



# CLASSIFICAÇÃO DE ELETROCARDIOGRAMA UTILIZANDO *EXTREME LEARNING MACHINE*

Trabalho de Conclusão de Curso  
Engenharia da Computação

**Carlos Eduardo Castor de Melo**

**Orientador:** Prof. Dr. Mêuser Jorge Silva Valença



**Carlos Eduardo Castor de Melo**

***CLASSIFICAÇÃO DE ELETROCARDIOGRAMA  
UTILIZANDO EXTREME LEARNING MACHINE***

Monografia apresentada como requisito parcial  
para obtenção do diploma de Bacharel em En-  
genharia de Computação pela Escola Politéc-  
nica de Pernambuco - Universidade de Pernam-  
buco

Orientador:

Prof. Dr. Mêuser Jorge Silva Valença

UNIVERSIDADE DE PERNAMBUCO  
ESCOLA POLITÉCNICA DE PERNAMBUCO  
GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Recife - PE, Brasil

28 de novembro de 2012

## MONOGRAFIA DE FINAL DE CURSO

### Avaliação Final (para o presidente da banca)\*

No dia 13 de 12 de 2012, às 10:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente CARLOS EDUARDO CASTOR DE MELO, orientado pelo professor Mêuser Jorge Silva Valença, sob título Classificação de eletrocardiograma utilizando extreme learning machine, a banca composta pelos professores:

**Sérgio Galdino**

**Mêuser Jorge Silva Valença**

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada       Aprovada com Restrições\*       Reprovada

e foi-lhe atribuída nota: 9,5 (nove e meio)

\*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 7 dias para entrega da versão final da monografia a contar da data deste documento.

*Sérgio Galdino*

SÉRGIO GALDINO

*Mêuser Jorge Silva Valença*

MÊUSER JORGE SILVA VALENÇA

Dedico esse trabalho à minha mãe, Maria Neuma Dias Castor.

# *Agradecimentos*

Primeiramente agradeço à minha mãe, Maria Neuma Dias Castor, pela dedicação e exemplo de força e coragem que sempre me inspiraram a superar todos os desafios impostos pela vida. Agradeço às minhas irmãs, Neumara Castor e Lucimara Castor, por todos os momentos de alegria e dificuldades compartilhados. Agradeço aos meus avós, tios, e primos por todo o suporte dado para que eu pudesse alcançar essa conquista.

Agradeço à minha namorada, Laurimeire Fernandes Ribeiro, por toda a paciência, compreensão e ajuda durante o desenvolvimento deste trabalho e todo o amor, carinho, companheirismo e atenção que me ajudaram muito desde o ingresso na faculdade até a sua conclusão.

Agradeço aos meus amigos Bruna Galle, Lara Dantas, Breno Menezes, Marcos Vinícius Souza, Adson Matoso e Afif Fikani, por todos os momentos compartilhados nesses últimos 5 anos.

Agradeço aos meus colegas, Antônio Nascimento e Everton Lacerda, por me auxiliarem neste trabalho.

Agradeço ao Professor Mêuser Jorge Silva Valença, pela orientação, apoio e paciência durante o desenvolvimento desse trabalho.

E finalmente agradeço a todos os professores do curso de Engenharia da Computação da Escola Politécnica de Pernambuco, pela boa formação profissional .

# *Resumo*

Doenças cardiovasculares são as principais causas de morte pelo mundo. Isso faz com que seja imperativa a busca por formas de diagnósticos precoces para a prevenção e tratamento dessas doenças. O eletrocardiograma é um exame muito utilizado nessa área, contudo sua análise ainda é principalmente realizada por uma avaliação visual de profissionais de saúde. Esse trabalho propõe o uso de uma rede neural artificial *Multi Layer Perceptron* treinada com a técnica *Extreme Learning Machine* para auxiliar o diagnóstico de arritmias cardíacas utilizando registros provindo do Eletrocardiograma. Foi utilizado um algoritmos para a seleção da melhor quantidade de neurônios da camada escondida, aquela que retornasse o menor erro de classificação. Os experimentos utilizando esse algoritmo foram feitos com diferentes limites para a geração de pesos entre a camada de entrada e a camada escondida e diferentes funções de ativação. O resultado encontrado foi uma rede com desempenho semelhante ao de técnicas clássicas de aprendizado de máquina, como a MLP treinada com o *Backpropagation* e o *Naive Bayes*.

**Palavras-chave:** Redes Neurais Artificiais, Extreme Learning Machine, Eletrocardiograma, Arritmia Cardíaca.

# *Abstract*

Cardiovascular diseases are the leading causes of death worldwide. This makes it imperative to search for ways to early diagnoses for the prevention and treatment of these diseases. The electrocardiogram is a test widely used in this area, but his analysis is still mainly performed by a visual assessment of health professionals. This work proposes the use of a Multi Layer Perceptron artificial neural network trained with Extreme Learning Machine technique to aid diagnosis of cardiac arrhythmias using records coming from the electrocardiogram. We used an algorithm for selecting the best number of hidden layer neurons, one that returned the lowest classification error. The experiments were done using this algorithm with different thresholds for generating weights between the input layer and hidden layer and different activation functions. The result found was a network with similar performance to the classical techniques of machine learning, such as the MLP trained with the Backpropagation and Naive Bayes.

