

**UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI**  
**Programa de Pós-Graduação em Tecnologia, Ambiente e Sociedade**  
**Wallinson Oliveira Schutte**

**DETECÇÃO DE ISQUEMIA CARDÍACA EM DIFERENTES DERIVAÇÕES**  
**UTILIZANDO REDES NEURAS ARTIFICIAIS E UM**  
**CLASSIFICADOR HÍBRIDO GAUSSIANO E BAYESIANO**

**Teófilo Otoni**

**2017**

**Wallinson Oliveira Schutte**

**DETECÇÃO DE ISQUEMIA CARDÍACA EM DIFERENTES DERIVAÇÕES  
UTILIZANDO REDES NEURAS ARTIFICIAIS E UM  
CLASSIFICADOR HÍBRIDO GAUSSIANO E BAYESIANO**

Dissertação apresentada ao PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA, AMBIENTE E SOCIEDADE – STRICTO SENSU, nível de MESTRADO como parte dos requisitos para obtenção do título de MAGISTER SCIENTIAE EM TECNOLOGIA, AMBIENTE E SOCIEDADE.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Lorena Sophia Campos de Oliveira

**Teófilo Otoni**

**2017**

Ficha Catalográfica  
Preparada pelo Serviço de Biblioteca/UFVJM  
Bibliotecário: Raniere Barros Barreto – CRB6 nº ES000861/0

S396d Schutte, Wallinson Oliveira.  
2017 Detecção de Isquemia Cardíaca em diferentes derivações utilizando redes neurais artificiais e um classificador híbrido Gaussiano e Bayesiano. / Wallinson Oliveira Schutte – Teófilo Otoni: UFVJM, 2017.  
58f. ; il.

Dissertação (Mestrado) – Universidade Federal dos Vales do Jequitinhonha e Mucuri. Programa de Pós-Graduação em Tecnologia, Ambiente e Sociedade, 2017.

Orientador: Prof.<sup>a</sup> Dr.<sup>a</sup> Lorena Sophia Campos de Oliveira

1. Eletrocardiograma. 2. Batimento Cardíaco. 3. Distribuição Normal. 4. Redes Neurais. I. Título.

**CDD: 006.3**

WALLINSON OLIVEIRA SCHUTTE

**DETECÇÃO DE ISQUEMIA CARDÍACA EM DIFERENTES DERIVAÇÕES  
UTILIZANDO REDES NEURAS ARTIFICIAIS E UM CLASSIFICADOR  
HÍBRIDO GAUSSIANO E BAYESIANO**

Dissertação apresentada ao  
PROGRAMA DE PÓS-GRADUAÇÃO  
EM TECNOLOGIA, AMBIENTE E  
SOCIEDADE - STRICTO SENSU,  
nível de MESTRADO como parte dos  
requisitos para obtenção do título de  
MAGISTER SCIENTIAE EM  
TECNOLOGIA, AMBIENTE E  
SOCIEDADE

Orientador : Prof.ª Dr.ª Lorena Sophia  
Campos De Oliveira

Data da aprovação : 29/11/2017

  
Prof.Dr. ANTONIO LUIZ PINHO RIBEIRO - UFMG

  
Prof.Dr. PATRICK WANDER ENDLICH - UFVJM

  
Prof.Dr. WEDERSON MARCOS ALVES - UFVJM

  
Prof.ª Dr.ª LORENA SOPHIA CAMPOS DE OLIVEIRA - UFVJM

TEÓFILO OTONI

**Trabalho dedicado a**

*Minha amada esposa Andréia,  
que sempre me apoiou em todos os momentos da minha vida.*

*Aos meus pais, Guilherme e Élia,  
que me incentivaram e mostraram a importância dos estudos.*

*Ao meu irmão Wesly e a toda minha família,  
que de alguma forma contribuíram para minha educação.*

## **AGRADECIMENTOS**

Primeiramente a Deus, que me concedeu a vida e oportunidade de estar aqui;

A minha esposa e meus pais, pelo apoio e incentivo constante;

A Profa. Lorena pela orientação, idéias e oportunidade de fazer parte de um projeto tão valoroso e interessante;

O André R. B. A. Maciel pela ajuda e orientação na fase inicial;

Aos amigos do DTI/CM pelo apoio de sempre.

## RESUMO

O presente estudo propõe o desenvolvimento de duas ferramentas para se fazer a classificação de batimentos cardíacos a fim de detectar a Isquemia Cardíaca. Uma baseada em propriedades da Distribuição Normal e Teorema de Bayes e a outra baseada em Redes Neurais Artificiais. Utilizando o banco de dados *Long-Term ST Database*, foi efetuado um filtro de dados, que foram agrupados pelas seguintes derivações: A-S, E-S, A-I, ML2, MV2, ML3, V4 e V5. Por meio dos algoritmos propostos, implementados por intermédio da Linguagem de Programação PHP, pôde-se verificar a derivação mais propícia a se detectar essa doença. Foi possível observar as derivações V5 e A-S com melhores resultados utilizando-se o algoritmo híbrido. Na V5, foi obtido Sensibilidade de 100%, Especificidade de 97%, Valor Preditivo Positivo de 95.89% e Valor Preditivo Negativo de 100% e, na A-S, valores de 99.22%, 99.99%, 99.99% e 99.61% para Sensibilidade, Especificidade, Valor Preditivo Positivo e Valor Preditivo Negativo. O algoritmo de Redes Neurais Artificiais apresentou o melhor resultado para derivação A-S com 99.98%, 100%, 100% e 99.99% para Sensibilidade, Especificidade, Valor Preditivo Positivo e Valor Preditivo Negativo respectivamente. Também foi calculado o intervalo de confiança para proporções populacionais com 95% de confiança, a fim de se estabelecer níveis de precisão das bases utilizadas.

**Palavras-chave:** Eletrocardiograma. Batimento Cardíaco. Distribuição Normal. Redes Neurais.

## ABSTRACT

The present study proposes the development of two tools to classify heart beats in order to detect cardiac ischemia. One based on properties of Normal Distribution and Bayes' Theorem and the other based on Artificial Neural Networks. Using the Long-Term ST Database, a data filter was performed, which was grouped by the following derivations: A-S, E-S, A-I, ML2, MV2, ML3 and V4, V5. By means of the algorithms proposed, implemented through the PHP Programming Language, we could verify the most favorable derivation to detect this disease. It was possible to observe the V5 and A-S leads with better results using the hybrid algorithm. In V5, 100% sensitivity, 97% specificity, 95.89% positive predictive value and 100% negative predictive value were obtained, and in A-S, values of 99.22%, 99.99%, 99.99% and 99.61% for sensitivity, specificity, positive predictive value, and negative predictive value. The algorithm of Artificial Neural Networks presented the best result for A-S derivation with 99.98%, 100%, 100% and 99.99% for sensitivity, specificity, positive predictive value and negative predictive value respectively. We also calculated the confidence interval for population proportions with 95% confidence in order to establish precision levels of the bases used.

**Keywords:** Electrocardiogram. Heartbeat. Bayes. Distribution Normal. Neural Networks.



## RESUMEN

El presente estudio propone el desarrollo de dos herramientas para la clasificación de los latidos del corazón con el fin de detectar la isquemia cardíaca. Uno basado en las propiedades de la Distribución Normal y el Teorema de Bayes y el otro basado en las Redes Neuronales Artificiales. Utilizando la base de datos Long-Term ST Database, se llevó a cabo un filtro de datos, que fueron agrupados según las siguientes derivaciones: a-S, E, S, A-I, ML2, MV2, ML3, V4 y V5. Por medio de los algoritmos propuestos, implementado a través del lenguaje de programación PHP, pudimos comprobar la derivación más favorable para detectar esta enfermedad. Fue posible observar las derivaciones V5 y A-S con mejores resultados utilizando el algoritmo híbrido. En V5 se obtuvo una sensibilidad del 100%, una especificidad del 97%, 95.89% de valor predictivo positivo y 100% de valor predictivo negativo y en A-S, valores de 99.22%, 99.99%, 99.99% y 99.61% para la sensibilidad, especificidad, valor predictivo positivo, y valor predictivo negativo respectivamente. El algoritmo de Redes Neuronales Artificiales presentó el mejor resultado para la derivación A-S con 99.98%, 100%, 100% y 99.99% para la sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo, respectivamente. También fue calculado el intervalo de confianza para proporciones de las poblaciones con 95% de confianza con el fin de establecer los niveles de confianza precisión de las bases utilizadas.

**Palabras clave:** Electrocardiograma. Latido del corazón. Bayes. Distribución normal. Redes Neuronales.

## LISTA DE ILUSTRAÇÕES

Figura 1 - As 10 principais causas de mortes no mundo em 2015 .....	14
Figura 2 - Representação clássica do Eletrocardiograma e seus segmentos .....	17
Figura 3 - Representação de diferentes episódios isquêmicos no ECG .....	17
Figura 4 – Posição dos eletrodos para um ECG de 12 derivações .....	19
Figura 5 - Exemplos das 12 derivações .....	19
Figura 6 – Modelo Padrão Modificado .....	20
Figura 7 – Posição dos eletrodos EASI .....	21
Figura 8 - Curva “normal” .....	23
Figura 9 - Modelo de um neurônio biológico .....	29
Figura 10 - Modelo de um neurônio artificial .....	30
Figura 11 - Diagrama de blocos ressaltando a aprendizagem por correção de erro .....	32
Figura 12 - Perceptron bidimensional .....	33
Figura 13 – Grafo de fluxo de sinal do perceptron .....	34
Figura 14 – Parte do arquivo A gerado do paciente S20161.....	35
Figura 15 - Plataforma ATM .....	36
Figura 16 – Parte do arquivo B gerado do paciente S20161 .....	36
Figura 17 – Parte do arquivo C gerado do paciente S20161 .....	37
Figura 18 - Episódio Significante (ST) relacionado à IC .....	38
Figura 19 - Parte do arquivo D, pertencente ao paciente s20161 .....	39
Figura 20 - Separação linear parcial de duas classes não linearmente separáveis .....	42
Gráfico 1 - Gráfico gerado com testes na derivação V4 .....	47

## LISTA DE TABELAS

Tabela 1 – Tabela de Contingência .....	43
Tabela 2 – Melhores resultados de cada derivação utilizando algoritmo híbrido .....	46
Tabela 3 – Melhores resultados em cada derivação usando RNAs .....	47
Tabela 4 – Resultados de outros trabalhos aplicando a base LTST-DB .....	48
Tabela 5 – Resultados de outros trabalhos aplicando outras bases .....	49

## LISTA DE ABREVIATURAS E SIGLAS

ECG	Eletrocardiograma
$E_s$	Especificidade
$E_v$	Evidência
IC	Isquemia Cardíaca
KLT	Transformada de Karhunen-Loeve
KLN	Conhecimento e Aprendizagem por exemplo em Redes recorrentes
$L_d$	Limiar de Decisão
LTST-DB	Long-Term ST Database
PHP	Personal Home Page Tools
RNA	Redes Neurais Artificiais
$S_e$	Sensibilidade
ST	Segmento
VPP	Valor Preditivo Positivo
VPN	Valor Preditivo Negativo

## LISTA DE SÍMBOLOS

$\delta$  Delta

$\Delta$  Delta

$\eta$  Eta

$\mu$  Mi

$\varphi$  Phi

$\prod$  Produtório

$\Sigma$  Somatório

$\sigma$  Sigma

$\upsilon$  Upsilon

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>14</b>
<b>2 ELETROCARDIOGRAMA .....</b>	<b>16</b>
<b>2.1 Derivações Padrão ou Mason-Likar .....</b>	<b>18</b>
<b>2.2 Sistema EASI .....</b>	<b>21</b>
<b>3 DISTRIBUIÇÃO NORMAL .....</b>	<b>23</b>
<b>4 TEOREMA DE BAYES .....</b>	<b>25</b>
<b>5 DISTRIBUIÇÃO NORMAL E TEOREMA DE BAYES .....</b>	<b>26</b>
<b>6 REDES NEURAIS ARTIFICIAIS .....</b>	<b>28</b>
<b>6.1 Neurônio biológico .....</b>	<b>28</b>
<b>6.2 Neurônio artificial .....</b>	<b>29</b>
<b>6.3 Função de ativação .....</b>	<b>30</b>
<b>6.4 Processos de aprendizagem .....</b>	<b>31</b>
<b>6.5 Perceptrons de Camada Única .....</b>	<b>33</b>
<b>7 METODOLOGIA .....</b>	<b>35</b>
<b>7.1 Importação dos dados .....</b>	<b>35</b>
<b>7.2 Classificação com algoritmo híbrido .....</b>	<b>39</b>
<b>7.3 Classificação usando Redes Neurais Artificiais .....</b>	<b>41</b>
<b>7.4 Avaliação .....</b>	<b>43</b>
<b>8 RESULTADOS E DISCUSSÕES .....</b>	<b>46</b>
<b>8.1 Algoritmo híbrido .....</b>	<b>46</b>
<b>8.2 Algoritmo aplicando RNAs .....</b>	<b>47</b>
<b>8.3 Resultados de outros trabalhos .....</b>	<b>48</b>
<b>9 CONCLUSÕES .....</b>	<b>51</b>

**REFERÊNCIAS** .....52

**GLOSSÁRIO** .....57

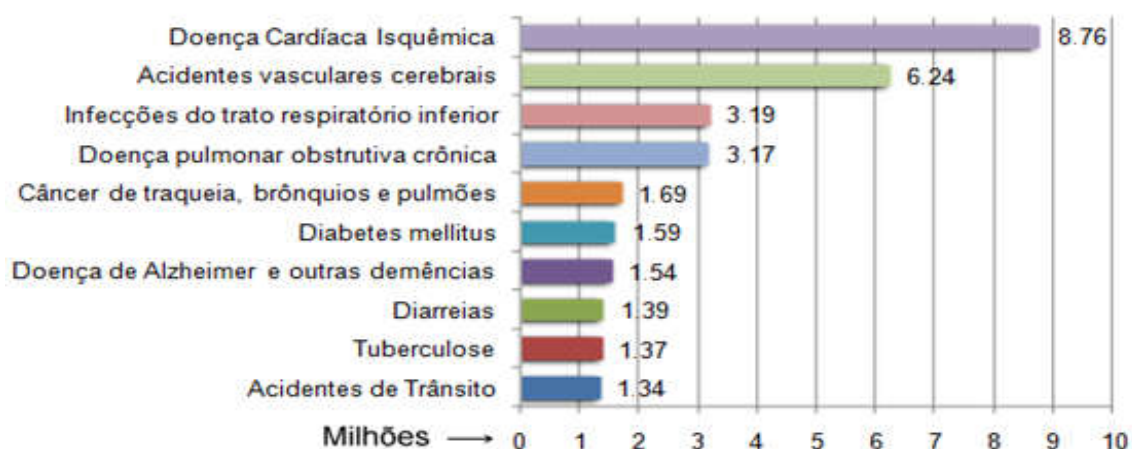
## 1 INTRODUÇÃO

Desde as primeiras semanas de vida, o músculo cardíaco, mais conhecido como coração, já começa a bater e se desenvolver. Localizado na região do mediastino, sua principal função é bombear sangue por todo corpo por meio de vasos sanguíneos (PAOLA; BARBOSA; GUIMARÃES, 2011). Portanto, é de extrema importância, que funcione da forma correta, visto que qualquer falha causará impacto direto na saúde do indivíduo.

