação 2.32, têm-se como resultado a Equação 2.33.

$$\frac{2(x_1 - x_2)}{\|w\| \|x_1 - x_2\|} \quad (2.33)$$

Uma vez que se deseja o comprimento deste vetor projetado, obtêm-se $\frac{2}{\|w\|}$. Distância esta, entre H_1 e H_2 , visto que não existem pontos entre estes, $\frac{1}{\|w\|}$ é a distância mínima entre o

plano e os dados, definida como margem geométrica do classificador linear. Para maximizar a separação destes planos em relação ao hiperplano, deve-se minimizar $\|w\|$. Logo, considera-se o problema de minimização que obtenha este resultado, a partir da Equação 2.34.

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (2.34)$$

Sujeito à restrição $y_i(w \cdot x_i + b) - 1 \leq 0, \forall i = 1, \dots, n$, não havendo dados entre as margens, de onde gera-se o nome, margens rígidas.

Considerando-se que o problema estudado tem uma função objetiva convexa com pontos do conjunto convexo, leva-se a apenas um resultado. Para resolução desta minimização, pode-se utilizar uma função Lagrangiana, associando esta equação com multiplicadores de Lagrange α_i que resultará na Equação 2.35.

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w \cdot x_i + b) - 1) \quad (2.35)$$

Para minimizar esta função, assim como desejado, deve-se maximizar α_i e minimizar w, b , levando-se ao ponto de sela $\frac{\partial L}{\partial b} = 0$ e $\frac{\partial L}{\partial w} = 0$, que resultam na Equação 2.36 e Equação 2.37

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \quad (2.36)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (2.37)$$

onde,

$$\frac{\partial L}{\partial w} = \left(\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \dots, \frac{\partial L}{\partial w_n} \right) \quad (2.38)$$

Considerando a equação com multiplicadores de Lagrange, obtêm-se a Equação 2.39.

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2.39)$$

Como restrição a esta maximização, tem-se $\alpha_i \geq 0, \forall i = 1, \dots, n$ e $\sum_{i=1}^n \alpha_i y_i = 0$.

O problema apresentado anteriormente de minimização é denominado forma primal, com w^* e b^* suas soluções, onde b^* é definido por α^* e condições de Kuhn-Tucker. A maximização apresentada na Equação 2.39 é chamada forma dual, permitindo uma forma conveniente de trabalhar-se com as constantes e escrevendo os termos em função de produtos vetoriais, o que permite o uso de funções kernel, sendo α_i^* sua solução, nestes problemas, tem-se a Equação 2.40.

$$\alpha_i^*(y_i(w^* \cdot x_i + b^*) - 1) = 0, \forall i = 1, \dots, n \quad (2.40)$$

Logo, nota-se que α^* pode ser diferente de zero, satisfazendo a equação com sinal de igualdade. Por este motivo, visto que a maioria dos α^* são nulos, w é uma combinação linear dos pontos x_i . A equação será diferente de zero, quando os dados estão sobre o Hiperplano. Para dados com $\alpha^* > 0$ denomina-se vetores de suporte (SV), uma vez que são os pontos mais próximos do plano, sendo responsável por determiná-lo com a Equação 2.41.

$$b^* = \frac{1}{n_{SV}} \sum_{x_j \in SV} \frac{1}{y_j} - w^* \cdot x_j \quad (2.41)$$

Onde, n_{SV} é a quantidade de SVs, e SV o conjunto dos SVs. Visto que $w = \sum_{i=1}^n \alpha_i y_i x_i = 0$, tem-se então 2.42

$$b^* = \frac{1}{n_{SV}} \sum_{x_j \in SV} \left(\frac{1}{y_j} - \sum_{x_i \in SV} \alpha_i^* y_i x_i \cdot x_j \right) \quad (2.42)$$

Pode-se representar o Hiperplano por uma função linear, dada na Equação 2.43, responsável por separar os dados de maior margem.

$$g(x) = \text{sinal}(f(x)) = \text{sinal}\left(\sum_{x_i \in SV} \alpha_i^* y_i x_i \cdot x + b^*\right) \quad (2.43)$$

No caso de conjuntos de dados não linearmente separáveis, que não formam um hiperplano separador, o que pode acontecer com a adição de ruídos ou a natureza do problema analisado, a análise pode ser generalizada pela introdução de N variáveis não negativas $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, chamadas variáveis de folga, para relaxar as restrições, tornando-se a Equação 2.44.

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, n \quad (2.44)$$

O objetivo da utilização de x_i é permitir que valores estejam entre os Hiperplanos. Assim, se x_i satisfaz a inequação, logo ela será zero e ξ dado como nulo, caso contrário, o termo $-\xi_i$ é adicionado do lado direito para obter a inequação. Estes casos são chamados de SVM com margens suaves. A generalização ótima do Hiperplano é considerada, para estes casos, como a Equação 2.45.

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^n \xi_i \right) \quad (2.45)$$

O termo extra $C \sum \xi_i$ mantém o controle do número de erros na classificação e torna a generalização ótima do hiperplano menos sensível. Semelhantemente a SVM de margens rígidas, pode-se transformar em forma dual, da mesma forma, por uma função Lagrangiana e derivadas parciais nulas, Equação 2.46.

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2.46)$$

Como restrição, tem-se $0 \leq \alpha_i \leq C, \forall i = 1, \dots, n$ e $\sum_{i=1}^n \alpha_i y_i = 0$.

Analogamente ao SVM de margens rígidas, para a solução dual α^* e w^* , b^* e ξ^* soluções para primal, w^* pode ser calculado da mesma maneira, enquanto ξ^* , através da Equação 2.47 .

$$\xi_i^* = \max(0, 1 - y_i \sum_{j=1}^n y_j \alpha_j^* x_j \cdot x_i + b^*) \quad (2.47)$$

b^* , advindo de α^* e de condições de Kuhn-Tucker 2.48 e 2.49.

$$\alpha_i^* (y_i (w^* \cdot x_i + b^*) - 1 + \xi_i^*) = 0 \quad (2.48)$$

$$(C - \alpha_i^*) \xi_i^* = 0 \quad (2.49)$$

Da mesma forma dos SVM com margens rígidas, os pontos x_i com $\alpha_i^* > 0$ são chamados de vetores de suporte, porém para estes pontos com $\alpha_i^* < C$, tem-se $\xi_i^* = 0$, logo estes vetores de suporte estão a uma distância $1/w^*$ do hiperplano ótimo de separação, obtido na Equação 2.48 e Equação 2.49. Os vetores de suporte com $\alpha_i^* = C$ são definidos como pontos de

erro de classificação, neste caso, se $\xi_i^* > 1$, determina-se como pontos corretamente classificados, para $0 < \xi_i^* \leq 1$, entre margens e $\xi_i^* = 0$ sobre as margens, chamados de limitados.

Toda esta construção também pode ser estendida para incluir conjuntos não lineares de solução ou com distribuição aproximadamente linear, considerando-se que nem todos os casos podem ser resolvidos satisfatoriamente, pois para uma dada classificação, os dados podem ser mais bem separados por curvas, ao invés de um hiperplano.

A partir de um conjunto de treinamento o SVM constrói um novo espaço com uma maior dimensão, onde $\Phi : X \rightarrow F$, sendo X o espaço de entrada, F o espaço de maior dimensão, chamado de espaço de características.

Deseja-se obter um espaço de características com alta probabilidade de se terem dados linearmente separáveis, através de um conjunto de entrada não linear, isto pode ser obtido através do teorema de Cover, pois a transformação seja não linear e a dimensão deste espaço de características suficientemente alta. Desta forma, pode-se encontrar um hiperplano neste novo espaço, Φ aplicando SVM linear, com maior margem de separação. Para isto, aplica-se Φ ao problema de maximização apresentado anteriormente, resultando na Equação 2.50.

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\Phi(x_i) \cdot \Phi(x_j)) \quad (2.50)$$

da mesma forma, o classificador e o b^* tornam-se a Equação 2.51 e Equação 2.52 respectivamente.

$$b^* = \frac{1}{n_{SV:\alpha^* < C}} \sum_{x_j \in SV:\alpha^* < C} \left(\frac{1}{y_j} - \sum_{x_i \in SV} \alpha_i^* y_i \Phi(x_i) \cdot \Phi(x_j) \right) \quad (2.51)$$

$$g(x) = \text{sinal}(f(x)) = \text{sinal} \left(\sum_{x_i \in SV} \alpha_i^* y_i \Phi(x_i) \cdot \Phi(x) + b^* \right) \quad (2.52)$$

Apesar de constituir-se um processo custoso, devido à alta dimensionalidade do espaço gerado, pode-se perceber que as únicas informações necessárias para o mapeamento são os cálculos do produto escalar $\Phi(x_i) \cdot \Phi(x_j)$, que pode ser obtido por Kernels, funções que recebem os pontos x_i, x_j e retornam o produto escalar destes dados no espaço desejado, resultando na Equação 2.53:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (2.53)$$

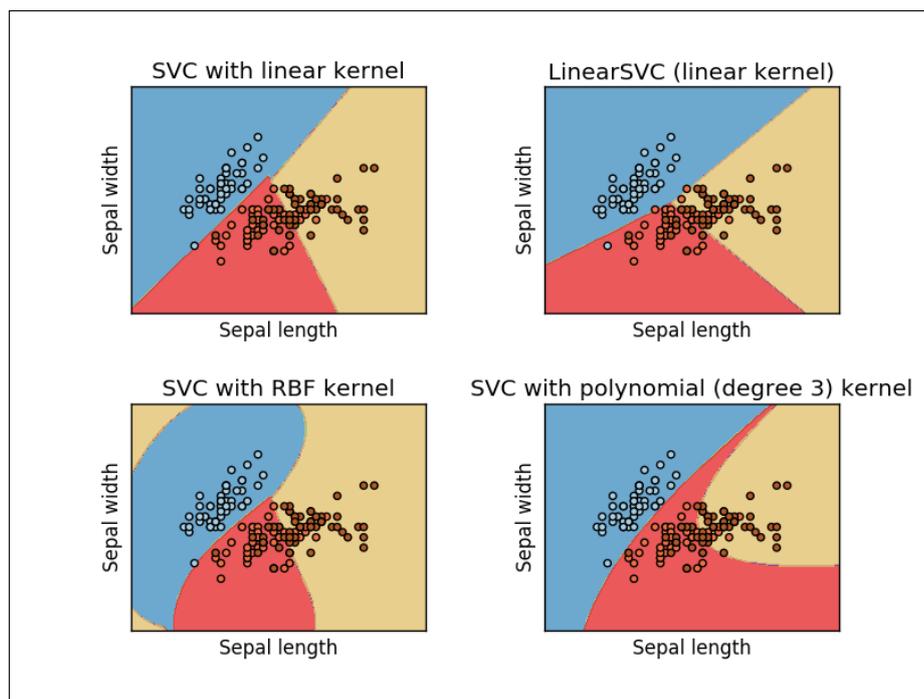
O Kernel pode ser utilizado sem conhecimento prévio do mapeamento de Φ , simplificando os cálculos e representação dos espaços criados. Este kernel é responsável por originar matrizes positivas semi-definidas K , onde um dado elemento K_{ij} pode ser definido por $K_{ij} = K(x_i, x_j)$, para $i, j = 1, \dots, n$.

Como exemplos de Kernel mais utilizados têm-se os Lineares, Polinomiais, Gaussiano e Sigmoidal. Cada um destes apresenta seus próprios parâmetros.

Linear: (x, x')
 Polinomial: $(\gamma(x, x') + r^d)$
 Gaussiano: $\exp(-\gamma||x - x'||^2)$
 Sigmoidal: $(\tanh(\gamma(x, x') + r))$

onde \tanh é a tangente Hiperbólica.

Figura 12 – Diferentes configurações de Kernel para Classificação de pétalas.



Fonte: Scikit-learn (2011)

Na Figura 12, pode-se observar as diferenças de classificação em cada um dos Kernels, retornando diferentes resultados obtidos.

Por esta técnica tem-se a separação de um espaço de dados por meio do hiperplanos. A partir destes pode-se separar classes de um determinado problema, como, por exemplo, as anomalias existentes em um ECG.

Uma outra técnica muito utilizada, porém com grandes diferenças das anteriormente apresentadas é a de Redes Neurais Artificiais, que visa aprender, por pesos e sinapses, o padrão de um conjunto de dados e suas classes.

2.9 REDES NEURAIAS ARTIFICIAIS

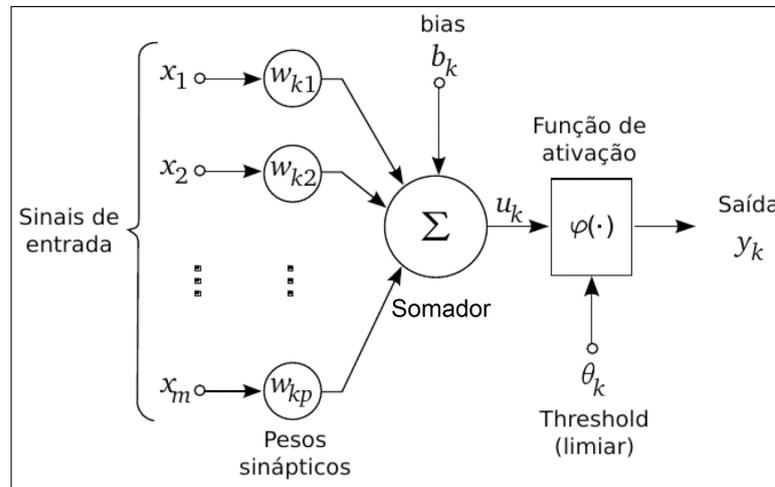
A motivação inicial para surgimento das técnicas relacionadas às redes neurais artificiais (RNA) é o reconhecimento de que os computadores digitais processam informações de forma totalmente diferente do cérebro humano, que é um sistema de processamento altamente complexo, não linear e paralelo, organizando os neurônios de forma a processar informações de forma muito mais rápida que um computador digital. Este sistema, pode, por exemplo, realizar um reconhecimento perceptivo, rosto familiar em uma cena não-familiar, em aproximadamente 100 à 200 ms, com baixas complexidades, enquanto esta tarefa pode levar dias em um computador convencional (HAYKIN, 2007). Apesar de os neurônios biológicos serem seis ordens de grandeza mais lentas que portas lógicas digitais, o cérebro compensa esta taxa com a quantidade de neurônios e conexões ativas, cerca de 10 bilhões de neurônios e 60 trilhões de sinapses.