**Keywords:** Artificial Neural Networks, Extreme Learning Machine, Electrocardiogram, Cardiac Arrhythmia.

# Conteúdo

<b>Lista de Figuras</b>	p. ix
<b>Lista de Tabelas</b>	p. x
<b>Lista de Abreviaturas e Siglas</b>	p. xi
<b>1 Introdução</b>	p. 12
1.1 Motivação . . . . .	p. 12
1.2 Objetivos . . . . .	p. 13
1.2.1 Objetivos Gerais . . . . .	p. 13
1.2.2 Objetivos Específicos . . . . .	p. 13
1.3 Estrutura da Monografia . . . . .	p. 14
<b>2 Referencial Teórico</b>	p. 15
2.1 O Sistema Elétrico do Coração e o Eletrocardiograma . . . . .	p. 15
2.1.1 O coração . . . . .	p. 15
2.1.2 Sistema Elétrico Cardíaco . . . . .	p. 15
2.1.3 Eletrocardiograma . . . . .	p. 17
2.1.4 Derivações do Eletrocardiograma . . . . .	p. 18
2.2 Redes Neurais Artificiais . . . . .	p. 20
2.2.1 <i>Multi Layer Perceptron</i> . . . . .	p. 20
2.3 <i>Extreme Learning Machine</i> . . . . .	p. 21
2.3.1 Modelagem de Redes SLFN . . . . .	p. 21



2.3.2	Treinamento do ELM . . . . .	p. 22
2.3.3	Pseudocódigo do ELM . . . . .	p. 23
2.3.4	Observações Conclusivas . . . . .	p. 23
<b>3</b>	<b>Metodologia</b>	p. 24
3.1	Base de Dados . . . . .	p. 24
3.2	Pré-processamento . . . . .	p. 25
3.3	Modelagem dos dados para classificação . . . . .	p. 26
3.4	Seleção de Parâmetro . . . . .	p. 26
3.5	Avaliação de Desempenho . . . . .	p. 28
3.5.1	Validação Cruzada . . . . .	p. 28
3.5.2	Métricas . . . . .	p. 28
3.5.3	Testes de Hipóteses . . . . .	p. 29
<b>4</b>	<b>Resultados</b>	p. 32
4.1	Análise dos Resultados . . . . .	p. 32
4.2	Comparação entre os experimentos . . . . .	p. 36
4.3	Comparação com outras técnicas . . . . .	p. 37
<b>5</b>	<b>Considerações Finais</b>	p. 39
5.1	Conclusões . . . . .	p. 39
5.2	Trabalhos Futuros . . . . .	p. 39
	<b>Referências</b>	p. 40

# *Lista de Figuras*

2.1	Anatomia do Coração . . . . .	p. 16
2.2	Sistema Elétrico Cardíaco . . . . .	p. 17
2.3	Eletrocardiograma Normal . . . . .	p. 18
2.4	Eletrodos posicionados para captação de DI,DII e DIII . . . . .	p. 19
2.5	Eletrodos posicionados para captação de V1,V2,V3,V4,V5 E V6 . . . . .	p. 19
2.6	Rede MLP com uma camada escondida . . . . .	p. 21
3.1	Tangente Hiperbólica. . . . .	p. 27
3.2	Sigmóide Logística. . . . .	p. 27
3.3	Validação Cruzada 5-fold. . . . .	p. 29
4.1	Busca ELM1 . . . . .	p. 33
4.2	Busca ELM2 . . . . .	p. 33
4.3	Busca ELM3 . . . . .	p. 34
4.4	Busca ELM4 . . . . .	p. 35
4.5	Busca ELM5 . . . . .	p. 35
4.6	Busca ELM6 . . . . .	p. 36

## *Lista de Tabelas*

3.1	Distribuição das classes . . . . .	p. 24
3.2	Atributos Comuns às Derivações . . . . .	p. 25
3.3	Atributos específicos por derivação . . . . .	p. 25
3.4	Modelagem das Classes . . . . .	p. 26
4.1	Experimentos Realizados . . . . .	p. 32
4.2	Resultado das buscas . . . . .	p. 34
4.3	Shapiro-Wilk para os experimentos com ELM . . . . .	p. 36
4.4	Teste F entre os experimentos com ELM . . . . .	p. 37
4.5	t-Student entre os experimentos com ELM . . . . .	p. 37
4.6	t-Student unicaudal entre os experimentos com ELM . . . . .	p. 37
4.7	Média de erros de MLP-BP e <i>Naive Bayes</i> . . . . .	p. 38
4.8	Teste Shapiro-Wilk para MLP-BP e <i>Naive Bayes</i> . . . . .	p. 38
4.9	Teste F entre ELM e MLP-BP e <i>Naive Bayes</i> . . . . .	p. 38
4.10	Soma dos Postos de Wilcoxon entre ELM e MLP-BP e <i>Naive Bayes</i> . . . . .	p. 38