Esse órgão é composto por quatro cavidades: átrio direito, ventrículo direito, átrio esquerdo e ventrículo esquerdo. Por meio da contração, que ocorre em decorrência de impulsos elétricos causados pela despolarização das células musculares, o coração bombeia o sangue que passa por essas quatro cavidades (COSTA; CAMANHO, 2012; PAOLA; BARBOSA; GUIMARÃES, 2011).

Um dos males que acomete o coração e compromete grande parte da população é a Isquemia Cardíaca (IC), que é desencadeada quando há um déficit na oferta e demanda de oxigênio no músculo cardíaco. Esforço físico, estresse emocional, taquicardia ou hipertensão arterial, associados à obstrução coronária, alteram não só a demanda como a oferta de oxigênio, o que desencadeia esse quadro (CARVALHO; SOUSA, 2001). Responsável pela maior quantidade de mortes anualmente, a IC, segundo relatório da Organização Mundial de Saúde, em 2015, foi responsável por 8.76 milhões de óbitos (Figura 1), colocando-a no primeiro lugar da lista das doenças que mais causam mortes anualmente, superando outras enfermidades como Acidentes Vasculares Cerebrais, Câncer, Tuberculose, entre outras.

**Figura 1 - As 10 principais causas de mortes no mundo em 2015**



Fonte: Organização Mundial de Saúde (WHO) (2017). Adaptado.



De acordo com o exposto anteriormente, este trabalho propõe dois sistemas de classificação, implementados por meio de um algoritmo baseado na Distribuição Gaussiana e Teorema de Bayes, e outro baseado em Redes Neurais Artificiais, com o intuito de identificar batimentos que apresentam ou não a IC.

Através dos classificadores propostos, aplicados a registros de exames de Eletrocardiograma, analisou-se a forma mais eficiente para fazer esta identificação e quais derivações cardíacas estão mais sensíveis a expor o problema.

## 2 ELETROCARDIOGRAMA

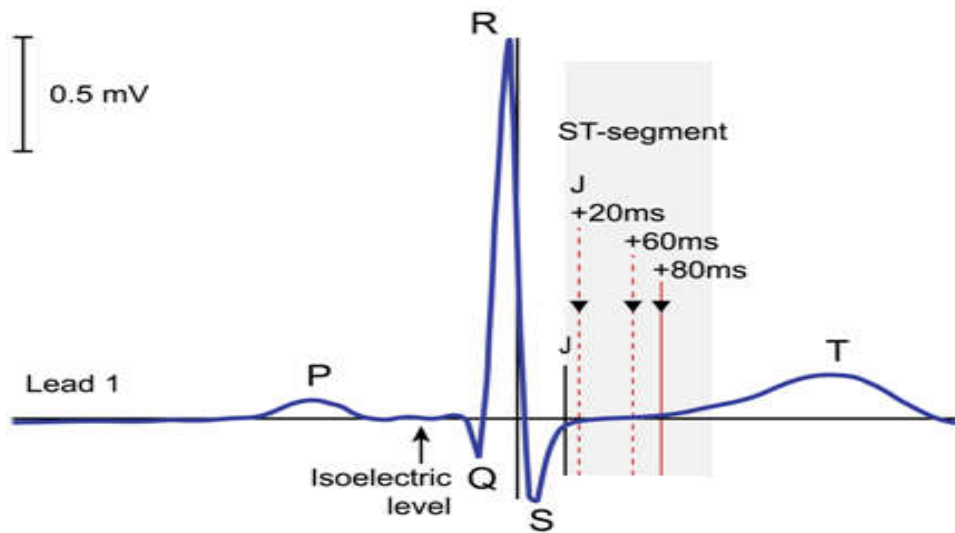
Criado por Willem Einthoven há mais de 100 anos, o Eletrocardiograma (ECG), está entre os dispositivos de diagnósticos mais usados hoje. Sua invenção evoluiu a partir de observações e inovações anteriores e da descoberta do movimento muscular por impulso elétrico através do desenvolvimento do eletrômetro capilar. Com as contribuições de Einthoven, o ECG se tornou uma ferramenta prática para o diagnóstico de anormalidades cardíacas (RIVERA-RUIZ; CAJAVILCA; VARON J, 2008).

De acordo com Oliveira (2010), ele é o principal exame para análise e diagnóstico de patologias relacionadas ao coração, em que, por meio de eletrodos fixados no corpo do paciente, é possível captar sinais elétricos relacionados ao batimento cardíaco e, assim, analisar seu funcionamento. Existem diversos meios de utilização do ECG na prática clínica e através dele é possível captar alterações primárias ou secundárias aos processos do miocárdio, como ocorrências de doenças das artérias coronárias, hipertensão arterial, cardiomiopatias, doenças metabólicas e alterações eletrolíticas, assim como efeitos tóxicos ou terapêuticos das drogas e próteses (GUIMARÃES et al., 2003).

Segundo Dranca, Goñi e Lllarramendi (2009), o ECG é o método menos invasivo e, por conseguinte um dos mais utilizados para extrair informações sobre a gravidade e duração de episódios Isquêmicos Cardíacos, que pode gerar problemas graves como infarto ou angina de peito. Ele é considerado padrão ouro para o diagnóstico não invasivo das arritmias e distúrbios de condução, além de ser muito importante nos quadros isquêmicos coronarianos (GUIMARÃES et al., 2003).

Uma representação clássica do resultado de um ECG e alguns dos segmentos gerados pode ser observada na Figura 2. O registro da disseminação da despolarização pelo miocárdio atrial do início ao fim é representado pela onda P. Já o complexo QRS representa a despolarização do miocárdio ventricular, que produz a contração ventricular. A dimensão do complexo QRS é bem maior do que a da onda P, porque os ventrículos têm muito mais massa muscular do que os átrios. A onda T representa a repolarização ventricular, sendo um processo mais lento que a despolarização ventricular, o que torna esta onda mais larga que o complexo QRS. O segmento ST tem início no ponto J e se trata da linha reta que conecta o final do complexo QRS com o começo da onda T, medindo o desfecho da despolarização ventricular até o início da repolarização ventricular (THALER, 2013).

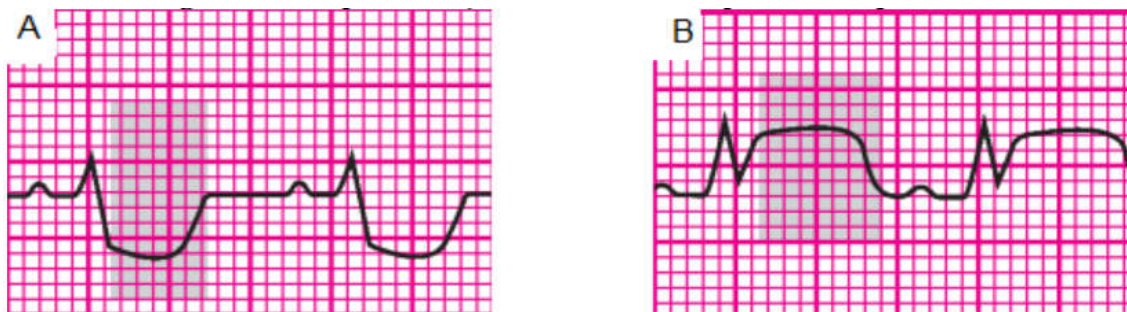
**Figura 2 - Representação clássica do Eletrocardiograma e seus segmentos**



Fonte: Amon e Jager (2016). Adaptado.

O processo isquêmico altera o Eletrocardiograma, sendo que o tecido afetado e a gravidade do estado do paciente influenciarão com maior intensidade ou não no resultado do exame. Esse mal causa a depressão ou elevação do segmento ST (Figura 3), assim como a inversão e mudanças na forma da onda T (CARVALHO; SOUSA, 2001; COSTA; CAMANHO, 2012).

**Figura 3 - Representação de diferentes episódios isquêmicos no ECG**



Fonte: Springhouse (2010).

Com isso, é possível fazer o diagnóstico de um processo isquêmico, analisando a forma que o ponto J, segmento ST e onda T se apresentam (SMRDEL; JAGER, 2004). Dessa forma o segmento ST é uma das principais informações utilizadas para fazer o diagnóstico. Para uma análise mais detalhada, este trabalho decompõe este segmento, que é dividido em pontos, separados por uma distância de tempo de 20 milissegundos (Figura 2). Assim, tem-se

o ponto correspondente ao início do segmento ST ( $J+0$ ) e posteriormente os pontos  $J+20$ ,  $J+40$ ,  $J+60$ ,  $J+80$ ,  $J+100$  e  $J+120$ , representados por  $(J_i, \dots, J_n)$ , com  $n=7$ . Por exemplo, o ponto  $J+60$  ( $J_4$ ) está a uma distância de 60 milissegundos do ponto  $J+0$  ( $J_1$ ) (Figura 2).

Existem vários repositórios contendo diversos resultados de ECG com longos períodos de monitoramento. Podem ser encontrados no *Physiobank*, que é uma base de dados livre para grandes coleções de sinais gravados fisiologicamente, sendo disponibilizados sob a licença *ODC Public Domain Dedication and License v1.0* (GOLDBERGER et al., 2000).

Contendo 86 gravações longas de 80 indivíduos humanos, o *Long-Term ST Database (LTST-DB)* é um desses bancos de dados. Os pacientes foram escolhidos para expor uma variedade de eventos de alterações do segmento ST, incluindo episódios ST isquêmicos, não isquêmicos e episódios que contêm misturas desses fenômenos. O banco de dados foi criado para apoiar o desenvolvimento e avaliação de algoritmos capazes de diferenciação precisa de eventos isquêmicos e não isquêmicos, bem como a pesquisa básica em mecanismos e dinâmicas de isquemia miocárdica (JAGER et al., 2003).

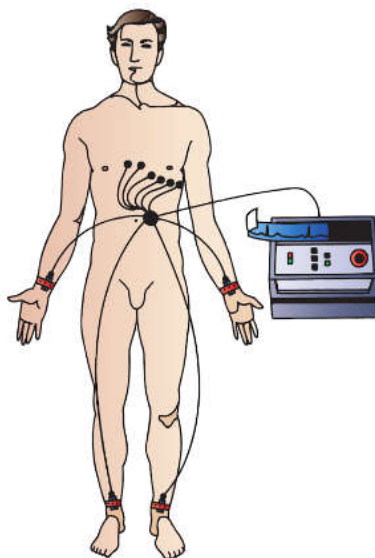
Cada um desses 80 registros de indivíduos contém doze arquivos mostrando uma série de informações de batimentos, incluindo os valores dos pontos do segmento ST (Figura 2), e as características específicas, que segundo Jager et al. (2016), podem ser encontradas na documentação da base de dados.

Por meio desses repositórios, é possível se criar métodos para desenvolver um estudo mais aprofundado usando a tecnologia de informações que, em saúde pública, tem sido indispensável para ajudar nas tomadas de decisão.

## **2.1 Derivações Padrão ou Mason-Likar**

Na prática clínica, é possível efetuar o ECG de doze derivações, geradas através de eletrodos colocados no corpo humano, captando, assim, diferentes sinais elétricos. Para se efetuar um ECG de doze derivações, são colocados dois eletrodos nos braços e dois nas pernas (Figura 4). Através deles são geradas as seis derivações dos membros, que incluem as três derivações clássicas ou padrão (I, II e III) e as três derivações aumentadas (aVR, aVL e aVF). Seis eletrodos também são colocados no tórax, formando as seis derivações unipolares precordiais: V1, V2, V3, V4, V5 e V6 (Figura 5b-c) (OLIVEIRA, 2010; THALER, 2013).

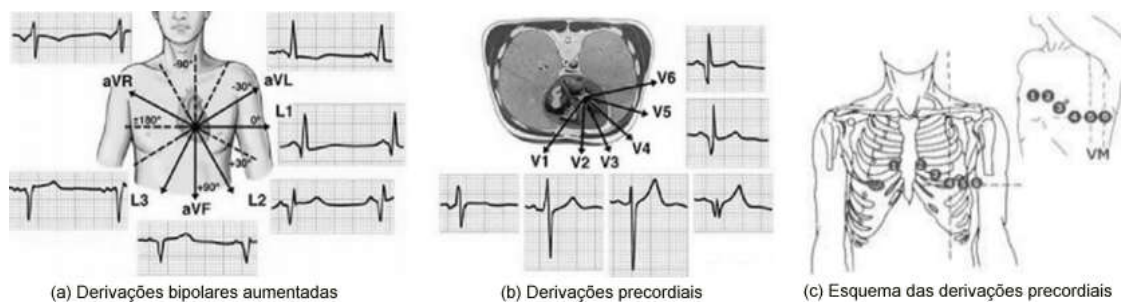
**Figura 4 – Posição dos eletrodos para um ECG de 12 derivações**



Fonte: Thaler (2013).

Um ângulo único do coração é visto através de cada derivação, aumentando a sua sensibilidade para uma região particular deste músculo. As derivações dos membros vêm o coração em um plano vertical, chamado de plano frontal, podendo ser visto como um grande círculo sobreposto ao corpo do paciente (Figura 5-a). Estas derivações vêm as forças elétricas se movendo para cima e para baixo, para esquerda e para direita neste círculo (THALER, 2013).

**Figura 5 - Exemplos das 12 derivações**



Fonte: Guilhon (2006). Adaptado.

Cada derivação tem a sua visão específica do coração, ou ângulo de orientação. Determina-se o ângulo de cada derivação desenhando-se uma linha do eletrodo negativo ao eletrodo positivo. Portanto, o ângulo resultante é expresso em graus pela sua sobreposição no círculo de 360° do plano frontal (THALER, 2013).