Estas estruturas biológicas acumulam experiências e formam regras de acordo com a vivência no ambiente inserido, sendo capaz de se adaptar. O propósito de redes neurais artificiais é apresentar um modelo matemático capaz de representar organismos inteligentes, que adquirem este conhecimento pela experiência, e a conexão destes neurônios, pesos sinápticos, armazenam o conhecimento adquirido. Este procedimento é chamado de processo de aprendizagem.

As redes neurais têm alto poder de generalização, de forma paralela e distribuída, dando-lhe um aspecto de não linearidade. De acordo com o conjunto de entrada na rede, tem-se o cálculo dos pesos relacionados a esta entrada e o conjunto de valores armazenados devido à experiência obtida, retornando um valor de saída, o que lhes concede alta capacidade de adaptar-se, pois estes pesos podem ser recalculados de acordo com as amostras de entrada.

A estrutura básica de uma rede neural é o neurônio, que é a unidade de processamento fundamental. Esta estrutura é formada por três elementos básicos, um conjunto de sinapses, um somador e a função de ativação, como ilustrado na Figura 13.

O conjunto de sinapse é caracterizado pelo peso relacionado a ele, onde um sinal x_j de entrada através da sinapse j que está conectada a um neurônio k é multiplicado por este peso sináptico w_{kj} , que pode ser positivo ou negativo. O somador tem como função somar os sinais resultantes da ponderação dos sinais de entrada com os pesos sinápticos, constituindo uma

Figura 13 – Modelo de neurônio artificial não-linear.

Fonte: Elaborada pelo Autor

combinação linear. A função de ativação restringe a amplitude do sinal resultante do neurônio, normalmente representado entre o intervalo fechado $[0,1]$ ou $[-1,1]$.

Outro componente importante na arquitetura do neurônio é a *bias* b_k , que funciona como uma variável de ajuste, capaz de aumentar ou diminuir a entrada da função de ativação.

Um neurônio k pode ser descrito matematicamente por 2.54 e 2.55.

$$y_k = \phi(u_k + b_k) \quad (2.54)$$

Onde:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2.55)$$

sendo x_j o sinal de entrada, w_{kj} o peso sináptico j do neurônio k , u_k a saída resultante do combinador linear, b_k o bias, ϕ a função de ativação utilizada e y o sinal de saída.

2.9.1 Função de Ativação

A função de ativação ϕ pode ser definida de diferentes maneiras, entre as mais utilizadas, tem-se a função de Limiar, Limiar por pares, Sigmoid e Tangente hiperbólica.

A função Limiar pode ser descrita na Equação 2.56, é também conhecida como modelo McCulloch-Pits, a saída do neurônio assume valores 1 ou 0, para resultados não negativos

ou negativos, respectivamente.

$$\phi(u) = \begin{cases} 1 & \text{se } u \geq 0 \\ 0 & \text{se } u < 0 \end{cases} \quad (2.56)$$

Para função linear por pares, assume-se um valor dentro de uma região linear e delimitada por uma unidade s , expressa na Equação 2.57.

$$\phi(u) = \begin{cases} 1 & \text{se } u \geq +s \\ u & \text{se } -s > u > s \\ 0 & \text{se } u \leq 0 \end{cases} \quad (2.57)$$

Uma das mais utilizadas funções é a sigmoide e é definida por uma função estritamente crescente com um balanceamento entre um comportamento linear e não-linear. Esta tem o formato de um s . Como exemplo de função sigmoide, tem-se a função logística, expressa pela Equação 2.58.

$$\phi(u) = \frac{1}{1 + e^{-u_k/t}} \quad (2.58)$$

onde T determina a inclinação da sigmoide, quanto T tende a zero, a função assume características da função linear. Os valores obtidos a partir desta função estão compreendidos entre um intervalo contínuo de valores entre 0 e 1.

Caso desejável, a função de ativação pode ter uma forma antissimétrica em relação à origem, entre os valores $[-1,+1]$, utilizando por exemplo a função tangente hiperbólica.

Para realizar o aprendizado da rede, têm-se definidas diferentes regras de aprendizagem, as quais definem como e quanto uma rede neural deve aprender.

2.9.2 Regras de Aprendizagem

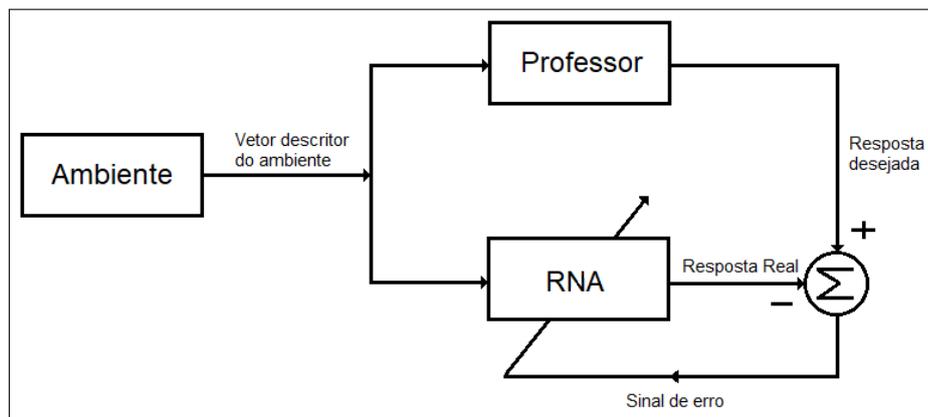
O processo de aprendizagem de uma rede neural está ligado à capacidade de esta rede aprender por meio do ambiente em que está inserida. Com a utilização das técnicas de aprendizagem, a rede neural, ajusta os pesos sinápticos e o nível bias para melhorar sua taxa de acerto diante de certo problema.

Existem várias técnicas de aprendizado para redes neurais, o conjunto de todas estas regras é chamado de algoritmo de aprendizagem. Para cada um dos modelos existentes,

existem diferenças em relação ao ajuste dos pesos sinápticos. Para construir este processo de aprendizagem, faz-se necessário conhecer as informações e como estes pesos serão atualizados. Existem três paradigmas de aprendizagem:

1. **Aprendizado supervisionado:** neste paradigma, utiliza-se um conjunto de dados em pares, classificados de entrada e saída, os quais serão utilizados como modelos de aprendizagem, ou seja, este modelo considera existir um professor que contém as informações sobre o ambiente, porém este é desconhecido pela rede neural. Este aprendizado ocorre ao comparar-se o resultado de saída, utilizando um conjunto de pesos ao padrão conhecido pela resposta desejada do professor. Se a saída é diferente do padrão desejado, o algoritmo deve ajustar os pesos da rede para reduzir o erro. Repete-se esta dinâmica para todo o conjunto de dados até que o resultado obtido atinja uma taxa considerada satisfatória. A Figura 14 demonstra o diagrama do processo de aprendizagem supervisionado.

Figura 14 – Diagrama de aprendizagem supervisionada.



Fonte: Elaborada pelo Autor

2. **Aprendizagem não-supervisionada:** Também conhecido como aprendizado auto supervisionado, classifica os pares sem o conjunto de pares ou professor, utilizando apenas os valores de entrada para treinamento da rede, classificando segundo critérios de semelhança. Estas redes utilizam neurônios como classificadores e os dados de entrada como elementos de classificação.
3. **Aprendizado híbrido:** Utiliza-se das duas técnicas anteriormente apresentadas, onde parte dos pesos são definidos pelo aprendizado supervisionado e outra parte por aprendizado não-supervisionado.

Outro fator que se deve considerar é a maneira que a informação e o ambiente se relacionam com a RNA e suas regras básicas de aprendizagem. Dentre estas regras, podemos

citar os tipos básicos:

1. **Aprendizagem por correção de erro:** Neste modelo, por comparação entre os valores de saída atual $y_k(n)$ e a saída desejada $d_k(n)$, em um mesmo instante de tempo, efetua-se a correção dos pesos sinápticos. O erro $e_k(n)$ é definido como a diferença entre o valor de saída e a saída desejada como ilustrado na Equação 2.59.

$$e_k(n) = d_k(n) - y_k(n) \quad (2.59)$$

Por meio do erro calculado, aplicam-se ajustes corretivos aos pesos sinápticos, com a intenção de aproximar o sinal de saída da resposta desejada. Para isto, deve-se minimizar uma função custo baseada no erro, dada Equação 2.60, que resulta no valor instantâneo do erro da rede $\xi(n)$. Este processo continua até que o estado da rede esteja estabilizado.

$$\xi(n) = \frac{1}{2} e_k^2(n) \quad (2.60)$$

Esta minimização resulta na regra de aprendizagem chamada regra Delta, onde tendo um valor $w_{ki}(n)$ que representa o peso sináptico de um neurônio k, ativado por um dado elemento $x_j(n)$ do vetor de sinais, em um dado tempo n. Por esta regra, o ajuste $\Delta w_{kj}(n)$ aplicado em w_{kj} no tempo n é dado na Equação 2.61.

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n) \quad (2.61)$$

sendo η uma constante que determina a taxa de aprendizagem positiva. Logo o novo valor do peso sináptico é dado na Equação 2.62.

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (2.62)$$

Observa-se que a rede aprende apenas nos momentos em que ocorrem erros, um valor pequeno destes resulta em uma aprendizagem mais lenta, porém uma taxa alta pode comprometer a estabilidade do sistema.

2. **Aprendizagem Hebbiana:** Por meio dos estudos sobre neuropsicologia, o neuropsicólogo Hebb, propôs um mecanismo de modificação de aprendizagem. Segundo este mecanismo, quando uma dada célula se aproxima o suficiente para excitar uma célula B e participa

de seu disparo, uma modificação metabólica ocorre em uma ou ambas, ou seja, se dois neurônios em uma sinapse ativam-se simultaneamente, a força desta sinapse é seletivamente aumentada, caso contrário, ativada assincronamente, a sinapse é seletivamente enfraquecida ou eliminada.

Neste modelo, o aprendizado é realizado de forma local, a mudança de pesos depende somente dos dois neurônios que estão em atividade e conectados. A forma mais simples de aprendizagem Hebbiana é dada pela Equação 2.63

$$\Delta w_{kj}(n) = \eta y_k(n) x_j(n) \quad (2.63)$$

onde η é uma constante, positiva, que determina a taxa de aprendizagem w_{kj} o neurônio k com sinais pré-sinápticos e pós-sinápticos representados por x_j e y_k .

3. **Competitiva:** Neste modelo, os neurônios competem pelo direito de tornarem-se ativos, uma vez que somente um neurônio responde a um determinado estímulo do ambiente. Três componentes básicos podem ser enumerados. O primeiro componente são os conjuntos de neurônios que estão interligados e com diferentes pesos sinápticos, onde cada um destes têm respostas únicas em relação a um determinado sinal de entrada. O segundo elemento, um limite de força para cada neurônio. E o terceiro elemento, um mecanismo que permite que os neurônios possam competir entre si.

Neste tipo de aprendizagem, os neurônios tornam-se especialistas em reconhecer certo padrão de entrada, tornando-se adequados para descobrir características estatísticas salientes. Para ser considerado como vencedor, o campo local induzido, ou seja, a ação combinada de todas as entradas em um neurônio k , em v_k para uma entrada x deve ser maior que os demais neurônios da rede. O neurônio deve então deslocar os pesos entre os neurônios. Onde não ocorre resposta a um determinado estímulo, também não existe aprendizado. Ao vencer a competição, cada nó de entrada libera um peso que é distribuído entre os nós de entrada ativo. A variação Δw_{kj} em um neurônio w_{kj} é calculada na Equação 2.64.

$$\Delta w_{kj} = \begin{cases} \eta(x_j - w_{kj}) & \text{se } k \text{ vence} \\ 0 & \text{se } k \text{ perde} \end{cases} \quad (2.64)$$

sendo k um neurônio pertencente a rede e η a taxa de aprendizagem.

4. **Aprendizagem de Boltzmann:** Construído sobre a teoria de Boltzmann, os neurônios atuam de maneira binária, ligado ou desligado. Como característica, tem-se a função de

energia, definida pelo estado dos neurônios individuais, dado na Equação 2.65:

$$E = -\frac{1}{2} \sum_j \sum_k w_{kj} x_k x_j \quad , \quad j \neq k \quad (2.65)$$

Onde x_j consiste no estado do neurônio j e w_{kj} o peso de uma conexão entre j e k , onde não há retroalimentação. A partir de uma escolha aleatória, um neurônio k troca seu estado de x_k para $-x_k$ em uma temperatura T , com a probabilidade calculada por 2.66.

$$P(x_k \rightarrow -x_k) = \frac{1}{1 + \exp(-\Delta E_k/T)} \quad (2.66)$$

Tendo E_k como variação da energia resultante para aquela troca. Desta forma, atingindo o equilíbrio térmico.