## *Lista de Abreviaturas e Siglas*

RNA	<i>Redes Neurais Artificiais</i>
MLP	<i>Multi Layer Perceptron</i>
BP	<i>Backpropagation</i>
ELM	<i>Extreme Learning Machine</i>
SLFN	<i>Single Layer Feedforward Network</i>
DCV	<i>Doenças Cardiovasculares</i>
OMS	<i>Organização Mundial de Saúde</i>
ECG	<i>Eletrocardiograma</i>

# ***1 Introdução***

Esse capítulo se inicia com a descrição da motivação para o desenvolvimento deste trabalho. Em seguida são expostos os objetivos gerais e objetivos específicos a serem atingidos. Finalmente, seu encerramento enumera o conteúdo abordado nos capítulos seguintes.

## **1.1 Motivação**

O sangue é transportado pelo organismo através do sistema circulatório. Esse sistema é composto por vasos, artérias e veias que distribuem o sangue por todo o corpo. O coração é o órgão responsável pelo controle do fluxo por esses canais, por meio de contrações (sístole) e relaxamentos (diástole). Essa importância do coração gera uma grande necessidade em prevenir e tratar doenças que possam afetar o seu funcionamento.

Doenças Cardiovasculares (DCV) são a principal causa de morte no mundo. A Organização Mundial da Saúde (OMS) estima que em 2008 17,3 milhões de pessoas foram a óbito devido a DVC's e é esperado que em 2030 esse número chegue a 23,6 milhões (World Health Organization, 2012).

Esses números tornam imperativa a busca por métodos de diagnóstico precoce, para que seja possível fazer um tratamento mais efetivo. Entre os métodos existentes, um que ganha destaque é a análise do Eletrocardiograma (ECG). O ECG é um registro de ondas que refletem a atividade elétrica do coração, detectado via eletrodos posicionados pelo corpo do paciente. Ele é composto por ondas que representam etapas do processo de polarização e despolarização do músculo cardíaco. A análise dessas ondas possibilita ao profissional de saúde o diagnóstico de DVC's como arritmias, bloqueios e infarto do miocárdio.

A maior parte dos diagnósticos ainda é realizada a partir de uma avaliação visual feita por profissionais de saúde, o que pode ter seu desempenho prejudicado devido a inexperiência do profissional, cansaço ou até mesmo por descaso.

Alguns trabalhos já foram realizados aplicando-se algoritmos de aprendizado de máquina

para suporte a diagnóstico de ECG, como Redes Neurais *Multi Layer Perceptron* (BARROSO; ESPACIAIS; CALDAS, 2008), redes neurais modulares (MNN) (JADHAV; NALBALWAR; GHATOL, 2011) e KNN (ZUO; LU, 2008).

Esse trabalho propõe o uso de uma Rede Neural Artificial (RNA) Multilayer Perceptron (MLP) (VALENÇA, 2010) treinada com o algoritmo *Extreme Learning Machine* (ELM) (HUANG; ZHU; SIEW, 2006) para a classificação de arritmias cardíacas baseada em dados de ECG.

O ELM apresenta um tempo de treinamento muito menor do que as técnicas tradicionais de treinamento de MLP baseadas no gradiente-descendente, como o *Backpropagation* e possui uma melhor performance de generalização comparada a essas mesmas técnicas (HUANG; ZHU; SIEW, 2006). Essas características tornam o uso dessa técnica interessante para o problema em questão, visto que não se pode perder muito tempo para se fazer um diagnóstico de um paciente que possa estar apresentando alguma DVC.

## 1.2 Objetivos

Essa seção apresenta, primeiramente, o objetivo geral e, em seguida, os objetivos específicos desse trabalho.

### 1.2.1 Objetivos Gerais

O objetivo principal desse trabalho é desenvolver uma ferramenta que utiliza uma *Multi Layer Perceptron* treinada com o algoritmo *Extreme Learning Machine* para a classificação de arritmias cardíacas baseado em sinais de ECG.

### 1.2.2 Objetivos Específicos

- Usar uma Rede Neural Artificial *Multi Layer Perceptron* treinada com o algoritmo *Extreme Learning Machine*, que receba dados relacionados ao ECG de uma paciente e classificá-los com relação aos tipo de arritmia cardíacas.
- Usar um algoritmo de busca para a seleção da quantidade de neurônios da camada escondida da RNA.
- Executar uma comparação de desempenho com outras técnicas de aprendizado de máquina.

## 1.3 Estrutura da Monografia

Os Capítulos 2 abordará os conteúdos teóricos necessários para o entendimento do trabalho proposto, o Eletrocardiograma e o *Extreme Learning Machine* . Em seguida o Capítulo 3 mostra a metodologia adotada para se alcançar o objetivo geral desse projeto. O Capítulo 4 traz os resultados obtidos das simulações realizadas com a técnica proposta. Por fim, o Capítulo 5 resume o trabalho, elencando conclusões obtidas e possíveis melhoras e trabalhos futuros.