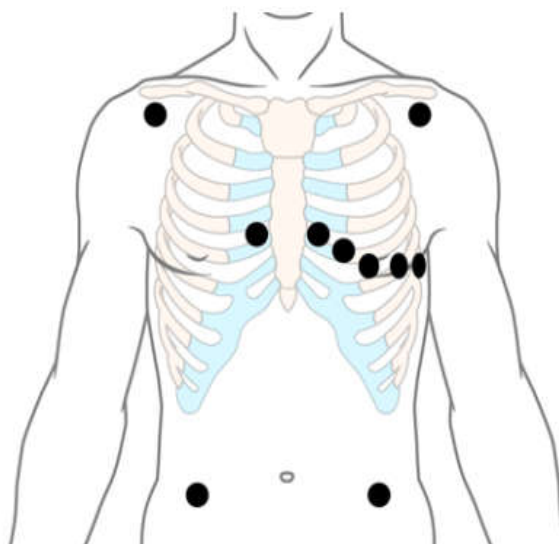
As derivações II, III e aVF são chamadas derivações inferiores porque elas vêm de maneira mais eficaz a superfície inferior do coração. As derivações I e aVL frequentemente são chamadas de derivações laterais esquerdas porque elas tem a melhor vista da parede lateral esquerda do coração. A derivação aVR é a única do lado direito (THALER, 2013).

Assim como nas derivações dos membros, cada derivação precordial vê o coração de uma forma diferente. Estas seis derivações (Figura 5-c) são criadas através de eletrodos fixados no tórax da seguinte forma (THALER, 2013):

- V1 é colocado no quarto espaço intercostal à direita do esterno;
- V2 é colocado no quarto espaço intercostal à esquerda do esterno;
- V3 é colocado entre a V2 e V4;
- V4 é colocado no quarto espaço intercostal, na linha mesoclavicular;
- V5 é colocado entre V4 e V6 e
- V6 é colocado no quinto espaço intercostal, na linha axilar média.

É possível modificar alguns eletrodos do modelo Mason-Likar. Nesta colocação de ECG de 12 derivações modificada, os eletrodos de três membros e o eletrodo de terra são colocados perto dos ombros e do abdômen inferior (Figura 6). Assim, os eletrodos de membros são colocados nos mesmos locais usados para o monitoramento contínuo de Mason-Likar (PHILIPS, 2007).

**Figura 6: Modelo Padrão Modificado**

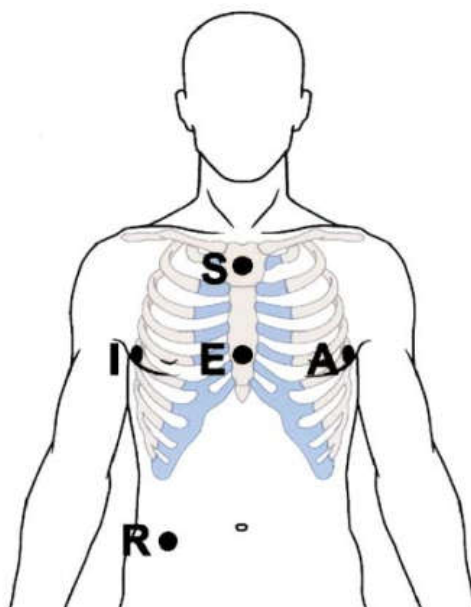


## 2.2 Sistema EASI

Em 1988, Dower e sua equipe apresentaram o sistema ECG EASI. Este, deriva o Eletrocardiograma padrão de 12 derivações usando 5 eletrodos (DOWER et al., 1988).

O eletrodo E está no esterno inferior, enquanto os eletrodos A e I estão nas linhas auxiliares médias esquerda e direita, respectivamente. O eletrodo S está no manúbrio esternal. O quinto eletrodo é um “terra” (Figura 7). (OLEKSY; TKACZ; BUDZIANOWSKI, 2012).

**Figura 7 – Posição dos eletrodos EASI**



Fonte: Wegr et al., 2006.

De acordo com Oleksy (2012), EASI demonstrou ter alta correlação com um ECG padrão de 12 derivações, assim como o ECG Mason-Likar ou Padrão (WELINDER et al., 2004) de 12 derivações.

Ainda, segundo ele, esse sistema é menos suscetível a artefatos, aumenta a mobilidade de pacientes e também é mais fácil e rápido de usar por causa de menor número de eletrodos, reduzindo também o custo de um dispositivo. Os eletrodos estão posicionados sobre pontos facilmente identificados que podem ser localizados com um mínimo de variabilidade independente do físico do paciente. A posição dos eletrodos facilita ainda o exame físico e imagético do coração e pulmões, pois geralmente não precisam ser removidos.

Na abordagem clássica introduzida por Dower (1998), usando a configuração EASI, três sinais cardiográficos vetoriais modificados são registrados a partir do eletrodo bipolar:

- 1) A-I ( $x$ , ou componente vetorial horizontal);
- 2) E-S ( $y$ , ou componente vetorial vertical) e
- 3) A-S ( $x$ ,  $y$  e  $z$ , o componente anterior posterior).

Vários estudos clínicos mostraram os benefícios do monitoramento utilizando EASI em Unidades Coronarianas, demonstrando que esse sistema é comparável ao padrão ECG de 12 derivações para detecção de anormalidades do ritmo cardíaco (DREW et al., 1999) e para Isquemia aguda do miocárdio e avaliação do infarto do miocárdio (JERNBERG; LINDAHL; WALLENTIN, 1999).



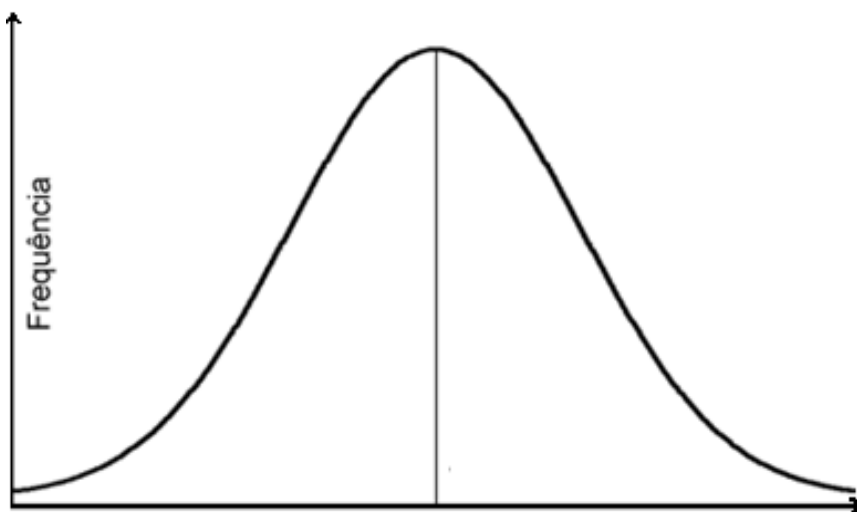
### 3 DISTRIBUIÇÃO NORMAL

Há muito tempo tem-se utilizado diversas técnicas da área de Estatística na criação de aplicações direcionadas à Inteligência Artificial, com o intuito de auxiliar vários setores da indústria, do comércio, da medicina, entre outros. Assim, no estudo de probabilidades, uma área muito difundida e usada é a Distribuição Normal.

Recebendo contribuições de matemáticos importantes, como De Moivre, Laplace e Gauss, a Distribuição Normal é uma distribuição contínua de probabilidades, em que a apresentação da distribuição de frequências  $f(x)$  de uma variável quantitativa  $x$  costuma apresentar-se em forma de sino e simétrica em relação à média (BRUNI, 2011).

No gráfico de distribuição de frequência (Figura 8), quanto mais próximo do valor central (média), maior será a concentração de frequências ou maior a probabilidade de ocorrências. Quanto mais se afasta do valor central, menor será a frequência e, com isso, menor será a probabilidade (BRUNI, 2011).

**Figura 8 - Curva “normal”**



Dessa forma, pode-se calcular a probabilidade de uma determinada variável ocorrer, analisando-se qual é a sua distância da média. A curva representada na Figura 8 varia de acordo com a amostra analisada, que dará origem a diferentes valores para a média e desvio-padrão. A média é calculada pela divisão da soma dos valores pelo número de observações, e o desvio-padrão (medida de dispersão) corresponde à raiz quadrada da variância. A variância de um conjunto de observações é uma média dos quadrados dos desvios das observações a partir de sua média (MOORE, 2011).

A média e o desvio-padrão da variável analisada  $x$  influenciam matematicamente a distribuição de frequência dos valores  $f(x)$ , cuja fórmula algébrica é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (1)$$

em que  $x$  é a variável normalmente distribuída;  $\sigma$  representa o desvio padrão e  $\mu$  é a média (BRUNI, 2011). Por meio dessas informações, é possível calcular a frequência/densidade de uma variável ou evento ocorrer, dada a média e o desvio-padrão da amostra, que são inseridos na Fórmula 1.

#### 4 TEOREMA DE BAYES

Também chamado de teorema da probabilidade *a posteriori*, é demonstrado por meio da fórmula:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}, \quad (2)$$

em que  $P(A)$  e  $P(B)$  são probabilidades *a priori*, tratando-se de dados já conhecidos. Assim, o teorema possibilita calcular a probabilidade de um evento A ocorrer visto que outro evento B já tenha ocorrido (MORETTIN, 2009; ZHANG; WEI, 2006).

As Redes Bayesianas são oriundas deste teorema e sua probabilidade pode ser calculada através da seguinte fórmula:

$$P(C_k|x) = \frac{P(x|C_k) \cdot P(C_k)}{P(x)}, \quad (3)$$

gerada através de uma inferência aplicada na Fórmula 2, onde  $P(C|x)$  é a probabilidade *posteriori* de C,  $P(x|C)$  é a probabilidade *a posteriori* de x,  $P(C)$  é a probabilidade *a priori* da hipótese C e  $P(x)$  é a probabilidade *a priori* de x (PEAR, 1991; OLIVEIRA, 2010).

Sendo assim, o teorema de Bayes relaciona causa e efeito de modo que, compreendendo o efeito, podemos descobrir a probabilidade de suas causas. Como resultado, essa proposição é importante tanto para determinar as causas de doenças, como o câncer, quanto é útil para determinar os efeitos de uma medição em particular sobre essa doença (LUGER, 2013).

## 5 DISTRIBUIÇÃO NORMAL E TEOREMA DE BAYES

Foi utilizado um classificador (KUKAR et al. 1997), associando dados obtidos através da Fórmula 1, aplicados na Fórmula 3. A estrutura do numerador desta última fórmula é equivalente ao modelo de probabilidade conjunta (DEGROOT; SCHERVISH, 2012), de forma que:

$$P(x|C_k).P(C_k) = P(C_k, x_1, \dots, x_n). \quad (4)$$

Aplicando-se a regra da cadeia (SCHUM, 1994), podemos escrever a fórmula acima da seguinte maneira:

$$\begin{aligned} P(C_k, x_1, \dots, x_n) &= \\ &= P(x_1, \dots, x_n, C_k) \\ &= P(x_1|x_2, \dots, x_n, C_k).P(x_2, \dots, x_n, C_k) \\ &= P(x_1|x_2, \dots, x_n, C_k).P(x_2|x_3, \dots, x_n, C_k) \dots \\ &P(x_{n-1}|x_n, C_k).P(x_n|C_k).P(C_k). \end{aligned} \quad (5)$$

Supondo que cada característica  $x_j$  é condicionalmente independente de qualquer outra característica  $x_i$ , para  $i \neq j$ . Seguindo pressupostos de independência condicional (DEGROOT; SCHERVISH, 2012), significa que:

$$P(x_i|x_{i+1}, \dots, x_n, C_k) = P(x_i|C_k). \quad (6)$$

O modelo conjunto pode ser expresso como:

$$\begin{aligned} P(C_k|x_1, \dots, x_n) &\propto P(C_k, x_1, \dots, x_n) \\ &\propto P(C_k).P(x_1|C_k).P(x_2|C_k).P(x_3|C_k) \dots \\ &\propto P(C_k). \prod_{i=1}^n P(x_i|C_k). \end{aligned} \quad (7)$$

A distribuição condicional (DEGROOT; SCHERVISH, 2012), sob a variável de classe C, sob pressupostos da independência acima é:

$$P(C_k|x_1, \dots, x_n) = \frac{1}{Z} P(C_k) \prod_{i=1}^n P(x_i|C_k). \quad (8)$$

Caso os valores das características  $x_i$  forem conhecidas, é dado uma constante  $Z$ , que, seguindo a terminologia da probabilidade bayesiana, pode ser tratada como a evidência  $Ev$ , que equivale a  $P(x)$  na Fórmula 3.

Ao lidar com dados contínuos, uma suposição típica é que os valores contínuos associados a cada classe são distribuídos de acordo com uma distribuição Gaussiana. Por exemplo, suponha que os dados de treinamento contiverem um atributo contínuo  $x$ . Primeiro, segmenta-se os dados pela classe e, em seguida, calculamos a média e variância de cada classe  $C_k$ . Suponhamos que  $y$  seja um valor de observação. Então, segundo Demichelis et al. (2006) a distribuição de probabilidade de  $y$ , pode ser computado utilizando a equação para distribuição de frequência (Fórmula 1) da seguinte forma:

$$P(x = y|C_k) = \frac{1}{\sqrt{2\pi\sigma_{C_k}}} e^{-\frac{1}{2}\left(\frac{y-\mu_{C_k}}{\sigma_{C_k}}\right)^2}. \quad (9)$$

## 6 REDES NEURAIIS ARTIFICIAIS

Possuindo diversas ramificações em diversas disciplinas como: Matemática, Estatística, Física, Ciências da Computação e Engenharia, as Redes Neurais Artificiais (RNAs) recebem este nome devido à sua inspiração biológica, onde o aprendizado está relacionado com a formação de conexões sinápticas entre os neurônios do sistema nervoso. Possuem aplicações em várias áreas, cuja habilidade de aprender a partir de dados de entrada, possibilita análise de séries temporais, modelagem, reconhecimento de padrões, processamentos de sinais e controle, entre outros (HAYKIN, 2001; FRENKEL; NADAL, 2000).