Considerando que ρ_{kj}^+ é a correlação de j e k em que todos os neurônios estão presos a um estado do ambiente e ρ_{kj}^- a relação destes neurônios em um estado que podem operar livremente, então a variação Δw_{kj} aplicada a w_{kj} de um determinado neurônio k em um j pode ser dada na Equação 2.67

$$\Delta w_{kj} = \eta (\rho_{kj}^+ - \rho_{kj}^-) \quad , \quad j \neq k \quad (2.67)$$

sendo η a taxa de aprendizagem.

Até o momento, verificaram-se várias técnicas de aprendizagem para as redes neurais, porém também tem-se as conhecidas redes de aprendizagem profunda que realizam a computação de uma rede neural em múltiplas camadas de representação.

2.9.3 Aprendizagem Profunda

Em Aprendizagem Profunda, a computação e o aprendizado são realizados em múltiplas camadas de representação do conhecimento, por múltiplos níveis de abstração. É difícil extrair estes conhecimentos de alto nível dos dados iniciais, pois muitos destes fatores estão próximos ao nível de entendimento humano. Assim, a aprendizagem profunda resolve este problema pela introdução de representações expressas em termo de outras mais simples. Entre os modelos mais conhecidos de aprendizagem profunda, tem-se o Perceptron de multicamadas (MLP), as Redes Neurais Convolucionais (CNN) e as Redes Neurais Recorrentes (RNN).

2.9.3.1 Perceptrons de Múltiplas Camadas (MLP)

O perceptron de multicamadas (MLP), também conhecido por Deep feedward networks ou feedforward neural networks, é a generalização do perceptron de camada única concebido em torno de neurônios não-lineares, o mais tradicional modelo de Aprendizagem profunda (GOODFELLOW *et al.*, 2016). No caso do perceptron de camada única, deseja-se classificar duas classes a partir da combinação linear dos estímulos, x_1, x_2, \dots, x_m , pela adequação dos pesos sinápticos, w_1, w_2, \dots, w_m , por meio do algoritmo de aprendizagem por correção de erro. Na forma mais simples, separa-se a região por um hiperplano definido na Equação 2.68, para um espaço m -dimensional, das m variáveis de entrada.

$$\sum_{i=1}^m w_i x_i + b = 0 \quad (2.68)$$

Considerando o perceptron de camada única e que o bias é tratado como peso sináptico, o vetor de entrada pode ser definido por $x(n) = [+1, x_1(n), x_2(n), \dots, x_m(n)]^T$, sendo n o passo de interação. Desta forma, o vetor de peso pode também ser definido por $w(n) = [b(n), w_1(n), w_2(n), \dots, w_m(n)]^T$. Com isto, a saída resultante, $v(n)$, pode ser definida na Equação 2.69.

$$v(n) = \sum_{i=0}^m w_i(n) x_i(n) = w^T(n) x(n) \quad (2.69)$$

Com n fixo, tem-se $w^T x = 0$. Porém, com dois conjuntos de treinamento que pertencem a duas classes distintas, deseja-se definir w de forma que as classes sejam separáveis e satisfaçam as equações, onde $w^T x > 0$, se o vetor x pertence a primeira classe e $w^T x \leq 0$ se o vetor x pertence a segunda classe. Para realizar a adaptação dos pesos, se um elemento de $x(n)$ for corretamente classificado pelo vetor de pesos $w(n)$, não é necessário correção dos pesos, logo $w(n+1) = w(n)$. Porém em casos de erro, atualizam-se os pesos por $w(n+1) = w(n) - \eta(n)x(n)$ ou $w(n+1) = w(n) + \eta(n)x(n)$ nos casos de $w^T(n)x(n) > 0$ ou $w^T x \leq 0$ respectivamente, de acordo com a classe pertencente. Onde η é a taxa de aprendizagem.

Em uma MLP, o modelo utilizado para função de ativação para cada um dos neurônios é não-linear suave, como por exemplo a sigmoide da função logística. Esta rede também deve conter uma ou mais camadas de neurônios ocultos, dando possibilidade do aprendizado de tarefas complexas. Também contém alto grau de conectividade.

Um algoritmo muito utilizado nas MLP, o Retropropagação de Erro (error back-propagation), é baseado neste algoritmo de aprendizagem por correção de erro apresentado. O algoritmo Retropropagação de Erro consiste em um passo de propagação, ou passo para frente, onde o vetor de entrada é aplicado à rede e seu efeito é propagado por toda a rede, gerando uma saída para este impulso. No passo de retropropagação, ou passo para trás, os pesos sinápticos são corrigidos por um algoritmo para correção de erro, gerado pela resposta real com a resposta desejada. Este erro é propagado contra a direção das conexões, ou seja, para trás, ajustando os pesos para aproximar a resposta real da desejada.

Para o aprendizado da rede, na primeira época, ou seja, no primeiro conjunto de treinamento apresentado à rede, é realizado o passo de propagação e retropropagação, que será repetido em múltiplas épocas, até que a rede atinja o estado desejado.

Na fase de propagação, com a apresentação de $x(n)$, vetor de entrada e $d(n)$, vetor de respostas desejadas, calculam os campos locais induzidos e os sinais funcionais da rede para cada camada, para frente. Este campo $v_j^l(n)$ de um dado neurônio j em uma camada l , e m o número total de entradas, é dado na Equação 2.70.

$$v_j^l(n) = \sum_{i=0}^m w_{ji}^l(n) y_i^{(l-1)}(n) \quad (2.70)$$

Sendo $y_i^{(l-1)}(n)$ o sinal função da saída no neurônio i da camada $l - 1$, na interação n . E w_{ji}^l o peso sináptico em um neurônio j na camada l com alimentação de um neurônio i na camada $l - 1$. No caso de $i = 0$, tem-se $y_0^{l-1}(n) = 1$ e $w_{ji}^l(n) = b_j^l(n)$, onde b é o bias no neurônio j da camada l . O sinal de saída de um neurônio j em uma camada l , considerando a utilização de uma função sigmoide é dada na Equação 2.71.

$$y_j^l = \phi_j(v_j(n)) \quad (2.71)$$

Onde ϕ é a função de ativação. No caso de $l = 1$, onde o neurônio j localiza-se na primeira camada, tem-se 2.72

$$y_j^l(n) = x_j(n) \quad (2.72)$$

Sendo $x_j(n)$ o j -ésimo elemento do vetor de entrada. No caso de $l = L$, onde L é a

profundidade da rede, tem-se 2.73

$$y_j^l(n) = o_j(n) \quad (2.73)$$

O sinal de erro para esta rede pode ser calculado por 2.74.

$$e_j(n) = d_j(n) - o_j(n) \quad (2.74)$$

sendo $d_j(n)$ o j -ésimo elemento no vetor de respostas desejadas e $o_j(n)$ o j -ésimo elemento do vetor de saída.

Para a fase de retropropagação, calculam-se os gradientes locais, na Equação 2.75.

$$\delta_j^l(n) = \begin{cases} e_j^l(n) \phi_j'(v_j^l(n)) & j \text{ na camada de saída } L \\ \phi_j'(v_j^l(n))' \sum_k \delta_k^{l+1}(n) w_{kj}^{l+1}(n) & j \text{ na camada de oculta } l \end{cases} \quad (2.75)$$

sendo "''" a diferenciação em relação ao argumento. Os pesos são ajustados de acordo com a regra delta, Equação 2.76

$$w_{ji}^l(n+1) = w_{ji}^l(n) + \alpha [w_{ji}^l(n-1)] + \eta \delta_j^l(n) y_i^{l-1}(n) \quad (2.76)$$

Novamente η é o parâmetro da taxa de aprendizagem, enquanto δ a constante de momento. Este procedimento durará até que o critério de parada seja alcançado. Não existem critérios bem definidos, mas algumas das formulações mais utilizadas são os de Kramer e Sargiovanni, que consideram a convergência da rede quando a norma euclidiana do vetor gradiente alcança um pequeno limiar. Também se utiliza um critério baseado no erro, onde a rede é considerada convergida, quando a taxa absoluta de variação do erro médio quadrado em cada época for pequena o suficiente.

Entre as funções de ativação mais utilizadas em MLP, essenciais para a capacidade de representação das redes neurais, estão as funções identidade, tangente, hiperbólica e a linear retificada. Estas funções introduzem um componente não linear, pois com sua remoção tem-se uma rede sujeita as mesmas restrições dos modelos lineares.

1. **Identidade:** A função de ativação da identidade, pode ser definida por 2.77:

$$\phi(x) = px \quad (2.77)$$

$$\phi'(x) = p \quad (2.78)$$

Observa-se que a derivada desta função, Equação 2.78, é linear e constante e que não depende do valor de entrada x , ou seja, na fase de retropropagação, o gradiente mantém-se o mesmo, não melhorando consideravelmente o erro. Por este motivo, esta função é utilizada nas unidades que compõem a camada de saída, visto que para saídas com $p=1$, por exemplo, o sinal que entra é apenas repetido em sua saída.

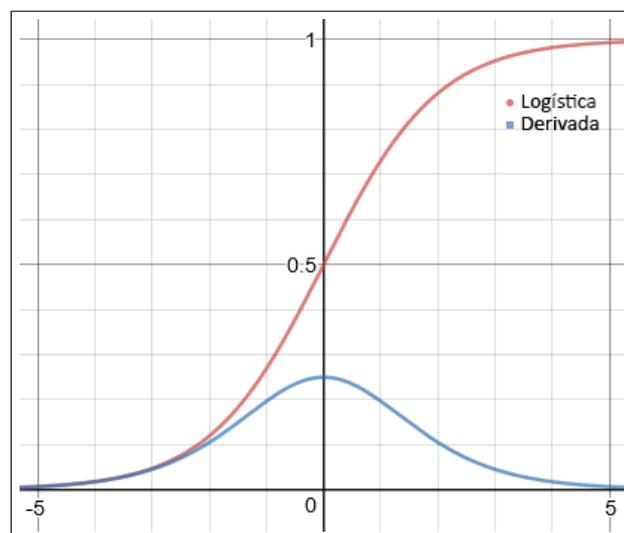
2. **Função Logística:** Esta função e sua derivada é representada por 2.79 e 2.80, respectivamente.

$$\phi(x) = \frac{1}{1 + e^{-x}} \quad (2.79)$$

$$\phi'(x) = \phi(x)(1 - \phi(x)) \quad (2.80)$$

Assim como os neurônios biológicos, esta função é binária, com formato sigmóide. Os pesos sinápticos dos neurônios que estão localizados no meio do intervalo desta função, são mais intensamente modificados, como observado no gráfico da função logística e sua derivada na Figura 15.

Figura 15 – Gráfico da função Logística e sua derivada.



Fonte: Elaborada pelo Autor

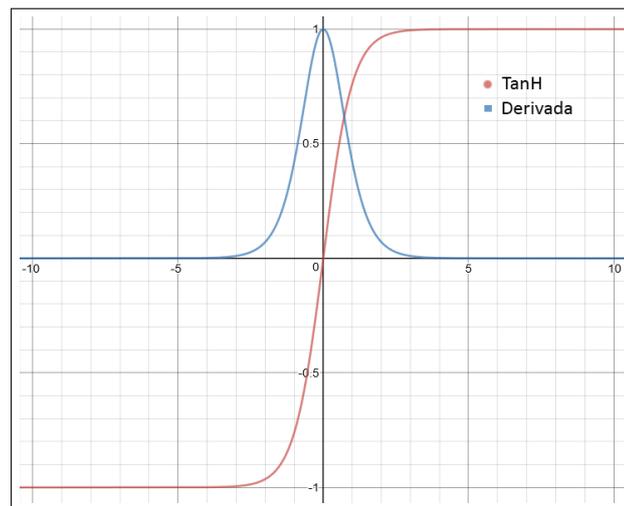
3. **TanH**: Assim como a função logística, a função tangente hiperbólica tem um formato sigmoide, porém sua variação é de -1 a 1, ao contrário de 0 a 1 como na função logística. Esta e sua derivada, são definidas na Equação 2.81 e Equação 2.82.

$$\tanh(x) = 2(\text{Logstica})(2x) - 1 \quad (2.81)$$

$$\tanh'(x) = 1 - \tanh^2(x) \quad (2.82)$$

O gráfico resultante desta função pode ser observado na Figura 16, observando que a derivada é maior, chegando ao máximo de 1 no momento que $x = 0$.

Figura 16 – Gráfico da função TanH e sua derivada.



Fonte: Elaborada pelo Autor

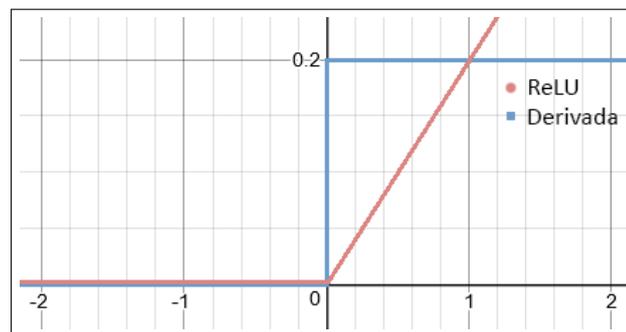
4. **ReLU**: A função de ativação retificada (ReLU) e sua derivada são apresentadas na Equação 2.83 e Equação 2.84 respectivamente.

$$\text{ReLU}(x) = \max\{0, x\} \quad (2.83)$$

$$\text{ReLU}'(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases} \quad (2.84)$$

Esta função é facilmente otimizada, uma vez que é parecida com a função identidade, porém, produz zero em metade do seu domínio. Desta forma as derivadas se mantêm altas quando a unidade está ativa. Uma de suas vantagens é a não ativação de todos os neurônios. Se a entrada for negativa, ela será convertida para zero, não sendo ativado o neurônio, mantendo a rede esparsa e eficiente para a computação. O gráfico para a função de ativação ReLU pode ser observado na Figura 17.