## **2    *Referencial Teórico***

### **2.1    O Sistema Elétrico do Coração e o Eletrocardiograma**

Essa seção objetiva elucidar a importância e as formas de se obter os registros do ECG. Ele inicia falando sobre o coração, sua importância e anatomia. Em seguida é abordado o sistema elétrico do coração, responsável pelo ritmo cardíaco. E a última seção aborda o eletrocardiograma.

#### **2.1.1    O coração**

O coração é o órgão responsável por bombear o sangue por todo o organismo através do sistema circulatório. Ele é formado por duas bombas distintas, o "coração direito" e o "coração esquerdo" (GUYTON; HALL, 2006). Por sua vez, cada uma dessas bombas é formada por um átrio e um ventrículo, como mostrado na Figura 2.1.

Os átrios são bombas fracas que enviam o sangue para os ventrículos, que por sua vez são os responsáveis em propeler o sangue para fora do coração. O ventrículo direito envia o sangue para a circulação pulmonar e o ventrículo esquerdo envia o sangue para a circulação periférica.

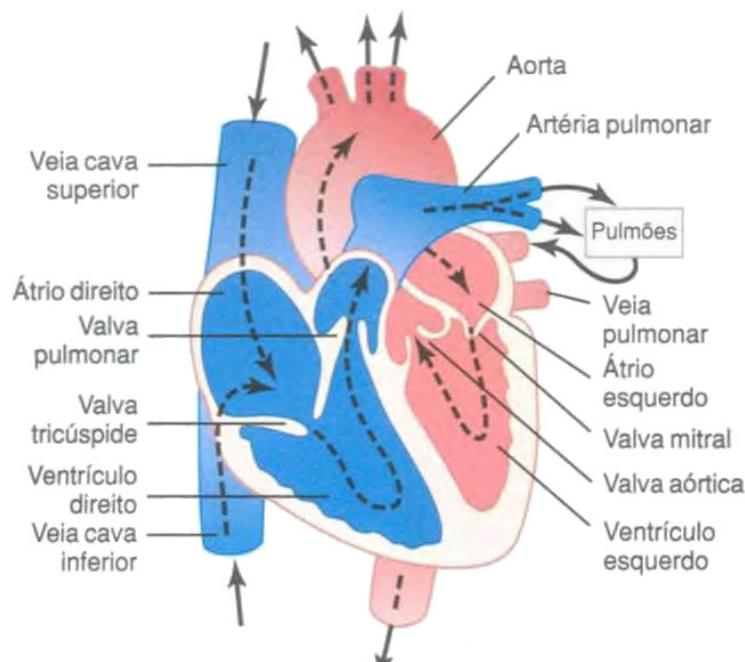
O coração conta com um sistema elétrico próprio que promove uma sucessão contínua de contrações, chamadas de ciclo cardíaco, ao transmitir potenciais de ação pelo miocárdio, causando os batimentos rítmicos do coração.

#### **2.1.2    Sistema Elétrico Cardíaco**

O ciclo cardíaco consiste no período de relaxamento, denominado diástole, durante o qual o coração se enche de sangue, seguido pelo período de contração, chamado sístole.

Para que esse ciclo funcione devidamente, o coração conta com um sistema que gera impulsos elétricos que causam contrações rítmicas e conduz esses impulsos rapidamente por todo o





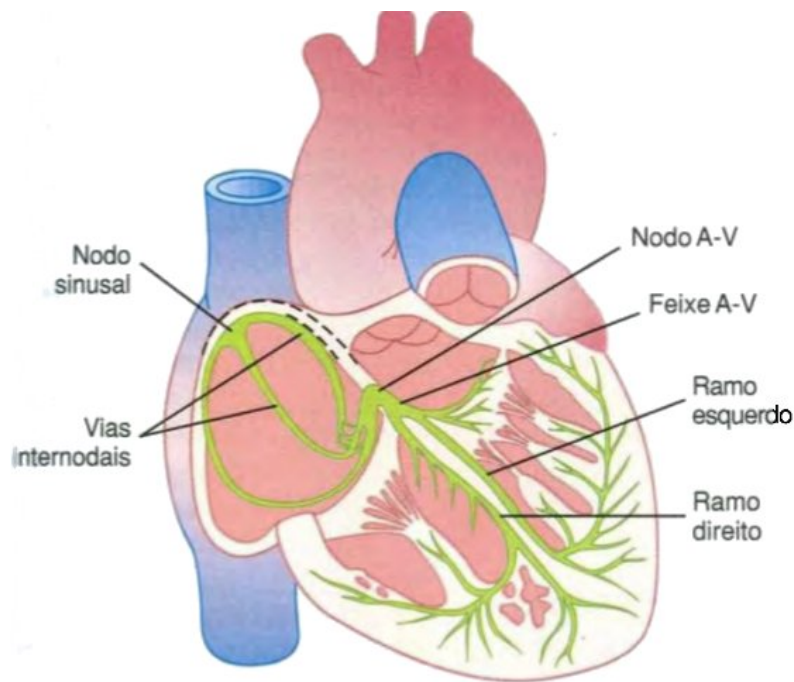
**Figura 2.1: Anatomia do Coração**

coração. Quando esse sistema funciona normalmente, os átrios se contraem aproximadamente um sexto de segundo antes da contração ventricular, o que permite o enchimento dos ventrículos antes que eles bombeiem o sangue para os pulmões e para a circulação periférica (GUYTON; HALL, 2006).