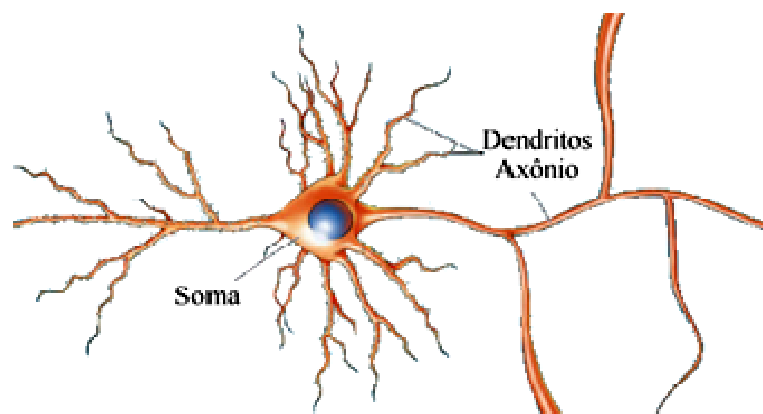
As RNAs tratam-se de uma imitação matemática de uma rede biológica de neurônios, baseado em suas interconexões sinápticas entre unidades fundamentais de informação (neurônios artificiais) (HAYKIN, 2001). Possuem características como: capacidade de aprendizado, capacidade de generalizar o conhecimento, armazenar o conhecimento de forma distribuída e tolerância a falhas, entre outras (SOUZA, 2012). Sendo assim, a possibilidade de aplicação das RNAs se torna muito abrangente e pode ser adaptada para resolver diversos problemas.

### 6.1 Neurônio biológico

Para estudar as RNAs e seu funcionamento, deve-se conhecer o funcionamento de um neurônio biológico, já que possui uma relação de semelhança com o neurônio artificial. No século 19, Ramón e Cajal, identificaram a célula nervosa (neurônio) anatomicamente e a descreveram com notável detalhe. Esta célula é delimitada por uma fina membrana celular que além da sua função biológica normal, possui determinadas propriedades que são essenciais para o seu funcionamento elétrico. A partir do corpo celular ou soma, o centro dos processos metabólicos da célula nervosa (Figura 9), projetam-se extensões filamentosas, os dendritos e os axônios (KOVÁCS, 2006).

Regularmente, os dendritos cobrem uma porção muitas vezes maior do que o próprio corpo celular e forma uma árvore dendrital. O axônio, considerada a outra projeção filamentar do corpo celular, serve para conectar a célula nervosa a outras do sistema nervoso. Geralmente o neurônio possui um único axônio, embora esse possa apresentar algumas ramificações e em alguns casos prolonga-se por distâncias comparáveis às dimensões do organismo podendo chegar a vários metros (KOVÁCS, 2006).

**Figura 9 – Modelo de um neurônio biológico**



Fonte: Bear, Connors, Paradiso, 2017. Adaptado.

Os impulsos nervosos transmitidos pelos neurônios, segundo Lorencetti (2011), ocorrem por meio de sinapses, que conduzem os impulsos do corpo para as células nervosas. As sinapses são capazes de gerenciar a transmissão destes sinais entre os nodos da rede.

A cada segundo de vida, o encéfalo executa bilhões de computações neurais, que só é possível graças aos 85 bilhões de neurônios que constituem o encéfalo humano. A área da superfície desta quantidade de neurônios representa aproximadamente 21.250 m<sup>2</sup>, o que equivale ao tamanho de três campos de futebol (BEAR; CONNORS; PARADISO, 2017).

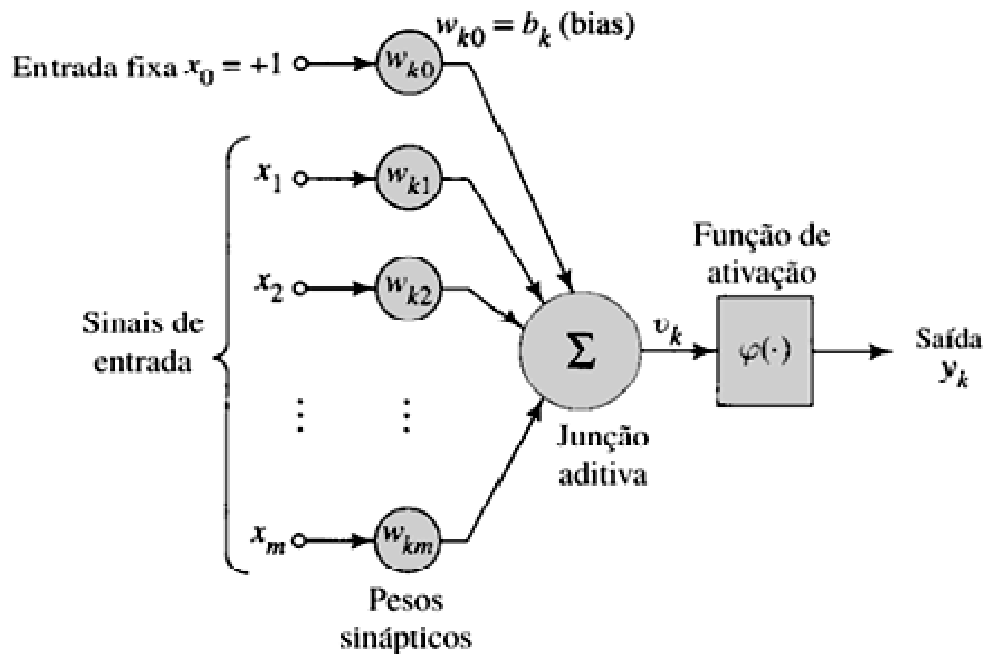
## 6.2 Neurônio artificial

Em 1943, Warren McCulloch, lembrado pelos anais da ciência como médico, filósofo, matemático e poeta, propôs, juntamente com o estatístico Walter Pitts o modelo de um neurônio artificial em uma publicação no *Bulletin of Mathematical Biophysics*, cujo artigo levou o título de “*A Logical Calculus of the Ideas Immanent in Nervous Activity*”. Este trabalho se tornou referência para a teoria das RNAs (KOVÁCS, 2006).

Matematicamente, o estímulo do neurônio artificial, é representado pelo conjunto de entradas, que são processados e convertidos em uma saída. Nele, são identificados cinco elementos básicos: um conjunto de sinapses, um somador, um limiar, uma função de ativação e uma saída (Figura 10). O bias  $b_k$  é um parâmetro externo do neurônio artificial  $k$  e sua função é transladar o valor de limiar da função.

Cada neurônio efetua um somatório da multiplicação dos dados de entrada pelos pesos, cujos valores já foram ajustados durante o processo de aprendizagem. A resposta ao estímulo corresponde ao valor de saída do neurônio (FRENKEL; NADAL, 2000).

**Figura 10 – Modelo de um neurônio artificial**



Fonte: Haykin, 2001.

A função  $v_k$  mostrada na Figura 10 é dada por:

$$v_k(n) = \sum_{j=1}^J w_{kj}(n)x_j(n) + w_{k0}(n), \quad (10)$$

onde  $x_1, x_2, \dots, x_j$ , são os dados de entrada,  $w_{k1}, w_{k2}, \dots, w_{kj}$ , são os pesos sinápticos e  $w_{k0}$  está relacionado com a correção de bias do k-ésimo neurônio (FRENKEL; NADAL, 2000).

### 6.3 Função de ativação

De acordo com Haykin (2001), a função de ativação, representada por  $\varphi(v)$ , define a saída de um neurônio em termos de campos local induzido  $v$ . Aqui, serão identificados três tipos básicos de funções de ativação:



1) função de limiar ou função de *Heaviside*, dada por:

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases}, \quad (11)$$

2) função linear por partes definida como:

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq +\frac{1}{2} \\ v, & +\frac{1}{2} > v > -\frac{1}{2} \\ 0, & \text{se } v \leq -\frac{1}{2} \end{cases} \quad (12)$$

3) função sigmóide ou função sinal é definida como:

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0, \\ -1 & \text{se } v < 0 \end{cases}, \quad (13)$$

onde, para dar forma a este função, pode-se utilizar a função tangente hiperbólica, definida como:

$$\varphi(v) = \tanh(v) \quad (14)$$

Desta forma o valor de  $v_k$  é submetido à função de ativação, produzindo uma saída  $y_k$ .

## 6.4 Processos de aprendizagem

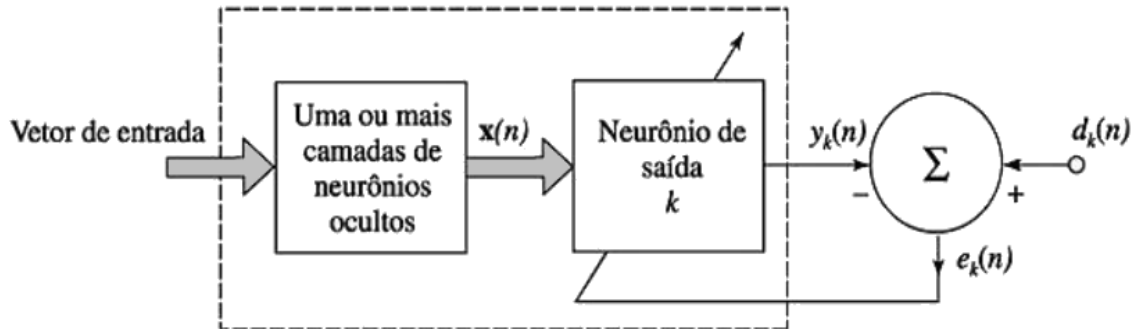
Segundo Hayking (2001), há vários processos de aprendizagem como: Baseada em Memória, Hebbiana, Competitiva, Boltzmann, com ou sem professor, Correção de Erro, entre outros.

Neste estudo será abordado o método de aprendizagem por Correção de Erro, onde a saída  $y_k$  (Figura 10) é usada para comparar com o resultado desejado ou saída-alvo (Figura

11). Na medida em que a saída desejada  $d_k(n)$  não é a esperada  $y_k(n)$ , o sinal de erro  $e_k(n)$  é calculado através da seguinte equação:

$$e_k(n) = d_k(n) - y_k(n). \quad (15)$$

**Figura 11 - Diagrama de blocos ressaltando a aprendizagem por correção de erro**



Fonte: Haykin, 2001

O sinal de erro  $e_k(n)$  aciona um mecanismo, cujo objetivo é aplicar uma sequência de ajustes nos pesos do neurônio  $k$ . Estes ajustes são feitos de forma que os pesos sejam projetados para que a saída encontrada se aproxime cada vez mais da saída desejada.

Utilizando-se a regra delta ou regra de Widrow-Hoff, cujo nome se dá em homenagem aos seus criadores (Widrow e Hoff, 1960), suponha que  $w_{kj}(n)$  represente o valor do peso sináptico  $w_{kj}$  do neurônio  $k$  excitado por um elemento  $x_j(n)$  do vetor de sinal  $x(n)$  no passo de tempo  $n$ . De acordo com esta regra, o ajuste  $\Delta W_{kj}(n)$  aplicado ao peso sináptico  $W_{kj}$  no passo de tempo  $n$  é definido por

$$\Delta W_{kj}(n) = \eta e_k(n) x_j(n), \quad (16)$$

onde  $\eta$  é uma constante positiva que determina a taxa de aprendizado quando avançamos em um passo no processo de aprendizagem e é portanto, denominado parâmetro taxa de aprendizagem (HAYKIN, 2001).

Desta forma, o peso atualizado  $W_{kj}$  é dado pela seguinte equação:

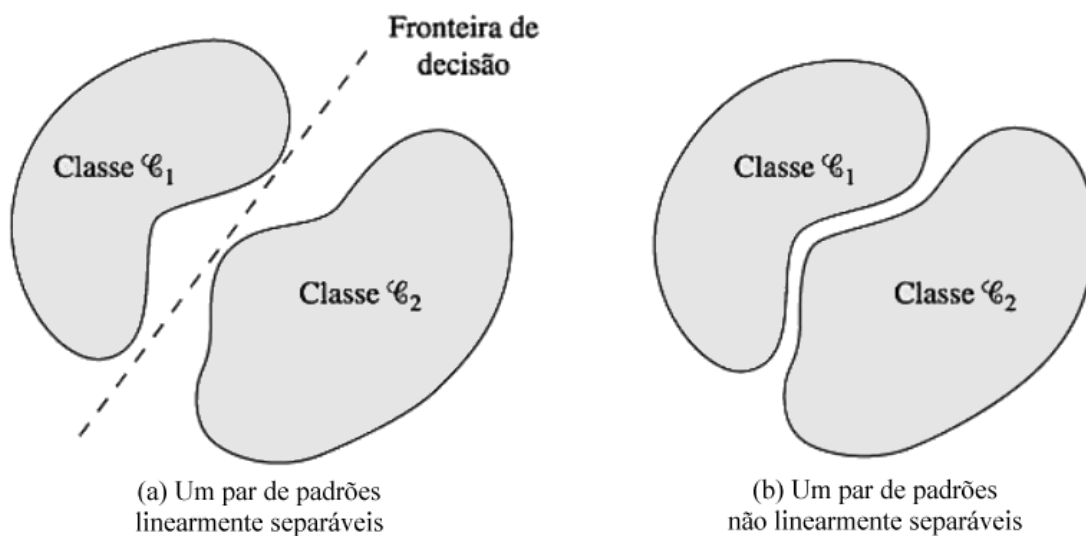
$$W_{kj}(n + 1) = W_{kj}(n) + \Delta W_{kj}(n), \quad (17)$$

onde o  $W_{kj}(n + 1)$  representa o peso atualizado e o  $W_{kj}(n)$  representa o peso antigo.

## 6.5 Perceptrons de Camada Única

O *perceptron* de camada única é a forma mais simplificada de uma rede neural, usada para a classificação de padrões ditos linearmente separáveis, ou seja, que estão em lados opostos de um hiperplano (Figura 12-a), consistindo em um único neurônio com pesos sinápticos ajustáveis e bias (HAYKIN, 2001).

Figura 12 – *Perceptron bidimensional*



Fonte: Haykin, 2001.

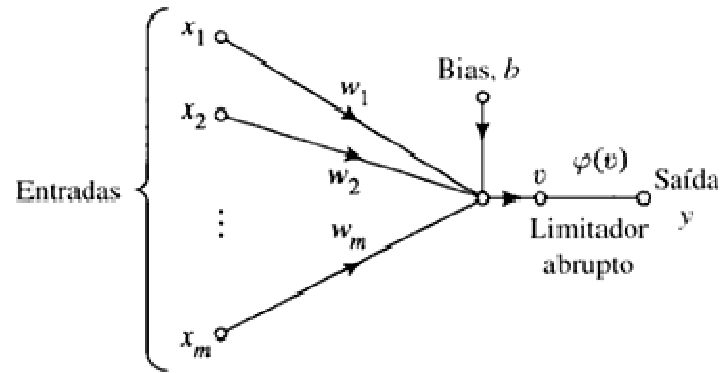
O uso de um único neurônio no *perceptron* propicia uma divisão linear de duas classes, no entanto é possível inserir novas classes neste modelo, desde que sejam linearmente separáveis, bastando para isso, adicionar novos neurônios na rede.