Figura 17 – Gráfico da função ReLU e sua derivada.



Fonte: Elaborada pelo Autor

Pela utilização das funções de ativação, estendem-se os modelos lineares em não lineares, obtendo um aprendizado não linear que é capaz de prover uma alta acurácia.

Uma das Redes Neurais de Aprendizagem Profunda mais conhecidas são as Redes Neurais Convolucionais (CNN). Estas redes são especialistas em processamento de dados com topologia em grade, como processamento de imagens e vídeos.

2.9.3.2 Rede Neural Convolucional

Também conhecidas como ConvNets, sua arquitetura é especializada no processamento de dados com topologias em grade, como por exemplo, sinais e sequências como a voz em uma dimensão, imagens de duas dimensões até mesmo vídeos ou imagens volumétricas.

Estas redes são estruturadas sobre três principais fases, a fase de convolução, fase de funções de ativação não lineares e a de Pooling. A fase de convolução equivale a integral do produto de uma das funções pela cópia deslocada e invertida de uma segunda função.

Uma função com um impulso de entrada possui uma resposta também em forma de impulso. Ao ser deslocado no tempo (k) e dado um vetor de escala a este impulso, a resposta também será deslocada e afetada por este mesmo fator de escala. Logo, a soma de um determinado número de entradas de impulsos é igual a soma das respostas destes. Esta soma

corresponde a integral, tanto na saída como na entrada deste sistema. Se esta entrada é uma função $f(t)$, a saída é dada como a integral de convolução.

Esta integral de convolução fornece um método matemático capaz de expressar a saída de um sistema com um sinal de entrada $x(t)$ e com uma dada resposta $h(t)$. Esta operação de convolução é expressa na Equação 2.85.

$$y(y) = \int x(k)h(t-k)dk \quad (2.85)$$

Esta operação também pode ser representada por um asterisco, Equação 2.86.

$$y(t) = (x * h)(t) \quad (2.86)$$

Uma vez que em sistemas computacionais, este sinal de entrada não é formado por uma entrada contínua, é necessário realizar a discretização do tempo, recebendo apenas períodos regulares de dados. Logo a convolução nestes cenários é definida na Equação 2.87.

$$y(y) = (x * h)(t) = \sum_{k=-\infty}^{\infty} x(k)h(t-k) \quad (2.87)$$

Em aplicações computacionais que utilizam CNN, tem-se geralmente como entrada uma matriz multidimensional de dados como primeiro argumento e o segundo argumento o kernel. Neste caso a função h é uma matriz multidimensional de parâmetros que serão adaptados pelos algoritmos de aprendizagem. Estas operações de convolução também podem ser utilizadas em mais de uma dimensão. Ao receber imagens 2d como entrada, imagem I , utiliza-se um kernel K de duas dimensões.

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i-m, j-n) \quad (2.88)$$

A convolução discreta pode ser vista como uma multiplicação de matrizes, onde o kernel, que muitas vezes é bem menor que a imagem de entrada, é multiplicada por cada linha e coluna, então deslocada em um elemento.

Na segunda fase das redes CNN, o sinal passa por funções de ativação não lineares, esta fase também é chamada de estágio de detecção. Após este processo, utiliza-se uma função pooling para modificar a saída da camada.

Esta função de pooling é responsável por substituir a saída da rede em certo ponto pela estatística sumária, por exemplo, pela utilização do Max Pooling retorna-se a saída máxima na área. Esta técnica ajuda a realizar uma representação aproximadamente invariante para pequenos valores de entrada, não alterando a saída para estes valores. Este mecanismo ajuda a rede a não sofrer grandes variações para pequenas mudanças. Ao determinarem-se as partes específicas em uma imagem, não é necessário saber a posição exata deste artefato, somente que ele se encontra em uma determinada região.

2.9.3.3 Rede Neural Recorrente (RNN)

As redes neurais recorrentes (RNN) são redes construídas para análise de dados sequenciais, capaz de verificar longas sequências. Estas redes são muito utilizadas em análise de textos e linguagem natural, sendo capaz de prever, por exemplo, o próximo carácter com boa precisão.

Existem várias maneiras de construir-se uma RNN, onde qualquer função que envolva recorrência pode ser considerada uma RNN. Um exemplo de função que pode ser utilizada em redes recorrentes pode ser observada na Equação 2.89.

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; 0) \quad (2.89)$$

Onde $x^{(t)}$ é o sinal de entrada.

As RNN podem gerar um resultado em cada passo entre as suas camadas. Em algumas arquiteturas, estas redes podem conter uma memória explícita, onde os elementos analisados passam por suas camadas ocultas e são armazenados em um vetor de estado que podem armazenar informações sobre a história de análise, chamadas de redes LSTM (long shor-term memory). Estas redes mostraram maior desempenho em relação às redes RNN tradicionais.

3 TRABALHOS RELACIONADOS

Neste capítulo serão descritos os principais trabalhos relacionados às fases envolvidas nesta pesquisa, a detecção dos pontos de importância em um ECG e a classificação de padrões em anomalias de ondas cardíacas. Várias destas estratégias foram desenvolvidas utilizando aprendizagem de máquinas, reconhecimentos de padrão, transformações em ondas, entre outros.

Yochum et. al (2016) descreve um método baseado em transformadas wavelets contínuas para detecção do complexo QRS, ondas P e T. O processo consiste em primeiramente encontrar o complexo QRS, visto que estes tem maior amplitude. Logo depois, encontrar as ondas T e P. Este processo é realizado por uma Transformada Wavelet Contínua que se utiliza de uma wavelet Daubechie como wavelet mãe, com uma escala variável.

O fator de escala corresponde ao máximo de correlação entre as ondas ECG e a wavelet. Utilizando-se deste fator, um limite é determinado automaticamente. Esse limite permite compartilhar corretamente os complexos QRS dos canais ECG. Após a detecção do complexo QRS o algoritmo é utilizado novamente para detecção da onda T. Para tentar obter um resultado mais confiável, utilizam-se 12 canais diferentes de ECG (Canais I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5 e V6) considerando as áreas com maior similaridade como áreas corretas.

Apesar dos bons resultados na detecção dos pontos de importância devido ao uso de múltiplos canais, têm-se maior custo computacional ao processar todos estes sinais em wavelets contínuas e compará-los.

Em Sahoo et. al (2016) aplica-se transformada de Hilbert com limite adaptativo para detecção do complexo QRS e picos R. Este trabalho também utiliza a transformada de wavelet para remoção de ruídos, assim como Yochum et. al (2016).

A transformada wavelet decompõe o sinal em uma série de escalas com diferentes componentes de frequências e analisa cada escala com certa resolução, obtendo as características do sinal. Utiliza-se Daubechies (db6) por se assemelhar morfológicamente ao sinal referente ao complexo QRS.

Após a remoção de ruídos do sinal através das wavelets, diferenciam-se os picos R da onda através de suas amplitudes e os pontos máximos são denominados ondas R. Este sinal é então processado por uma transformada de Hilbert, fornecendo uma região onde se encontra os verdadeiros complexos QRS. Finalmente, os verdadeiros picos R são detectados através de um limiar adaptativo.

Assim como no trabalho anterior, este trabalho utiliza transformadas em larga escala,

que pode acarretar em altos custos computacionais, aumentando assim a complexidade e tempo necessário para detecção dos pontos. Esta característica torna esta técnica imprópria para aplicações em sistemas de saúde em tempo real. A método também não se preocupa em detectar os demais pontos de importância, P, Q, S e T, somente os picos R.

O UNSW (KHAMIS *et al.*, 2016) é um método que gera um sinal de características contendo informações sobre a amplitude e derivações do ECG que será filtrado de acordo com sua frequência e um limite adaptativo.

As características utilizadas para extração da localização dos complexos QRS é a combinação do sinal derivado e de suas amplitudes. O sinal derivado é calculado aplicando um filtro de derivação de resposta ao impulso (FIR). As amplitudes são calculadas a partir da amplitude máxima e mínima do sinal. Calcula-se o valor da Transformada Discreta de Fourier para cada janela de 2 s. O objetivo desta aplicação é reduzir os efeitos de batidas atípicas no sinal ECG.

O UNSW tem bons resultados na detecção dos complexos QRS, porém além de não detectar todos os pontos de importância, foca apenas na detecção dos complexos QRS, também utiliza Transformada Discreta de Fourier, o que pode ocasionar em um maior tempo para detecção destes pontos.

Tem-se em Varanini *et. al* (2017) um método multicanal para detecção do complexo QRS de fetos através de batimentos anormais maternos. Consiste em elaborar um índice de qualidade responsável por sintetizar as características de pseudo-periodicidade e tempo do QRS para encontrar a combinação linear dos ECG, que maximize o índice de qualidade. A aproximação e cancelamento dos componentes maternos são realizados por Decomposição em Valores Singulares (SVD).

Esta abordagem foca na detecção do complexo QRS para fetos, priorizando a remoção do ECG materno, não se aplicando a todos os casos. Este algoritmo também não se preocupa em detectar todos os pontos de importância.

O método de detecção do complexo QRS, proposto por Elgendi *et. al* (2016), utiliza duas médias móveis, através de duas janelas de médias diferentes, para detecção destes complexos.

As fases deste algoritmo consistem em filtragem dos dados recebidos, através do filtro Butterworth. Após a filtragem, o sinal é elevado ao quadrado para então ser detectado os blocos de interesse, os quais são encontrados através de duas médias móveis. Ao detectar os blocos de interesse, tais blocos são comparados com um limite, para detectar pontos fora do

padrão.

Apesar de bons resultados para detecção, tanto em precisão e tempo, este algoritmo preocupa-se apenas na detecção dos blocos de interesse, que compreendem basicamente o complexo QRS.

Por Transformada de Wavelets, Shyu, Wu e Hu (2004) utilizam as informações obtidas da fase de detecção do QRS para classificar contrações ventriculares prematuras, utilizando uma rede neural feedforward (FNN) de cinco camadas adaptativas.

Ainda que o reuso das informações para a detecção de anomalias seja uma vantagem, este trabalho limita-se em detectar apenas contrações ventriculares prematuras.

Um sistema de monitoramento de pacientes é apresentado por Kiranyaz, Ince e Gabbouj (2016). Este trabalho utiliza uma Rede Neural Convolutiva de uma dimensão para classificação de doenças cardíacas através do ECG. Os dados são analisados separadamente através de redes CNN diferentes, criando assim uma rede personalizada para cada paciente.

Mesmo com bons resultados, este trabalho também limita-se na detecção de apenas duas anomalias cardíacas, batimento ventricular prematuro e contrações ventriculares prematuras.

Oresko et. al desenvolveu uma aplicação para smartphones capaz de classificar pontos de importância por doenças cardíacas. Para detecção destes pontos, este trabalho utilizou um algoritmo baseado em (HAMILTON; TOMPKINS, 1986) que utiliza um limiar adaptativo. Quanto a classificação destes sinais, utilizou-se uma Rede Perceptron de multicamadas (MLP).

Como entrada para rede MLP, tem-se um sinal de complexo QRS com 51 amostras de segmentos com tempos de 150 ms. A saída da rede MLP é então associada a um dos problemas cardíacos analisados no escopo do trabalho. Para a implementação desta rede utilizou-se o MATLAB e para validação dos resultados, utilizou-se a base de dados MIT-BIH.

Este trabalho também apresentou bons resultados para as anomalias analisadas, porém assim como os outros, tem uma limitação na quantidade destas anomalias, trabalhando apenas com quatro, que são: Bloqueio do ramo direito do feixe de His, contração ventricular prematura, batimento cardíaco estimulado por um marcapasso e fusão destes batimentos com batimentos normais.

A próxima seção será responsável por apresentar a proposta deste trabalho e como as técnicas foram utilizadas para detecção dos pontos de importância e classificação das 19 diferentes anomalias detectadas a partir do ECG.

A Tabela 1 explicita as principais diferenças dos trabalhos relacionados em relação a completude de detecção de pontos de fiduciais e o custo computacional, relacionando assim os

trabalhos apresentados e esta proposta.

Tabela 1 – Comparação entre os Trabalhos Relacionados.

Referências	Deteccão de Todos os 5 pontos Fiduciais	Custo Computacional
Yochum et. al (2016)	Sim	Alto
Sahoo et. al (2016)	Não	Alto
Khamis, Heba, et al. (2016)	Não	Alto
Varanini et. al (2017)	Não	Baixo
Elgendi et al. (2016)	Não	Alto
Shyu et. al (2004)	Sim	Alto
Kiranyaz et. al (2016)	Sim	Alto
Moreira et al.(2018)	Sim	Baixo

Fonte: Elaborada pelo autor

4 PROPOSTA

Este trabalho tem como proposta a construção de um sistema capaz de analisar um sinal da onda de Eletrocardiograma, extraindo os pontos fiduciais que serão então utilizados para classificação de 19 diferentes anomalias no ECG dos pacientes analisados. Como resultado deve-se sinalizar aos especialistas as possíveis anomalias presentes nos ECGs de entrada, como um sistema de suporte à decisão. Para alcançar este objetivo, o trabalho foi dividido em algumas etapas, são elas:

- **Filtragem da Onda e Detecção dos Pontos de Importância:**

Para monitoração de variações cardíacas, a técnica mais utilizada consiste no eletrocardiograma (ECG), que contém cinco pontos fiduciais que correspondem as ondas P, Q, R, S e T. Podemos dividir esta fase de detecção das características do ECG em duas principais etapas, a primeira concerne a filtragem da onda e a segunda com a aplicação dos algoritmos de detecção.