Outra característica desse sistema é que ele faz com que as diferentes porções do ventrículo se contraíam quase simultaneamente, o que é essencial para gerar pressão com o máximo de eficiência nas câmaras ventriculares.

A Figura 2.2 mostra o sistema elétrico do coração. Esse sistema é formado pelo: 1) nó sinoatrial (SA) (ou marcapasso), onde é gerado o impulso elétrico normal; 2) as fibras de condução atrial, que conduzem o impulso do nó SA para o nó atrioventricular (AV); 3) pelo nó AV, responsável por gerar um retardo no impulso antes de conduzi-lo para os ventrículos; 4) o feixe AV, que conduz o impulso dos átrios para os ventrículos; 5) pelos ramos direito e esquerdo das fibras de Purkinje, que conduzem o impulso cardíaco para todas as partes do ventrículo.

Esse sistema é suscetível a danos por doenças cardíacas, especialmente à isquemia dos tecidos cardíacos, por causa da circulação coronária deficiente. O resultado frequentemente é uma arritmia cardíaca, o que pode muitas vezes afetar gravemente a eficiência do bombeamento cardíaco, chegando até a causar morte (GUYTON; HALL, 2006).



**Figura 2.2: Sistema Elétrico Cardíaco**

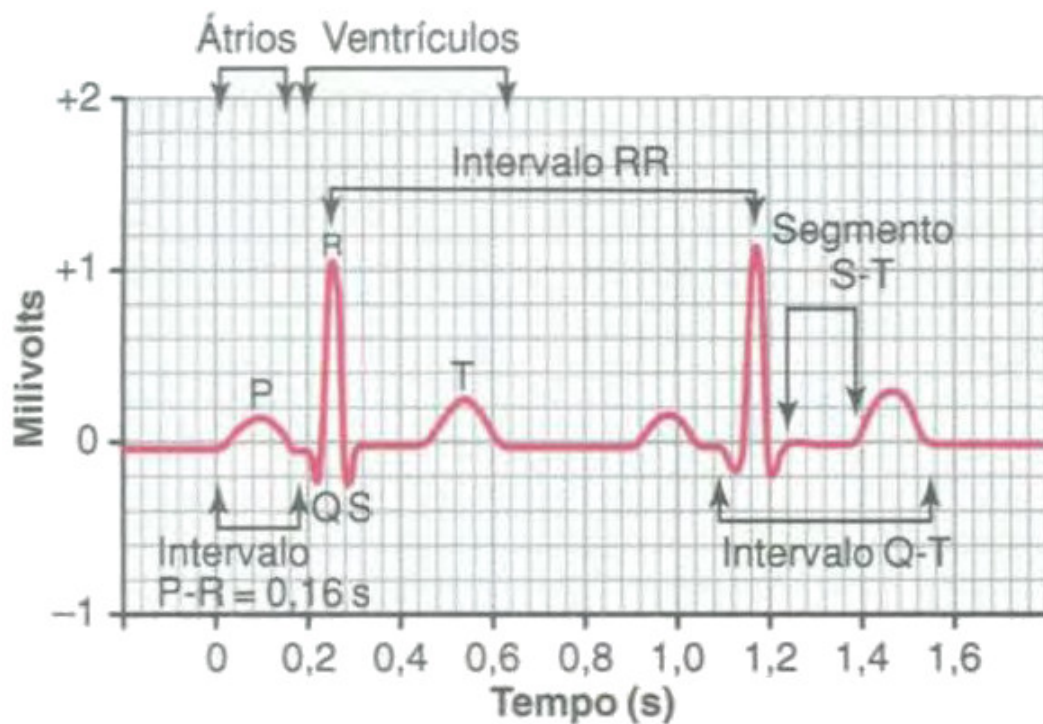
### 2.1.3 Eletrocardiograma

Quando o impulso cardíaco passa através do coração, uma corrente elétrica também se propaga do coração para os tecidos adjacentes e pequena parte dessa corrente se propaga até a superfície do corpo. Quando eletrodos são posicionados sobre a pele, em lados opostos do coração, será possível registrar os potenciais elétricos gerados por essa corrente. Esse registro é conhecido como Eletrocardiograma (ECG) (GUYTON; HALL, 2006).

O eletrocardiograma normal (Figura (2.3)) é composto pela onda P, pelo complexo QRS e pela onda T. O complexo QRS é composto por três ondas distintas: a onda Q, a onda R e a onda S.

A onda P é produzida pelos potenciais elétricos gerados quando os átrios se despolarizam, logo antes da contração atrial começar. O complexo QRS é produzido pelos potenciais gerados quando os ventrículos se despolarizam antes de sua contração, i.e, enquanto a onda de despolarização se propaga pelos ventrículos. Portanto tanto a onda P como os componentes do complexo QRS são ondas de despolarização.