Este modelo de neurônio consiste de um combinador linear seguido por um limitador abrupto (função sinal) como mostra a Figura 13. O neurônio então produz uma saída igual a  $+1$  se a entrada do limitador abrupto for positiva e  $-1$  se for negativa. O objetivo deste *perceptron* é que, através de um conjunto de entradas:  $x_1, x_2, \dots, x_m$ , possa ser feita a classificação de duas classes. Se  $y$  for  $+1$  estará vinculada à classe  $\mathcal{C}_1$  e se  $-1$  estará vinculada à classe  $\mathcal{C}_2$ .

Deste modo, a função sinal em relação à resposta desejada  $d(n)$  é dada por:

$$d(n) = \begin{cases} +1 & \text{se } x(n) \text{ pertence à classe } C_1 \\ -1 & \text{se } x(n) \text{ pertence à classe } C_2 \end{cases} \quad (18)$$

Figura 13 – Grafo de fluxo de sinal do perceptron



Fonte: Haykin, 2001.

## 7 METODOLOGIA

### 7.1 Importação dos dados

Um passo importante em todo o processo é a importação dos registros do banco, que contém os resultados dos ECG, diagnósticos, descrição dos pacientes, entre outros. Para isso, foi usada a plataforma ATM da *Phisiobank*, disponibilizada por Goldberger et al., (2000), que possibilita a exportação e visualização de informação por meio do navegador web. Ao acessá-la, toda a documentação estará disponível, assim como os arquivos de cada paciente.

Neste estudo, o banco usado foi o *Long-Term ST Database (LTST-DB)*, e as demais informações foram preenchidas conforme Figura 15. Foram importados três arquivos de cada registro. O primeiro nomeado de arquivo A (Figura 14), contém detalhes do paciente e pode ser selecionado na plataforma por meio do item *Toolbox*, na opção *Describe Record*. Ao escolher esse item aparecerão informações como: sexo, idade, diagnósticos, histórico do paciente, entre outros.

**Figura 14 – Parte do arquivo A gerado do paciente s20161**

```
Record ltstdb/s20161
Notes
=====
Age: 65 Sex: M
Comments:
  This patient has Prinzmetal's angina, with episodes of
  major ST elevation in lead 0. In lead 1 the corresponding
  changes are ST depressions. Of interest, the QRS morphology
  changes slightly during the ischemic episodes, possibly
  due to transient chamber enlargement (Bradie effect). The
  time course of the morphology change is gradual, as opposed
  to the sudden axis shifts characteristic of postural
  changes.
  An excerpt of this recording is included in the European
  ST-T Database (record e0105).
Symptoms during Holter recording: No data
Diagnoses:
  Prinzmetal's angina
  Coronary artery disease
Treatment:
  Medications:
    Verapamil
  Balloon Angioplasty: No data
  Coronary Artery bypass Grafting: No data
History:
  Hypertriglyceridemia, coronary artery disease, previous myocardial infarction
  Hypertension: No
  Left ventricular hypertrophy: No
  Cardiomyopathy: No
  Valve disease: No
  Electrolyte abnormalities: No
  Hypercapnia, anemia, hypotension, hyperventilation: No
```

Fonte: Goldberger et al., (2000). Adaptado.

Figura 15 - Plataforma ATM

The screenshot shows the ATM platform interface with the following sections:

- Input:** Database: Long Term ST Database (Itsddb); Record: s20011; Signals: all; Annotations: automatically-generated, manually-corrected ST-segment measurements (16a).
- Output:** Length: 10 sec, 1 min, 1 hour, 12 hours, to end (selected); Time format: time/date, elapsed time, hours, minutes, seconds (selected), samples; Data format: standard (selected), high precision, raw ADC units.
- Toolbox:** Describe record.
- Navigation:** Navigation buttons (|<<, <<, <, \*, >, >>, >>|) and Previous record, Next record buttons.
- Help/About ATM:** Buttons at the bottom.

Fonte: Goldberger et al., (2000).

Para importação do arquivo B, foi necessário selecionar, no item *Toolbox*, a opção *Show annotations as text* e no item *Annotations* a opção *automatically-generated, manually-corrected ST-segment measurements (16a)* (Medições do Segmento ST corrigidas manualmente), em que serão apresentadas informações através dos campos: *seconds* (segundos), *chan* (canal), *Aux* (auxiliar), entre outros (Figura 16). No campo “auxiliar”, serão encontrados os valores dos pontos ( $J_1, \dots, J_7$ ), e a sintaxe desse campo pode ser vista na documentação do banco (JAGER et al. 2016).

Figura 16 – Parte do arquivo B gerado do paciente s20161

Seconds	Minutes	Hours	Type	Sub	Chan	Num	Aux
831.264	13.85440	0.2309067	s	0	1	0	ST1-53,+2,-30,-42,-58,-53,-47,-43,64,64,144,7,3
832.216	13.87027	0.2311711	s	0	0	0	ST0+223,+79,+134,+153,+182,+223,+268,+336,68,52,132,7,3
832.216	13.87027	0.2311711	s	0	1	0	ST1-52,+7,-28,-34,-45,-52,-43,-44,64,64,144,7,3
833.144	13.88573	0.2314289	s	0	0	0	ST0+231,+97,+139,+156,+191,+231,+281,+348,68,56,136,8,2
833.144	13.88573	0.2314289	s	0	1	0	ST1-52,+7,-28,-34,-45,-52,-43,-44,64,64,144,8,2
837.736	13.96227	0.2327044	s	0	0	0	ST0+237,+95,+145,+172,+199,+237,+284,+350,68,56,136,4,2
837.736	13.96227	0.2327044	s	0	1	0	ST1-40,+22,-26,-26,-31,-40,-33,-28,64,64,144,4,2
839.548	13.99247	0.2332078	s	0	0	0	ST0+227,+79,+123,+153,+190,+227,+278,+346,68,56,136,3,2
839.548	13.99247	0.2332078	s	0	1	0	ST1-45,+12,-37,-33,-36,-45,-37,-29,64,64,144,3,2
843.196	14.05327	0.2342211	s	0	0	0	ST0+427,+198,+267,+318,+379,+427,+490,+551,68,56,136,2,5
843.196	14.05327	0.2342211	s	0	1	0	ST1-9,+24,-11,-2,-2,-9,-5,+5,64,64,144,2,5
846.716	14.11193	0.2351989	s	0	0	0	ST0+447,+207,+282,+335,+398,+447,+521,+593,68,56,136,2,7
846.716	14.11193	0.2351989	s	0	1	0	ST1-7,+8,-17,-8,-2,-7,-8,+2,64,64,144,2,7
847.592	14.12653	0.2354422	s	0	0	0	ST0+474,+212,+290,+352,+424,+474,+549,+622,68,56,136,2,6
847.592	14.12653	0.2354422	s	0	1	0	ST1-6,+5,-15,-11,-7,-6,-4,+6,64,64,144,2,6

Fonte: Goldberger et al., (2000). Adaptado.

Por conter mais de 24 horas de dados, não será possível a visualização de todo o conteúdo no navegador e, por isso, é necessário o uso do programa “rdann”, contido no pacote do Software WFDB, que faz parte de um conjunto de ferramentas (SEMIA), criado para desenvolver e visualizar o banco em questão (GOLDBERGER, 2000; JAGER et al., 2016).

O arquivo C (Figura 17) é importado selecionando-se, no item *Annotations*, a opção *ST-segment episode annotations, Vmin=75 microV, Tmin=30s (sta)* (Anotações de segmentos ST). Por intermédio desta opção é possível averiguar a ocorrência de episódios isquêmicos e em que momentos ocorreram.

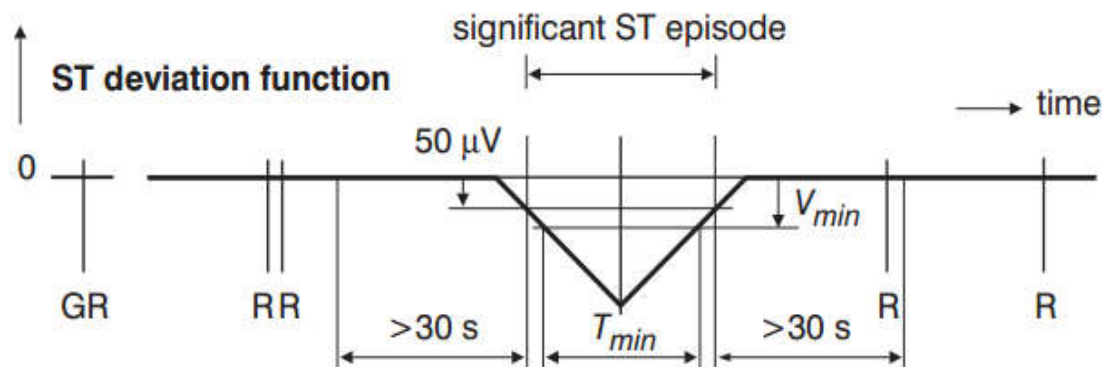
**Figura 17 – Parte do arquivo C gerado do paciente s20161**

Seconds	Minutes	Hours	Type	Sub	Chan	Num	Aux
...							
595.000	9.91667	0.1652778	s	0	1	0	LRST1+26
597.000	9.95000	0.1658333	s	0	0	0	LRST0+81
609.000	10.15000	0.1691667	s	0	0	0	sst0
621.000	10.35000	0.1725000	s	0	0	0	LRST0-7
629.000	10.48333	0.1747222	s	0	1	0	LRST1+11
763.000	12.71667	0.2119444	s	0	0	0	LRST0-8
765.000	12.75000	0.2125000	s	0	1	0	LRST1-11
783.000	13.05000	0.2175000	s	0	0	0	sst0
807.000	13.45000	0.2241667	s	0	0	0	LRST0+84
829.000	13.81667	0.2302778	s	0	0	0	LRST0+45
829.000	13.81667	0.2302778	s	0	1	0	LRST1-9
839.000	13.98333	0.2330556	s	0	0	0	(st0+72
1059.000	17.65000	0.2941667	s	0	0	0	ast0+553
1139.000	18.98333	0.3163889	s	0	0	0	st0+44)
1173.000	19.55000	0.3258333	s	0	0	0	LRST0+36
1173.000	19.55000	0.3258333	s	0	1	0	LRST1-2
1211.000	20.18333	0.3363889	s	0	0	0	LRST0+1
1213.000	20.21667	0.3369444	s	0	1	0	LRST1-17
...							

Fonte: Goldberger et al., (2000). Adaptado.

Assim como no arquivo B, os mesmos campos foram utilizados, com a diferença de que na coluna auxiliar, há outros tipos de informações, como nomenclaturas e siglas, que estão relacionadas ao diagnóstico do batimento da linha de tempo em questão (Figura 18). Um episódio significativo de IC é representado pela sigla ST (JAGER et al., 2003). Por exemplo, na Figura 17, o intervalo de tempo entre 839 segundos e 1139 segundos corresponde a um período significativo com relação à isquemia, visto que a nomenclatura “(st0+72” representa o início e a “st0+44)” representa o fim o ciclo (JAGER et al., 2003).

**Figura 18 - Episódio Significante (ST) relacionado à IC**



Fonte: Jager et al. (2003). Adaptado.

A importação desses três arquivos deve ser feita para cada indivíduo, que pode ser selecionado por meio do campo *Record* (Figura 15). Outra informação também de grande importância é a do campo *Signals*, em que são mostradas as derivações registradas daquele paciente em questão, representada nos arquivos importados através da coluna *Chan* (0, 1 ou 2). Em alguns casos, esse campo mostra a sigla ECG, significando que não foi possível ou não foi registrado no momento do exame a derivação utilizada. Varias derivações foram registradas, sendo apresentadas as seguintes: ML2, ML3, MV2, V2, V3, V4, V5, V6, A-S, E-S e A-I, em que as iniciadas com “M” são as que foram modificadas. Ambas foram geradas pelo equipamento “Philips ZymedHolter”, usado para fazer os exames e registros (JAGER et al., 2003).

Para a utilização dos dados nos algoritmos propostos é necessária a utilização de um único arquivo com todos os dados necessários e por isso foi realizada a junção de algumas informações do arquivo B – tempo, canal e pontos ( $J_1, \dots, J_7$ ) –, com algumas do arquivo C – tempo, canal, diagnóstico da linha de tempo –, gerando, assim, o arquivo D (Figura 19), com a seguinte sintaxe: segundos, canal, J+0 ( $J_1$ ), J+20 ( $J_2$ ), J+40 ( $J_3$ ), J+60 ( $J_4$ ), J+80 ( $J_5$ ), J+100 ( $J_6$ ), J+120 ( $J_7$ ) e resultado. O resultado possui o valor de 1, significando que aquela linha de tempo e pontos ( $J_1, \dots, J_7$ ) correspondem a um batimento isquêmico e o valor de 2 para representar um batimento não isquêmico. Fazer essa combinação dos arquivos B e C manualmente demanda muito esforço e, portanto, foi confeccionado um algoritmo apenas para esse fim. Para isso, foi utilizada a linguagem de programação PHP (*Personal Home Page Tools*), criada no ano de 1994 por Rasmus Lerdof, distribuída sob a licença PHP v3.01, sendo a mais utilizada para desenvolvimento web (DALL’OGLIO, 2016).



**Figura 19 - Parte do arquivo D, pertencente ao paciente s20161**

```

...
826.540,0,+57,+116,+140,+173,+220,+279,+363,68,52,132,2
827.452,0,+56,+116,+140,+173,+218,+276,+357,68,52,132,2
828.380,0,+51,+111,+133,+167,+212,+270,+352,68,52,132,2
829.348,0,+55,+109,+130,+166,+214,+274,+357,68,52,132,2
830.304,0,+53,+112,+138,+177,+228,+284,+360,68,52,132,2
831.264,0,+69,+127,+146,+184,+231,+279,+350,68,52,132,2
832.216,0,+79,+134,+153,+182,+223,+268,+336,68,52,132,2
833.144,0,+97,+139,+156,+191,+231,+281,+348,68,56,136,2
837.736,0,+95,+145,+172,+199,+237,+284,+350,68,56,136,2
839.548,0,+79,+123,+153,+190,+227,+278,+346,68,56,136,1
843.196,0,+198,+267,+318,+379,+427,+490,+551,68,56,136,1
846.716,0,+207,+282,+335,+398,+447,+521,+593,68,56,136,1
847.592,0,+212,+290,+352,+424,+474,+549,+622,68,56,136,1
848.484,0,+212,+290,+352,+424,+474,+549,+622,68,56,136,1
849.384,0,+209,+288,+344,+413,+467,+544,+618,68,56,136,1
850.256,0,+209,+288,+344,+413,+467,+544,+618,68,56,136,1
852.044,0,+217,+299,+357,+433,+494,+571,+651,68,56,136,1
852.944,0,+217,+299,+357,+433,+494,+571,+651,68,56,136,1
...