- **Filtragem da Onda ECG:**

Para detectar os pontos de importância em um ECG devemos primeiramente filtrar o sinal original, uma vez que ruídos são introduzidos no ECG pelo circuito de alta impedância de entrada e alto ganho, necessário para detecção das ondas de baixa tensão produzidas pelo coração (METSHEIN *et al.*, 2017). Além do ruído da rede elétrica local e da atividade muscular, a respiração é uma das principais causas da variação da linha de base no ECG, que assim como os ruídos, devem ser removidas.

- **Detecção dos Pontos Fiduciais:**

As anomalias na posição e valores destes pontos podem indicar doenças cardíacas. Para isto, propõe-se um método de detecção dos pontos fiduciais de um ECG em tempo real utilizando médias móveis e heurísticas.

- **Classificação dos Dados:** Esta etapa também pode ser executada e dividida de duas formas diferentes. Pode-se aplicar os algoritmos de aprendizagem de máquina em toda a banco de dados ou especializar as máquinas em somente um paciente por vez.

- **Treinamento e classificação com Base de Dados Completa:** Neste caso, os algoritmos de aprendizagem de máquina podem ser treinados a partir de todos os dados disponíveis. Com esta abordagem, tem-se uma máquina genérica, capaz de verificar a existência de doenças em diversos pacientes. Aplica-se este método em pacientes sem prévia das possíveis doenças existentes, alertando assim, através de um diagnóstico,

as anomalias detectadas no exame ECG.

- **Treinamento e classificação por paciente:** Pode-se também aplicar os algoritmos de aprendizagem de máquina para aprender e classificar os dados de um único paciente. Neste cenário, espera-se uma máquina especialista nas doenças do paciente de entrada, desta forma tem-se resultados bem melhores de detecção das anomalias. Esta abordagem é utilizada quando deseja-se um ambiente de monitoramento de um paciente com doenças já conhecidas, para evitar e alertar atividades cardíacas bruscas ou realizar estudos de gravidade da doença.

Para realizar a análise de cada uma destas etapas mencionadas anteriormente, este capítulo é dividido de forma que primeiramente será apresentado o mecanismo utilizado para filtragem e detecção dos pontos de importância, para então expor como estes resultados podem ser utilizados nas fases de particionamento e classificação dos pacientes por doenças cardíacas. Analisaremos cada uma destas etapas mais a fundo nas seguintes seções.

Em relação aos dados recebidos no sistema, frequentemente são obtidos através dos objetos que estão introduzidos na rede. Geralmente tais objetos são dotados de sensores que são responsáveis pelo sensoriamento, sendo capazes de receber diversos estímulos, como temperatura, pressão, umidade. Alguns outros também são capazes de atuar no ambiente, através dos atuadores. Possibilitando assim que tais objetos não só percebam, mas também reajam a estímulos específicos.

Em redes IoT e-health, tais sensores e atuadores são responsáveis por obter estímulos e alterações no corpo humano. Como exemplo, temos sensores de temperatura corporal, pressão, glicemia, taxa respiratória, ECG. Assim como nas redes IoT tradicionais, tal rede pode também atuar caso necessário. A Figura 1 demonstra um exemplo de plataforma e-health completa, capaz de realizar as funções citadas. Após tais dados serem recolhidos podem então ser enviados através da rede IoT.

Para validação e testes em pesquisas realizadas na área de e-health em ECG, a base de dados MIT-BIH é amplamente utilizado. A base de dados de arritmia MIT-BIH consiste em um trecho de 48 horas de registros em ECG de ambulatórios em dois canais, obtidos através de 47 pessoas estudadas pelo Instituto de Arritmias BIH entre 1975 e 1979. Um registro de ECG de 24 horas retirado de uma população aleatória e mista de pacientes internados (cerca de 60%) e pacientes ambulatoriais (cerca de 40%) no hospital Beth Israel.

A gravação foi digitalizada com 360 amostras por segundo em cada canal e digitalizada com resolução de 11 bits na faixa de 10 mV. Mais de um cardiologista anotou cada registro

Figura 18 – Exemplo de plataforma de Sensores IoT eHealth



Fonte: Oresko (2010)

de forma independente, cerca de 110.000 anotações contidas na base de dados.

4.1 FILTRAGEM DA ONDA E DETECÇÃO DOS PONTOS DE IMPORTÂNCIA.

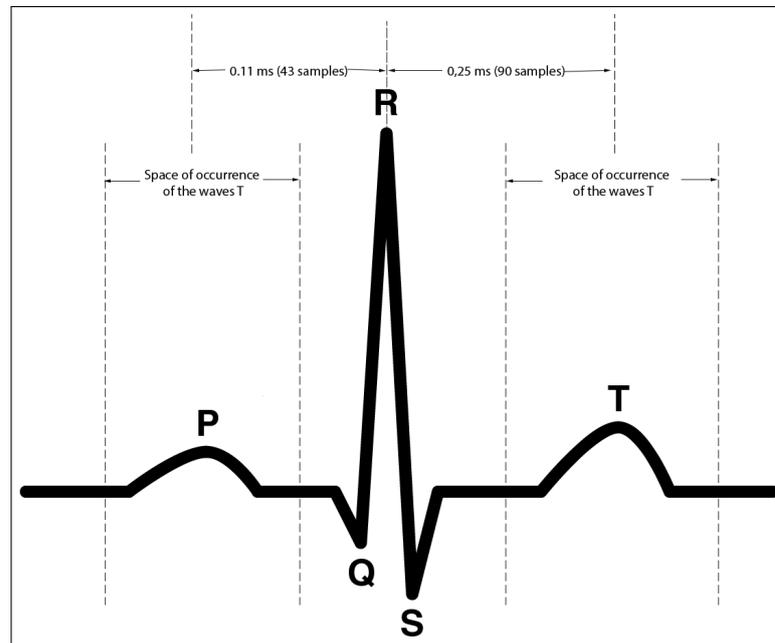
O sinal ECG é formado por várias ondas que são correspondentes aos eventos relacionados ao ciclo cardíaco, chamados de pontos fiduciais, como mostrado na Figura 19. O ciclo cardíaco é formado pelas Ondas P, que representam a despolarização atrial; O Complexo QRS, que expressa a despolarização ventricular; e a onda T, correspondente a repolarização ventricular (CARNEVALE *et al.*, 2017). Também podemos observar uma distância padrão entre estas ondas, a média entre os picos P ao S é de 0.11 ms enquanto do R ao T, 0.24 ms, ambos em pacientes saudáveis de acordo com os padrões de ECG (MAO *et al.*, 2003).

Neste trabalho propomos o fluxo que pode ser usado para detectar os pontos fiduciais em um sinal de ECG, como mostrado na Figura 20

A fim de analisar o ECG para detecção dos pontos de importância devemos iniciar através da filtragem dos dados recolhidos, uma vez que ruídos são facilmente introduzidos em um ECG, visto que os circuitos de um Eletrocardiógrafo tem uma alta impedância de entrada e é necessário um alto ganho para detecção de ondas de baixas voltagens produzidas pelo coração (METSHEIN *et al.*, 2017). Temos também o ruído introduzido através da rede elétrica local, da atividade muscular e da respiração que é uma das principais causas da variação da linha de base no ECG.

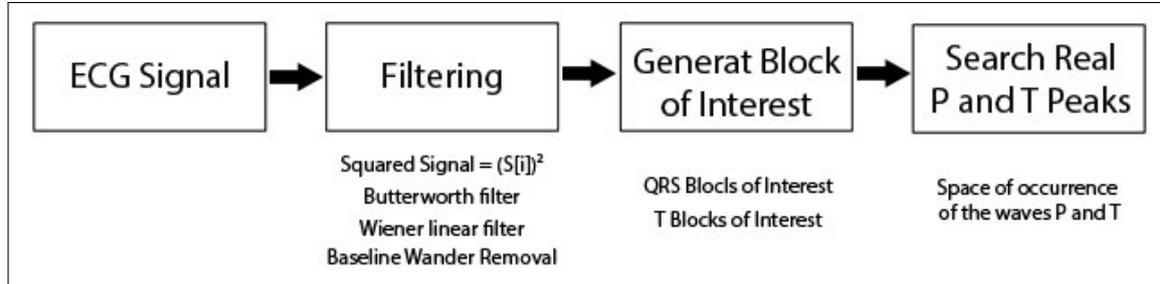
O processo de filtragem é realizado em três fases, sendo a primeira delas a remoção da variação da linha de base, utilizando wavelet. Para este trabalho, implementou-se um algoritmo

Figura 19 – Pontos de Importância no sinal ECG e Distâncias entre ondas.



Fonte: Elaborada pelo Autor

Figura 20 – Fluxo de Detecção dos Pontos de Importância.



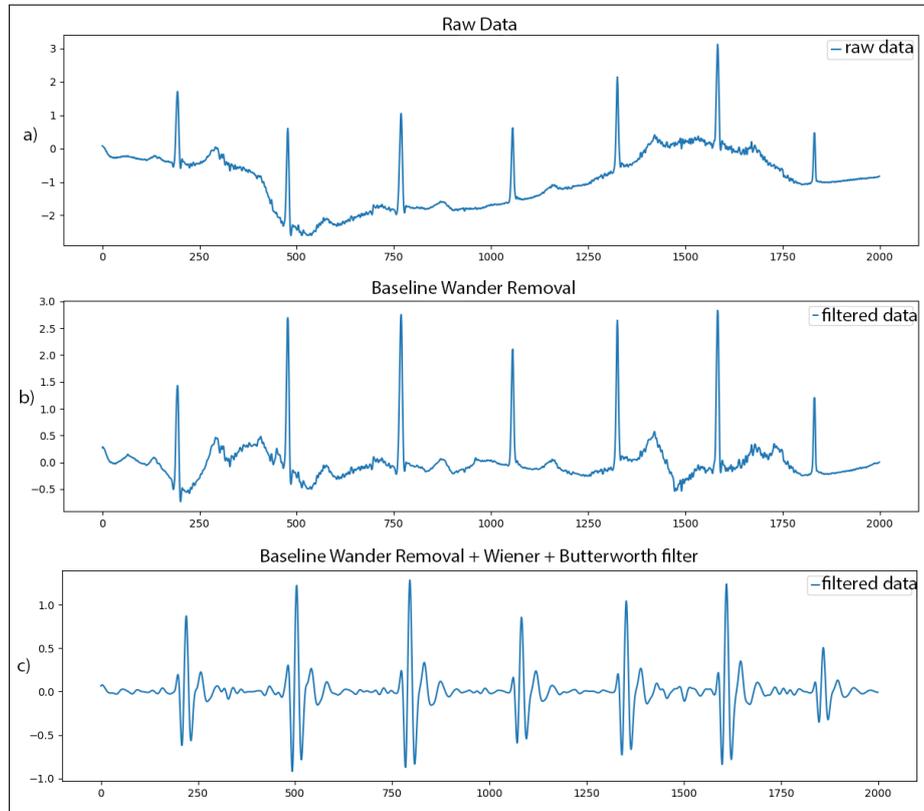
Fonte: Elaborada pelo Autor

baseado em wavelet proposto por Sargolzaei et al. , que foi aplicado ao sinal original recolhido. A Figura 21 mostra o sinal ECG antes e após a remoção da variação da linha de base. O sinal Original corresponde a imagem a), e o sinal filtrado após o uso do algoritmo exibido em b).

Após remover a variação da linha de base, dois filtros são aplicados ao sinal ECG, com o objetivo de melhorar a qualidade do sinal recebido, o filtro passa-faixa Butterworth e o filtro linear Wiener, ambos disponíveis na biblioteca Python de código aberto SciPy (Jones et al., 2001) .

O filtro de passagem de banda Butterworth é aplicado na redução de ruído, atribuindo uma forma de onda de ECG mais plana, com frequências de corte baixas e altas, enquanto o filtro Wiener é então aplicado ao sinal filtrado Butterworth, como um suporte para redução de

Figura 21 – a) Sinal ECG Original. b) Sinal após a remoção da variação da linha de base. c) Sinal após a remoção da variação da linha de base e aplicação dos filtros Butterworth e Wiener



Fonte: Elaborada pelo Autor

ruído. A Figura 21 mostra um ECG típico após a remoção da variação da linha de base em b) e os filtros Butterworth e Wiener aplicados em c).

Como uma fase de melhoria, a onda inteira é elevada ao quadrado, ponto a ponto, a fim de destacar os grandes valores e aumentar as altas frequências.

Após o sinal de ECG ter sido filtrado, o próximo passo é gerar os blocos de interesse, cada bloco é composto por um complexo QRS e uma onda T. O complexo QRS é um componente significativo na onda do ECG para permitir a detecção de doenças cardíacas. Para isso, usamos um esquema de médias móveis para gerar os blocos de interesse do complexo QRS de acordo com Elgendi (2016).

O esquema de média móvel usa dois períodos de tempo: W_1 , que é a duração aproximada do evento de interesse, e W_2 , a duração aproximada do ciclo (batimento cardíaco, no caso de ECGs).