A onda T é produzida pelos potenciais gerados enquanto os ventrículos se reestabelecem do estado de despolarização. Esse processo no músculo ventricular normalmente ocorre 0,25 a 0,35 segundo após a sua despolarização e a onda T é conhecida como onda de repolarização



**Figura 2.3: Eletrocardiograma Normal**

(GUYTON; HALL, 2006).

A frequência dos batimentos cardíaco também pode ser determinada no ECG, visto que a frequência cardíaca corresponde ao inverso do intervalo entre dois batimentos cardíacos sucessivos.

Para se obter o registro do ECG, se faz uso de um aparelho denominado Eletrocardiógrafo. Esse aparelho capta os sinais do eletrocardiograma através de eletrodos posicionados no corpo do paciente.

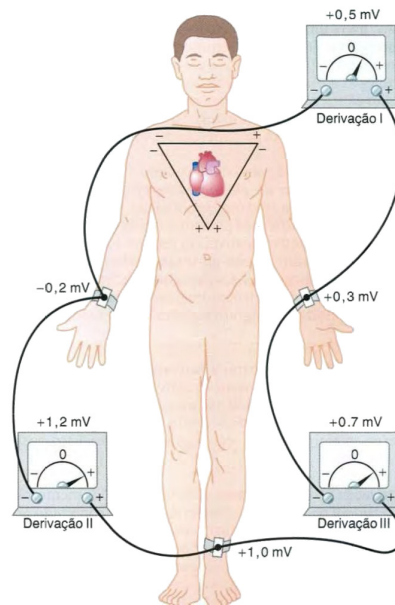
O Eletrocardiograma é um exame muito importante para detecção de comportamentos anômalos do sistema elétrico do coração. Ele pode ser utilizado para o diagnóstico de arritmias cardíacas, o que permite ao paciente se submeter a um tratamento em tempo hábil.

### 2.1.4 Derivações do Eletrocardiograma

O ECG pode ser captado por 12 derivações, sendo elas:

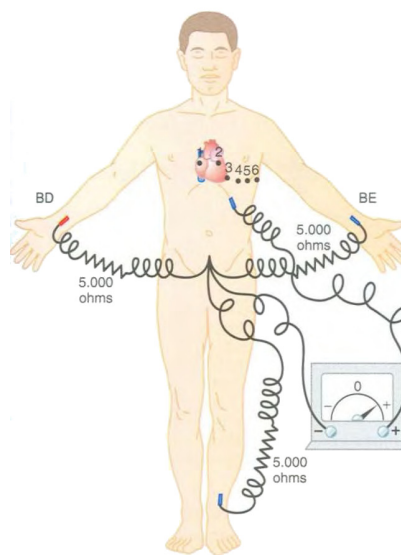
- Derivações Bipolares dos membros (DI, DII e DIII): o registro do eletrocardiograma é captado por dois eletrodos conectando dois membros do paciente ao terminal positivo do

eletrocardiógrafo e um eletrodo conectando o terceiro membro do paciente ao terminal negativo do eletrocardiógrafo. Como visto na Figura 2.4.



**Figura 2.4: Eletrodos posicionados para captação de DI,DII e DIII**

- Derivações torácicas (V1,V2,V3,V4,V5 e V6): para captar esses sinais, um eletrodo, o eletrodo torácico, é conectado ao terminal positivo do eletrocardiógrafo e um eletrodo negativo é conectado, simultaneamente, ao braço direito, ao braço esquerdo e à perna esquerda. As seis derivações são captadas de forma sequencial nos seis pontos mostrados na Figura 2.5.



**Figura 2.5: Eletrodos posicionados para captação de V1,V2,V3,V4,V5 E V6**

- Derivações Unipolares Aumentadas dos Membros (aVR,aVL,aVF): nesse tipo de registro, dois dos membros são conectados ao terminal negativo do eletrocardiógrafo e o terceiro membro é conectado ao terminal positivo.

## 2.2 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são um grupo de técnicas de computação inteligente que possuem arquitetura e funcionamento baseados nas teorias sobre o funcionamento do cérebro humano (VALENÇA, 2010). Elas são formadas por unidades de processamento interconectadas, denominadas neurônios artificiais, que se baseiam no modelo matemático do neurônio biológico proposto por McCulloch e Pitts em (MCCULLOCH; PITTS, 1943).

Esse modelo faz uso de uma regra de propagação e uma função de ativação para representar de maneira bastante simples o neurônio biológico. A emissão do sinal por esse modelo é feita baseando-se em um limiar  $\theta$ , como mostrado na Equação 2.1.