```

Após a importação, os 86 registros foram submetidos a um filtro de forma que foram descartados os que não possuíam derivações conhecidas, representados pela sigla ECG no campo *Signals*. Usando o arquivo que descrevia os detalhes de cada paciente (arquivo A), foram descartados, também, os que não apresentavam registros significantes de Isquemia Cardíaca em sua base.

A partir dessas informações, os registros foram separados por derivações, resultando num banco com 12 registros de derivação A-I, 12 registros com A-S, 12 registros com E-S, 4 registros com ML2, 14 registros com ML3, 4 registros com MV2, 10 registros com V4 e 3 registros com V5.

O estudo foi feito para analisar cada derivação separadamente; portanto, os dados produzidos de cada uma foram separados em dois arquivos: um para efetuar o aprendizado (treinamento), e outro para efetuar o teste (classificação), cuja proporção foi de 70% e 30% respectivamente.

## 7.2 Classificação com algoritmo híbrido

O processo de aprendizagem é feito separando-se a base de aprendizado em duas classes  $\mathcal{C}$ : uma de registros com batimentos isquêmicos e outra com não isquêmicos. Para cada uma das classes, foi calculada a média  $\mu_{C_k}$  e o desvio-padrão  $\sigma_{C_k}$  dos pontos  $(J_1, \dots, J_7)$

separadamente, de forma que se obteve média dos pontos  $J_1$  isquêmicos, média dos pontos  $J_1$  não isquêmicos, desvio-padrão dos pontos  $J_1$  isquêmicos, desvio padrão dos pontos  $J_1$  não isquêmicos e assim subsequentemente com todos os pontos  $J_i$ , cujos cálculos foram feitos por um algoritmo desenvolvido utilizando-se PHP como linguagem de programação. Também foi obtida a probabilidade *a priori*  $P(C_k)$ , que consiste na percentagem de registros rotulados como isquêmicos ( $k = 1$ ) e na de registros não isquêmicos ( $k = 2$ ) da base de treinamento.

Com os dados do aprendizado prontos, o próximo passo é fazer a classificação da base de testes. Nesse momento, é calculada, por meio da fórmula de Distribuição Normal (Fórmula 9), a probabilidade de distribuição de frequência (densidade) de cada ponto do segmento ST, em que  $i$  assumirá valores de 1 a 7 e  $k$  valores 1 e 2, ficando da seguinte forma:

$$P(x = J_i | C_k) = \frac{1}{\sqrt{2\pi\sigma_{C_k}}} e^{-\frac{1}{2}\left(\frac{J_i - \mu_{C_k}}{\sigma_{C_k}}\right)^2}. \quad (19)$$

Os resultados acima, são aplicados então na Fórmula 8, que após adaptação é apresentada como:

$$P(C_k | J_1, \dots, J_7) = \frac{P(C_k) \prod_{i=1}^7 P(J_i | C_k)}{Ev}, \quad (20)$$

onde  $Ev$  (evidência) neste caso é dada por:

$$Ev = \sum_{k=1}^2 P(C_k) \prod_{i=1}^7 P(J_i | C_k). \quad (21)$$

Após a aplicação proposta, serão obtidas duas probabilidades *posteriori*: a do batimento em análise ser isquêmico e a de ser não isquêmico, em que a soma das duas será 100%.

Seguindo a ideia de Duda, Hart e Stork (2001), foi necessário estabelecer um Limiar de Decisão ( $L_d$ ) para definir a partir de qual probabilidade será definida a classificação. Por exemplo: suponha que a Fórmula 20 apresente a probabilidade de um determinado batimento ser isquêmico de  $P_{(Isquêmico)} = 70\%$  e  $P_{(Não Isquêmico)} = 30\%$ . Se a condição de classificação para batimento isquêmico for  $P_{(Isquêmico)} > L_d$  e  $L_d = 40\%$ , então,

a classificação será dada como isquêmico, visto que a probabilidade do batimento ser isquêmico é maior do que o Limiar de Decisão ( $70\% > 40\%$ ). No entanto, se  $L_d = 80\%$ , o resultado será não isquêmico, visto que a probabilidade de ser isquêmico neste caso será menor do que o Limiar de Decisão ( $70\% < 80\%$ ). Todo esse processo tem que ser feito com cada linha de tempo (batimento) contido no banco de dados para teste e, por isso, foi necessário criar um algoritmo para automatizar o processo.

### 7.3 Classificação usando Redes Neurais Artificiais

Para este trabalho usaremos o modelo do perceptron de Rosenblatt (*perceptron*), onde se adotou o uso de um neurônio com 7 entradas correspondentes aos pontos do segmento ST ( $J_i, \dots, J_n$ ), com  $n=7$  e o valor do bias como sendo 1, gerando uma única saída (7-1-1). A função de ativação adotada é a do tipo limiar ou função de *Heaviside* e o valor da taxa de aprendizagem  $\eta$  é 0.05. Os valores dos pesos iniciais  $w_k(n)$  foram determinados de forma aleatória com o valor de 0.3.

O teorema de convergência do *perceptron*, derivado do algoritmo de aprendizagem por Correção de Erro, resulta na fórmula:

$$w(n + 1) = w(n) + \eta[d(n) - y(n)].x(n), \quad (22)$$

onde  $d(n) - y(n)$  assume o papel de um sinal de erro (HAYKIN, 2001).

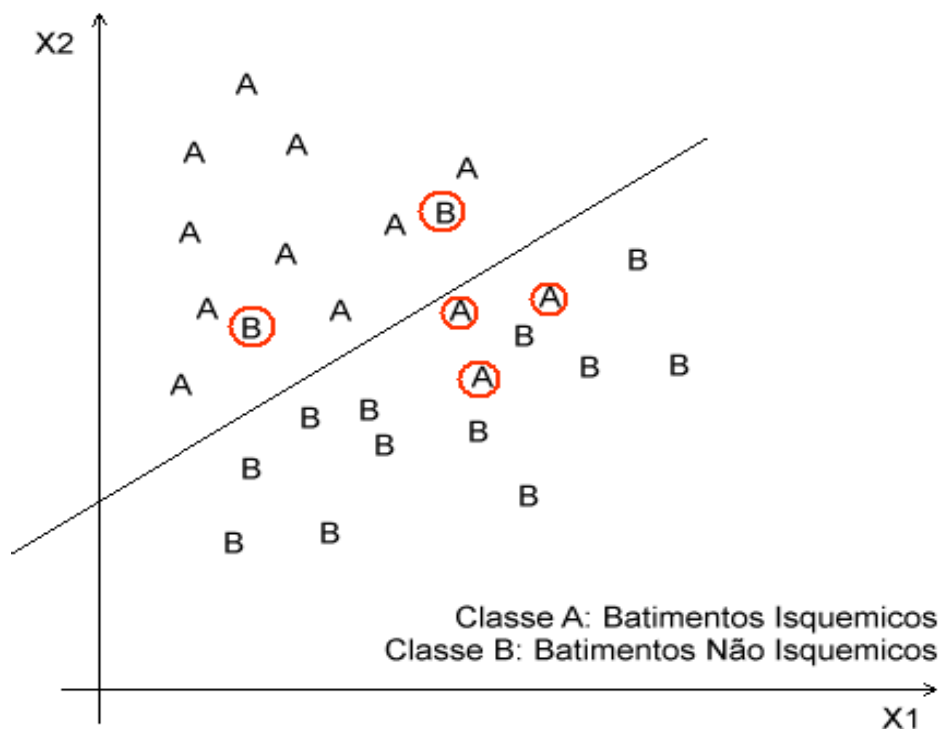
Desta forma, para cada arquivo de aprendizagem gerado de cada derivação, é feito o processo de treinamento, aplicando-se a Fórmula 22 para que o ajuste dos pesos seja feito no momento da interação em que o erro for encontrado.

O processo de ajuste dos pesos é interrompido no momento em que não seja encontrado algum erro em todo o banco do aprendizado. No entanto, observou-se que o banco não é totalmente linearmente separável. Portanto, o critério de parada não seria ativado e o algoritmo entraria num *loop* interminável.

Para sanar o problema foi feita uma adaptação do programa, para que, quando atingisse determinada quantidade de iterações ele parasse. Portanto, foi feito um processo de aprendizagem de 10.000 iterações em cada banco para treinamento. Para cada iteração  $i$ , foi registrada a quantidade de erros encontrados e foi determinado que os melhores pesos admitidos fossem os que correspondessem a iteração em que ocorressem menos falhas. Desta forma, puderam-se separar as classes (isquêmicos / não isquêmicos) de forma que alguns

erros fossem aceitos. A Figura 20 mostra, com pontos fictícios, como ocorre a separação das classes neste experimento, onde os pontos A e B com círculo correspondem às falhas encontradas pelo algoritmo. Deste modo, os pesos foram ajustados de forma que o mínimo de erros fosse encontrado durante o processo.

**Figura 20 - Separação linear parcial de duas classes não linearmente separáveis**



Após o processo de aprendizado, os melhores pesos para cada derivação foram armazenados e usados nas bases de testes. O processo de testes ocorre aplicando os valores dos pontos  $(J_1, \dots, J_7)$  como entradas do neurônio, adaptando-se a Fórmula 10 da seguinte forma:

$$y(n) = \sum_{i=1}^7 w_i(n) \cdot J_i(n) + Bias, \quad (23)$$

onde  $w_i$  corresponde aos pesos de cada entrada determinado no processo de aprendizagem. Com base na função de limiar, a saída do neurônio  $y$  na linha de tempo (batimento)  $n$  assumindo um valor maior ou igual a zero será vinculado à classe  $\mathcal{C}_1$ , que foi relacionada ao batimento isquêmico. Caso a saída  $y(n)$  seja menor que zero será vinculado à classe  $\mathcal{C}_2$ , que foi relacionada ao batimento não isquêmico.

Assim sendo, o algoritmo neural fez a classificação em cada base, de forma a calcular o resultado em cada linha de tempo (batimento) contido no banco de dados para teste e, por isso, foi necessário criar um algoritmo para automatizar o processo. Dessa forma, é possível classificar e comparar cada resultado de classificação com o resultado já rotulado no banco de dados e fazer uma avaliação do método usado.

#### 7.4 Avaliação

Dessa forma, é possível classificar e comparar cada resultado de classificação com o resultado já rotulado no banco de dados e fazer uma avaliação do método usado.

A avaliação proposta foi o uso da Tabela de Contingência (Tabela 1), de que se obtêm quatro variáveis (A, B, C e D) (OLIVEIRA; ANDREÃO; FILHO, 2016).

**Tabela 1 – Tabela de Contingência**

	Diagnóstico da Base de Dados		
Diagnóstico do Algoritmo	Isquêmico	Não Isquêmico	Total
Isquêmico	A Verdadeiro Positivo	B Falso Negativo	A + B Batimentos classificados como Isquêmicos
Não Isquêmico	C Falso Positivo	D Verdadeiro Negativo	C + D Batimentos classificados como não isquêmicos
Total	A + C Batimentos rotulados como isquêmicos	A + D Batimentos rotulados como não isquêmicos	A + B + C + D

Fonte: Oliveira, Andreão e Filho (2016). Adaptado.

Com base nessas variáveis, puderam-se calcular as seguintes medidas de desempenho:

1. Sensibilidade, dada pela fórmula:

$$S_e = \frac{A}{A + C}, \quad (24)$$

definida pela capacidade do sistema reconhecer verdadeiros positivos em relação ao total de batimentos isquêmicos. Corresponde a porcentagem dos batimentos classificados como isquêmicos em relação ao total de batimentos diagnosticados como isquêmicos;

2. Especificidade dada pela fórmula:

$$E_s = \frac{D}{B + D}, \quad (25)$$

é a probabilidade de o sistema classificar os verdadeiros negativos. Equivale à porcentagem dos batimentos classificados como não isquêmicos em relação ao total de batimentos diagnosticados como não isquêmicos;

3. Valor Preditivo Positivo dado pela fórmula:

$$VPP = \frac{A}{A + B}, \quad (26)$$

indicando o grau de certeza dos batimentos classificados como isquêmicos. Corresponde a porcentagem dos batimentos classificados como isquêmicos e diagnosticados como isquêmicos em relação ao total de batimentos classificados como isquêmicos;

4. Valor Preditivo Negativo, dado pela fórmula:

$$VPN = \frac{D}{C + D}, \quad (27)$$

indica o grau de certeza dos batimentos classificados como não isquêmicos. Trata-se da porcentagem dos batimentos classificados como não isquêmicos e diagnosticados como não isquêmicos em relação ao total de batimentos classificados como não isquêmicos (NASSAR, 2012). O que se procura nessas quatro medidas são porcentagens que se aproximem ao máximo de 100%.

Como o objetivo do trabalho é fazer uma comparação dos resultados em diferentes derivações, que foram agrupadas separadamente, é necessário estabelecer um diferencial de comparação levando-se em conta a quantidade de batimentos de cada grupo.

Para isto, utilizou-se a metodologia dada por Larson e Farber (2012), adotando para validação o intervalo de confiança (IF), que é uma medida da certeza de que o intervalo contém o parâmetro populacional. Para cada índice de desempenho obtido, será acrescentado um intervalo de confiança de 95% em torno da  $S_e$ ,  $E_s$ , VPP e VPN. Para atingir este intervalo é necessário usar o valor de  $Z_c$  igual a 1.96, significando que 95% da área sob a curva normal padrão está dentro de 1.96 desvios padrão da média (LARSON; FARBER, 2010). Um intervalo de confiança  $c$  para a proporção populacional  $p$  é  $p' - E < p < p' + E$ , onde

$$E = Z_c \sqrt{\frac{p'q'}{n}}, \quad (28)$$

em que  $E$  corresponde a margem de erro,  $q'$  é  $1 - p'$  e  $n$  é o total da amostra (LARSON; FARBER, 2012). Percebe-se, na Fórmula 28, que, quanto maior o número de  $n$ , que neste caso corresponde ao total de batimentos analisados, menor será a margem de erro o que origina um intervalo de confiança menor e mais preciso.

## 8 RESULTADOS E DISCUSSÕES

### 8.1 Algoritmo híbrido

Os resultados foram categorizados de acordo com as derivações. Para cada uma, foram testados os seguintes índices de  $L_d$ : 1, 0.999, 0.995, 0.99, 0.98, 0.97, 0.96, 0.95, 0.9, 0.85, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05, 0.01 e 0.001. Os mesmos foram escolhidos aleatoriamente, de forma a englobar percentagens entre 0 e 100%.