A partir dos tempos W_1 e W_2 , as médias móveis MA_{evento} e MA_{ciclo} podem ser

computadas, respectivamente, através da Equação 4.1 e Equação 4.2, onde n é o índice de cada ponto dos N pontos de dados e $y[n]$ o valor atual.

$$\text{MA}_{\text{evento}}[n] = \frac{1}{W_1} \left(y\left[n - \frac{W_1 - 1}{2}\right] + y\left[n - \frac{W_1 - 3}{2}\right] + \dots + y\left[n + \frac{W_1 - 3}{2}\right] + y\left[n + \frac{W_1 - 1}{2}\right] \right) \quad (4.1)$$

$$\text{MA}_{\text{ciclo}}[n] = \frac{1}{W_2} \left(y\left[n - \frac{W_2 - 1}{2}\right] + y\left[n - \frac{W_2 - 3}{2}\right] + \dots + y\left[n + \frac{W_2 - 3}{2}\right] + y\left[n + \frac{W_2 - 1}{2}\right] \right) \quad (4.2)$$

Utilizando os vetores de médias móveis $\text{MA}_{\text{evento}}$ e MA_{ciclo} , os pontos de interesse dos eventos correspondentes podem ser seleccionados, pois eles são os pontos de sinal $y[n]$ dentro da(s) região(ões) onde $\text{MA}_{\text{evento}} \geq \text{MA}_{\text{ciclo}}$.

Dependendo de quão ruidoso é o sinal, pode-se ajustar os limiares $\text{MA}_{\text{evento}}$ e MA_{ciclo} adicionando um deslocamento a eles. É uma prática comum considerar o deslocamento proporcional à média do sinal.

Para detectar o complexo QRS, foram realizados alguns testes para determinar os melhores valores para os períodos do complexo QRS de acordo com a base de dados MIT-BIH. Os valores são então usados para calcular a média móvel MA_{QRS} do evento e a média móvel MA_{beat} do ciclo, seguindo a Equação 4.1 e Equação 4.2, respectivamente. Visto que alguns outros eventos, como as ondas P e T, podem ser detectados junto com o complexo QRS, regiões menores que W_1 ms são descartadas.

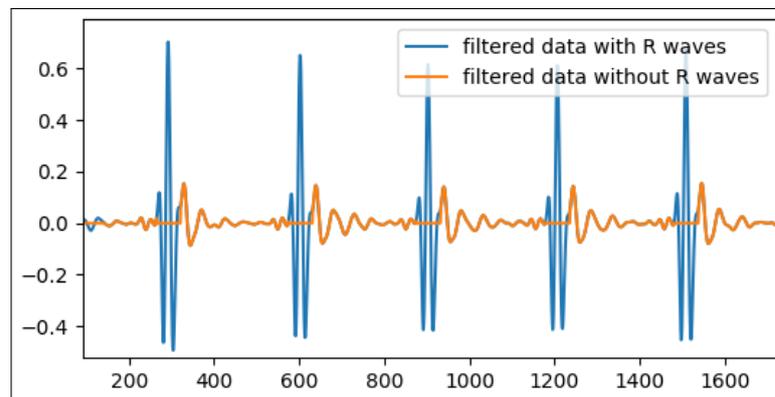
Após a detecção do complexo QRS, o método focará na detecção dos picos Q, R e S, como segue. O maior valor do vetor do complexo QRS é atribuído ao pico R uma vez que é a onda com a amplitude mais significativa em um sinal de ECG.

O valor máximo da onda à esquerda do pico R, considerando a onda elevada ao quadrado, é atribuído ao pico Q. O valor máximo da direita do pico R é atribuído como pico S.

Para prosseguir com a detecção dos pontos fiduciais P e T, o próximo passo refere-se à extração dos blocos de interesse já definidos. O complexo QRS detectado é removido do sinal, zerando os valores correspondentes no vetor de sinal. Após isto, a onda contém apenas os sinais ainda não identificados. Um resultado típico pode ser visto na Figura 22.

Similarmente ao que foi feito para detectar o complexo QRS, foram realizados testes para determinar os melhores valores para os intervalos W_1 e W_2 da onda T, levando em consideração a base de dados MIT-BIH. Estes intervalos são usados para calcular a média móvel

Figura 22 – Sinal após remoção do complexo QRS.



Fonte: Elaborada pelo Autor

relacionada ao evento MA_T e a média móvel relacionada ao ciclo das ondas MA_{beat} , aplicando a Equação 4.1 e Equação 4.2 para o sinal sem o complexo QRS.

Uma vez detectada a onda T, ela é removida do sinal deixando apenas a onda P e algum ruído. O maior valor à esquerda do pico R é considerado o pico P.

Para melhorar a detecção dos picos reais das ondas P e T, realizamos uma busca entre os pontos existentes com base no espaço de ocorrência de cada onda, mostrado na Figura 19, para determinar as ondas P e T reais, pelos valores regulares das distâncias entre as ondas PR e entre as ondas RT em um ECG. Em um alcance dentro dessa distância, podemos definir quais pontos são realmente reais. Pontos que estão fora deste intervalo são considerados pontos de erro. Esta busca é realizada somente a partir dos picos R, pois as ondas P e T podem ocorrer em pontos que não são detectados uma onda R, em ondas anômalas.

Como resultado desta fase, produz-se um sinal com todos os picos de importância detectados, visto na Figura 23, que serão então utilizados na fase de classificação.

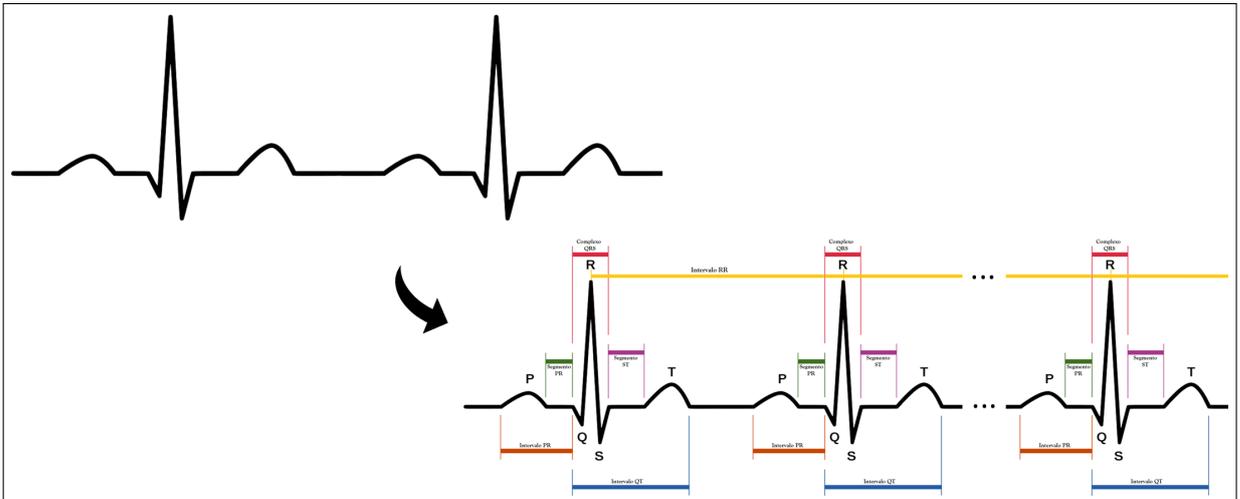
A próxima fase, após a detecção dos pontos fiduciais, é a utilização destes pontos para detecção de anomalias pela aplicação das máquinas de aprendizagem.

4.2 CLASSIFICAÇÃO DOS PACIENTES.

Para realizar a classificação de anomalias de acordo com os pontos de importância, dividiu-se esta etapa em outras duas fases, criação de blocos de interesse e Classificação das anomalias.

- **Blocos de interesse:** Nesta fase, constrói-se blocos com as características desejáveis para classificar um trecho do sinal. Cada bloco deve conter as principais características de um

Figura 23 – Onda com pontos fiduciais detectados.



Fonte: Elaborada pelo Autor

período da onda.

- **Classificação de anomalias:** Como última fase, os blocos são então classificados utilizando técnicas de aprendizagem de máquina. Define-se a classificação de cada bloco dentre 19 diferentes anomalias.

4.2.1 Blocos de interesse:

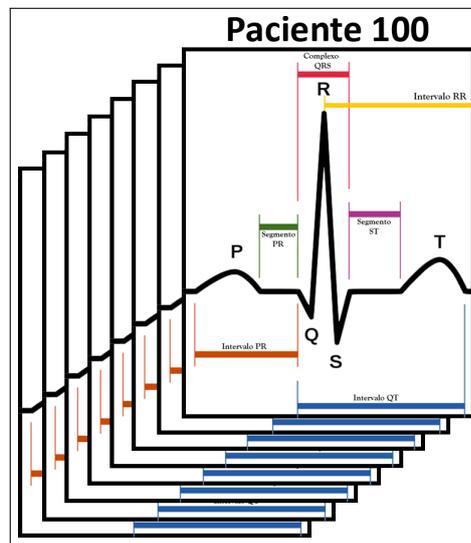
Para classificar a onda, definiu-se um sistema de blocos. Cada bloco é formado por um conjunto de pontos de importância, referentes a um período da onda. Entre os pontos de importância recebidos da etapa anterior, têm-se os pontos P, Q, R, S e T. Cada um destes pontos têm informações da amplitude da onda.

A partir dos pontos fiduciais, também são formados segmentos de importância, que representam a distância em amostras entre os pontos. Estes segmentos são definidos segundo os principais segmentos cardíacos existentes, segmento PR, RR, QT, ST e QRS.

Desta forma cada bloco pode ser representado por um período que compreende os pontos fiduciais e os segmentos cardíacos entre uma onda P à T, como pode ser observado na Figura 24.

O ECG pode ser representado por um conjunto de blocos de interesse. Cada bloco pode ser analisado separadamente ou em conjunto.

Figura 24 – Blocos formados pelos períodos e segmentos compreendidos entre uma onda P à T



Fonte: Elaborada pelo Autor

4.2.2 Classificação de Anomalias:

Para realizar a classificação das possíveis anomalias em um bloco, serão utilizados algoritmos de aprendizagem de máquinas, Árvores de Decisão, Florestas Aleatórias, Máquina de vetores de suporte e métodos de aprendizagem profunda, como Perceptron multicamadas.

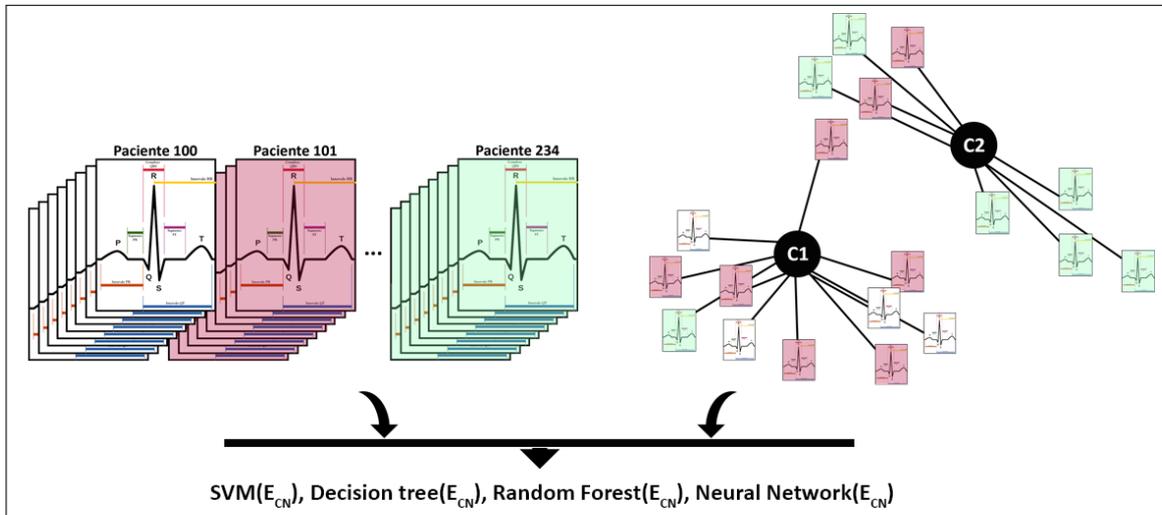
Como entrada para esta fase, pode-se receber um bloco isolado de ECG, o conjunto de blocos de um determinado paciente ou um conjunto de clusters formado por blocos agrupados de diferentes pacientes. Estes conjuntos de dados constituem a entrada para os algoritmos de aprendizagem, Figura 25.

No caso de receber somente um bloco, o objetivo do sistema é classifica-lo como um trecho saudável ou com anomalias, sendo utilizado para análises pontuais ou para reanálise de trechos específicos e pontuais.

Para um conjunto de blocos de um único paciente, classificam-se as anomalias deste paciente e se lista todos os blocos com irregularidades, desta forma, cria-se uma máquina especialista nos sinais e problemas do paciente, com isto, pode-se obter ótimos resultados em um exame contínuo de observação. Este cenário é indicado para pacientes que necessitam de uma longa observação, longos trechos de ECG são difíceis de observar, pelo método proposto, este trecho é facilmente classificado, gerando um retorno menor para observação do especialista.

Para entradas de um conjunto de blocos de diferentes pacientes, gera-se uma máquina

Figura 25 – Classificação dos blocos ECG pelos algoritmos de aprendizagem de máquina.



Fonte: Elaborada pelo Autor

com objetivos amplos, de forma que pode-se ter um melhor resultado na classificação de doenças em geral. Este cenário é indicado para detecção de novas anomalias em diferentes pacientes com uma grande variedade e quantidade de sinais.

No próximo capítulo serão apresentados os resultados obtidos através da aplicação da proposta aqui apresentada. Primeiramente verificar o processo de criação e implementação do sistema e depois os resultados quantitativos em comparação com outros trabalhos da área.