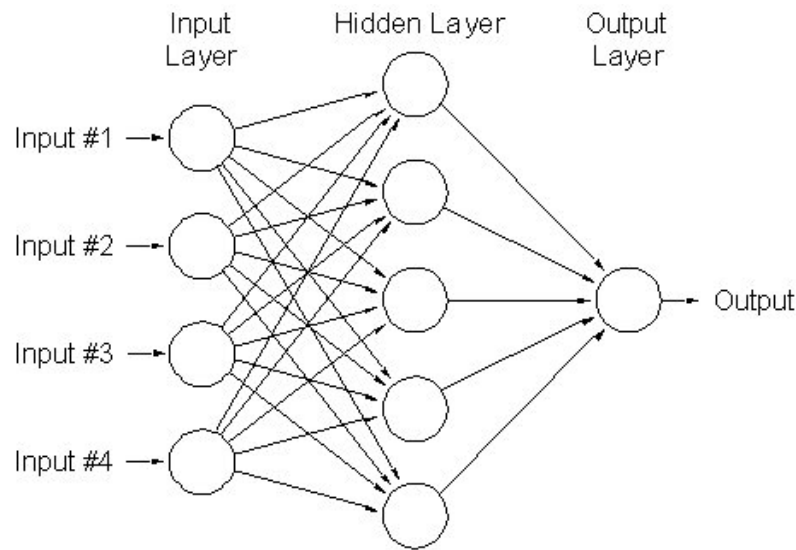
$$net_i = \sum_{j=0}^n w_{ij}x_j - \theta \quad (2.1)$$

Onde  $net_i$  é o sinal emitido,  $x_i$  são as entradas do neurônio e  $w_{ij}$  são os pesos sinápticos associados a cada entrada (VALENÇA, 2010).

### 2.2.1 Multi Layer Perceptron

MLP é um modelo de RNA que possui seus neurônios distribuídos em pelo menos três camadas: uma de entrada, uma ou mais camadas intermediárias e uma camada de saída. O surgimento dessa camada intermediária possibilitou às RNA trabalhar com problemas não-linearmente separáveis (VALENÇA, 2010), capacidade esta que os modelos predecessores (Perceptron e ADALINE) não têm.

Para ajustar os pesos entre os neurônios dessas camadas, se faz necessário um algoritmo de treinamento. Muitos algoritmos utilizados são baseados no gradiente descendente, como é o caso do Backpropagation (BP). Contudo, esses algoritmos tendem a ser bastante lentos e facilmente convergem para mínimos locais (BRAGA; CARVALHO; LUDERMIR, 2007) (HAYKIN, 2001).



**Figura 2.6: Rede MLP com uma camada escondida**

## 2.3 Extreme Learning Machine

O *Extreme Learning Machine* (ELM) é um algoritmo de treinamento de redes MLP proposto por Huang em (HUANG; ZHU; SIEW, 2006) que contorna os problemas encontrados por algoritmos baseados no gradiente descendente. Esse algoritmo foi desenvolvido para treinamento de MLPs com somente uma camada escondida, denominadas *Single Layer Feedforward Network* (SLFN).

No ELM, os pesos entre as camadas de entrada e escondida e os *bias* (limiares) da camada escondida são atribuídos aleatoriamente e não são ajustados no decorrer do treinamento. Já os pesos da camada escondida para a camada de saída são determinados analiticamente, através de operações de matrizes.

### 2.3.1 Modelagem de Redes SLFN

Para  $N$  amostras distintas  $(x_i, t_i)$ , onde  $x_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}]^T \in R^n$  e  $t_i = [t_{i1}, t_{i2}, t_{i3}, \dots, t_{im}]^T \in R^m$ , redes SLFN com  $\tilde{N}$  neurônios na camada intermediária e função de ativação  $g(x)$  nesses neurônios é matematicamente modelada como:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i x_i + b_i) = o_j, j = 1, 2, 3, \dots, N \quad (2.2)$$

Onde  $i$  é o índice do neurônio da camada escondida,  $w_i = [w_{i1}, w_{i2}, w_{i3}, \dots, w_{in}]^T$  é o vetor de

pesos conectando esse neurônio aos neurônios da camada de entrada,  $\beta_i = [\beta_{i1}, \beta_{i2}, \beta_{i3}, \dots, \beta_{in}]^T$  é o vetor de pesos conectando esse mesmo neurônio para a camada de saída,  $b_i$  é o limiar desse neurônio.

SLFN com  $\tilde{N}$  neurônios na cama escondida e com função de ativação  $g(x)$  podem aproximar esses  $N$  exemplos com erro zero,  $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0$ . Ou seja, existe  $\beta_i, w_i$  e  $b_i$  tal que:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i x_i + b_i) = t_j, j = 1, 2, 3, \dots, N \quad (2.3)$$

As  $N$  equações descritas acima podem ser escritas como:

$$H\beta_i = T \quad (2.4)$$

Onde:

$$H = \begin{bmatrix} g(w_1 x_1 + b_1) & \dots & g(w_{\tilde{N}} x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 x_N + b_1) & \dots & g(w_{\tilde{N}} x_N + b_{\tilde{N}}) \end{bmatrix} \quad (2.5)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_i^T \end{bmatrix}_{\tilde{N} \times m} \quad (2.6)$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \quad (2.7)$$