Os resultados foram agrupados na Tabela 2, onde são apresentados os melhores índices de cada derivação e os limiares de decisão que geraram estes valores. Com relação aos intervalos de confiança, considerar os índices apresentados para mais ou para menos.

**Tabela 2 – Melhores resultados de cada derivação utilizando algoritmo híbrido**

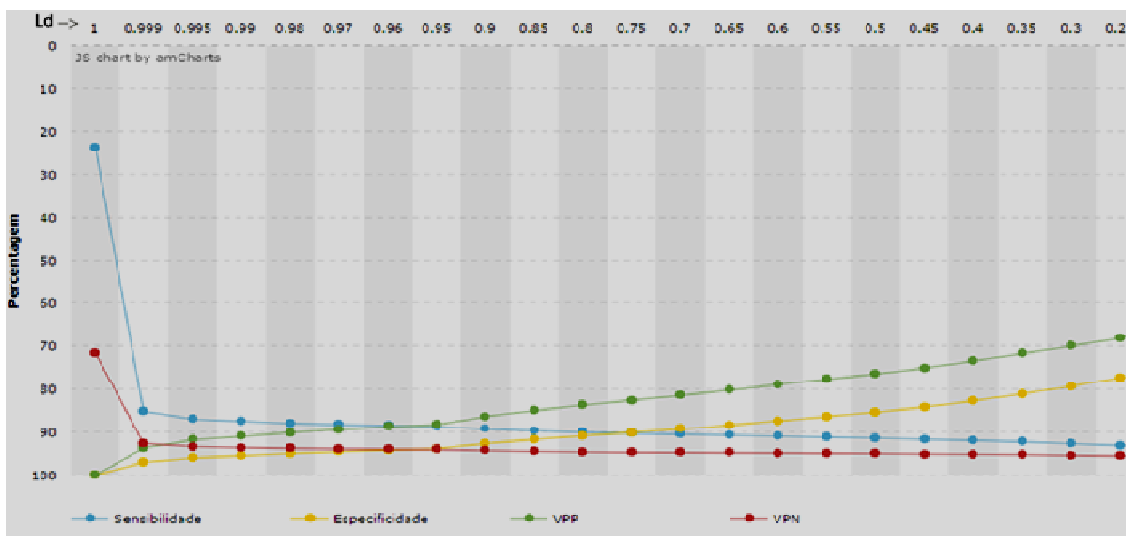
Derivações	<i>A-I</i>	<i>A-S</i>	<i>E-S</i>	<i>ML2</i>	<i>ML3</i>	<i>MV2</i>	<i>V4</i>	<i>V5</i>	
<b>Melhor <math>L_d</math></b>	0.15	0.005	0.05	0.001	0.005	0.995	0.95	0.01	0.05
$S_e$ (%)	64.97	99.22	94.36	99.91	98.02	79.20	88.69	100	99.9
IF - $S_e$ ( $\times 10^{-4}$ )	35.5	5.49	34.1	2.52	11.7	144	22.7	0	1.79
$E_s$ (%)	66.91	99.99	96.39	97.77	97.77	77.41	93.88	97.9	99.8
IF - $E_s$ ( $\times 10^{-4}$ )	35	0.625	27.6	12.4	12.4	148	17.2	12.9	3.90
VPP (%)	49.57	99.99	93.41	99.91	95.76	57.65	88.29	95.9	99.6
IF - VPP ( $\times 10^{-4}$ )	37.2	0.625	36.7	2.52	16.9	175	23	17.8	5.51
VPN (%)	66.91	99.61	96.92	97.77	98.97	90.55	94.10	100	99.9
IF - VPN ( $\times 10^{-4}$ )	35	3.89	25.5	12.4	8.46	104	16.9	0	1.27

A fim de automatizar o processo, o próprio algoritmo, ao final do processamento, gera um gráfico (Gráfico 1), em que é possível averiguar o melhor valor de  $L_d$ , observando-se as variações das medidas de  $S_e$ ,  $E_s$ , VPP e VPN para determinado conjunto de testes. Como é



possível visualizar, os resultados diferem, conforme for o valor do Limiar de Decisão. O gráfico abaixo torna mais compreensível a visualização dos resultados e a comparação dos mesmos. O cabeçalho do gráfico mostra os diferentes Limiares usados e em cada coluna são mostrados os valores dos quatro índices. O melhor  $L_d$  escolhido é aquele em que foram obtidos os melhores resultados, ou seja, quando os mesmos se aproximarem o máximo possível de 100%.

**Gráfico 1 - Gráfico gerado com testes na derivação V4**



## 8.2 Algoritmo aplicando RNAs

Os resultados foram categorizados de acordo com as Derivações e, para cada uma, foram usados os pesos que representaram os melhores resultados, que foram agrupados na Tabela 3, onde são apresentados os melhores índices de cada derivação com seus respectivos intervalos de confiança. Ao contrário do algoritmo híbrido, este não contém o fator de Limiar de Decisão, pois neste caso a classificação é feita apresentando como resposta positiva ou negativa para a IC e não como uma porcentagem de que seja ou não Isquemia.

**Tabela 3 - Melhores resultados em cada derivação usando RNAs**

Derivações	A-I	A-S	E-S	ML2	ML3	MV2	V4	V5
$S_e$ (%)	95.19	99.98	99.39	95.96	82.32	94.51	95.22	94.79
$IF - S_e$ ( $\times 10^{-4}$ )	15.8	0.878	11.44	16.44	31.79	80.28	15.21	19.78

$E_s$ (%)	98.19	100	96.26	99.51	90.31	89.75	40.52	77.77
IF - $E_s$ ( $\times 10^{-4}$ )	9.87	0	27.88	5.83	24.65	106.9	35	37
VPP (%)	96.35	100	93.51	99.15	81.34	78.16	45.44	70.21
IF - VPP ( $\times 10^{-4}$ )	13.89	0	36.20	7.66	32.46	145.6	35.5	40.7
VPN (%)	97.61	99.99	99.66	97.63	90.87	97.68	94.22	96.43
IF - VPN ( $\times 10^{-4}$ )	11.31	0.621	8.55	12.70	24	53	16.6	16.5

### 8.3 Resultados de outros trabalhos

Outros estudos foram realizados (Tabela 4), utilizando a mesma base LTST-DB, com o intuito de classificação de isquemias:

**Tabela 4 – Resultados de outros trabalhos aplicando a base LTST-DB**

Autor	Método	$S_e$ (%)	$E_s$ (%)	VPP (%)	VPN (%)
<b>Smrdel e Jager (2004)</b>	Transformada de Karhunen-Loeve (KLT)	78.9	-	80.7	-
<b>Dranca, Goñi e Lllarramendi (2009)</b>	Árvores de decisão	68.62	-	74.91	-
<b>Faganeli e Jager (2010)</b>	KLT	98.1	85.2	-	-
<b>Song et al. (2012)</b>	Lógica Fuzzy	75	85	78	87
<b>Amon e Jager (2016)</b>	KLT	91	85	-	-
<b>Smyrnakis e Evans (2007)</b>	Inferência Bayesiana	82,3	91.2	-	-

Observa-se, que na maioria desses estudos similares, foram usadas apenas duas medidas de desempenho, como:  $S_e$  e VPP em Smrdel e Jager (2004) e Dranca, Goñi e Lllarramendi (2009) ou  $S_e$  e  $E_s$  em Faganeli e Jager (2010), Amon e Jager (2016) e Smyrnakis e Evans (2007) e por último em Song et al., que, assim como neste trabalho, apresentou as quatro medidas de desempenho.

Em estudos utilizando outras bases (Tabela 5), podem-se observar métodos diversos usados a fim de detectar a IC. Nesta tabela, observou-se uma quantidade maior de trabalhos, mas, assim como na base LTST-DB, a maioria não apresenta os quatro índices de medições, com exceção apenas de Oliveira (2010).

**Tabela 5 – Resultados de outros trabalhos aplicando outras bases**

<b>Autor</b>	<b>Método</b>	<b><math>S_e</math> (%)</b>	<b><math>E_s</math> (%)</b>	<b>VPP (%)</b>	<b>VPN (%)</b>
<b>Kukar et al. (1997)</b>	<i>Naive Bayes</i>	91.7	80.1	-	-
<b>Papaloukas (2002)</b>	RNA	90	90	-	-
<b>Jager (1992)</b>	<i>KLT</i>	85	-	86	-
<b>Fotiades (2001)</b>	<i>Knowledge-based technique</i>	93.8	-	78.5	-
<b>Oliveira (2010)</b>	Redes Bayesianas	99.53	100	100	99.97
<b>Silipo e Marchesi (1998)</b>	<i>Threshold</i>	77	-	82	-
<b>Silipo e Marchesi (1998)</b>	<i>sANN</i>	77	-	83	-
<b>Silipo e Marchesi (1998)</b>	RNA	77	-	85	-
<b>Silipo e Marchesi (1998)</b>	<i>KLN</i>	71	-	66	-
<b>Silipo e Marchesi (1998)</b>	<i>sANN + PCA</i>	77	-	86	-
<b>Silipo, Taddei e Marchesi (1994)</b>	RNA	85	-	88	-
<b>Silipo, Taddei e Marchesi (1994)</b>	RNA	78	-	90	-
<b>Stamkopoulos (1992)</b>	RNA	84	-	79	-
<b>Taddei (1995)</b>	Próprio	84	-	85	-
<b>Vila et al. (1997)</b>	Software SUTIL	83	-	75	-

Nesta, podemos observar métodos similares ao usado neste trabalho, como por exemplo, o emprego das RNAs, utilizado por diversos autores. Com relação ao algoritmo Híbrido foi encontrado apenas o trabalho de Kukar et al. (1997), como método similar, que utilizou *Naive Bayes Classifier*, inclusive implementando também a idéia de Limiar de Decisão.

## 9 CONCLUSÕES

Foi possível observar com o estudo que a classificação proposta teve diferentes resultados em distintas derivações. Como foram usadas proporções iguais para o aprendizado e teste em todos os bancos preparados, pode-se dizer que a detecção da IC em algumas derivações teve resultados mais significativos que em outras.

Os melhores índices encontrados utilizando o classificador híbrido foram nas derivações A-S, E-S, ML2, ML3 e V5, tendo a primeira e a última atingindo os melhores índices, com a V5 atingindo 100% na  $S_e$  e VPN e intervalo de confiança de  $\pm 0$  para ambas. A derivação V4 apresentou resultados satisfatórios, com bons intervalos de confiança. Por fim, A-I e MV2 não apontaram bons resultados.

O classificador utilizando RNAs apresentou melhores resultados nas derivações ML2, A-I, E-S e A-S, sendo a última com os melhores valores, atingindo 99.98% na  $S_e$  e intervalo de confiança de  $\pm 0$  na Especificidade e Valor Preditivo Positivo. As demais derivações apontaram resultados regulares.

Pôde-se observar, também, que diferentes  $L_{dS}$  foram selecionados para cada derivação, indicando a melhor fronteira de separação entre as duas classes (isquêmico e não isquêmico), e outro estudo utilizando um banco de dados diferente seria necessário para determinar se esse valor sofreria alguma alteração ou seguiria o mesmo padrão.

Verificou-se também que o processo de aprendizado utilizando-se RNAs, exige um maior esforço de *hardware* computacional e por isso teve-se que limitar a quantidade de iterações. No entanto, caso houvesse um número maior delas, provavelmente seriam identificados melhores pesos e assim resultados ainda melhores.

Algumas derivações não foram incluídas pelo fato de o banco de dados LTST-DB não as conter. Portanto a análise de outra base com maior quantidade de registros poderia ter uma maior abrangência. Espera-se fazer mais testes com essas mesmas bases, utilizando os dois algoritmos em conjunto e analisar assim a possibilidade de melhorar os resultados.

As melhores derivações comuns entre os dois algoritmos foram A-S, E-S e ML2, cujos melhores resultados apresentados foram do algoritmo de RNA. No entanto, analisando de forma geral as melhores derivações encontradas para detecção da IC, utilizando o classificador híbrido proposto, foram a V5 e A-S e utilizando Redes Neurais Artificiais foi a A-S, tendo este resultado atingido os melhores níveis, mostrando que o uso destas ferramentas foi adequado e pode ser utilizada futuramente para auxiliar num diagnóstico mais rápido e preciso.

**REFERÊNCIAS**

- AMON, M.; JAGER F. **Electrocardiogram ST-Segment Morphology Delineation Method Using Orthogonal Transformations**. Califórnia: PlosOne, fev. 2016. DOI: 10.1371/journal.pone.0148814. ISSN 1932-6203.
- BEAR, M. F.; CONNORS, B. W.; PARADISO, M. A. **Neurociências: Desvendando o Sistema Nervoso**. Porto Alegre: Artmed, 4ª ed., 2017. ISBN: 978-85-8271-433-1.
- BRUNI, A.L. **Estatística Aplicada à Gestão Empresarial**. 2ª ed. São Paulo: Atlas S.A., 2011. 382p. ISBN: 978-85-224-6372-5.
- CARVALHO, A.C.C.; SOUSA, J.M.A. **Cardiopatia Isquêmica**. Revista Brasileira de Hipertensão, vol. 8, no. 3, pp. 297-305, ago. 2001. Disponível em: <<http://departamentos.cardiol.br/dha/revista/8-3/cardiopatia.pdf>>. Acesso em: 06 nov. 2017.
- COSTA, I.; CAMANHO, L. E. **Síndrome da Onda J: A Repolarização Precoce Pode não ser tão Benigna Quanto Parece**. Curitiba: Jornal Sobrac, Informativo da Sociedade Brasileira de Arritmias Cardíacas, XXIX Congresso Brasileiro de Arritmias Cardíacas, no. 26, pp. 18-19, 2012. Disponível em: <[http://departamentos.cardiol.br/sobrac/pdf/jornal\\_sobrac\\_26.pdf](http://departamentos.cardiol.br/sobrac/pdf/jornal_sobrac_26.pdf)>. Acesso em: 06 nov. 2017.
- DALL'OGGIO, P. **PHP: Programando com Orientação a Objetos**. 3ª ed. São Paulo: Novatec Editora LTDA, 2016. 552p. ISBN: 978-85-7522-465-6.
- DEGROT, M. H.; SCHERVISH, M. J. **Probability and Statistics**, Pearson Education, 4ª ed., 2012. 885p. ISBN 978-0-321-50046-5.
- DEMICHELI, F. **A hierarchical Naïve Bayes Model for handling sample heterogeneity in classification problems: an application to tissue microarrays**. BMC Bioinformatics, 7:514, 2006. <https://doi.org/10.1186/1471-2105-7-514>.
- DOWER G. E. et al. **Deriving the 12-lead electrocardiogram from four (EASI) electrodes**. Elsevier, Journal of Electrocardiol, vol. 21, pp. S182-S187, 1988. ISSN: 0022-0736. DOI: 10.1016/0022-0736(88)90090-8.
- DRANCA, L.; GOÑI, A.; LLLARRAMENDI, A. **Real-time detection of transient cardiac ischemic episodes from ECG signals**. Physiological Measurement, vol. 30, no. 9, ago. 2009. DOI: 10.1088/0967-3334/30/9/009. ISSN 1361-6579.
- DREW, B. J. et al. **Accuracy of the EASI 12-lead electrocardiogram compared to the standard 12-lead electrocardiogram for diagnosing multiple cardiac abnormalities**. Elsevier, Journal of Electrocardiology, vol. 32, pp. 38-47, 1999. ISSN: 0022-0736. DOI: 10.1016/S0022-0736(99)90033-X.
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**, New York: Wiley, 2ª ed., 2001. 680p. ISBN: 0-474-05669-3.
- FAGANELI, J.; JAGER, F. **Automatic classification of transient ischaemic and transient non-ischaemic heart-rate related ST segment deviation episodes in ambulatory ECG**

**records.** *Physiological Measurement*, vol. 31, no. 3, fev. 2010. DOI: 10.1088/0967-3334/31/3/004. ISSN: 1361-6579.

FOTIADIS, D. I. et al. **A knowledge-based technique for automated detection of ischaemic episodes in long duration electrocardiograms.** *Medical and Biological Engineering and Computing*, vol. 39, 1ª ed., pp. 105-112, 2001. ISSN. 1741-0444. DOI: 10.1007/BF02345273.