5 RESULTADOS

Os resultados obtidos neste trabalho podem ser divididos em dois grupos, os resultados relacionados com as fases de detecção dos pontos de importância e os resultados das classificações das anomalias dos pacientes:

- **Detecção dos Pontos de Importância:** Resultados concernentes a precisão do processo de detecção de pontos de importância em um ECG que podem ser separados em resultados das detecções do QRS e da detecção dos pontos P e T.
- **Classificação das Anomalias dos Pacientes:** Resultados da classificação dos pacientes utilizando os algoritmos de aprendizagem de máquina, divididos em classificação de um conjunto de blocos de um único paciente e classificação de entradas de um conjunto de blocos de diferentes pacientes.

Cada abordagem será apresentada a seguir.

5.1 DETECÇÃO DOS PONTOS DE IMPORTÂNCIA.

Na perspectiva de avaliar o desempenho do método proposto, bem como para validar os experimentos realizados, a base de dados de arritmias *MIT-BIH* foi utilizado. O MIT-BIH é composto por quarenta e oito trechos de hora de gravações de ECG de dois canais em ambulatório, foram obtidos de 47 indivíduos estudados pelo Laboratório de Arritmias do Hospital Beth Israel entre 1975 e 1979. Vinte e três registros foram selecionados aleatoriamente a partir de um conjunto de quatro mil registros de 24 horas, coletados de uma população mista de pacientes internados (cerca de 60%) e ambulatoriais (cerca de 40%) no Hospital Beth Israel, em Boston.

Os registros foram digitalizados em 360 amostras por segundo por canal com resolução de 11 bits em um intervalo de dez (10) mV. Dois ou mais cardiologistas anotaram, independentemente, cada registro. Discordâncias foram resolvidas para obter legibilidade por um computador de anotações de referência para cada batimento (aproximadamente 110.000 anotações), incluídas na base de dados.

Os especialistas realizaram anotações em cada ponto amostral do sinal ECG presente na base de dados com os rótulos que definem este sinal, R, P, T e N (Caso a amostra não faça parte de nenhum ponto de interesse). Estas anotações foram consideradas como valores verdadeiros para comparação com os encontrados pelo método. Tem-se um conjunto D de anotações definidas pelos especialistas e um conjunto A definidas no método. Pode-se definir então, para uma n -ésima amostra de ECG se está faz parte ou não do conjunto do sinal desejado.

Para as ondas R por exemplo, tem-se:

$$D_w(n) = \begin{cases} 1 & \text{se } w = R, \\ 0 & \text{caso contrário} \end{cases}$$

$$A_w(n) = \begin{cases} 1 & \text{se } w = R, \\ 0 & \text{caso contrário} \end{cases}$$

Onde n representa a n -ésima amostra. Para notação, caso o sinal faça parte da onda em questão no conjunto de dados dos especialistas, é representada por D , caso contrário, por \bar{D} . Da mesma forma, para o conjunto de dados classificados, uma onda pode ser considerada como presente no conjunto, representada por A , ou ausente neste conjunto, \bar{A} .

Usando essa notação, pode-se definir também as taxas de Verdadeiro Positivo (VP), Verdadeiro Negativo (VN), Falso Positivo (FP) e Falso Negativo (FN) pelas seguintes equações:

$$\begin{aligned} \text{VP}_w &= \sum_{n=1}^N A_w(n) \cdot D_w(n), \\ \text{VN}_w &= \sum_{n=1}^N \bar{A}_w(n) \cdot \bar{D}_w(n), \\ \text{FP}_w &= \sum_{n=1}^N A_w(n) \cdot \bar{D}_w(n), \\ \text{FN}_w &= \sum_{n=1}^N \bar{A}_w(n) \cdot D_w(n), \end{aligned}$$

para $w = R, P, T$ e N , onde n é o número total de amostras no sinal de ECG. Verdadeiro Positivo (VP) é o número de amostras com resultados identificados como positivo pelos especialistas e pelo método. Os Verdadeiros Negativos (VN) são resultados que os especialistas consideram pontos de interesse, mas o método não considera. O Falso Positivo (FP) lista os pontos encontrados como pontos possíveis, mas não citados como pontos pelos especialistas. E o Falso Negativo (FN), são os resultados que não foram encontrados pelo método, mas citado como pontos verdadeiros por especialistas.

Na tentativa de avaliar a robustez do método proposto definimos as seguintes métricas:

- Acurácia (Acc) mede a capacidade do método de obter a classificação correta:

$$Acc = \frac{\text{VP}_w + \text{VN}_w}{\text{VP}_w + \text{VN}_w + \text{FP}_w + \text{FN}_w}.$$

- Sensibilidade (Se) representa a proporção de resultados positivos corretamente identificados como tal, pelo método:

$$Se_w = \frac{VP_w}{VP_w + FN_w}.$$

Uma alta taxa de sensibilidade significa que a classificação errou alguns pontos positivos, no nosso caso, que alguns pontos fiduciais podem não ter sido identificados pelo método.

- Especificidade (Sp) representa a proporção de resultados negativos corretamente identificados pelo método de classificação:

$$Sp_w = \frac{VN_w}{VN_w + FP_w}.$$

No nosso caso, uma alta taxa de especificidade indica que as ondas P não foram classificadas como ondas T, ou ruídos não foram classificados como uma onda P.

- O Índice Youden (Y_w) (Fluss et al., 2005), como alguns métodos podem apresentar alta sensibilidade e baixa especificidade (ou vice-versa), este índice ajuda a identificar tais casos indesejáveis, combinando os resultados de Se e Sp

$$Y_w = Se_w + Sp_w - 1.$$

- Valor preditivo (P+), indica a proporção de resultados positivos, também é definida para comparação com outros resultados na literatura:

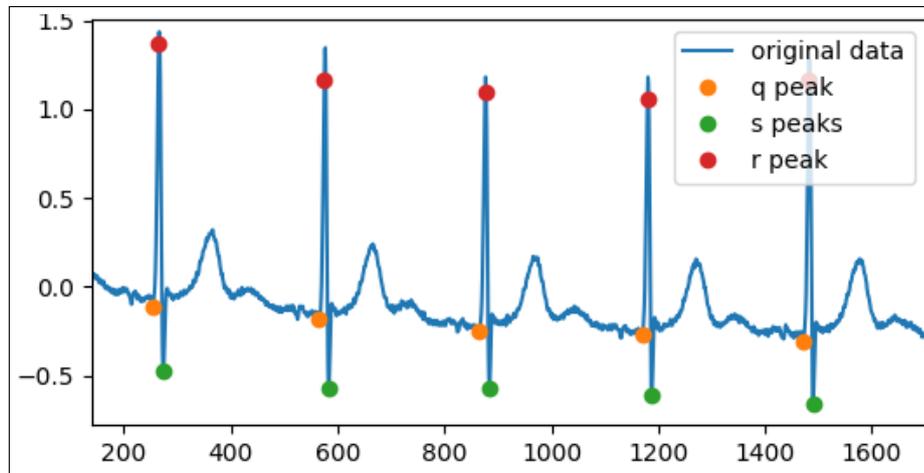
$$P+_w = \frac{VP_w}{VP_w + FP_w}.$$

Os resultados obtidos serão avaliados quanto à detecção do complexo QRS, a detecção de todos os pontos fiduciais e o custo computacional envolvido no processo. Os resultados utilizam o filtro passa-faixa Butterworth 5 th aplicado ao sinal com frequências baixas e altas de corte, respectivamente, de 8 Hz e 20 Hz.

Os valores ótimos para W_1 e W_2 para a detecção do complexo QRS são, respectivamente 35 e 240 e os valores ótimos para a detecção de ondas T são 20 e 240, para W_1 e W_2 respectivamente. Estes valores foram mensurados através das médias dos tempos de evento de interesse das ondas estudadas, ou seja, a média dos tempos de duração do complexo QRS tem 35 amostras, ondas T com tempo médio de 20 amostras e o tempo de um ciclo de ECG para os dois eventos é de 240 amostras.

A Figura 26 mostra a detecção dos pontos fiduciais Q, R e S no complexo QRS do registro 103 da base de dados MIT-BIH.

Figura 26 – Detecção dos pontos Q, R e S no registro 103 do MIT-BIH.



Fonte: Elaborada pelo Autor

Como mostrado na Tabela 2, o método proposto traz bons resultados ao detectar o complexo QRS, com diferenças mínimas dos resultados obtidos por Pan e Tompkins (1985), Yochum et al. (2016), e Elgendi (2016). Além disso, os resultados não são tão distintos dos obtidos por Elgendi (2016), uma vez que ambos usam a mesma técnica de médias móveis para realizar a detecção do complexo QRS.

Tabela 2 – Resultados da detecção do complexo QRS.

Detector de QRS	Se (%)	Sp (%)	Y (%)	Acc (%)	P+ (%)
Pan-Tompkins (1985)	99.75	N/R	N/R	N/R	99.53
Yochum et al. (2016)	99.85	98.42	98.29	98.64	99.48
TERMA (2016)	99.78	N/R	N/R	N/R	99.87
Moreira et al. (2018)	99.74	99.02	98.76	99.04	99.63

Fonte: Elaborada pelo Autor

A Tabela 3 mostra a performance do método na detecção dos pontos fiduciais P e T. Da comparação anterior, apenas os resultados de Yochum et al. (2016) são usados já que os outros dois trabalhos não detectam as ondas P e T.

O método proposto alcança altas taxas de acurácia na detecção de ondas P e T: 99.04% e 99.03%, respectivamente, superando os resultados obtidos em Yochum et al. (2016) em 6.60% e 4.20%, respectivamente. Também atingiu uma taxa de sensibilidade um pouco superior (em média 0.64%) ao obtido por Yochum et al. (2016) na detecção das ondas P e T.

Quanto às taxas de especificidade, o método proposto supera os resultados obtidos por Yochum et al. (2016) em 7.81% e 5.81% na detecção de ondas P e T, respectivamente. Alcançamos 99.02% de especificidade em ambos os casos, mostrando que este método é muito robusto ao ruído para detectar esses pontos fiduciais.

Tabela 3 – Resultados da detecção das ondas P e T

	Moreira et al. (2018)	Yochum et al. (2016)	
Onda P	Acc(%)	99.04	92.44
	Se(%)	99.82	99.06
	Sp(%)	99.06	91.21
	Y(%)	98.88	90.27
	P+(%)	99.85	83.22
Onda T	Acc(%)	99.03	94.83
	Se(%)	99.69	99.17
	Sp(%)	99.01	93.21
	Y(%)	98.70	91.38
	P+(%)	99.80	84.46

Fonte: Elaborada pelo Autor

Com taxas de sensibilidade e especificidade muito semelhantes, o método proposto alcança um alto índice Youden 98.88% e 98.70% para detecção de ondas P e T, respectivamente.

Os resultados sugerem que o método proposto é muito preciso, sensível e robusto ao ruído, tornando-o adequado para aplicações reais em sistemas de cuidados de saúde.

O impacto no custo computacional dos pontos fiduciais também é analisado, uma vez que pretende-se integrar a proposta em um sistema de assistência médica em tempo real.

Em um sinal composto por 650,000 amostras e 2314 complexos QRS, este método levou 53.23 segundos para detectar todos os complexos, e 99.89 segundos para todos os pontos fiduciais. Esses resultados revelam tempos médios de 23 ms para detecção de QRS e 43 ms para detectar cada conjunto de cinco pontos fiduciais. Assim, para detectar todos os picos que um coração humano adulto realiza em um minuto, levaria apenas 2,5s.

5.2 CLASSIFICAÇÃO DAS ANOMALIAS DOS PACIENTES.

Para avaliação da classificação dos pacientes em relação a anomalias cardíacas, recebe-se como entrada os blocos de importância do sinal ECG que foram definidos anteriormente. Cada bloco é então classificado utilizando os algoritmos de árvores de decisão, floresta aleatória, SVM e MLP. Estes blocos foram avaliados em dois principais cenários.

O primeiro cenário realiza a classificação de um conjunto de blocos de um único paciente, ou seja, todos os blocos de um deste paciente são recebidos e classificados de acordo com suas anomalias. Em segundo cenário os blocos de entrada fazem parte de um conjunto de blocos de diferentes pacientes.

Para o treinamento dos algoritmos de classificação utilizou-se a Validação Cruzada (Cross-validation), onde 33% dos dados são utilizados para o treinamento dos algoritmos e os 66% restantes dos dados como entrada para classificação.

Diferentes configurações foram utilizadas para múltiplos testes dos algoritmos de aprendizagem. A combinação das seguintes configurações, para cada um dos algoritmos, pode ser observada na Tabela 4.

Tabela 4 – Parâmetros dos Algoritmos de Aprendizagem.

Algoritmo de Aprendizagem	Parâmetros	Valores
Árvores de Decisão	Critério	Gini, Entropy
	Splitter	Best, Random
	Min. Samples Split	$\frac{mss}{10}; mss \in [1, 10], mss \in \mathbb{N}$
	Min. Samples Leaf	$\frac{mss}{10}; mss \in [1, 5], mss \in \mathbb{N}$
	N ^o Max. recursos	1,2,3,4,5,6,7,8,9,10, auto, sqrt, log2, None
Floresta Aleatória	Estimadores	10, 20
SVM	C	0.25, 0.50, 0.75, 1
	Kernel	linear, poly, sigmoid, rbf
	Grau	1, 2, 3
	coef0	0.25, 0.50, 0.75, 1
MLP	N ^o de Neurônios - Primeira Camada	128, 256, 512, 1024, 2048
	N ^o de Neurônios - Segunda Camada	128, 256, 512, 1024, 2048
	Taxa de Aprendizado	0.001
	Otimizador	Adam (KINGMA; BA, 2014), Gradiente Descendente

Fonte: Elaborada pelo Autor

Para árvores de decisão, os parâmetros utilizados foram o critério Gini e Entropia, medida para mensurar a qualidade da divisão dos ramos. Splitter define a estratégia de divisão em cada nó, melhor ou aleatória, ajustando a melhor divisão ou de modo aleatório, sem critério. O mínimo de amostras para divisão define o número mínimo de amostras que são requeridas para dividir um nó interno.