FRENKEL, D.; NADAL, J. **Comparação de métodos de representação do segmento ST na detecção automática de isquemias miocárdicas.** *Revista Brasileira de Engenharia Biomédica.* SBEB - Sociedade Brasileira de Engenharia Biomédica, v. 16, n. 3, p. 153-162, 2000. ISSN: 1517-3151.

GOLDBERGER, A.L. et al. **PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,** *Circulation*, vol. 101, no. 23, pp. 215-220, jun. 2000. Disponível em: <<http://circ.ahajournals.org/content/101/23/e215.full>>. Acesso em: 06 nov. 2017.

GOLDBERGER, A.L. et al. **PhysioBank ATM.** jun. 2000. Disponível em: <<https://physionet.org/cgi-bin/atm/ATM>>. Acesso em: 06 nov. 2017.

GUILHON, D.R.R. **Compressão de Sinais de Eletrocardiograma utilizando análise de componentes independentes.** 2006. 37p. Dissertação (Programa de Pós-Graduação em Engenharia de Eletricidade) - Universidade Federal do Maranhão, São Luís, 2006.

GUIMARÃES, J.I. et al. **Diretrizes de Interpretação de Eletrocardiograma de Repouso.** *Arquivos Brasileiros de Cardiologia, Realização da Sociedade Brasileira de Cardiologia*, vol. 80, (suplemento II), 2003. Disponível em: <http://www.scielo.br/pdf/abc/v80s2/a01v80s2.pdf>. Acesso em: 06 nov. 2017.

HAYKIN, S. **Redes Neurais: Princípios e prática.** Tradução: Paulo Martins Engel. 2ª ed. Porto Alegre: Bookman, 2001. 900p. ISBN: 978-85-7307-718-6.

JAGER F. et al. **Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loève transform.** *Proceedings of the IEEE Computers in Cardiology*, 1992. 691–4. ISSN: 0276-6547. ISBN: 0-8186-3552-5. DOI: 10.1109/CIC.1992.269339.

JAGER, F. et al. **Long-term ST database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia,** *Medical & Biological Engineering & Computing*, vol. 41, pp. 172-183, 2003. DOI: 10.1007/BF02344885. ISSN: 1741-0444.

JAGER, F. et al. **Long-term ST database.** 2016. DOI: 10.13026/C2G01T.

JERNBERG, T.; LINDAHL B.; WALLENTIN L. **ST-segment monitoring with continuous 12-lead ECG improves early risk stratification in patients with chest pain and ECG nondiagnostic of acute myocardial infarction.** *Journal of the American College of Cardiology*, 5 ed., vol. 34, 1999. ISSN: 1413–1419. DOI: 10.1016/S0735-1097(99)00370-8.

KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. São Paulo: Editora livraria da Física, 4ª ed., 2006. ISBN: 85-88325-14-4.

KUKAR, M. et al. **An application of machine learning in the diagnosis of ischaemic heart disease**. Slovenia: Tenth IEEE Symposium, Computer-Based Medical Systems, pp 70-75, 1997. ISSN: 1063-7125. ISBN: 0-8186-7928-X. DOI: 10.1109/CBMS.1997.596411.

LARSON, R.; FARBER, B. **Estatística Aplicada**. 4ª ed. São Paulo: Pearson Education do Brasil, Tradução de Luciane Paulete Viana, 2010. 656p. ISBN 978-85-7505-372-9.

LUGER, G. F. **Inteligência Artificial**. Editora Pearson, 6ª ed., University New Mexico, 2013. 632p. ISBN: 978-85-8143-550-3.

MOORE, D.S. **A Estatística Básica e Sua Prática**. 5ª ed. LTC, 2011. 584p. ISBN: 878-85-216-1790-7.

MORETTIN, L.G. **Estatística Básica: Probabilidade e Inferência**. São Paulo: Pearson Education do Brasil, vol. Único, 2009. 464p. ISBN: 978-85-7605-370-5.

NASSAR, S.M. **Tratamento de Incerteza: Sistemas Especialistas Probabilísticos**, 2012. 69p. Dissertação (Programa de Pós-Graduação em Ciência da Computação) - Universidade Federal de Santa Catarina, Santa Catarina, 2012. Disponível em: <[http://www.inf.ufsc.br/~silvia.nassar/disciplinas/sep/material\\_didatico/MaterialDidatico.pdf](http://www.inf.ufsc.br/~silvia.nassar/disciplinas/sep/material_didatico/MaterialDidatico.pdf)>. Acesso em: 06 nov. 2017.

OLEKSY, W.; TKACZ, E.; BUDZIANOWSKI, Z. **Improving EASI ECG Method Using Various Machine Learning and Regression Techniques to Obtain New EASI ECG Model**. International Journal of Computer and Communication Engineering, vol. 1, n. 3, pp. 287-289, 2012. ISSN: 2010-3743. DOI: 10.17706/IJCCE.

OLIVEIRA, L.S.C. de. **Redes Bayesianas Dinâmicas com definição no limiar aplicadas ao estudo de caso de detecção de extrassístole ventricular**. 2010. 121p. Dissertação (Pós-Graduação em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Vitória, 2010. Disponível em: <[http://portais4.ufes.br/posgrad/teses/tese\\_2821\\_TeseDoutoradoLorenaSophiaCamposdeOliveira.pdf](http://portais4.ufes.br/posgrad/teses/tese_2821_TeseDoutoradoLorenaSophiaCamposdeOliveira.pdf)>. Acesso em: 06 nov. 2017.

OLIVEIRA, L.S.C. de.; ANDREÃO, R.V.S, FILHO, M.S. **Bayesian Network with Decision Threshold for Heart Beat Classification**. IEEE Latin America Transactions, vol. 14, no. 3, pp. 1103-1108, mar. 2016. DOI: 10.1109/TLA.2016.7459585. ISSN: 1548-0992.

PAPALOUKAS, C. et al. **An ischemia detection method based on artificial neural networks**. Elsevier Science B. V., Artificial Intelligence in Medicine, vol. 24, 2ª ed., pp 167-178, 2002. ISSN: 0933-3657. DOI: 10.1016/S0933-3657(01)00100-2.

PAOLA, A.A.V.; BARBOSA, M.M.; GUIMARÃES, J.I. **Cardiologia: Livro-texto da Sociedade Brasileira de Cardiologia**, Editora Manole Interesse Geral, 1ª ed., 2011. 1700p. ISBN: 9788520431900.



PEARL, J. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference**, Revised Second Printing. San Mateo, CA: Morgan Kaufmann, 1993.

RIVERA-RUIZ M., CAJAVILCA C., VARON J. **Einthoven's String Galvanometer: The First Electrocardiograph**. Texas Heart Institute Journal, vol. 35, nº 2, pp. 174-178, 2008. PMID: PMC2435435. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2435435>. Acesso em: 06 nov. 2017.

SCHUM, D. A. **The Evidential Foundations of Probabilistic Reasoning**, Illinois: Northwestern University Press Evanston, 1994. ISBN: 0-8101-1821-1.

SILIPO, R.; MARCHESI, C. **Artificial neural networks for automatic ECG analysis**. **IEEE Transactions on Signal Processing**, vol. 46, 5ª ed., pp. 1417-1425, 1998. ISSN: 1053-587X.

SILIPO, R.; TADDEI, A.; MARCHESI, C. **Continuous monitoring and detection of ST-T changes in ischemic patients**. Proceedings of the IEEE Computers in Cardiology, 1994. ISSN: 0276-6547. ISBN: 0-8186-6570-X. DOI: 10.1109/CIC.1994.470209.

SMRDEL, A.; JAGER, F. **Automated detection of transient ST-segment episodes in 24 h electrocardiograms**. Medical & Biological Engineering & Computing, vol. 42, pp. 303-311, 2004. DOI: 10.1007/BF02344704. ISSN: 1741-0444.

SMYRNAKIS, M.G.; EVANS, D.J. **Classifying Ischemic Events Using a Bayesian Inference Multilayer Perceptron and Input Variable Evaluation Using Automatic Relevance Determination**. IEEE Conference Publications, Computers in Cardiology, Aston University, Birmingham United Kingdom, pp. 305-308, out. 2007. DOI: 10.1109/CIC.2007.4745482. ISSN: 2325-8853.

SONG, J. et al. **Application of fuzzy reasoning to myocardial ischemia judgment based on electrocardiogram ST-T complex**. Journal of Biomedical Engineering, China, 1ª ed., pp. 147-151, 156, fev. 2012. ISSN 1001-5515. Disponível em: [http://caod.oriprobe.com/articles/28723212/Application\\_of\\_Fuzzy\\_Reasoning\\_to\\_Myocardial\\_Ischemia\\_Judgment\\_Based\\_o.htm](http://caod.oriprobe.com/articles/28723212/Application_of_Fuzzy_Reasoning_to_Myocardial_Ischemia_Judgment_Based_o.htm). Acesso em: 06 nov. 2017.

SOUZA, F. A. A. de. **Análise de desempenho da rede neural artificial do tipo multilayer perceptron na era multicore**. 2012. 62p. Dissertação (Mestrado em Automação e Sistemas; Engenharia de Computação; Telecomunicações) - Universidade Federal do Rio Grande do Norte, Natal, 2012. Disponível em: <https://repositorio.ufrn.br/jspui/handle/123456789/15447>. Acesso em: 06 nov. 2017.

SPRINGHOUSE, **ECG Strip Ease: An Arrhythmia Interpretation Workbook**, Lippincott Williams & Wilkins, Wolters Kluwer Health, 2010. 384p. ISBN: 9781451105599.

STAMKOPOULOS, T. **One-lead ischemia detection using a new backpropagation algorithm and the European ST-T database**. Proceedings of the IEEE Computers in Cardiology, pp.663-666, 1992. ISSN: 0276-6541. ISBN: 0-8186-3552-5. DOI: 10.1109/CIC.1992.269346.

TADDEI, A. **A system for the detection of ischemic episodes in ambulatory ECG.** Proceedings of the IEEE Computers in Cardiology, pp 705-708, 1995; ISSN: 6547. ISBN: 0-7803-3053-6. DOI: 10.1109/CIC.1995.482762.

THALER, M. **ECG Essencial:** Eletrocardiograma na prática diária. Porto Alegre: Artmed, 7ª ed., 2013. ISBN: 9781451119053.

VILA, J. et al. **SUTIL:** Intelligent ischemia monitoring system. Elsevier, International Journal of Medical Informatics, vol. 47, 3ª ed.m pp. 193-214, 1997. ISSN: 0933-3657. DOI: 10.1016/S1386-5056(97)00095-6.

WEHR, G. et al. **A vector-based, 5-electrode, 12-lead monitoring ECG (EASI) is equivalent to conventional 12-lead ECG for diagnosis of acute coronary syndromes.** Elsevier, Journal of Electrocardiology, 1 ed. vol. 39, pp. 22–28, 2006. ISSN: 0022-0736. DOI: 10.1016/j.jelectrocard.2005.08.006.

WELINDER, A. et al. **Comparison of signal quality between EASI and standard Mason-Likar 12-lead electrocardiograms during physical activity.** American Journal of Critical Care, vol. 13, no. 3, pp. 228-234, 2004. ISSN: 1937-710X. Disponível em: <http://ajcc.aacnjournals.org/content/13/3/228.full.pdf+html?sid=07e5be8f-c9f6-4012-9bf1-b7a5b4f9f932>. Acesso em: 06 nov. 2017.

WORLD HEALTH ORGANIZATION (WHO), **The top 10 causes of death, Geneva, Switzerland.** Jan. 2017. Disponível em: <http://www.who.int/mediacentre/factsheets/fs310/en>>. Acesso em: 06 nov. 2017.

PHILIPS. **12-Lead ECG Monitoring with EASI™ Lead System:** IntelliVue Patient Monitor and Information Center, Application Note. Holanda: Koninklijke Philips Electronics N.V., 2007. Disponível em: [http://www.theonlinelearningcenter.com/Assets/PMDCBT/PIIC\\_Fundamentals\\_1.0/shell/viewer/swfs/assets/downloads/easi.pdf](http://www.theonlinelearningcenter.com/Assets/PMDCBT/PIIC_Fundamentals_1.0/shell/viewer/swfs/assets/downloads/easi.pdf). Acesso em: 06 nov. 2017.

ZHANG, Z.; WEI, D. **A New ECG Identification Method Using Bayes' Theorem.** IEEE Conference Publications, TENCON - IEEE Region 10 Conference, pp. 1-4, nov. 2006. DOI 10.1109/TENCON.2006.344146. ISSN: 2159-3442.

**GLOSSÁRIO**

SEMIA	Ferramenta de edição e visualização gráfica manual e semi-automática que foi usada no LTST-DB
aVR, aVL e aVF	Derivações unipolares dos membros
D1, D2, D3	Derivações clássicas ou padrão
V1, V2, V3, V4, V5 e V6	Derivações unipolares precordiais
A-I, A-S, E-S	Derivações do Sistema EASI
ML2, ML3 e MV2	Derivações Modificadas