A quantidade mínima de amostras nas folhas define a quantidade de amostras que são requeridas para ser um nó folha. Por último, o número máximo de recursos, é considerado ao procurar a melhor divisão, este valor pode variar em um inteiro de 1 a 10, que considera uma quantidade fixa de valores, SQRT, ou seja, a raiz quadrada da quantidade de características, log2, definindo o número máximo como o log2 e nenhum para o número máximo igual à quantidade de características.

Os parâmetros considerados para Floresta Aleatória compreendem o conjunto de parâmetros das árvores de decisão, com a adição do número de estimadores que define a quantidade de árvores em uma floresta, foram utilizadas 10 e 20 árvores.

No SVM, utilizou-se o parâmetro de erro C, definindo um valor para evitar classificar erroneamente cada exemplo de treinamento, através de uma margem no hiperplano gerado. Os kernels utilizados foram o linear, polinomial, sigmoide e rbf. Para definir o grau do kernel

polinomial, utiliza-se os de 1º, 2º e 3º grau. Como termo independente nas funções polinomiais e sigmóides, utilizou-se o parâmetro coef0, com os valores 0.25, 0.50, 0.75 e 1.

Nas MLP foram utilizadas duas camadas intermediárias com o número de neurônios igual à combinação dos valores 128, 256, 512, 1024, 2048. A taxa de aprendizado da rede foi definida como 0.001 e os otimizadores Adam e o Gradiente Descendente foram utilizados.

Para classificação dos sinais levou-se em conta 19 diferentes tipos de anomalias: Normal Beat, Left bundle branch block beat, Right bundle branch block beat, Aberrated atrial premature beat, Premature ventricular contraction, Fusion of ventricular and normal beat, Nodal (junction) escape beat, Atrial premature contraction, Premature or ectopic supraventricular beat, Ventricular escape beat, Nodal (junctional) escape beat, Paced beat, Ventricular flutter wave, Non-conducted P-wave (clocked APB), Fusion of paced and normal beat.

Na classificação do primeiro cenário, onde todos os blocos de treinamento e classificação pertencem apenas ao paciente que terá seus blocos classificados, os resultados obtidos podem ser observados na Tabela 5.

Tabela 5 – Resultado dos Algoritmos de Aprendizagem para as 19 diferentes anomalias por paciente

	Algoritmos de Aprendizagem	Acurácia
Este Trabalho	Árvore de Decisão	0.9530
	Florestas Aleatória	0.9502
	SVM	0.9818
	MLP	0.9878

Fonte: Elaborada pelo Autor

Este cenário pode ser utilizado, por exemplo, para monitoramento de pacientes separadamente, onde cada paciente tem seu próprio escopo e uma máquina especialista para classificá-lo, obtendo um resultado de até 98% de acurácia na classificação das anomalias cardíacas. Por meio desta classificação, pode-se por exemplo, realizar uma observação prolongada deste paciente e retornar apenas os períodos com maior quantidade de anomalias detectadas, com o objetivo de monitoramento ao invés de classificação de novas doenças.

Para o segundo cenário, de classificação dos blocos em um conjunto de outros blocos, não necessariamente de um mesmo pacientes, as médias dos resultados obtidos para estas anomalias podem ser observadas na Tabela 6.

Este cenário pode ser utilizado, por exemplo, em um grande conjunto de dados onde todos os pacientes fazem parte de um único escopo. Tem-se como vantagem a construção de uma máquina mais generalista, capaz de detectar novas anomalias em diferentes pacientes.

Tabela 6 – Resultado dos Algoritmos de Aprendizagem para as 19 diferentes anomalias por blocos

Algoritmos de Aprendizagem	Acurácia
Árvore de Decisão	0.7122
Florestas Aleatórias	0.7094
SVM	0.8820
MLP	0.9640

Fonte: Elaborada pelo Autor

Os melhores resultados foram obtidos através das configurações apresentadas na Tabela 7, para cada um dos algoritmos testados anteriormente.

Tabela 7 – Melhor conjunto de parâmetros para os Algoritmos de Aprendizagem.

Algoritmo de Aprendizagem	Parâmetros	Valores
Árvores de Decisão	Critério	Entropy
	Splitter	Best
	Min. Samples Split	0.1
	Min. Samples Leaf	0.1
	N ^o Max. recursos	8
Floresta Randomica	Critério	Entropy
	Splitter	Best
	Min. Samples Split	0.7
	Min. Samples Leaf	0.4
	N ^o Max. recursos	4
	Estimadores	20
SVM	C	1
	Kernel	rbf
	coef0	0.25
MLP	N ^o de Neurônios primeira Camada	1024
	N ^o de Neurônios segunda Camada	1024
	Taxa de Aprendizado	0.001
	Otimizador	Adam (KINGMA; BA, 2014)

Fonte: Elaborada pelo Autor

Para realizar a comparação dos resultados obtidos neste trabalho com os resultados de (ORESKO *et al.*, 2010), também utilizou-se os algoritmos de classificação em um escopo de apenas quatro doenças: Right Bundle Branch Block, Premature ventricular contraction, Paced beat, Fusion of paced and normal beat. Estes resultados são descritos na Tabela 8 e levam em consideração um cenário com blocos de diferentes pacientes.

Pode-se observar um ganho de até 4% na classificação das anomalias, se comparado apenas ao cenário de classificação com blocos de diferentes pacientes. Estes resultados devem-se a estratégia de divisão dos blocos de interesse e a utilização do otimizador Adam (KINGMA; BA, 2014) para aprendizagem profunda através da plataforma para aprendizagem profunda Tensor Flow (ABADI *et al.*, 2015).

Tabela 8 – Resultado dos Algoritmos de Aprendizagem para as 4 diferentes anomalias por bloco

Algoritmos de Aprendizagem		Acurácia
Este Trabalho	Árvore de Decisão	0.7927
	Florestas Aleatórias	0.7886
	SVM	0.9218
	MLP (Adam)	0.9825
(ORESKO <i>et al.</i> , 2010)	MLP	0.93

Fonte: Elaborada pelo Autor

Em relação ao tempo decorrido para classificação das anomalias dos pacientes, no pior cenário, com a classificação de todos os pacientes, em média 650,000 amostras e 2314 complexos QRS e utilizando o algoritmo com maior performance de classificação, uma rede MLP, tem-se um resultado em apenas 2.8 segundos para classificação de todos os blocos existentes. Este resultado demonstra a capacidade de tal sistema ser utilizado em situações de tempo real, visto que para gerar 2314 blocos, um ser humano leva em média 33 minutos.

Este capítulo apresentou os resultados obtidos tanto para a classificação dos pontos fiduciais de um ECG quanto a classificação destes pontos a partir de uma estratégia de separação por bloco. Cada bloco foi classificado em um conjunto de blocos do mesmo paciente ou de pacientes diferentes, apresentando assim as qualidades de cada abordagem. Também observou-se o tempo necessário para que o método seja utilizado em sistemas de tempo real. No próximo capítulo, apresenta-se as conclusões do trabalho e os trabalhos futuros.

6 CONCLUSÕES E TRABALHOS FUTUROS

Devido ao grande número de mortes por doenças cardíacas, tornou-se importante ter uma preocupação extra para detectar essas doenças. A maneira mais comum é pelo eletrocardiograma (ECG) dos pacientes e de maneira visual. Porém, com o avanço da Internet das Coisas, o monitoramento contínuo de um paciente tornou-se possível. Para isso, é crucial seguir alguns passos, desde a detecção dos pontos de importância de um ECG até a classificação destes pontos em relação as possíveis anomalias.

Em um cenário de monitoramento contínuo um único paciente gera em média 430 a 720 mil registros relacionados somente à despolarização miocárdica. Em um cenário hospitalar ou clínico estes dados podem chegar em uma alta quantidade de dados. Para isto, a utilização de um mecanismo de análise destes dados em um tempo hábil faz-se necessário.

Neste trabalho, propõe-se um sistema de suporte a decisão e-health. Este é responsável por detectar características de importância em um ECG e formar uma representação que é então utilizada para classificação de anomalias cardíacas por algoritmos de aprendizagem de máquina.

A proposta foi avaliada utilizando a base de dados MIT-BIH que contém trechos de exames ECG reais. A proposta tem um bom desempenho tanto na detecção dos pontos de importância quanto à classificação das anomalias em relação a outros trabalhos importantes nesse domínio.

Para a detecção dos pontos de importância, revela melhoria na detecção das ondas P (7.81%) e T (5.81%) em relação a outros métodos. O tempo computacional envolvido nesta técnica foi analisado e considerado viável para sistemas de saúde em tempo real.

Quanto às anomalias cardíacas, realizou-se a classificação em 19 diferentes anomalias listadas na base de dados, enquanto que os demais métodos listam apenas 4, listadas entre as 19 anteriormente comentadas.

Como trabalho futuro, pretende-se analisar a morfologia da onda a fim de melhorar os resultados obtidos na classificação. Para a classificação das anomalias, serão aplicados diferentes métodos de aprendizagem, entre eles, outros algoritmos de aprendizagem profunda, como CNN e RNN.

REFERÊNCIAS

- ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANÉ, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIÉGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>. Acesso em: 19 fev., 2019.
- BEGG, R.; LAI, D. T.; PALANISWAMI, M. **Computational intelligence in biomedical engineering**. [S.l.]: CRC Press, 2007.
- CARNEVALE, L.; CELESTI, A.; FAZIO, M.; BRAMANTI, P.; VILLARI, M. Heart disorder detection with menard algorithm on apache spark. **European Conference on Service-Oriented and Cloud Computing**, p. 229–237, 2017.
- DEMOGRÁFICO, I. C. Disponível em: <http://www.ibge.gov.br>. 2018. Acesso em: 19 fev., 2019.
- ELGENDI, M. Terma framework for biomedical signal analysis: An economic-inspired approach. **Biosensors**, v. 6, n. 4, p. 55, 2016.
- GANTZ, J.; REINSEL, D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. **IDC iView: IDC Analyze the future**, v. 2007, n. 2012, p. 1–16, 2012.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1.
- HAMILTON, P. S.; TOMPKINS, W. J. Quantitative investigation of qrs detection rules using the mit/bih arrhythmia database. **IEEE transactions on biomedical engineering**, IEEE, n. 12, p. 1157–1165, 1986.
- HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2007.
- JONES, E.; OLIPHANT, T.; PETERSON, P. *et al.* **SciPy: Open source scientific tools for Python**. 2001. Disponível em: <<http://www.scipy.org/>>. Acesso em: 19 fev., 2019.
- KHAMIS, H.; WEISS, R.; XIE, Y.; CHANG, C.-W.; LOVELL, N. H.; REDMOND, S. J. Qrs detection algorithm for telehealth electrocardiogram recordings. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 63, n. 7, p. 1377–1388, 2016.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.
- KIRANYAZ, S.; INCE, T.; GABBOUJ, M. Real-time patient-specific ecg classification by 1-d convolutional neural networks. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 63, n. 3, p. 664–675, 2016.

- MAO, S.; LU, B.; TAKASU, J.; OUDIZ, R. J.; BUDOFF, M. J. Measurement of the rt interval on ecg records during electron-beam ct. **Academic radiology**, Elsevier, v. 10, n. 6, p. 638–643, 2003.
- METSHEIN, M.; ANNUS, P.; LAND, R.; KRIVOŠEI, A.; OJARAND, J.; AABLOO, A.; MIN, M. Variation of cardiac and respiratory waveform on human thorax in the case of inductive coupling. In: **EMBEC & NBC 2017**. [S.l.]: Springer, 2017. p. 671–674.
- ORESKO, J. J.; JIN, Z.; CHENG, J.; HUANG, S.; SUN, Y.; DUSCHL, H.; CHENG, A. C. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. **IEEE Transactions on Information Technology in Biomedicine**, IEEE, v. 14, n. 3, p. 734–740, 2010.
- PAN, J.; TOMPKINS, W. J. A real-time QRS detection algorithm. **IEEE transactions on biomedical engineering**, IEEE, n. 3, p. 230–236, 1985.
- SAHOO, S.; BISWAL, P.; DAS, T.; SABUT, S. De-noising of ecg signal and qrs detection using hilbert transform and adaptive thresholding. **Procedia Technology**, Elsevier, v. 25, p. 68–75, 2016.
- SARGOLZAEI, A.; FAEZ, K.; SARGOLZAEI, S. A new robust wavelet based algorithm for baseline wandering cancellation in ECG signals. **IEEE International Conference on Signal and Image Processing Applications**, p. 33–38, Nov 2009.
- SHYU, L.-Y.; WU, Y.-H.; HU, W. Using wavelet transform and fuzzy neural network for vpc detection from the holter ecg. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 51, n. 7, p. 1269–1273, 2004.
- VARANINI, M.; TARTARISCO, G.; BALOCCHI, R.; MACERATA, A.; PIOGGIA, G.; BILLECI, L. A new method for qrs complex detection in multichannel ecg: Application to self-monitoring of fetal health. **Computers in biology and medicine**, Elsevier, v. 85, p. 125–134, 2017.
- YOCHUM, M.; RENAUD, C.; JACQUIR, S. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. **Biomedical Signal Processing and Control**, Elsevier, v. 25, p. 46–52, 2016.