### 2.3.2 Treinamento do ELM

O objetivo de treinar uma SLFN é encontrar valores  $\hat{w}_i$ ,  $\hat{b}_i$  e  $\hat{\beta}$  tais que:

$$\|H(\hat{w}_1, \dots, \hat{w}_{\tilde{N}}, \hat{b}_1, \dots, \hat{b}_{\tilde{N}})\hat{\beta} - T\| = \min_{w_i, b_i, \beta_i} \|H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}})\beta - T\| \quad (2.8)$$

Como no ELM os pesos ligando a camada de entrada à camada de saída são gerados aleatoriamente e permanecem inalterados, o treinamento da SLFN se resume a encontrar um solução de mínimos quadrados  $\beta$  que satisfaça a Equação 2.8. Se o número de neurônios escondidos for igual ao número de exemplos utilizados no treinamento,  $\tilde{N} = N$ , a matriz  $H$  é quadrada e

pode ser invertida por métodos clássicos e a SLFN pode aproximar esses exemplos de treino com zero de erro.

Contudo, na maior parte dos casos, o valor de  $\tilde{N}$  é muito menor que  $N$ . Com isso a matriz  $H$  não é quadrada e não é possível se calcular a sua inversa. Nesses casos, a solução para a equação 2.8 é dada por

$$\hat{\beta} = H^\dagger T \quad (2.9)$$

Onde  $H^\dagger$  é a pseudo-inversa de Moore-Penrose da matrix  $H$ .

### 2.3.3 Pseudocódigo do ELM

---

#### Algoritmo 1: Pseudocódigo do ELM

---

- 1 Dado um conjunto de treinamento  $C = \{(x_i, t_i) \mid x_i \in R^n, t_i \in R^m, i = 1, 2, \dots, N\}$ , função de ativação  $g(x)$  infinitamente diferenciável e número de neurônios na camada escondida  $\tilde{N}$
  - 2 Atribua aleatoriamente valores para  $w_i$  e  $b_i$
  - 3 Calcule a matriz,  $H$ , de saída da camada escondida
  - 4 Calcule a matriz de pesos de saída  $\beta$ , onde  $\beta = H^\dagger T$
- 

### 2.3.4 Observações Conclusivas

Em (HUANG; ZHU; SIEW, 2006) foram feitas comparações de desempenho entre ELM e outros algoritmos de treinamento de SLFN, como BP e Support Vector Machine (SVM), utilizando diversos Benchmarks. Em geral, foi observado que o ELM apresentou uma velocidade de treino muito menor a que a dos outros algoritmos, chegando a ordem de centenas essa diferença em alguns casos, e obteve valores semelhantes em relação a métricas de avaliação de desempenho, como o Erro Médio Quadrático e seu Desvio-Padrão.



## 3 Metodologia

Esse capítulo descreve como se deu o desenvolvimento do projeto de acordo com as teorias apresentadas no capítulo 2

### 3.1 Base de Dados

Para os experimentos desse projeto foi utilizada a base de arritmias cardíacas do repositório UCI (FRANK; ASUNCION, 2010). Essa base é formada por 452 amostras, distribuídas entre 13 classes, cada uma representando um tipo de arritmia cardíaca, como mostrado na Tabela 3.1.

Tipo de Arritmia	Quantidade de Amostras
Ausência de arritmia	245
Alterações isquêmicas (Doença Arterial Coronariana)	44
Infarto do Miocárdio Anterior	15
Infarto do Miocárdio Inferior	15
Taquicardia Sinusal	13
Braquicardia Sinusal	25
Contração Ventricular Prematura	3
Contração Supraventricular Prematura	2
Bloqueio do Ramo Esquerdo	9
Bloqueio do Ramo Direito	50
Hipertrofia do Ventrículo Esquerdo	4
Fibrilação Arterial	5
Outros	22

**Tabela 3.1: Distribuição das classes**

Para cada amostra existem 279 atributos, onde os quatro primeiros são a idade, o sexo, a altura e o peso da amostra. Os demais atributos são informações relacionadas ao ECG, sendo os atributos 5 a 15 dados comuns e compostos às 12 derivações do ECG e o restante dados específicos de cada derivação, como mostrado na Tabela 3.2 e Tabela 3.3.

As ondas  $\hat{R}$  e  $\hat{S}$  representam pequenos picos logo a após as onda R e S, respectivamente.

Índice do atributo	Descrição
5	Duração média do complexo QRS em milisegundos
6	Duração média entre o início da onda P e da onda Q em milisegundos
7	Duração média entre o início da onda Q e da onda T em milisegundos
8	Duração média da onda T, em milisegundos
9	Duração média da onda P, em milisegundos
10	Vetor Resultante no Plano Frontal QRS
11	Vetor Resultante no Plano Frontal T
12	Vetor Resultante no Plano Frontal P
13	Vetor Resultante no Plano Frontal QRST
14	Vetor Resultante no Plano Frontal J
15	Frequência Cardíaca

Tabela 3.2: Atributos Comuns às Derivações

Descrição	Índices dos atributos
Largura Média Onda Q	(16,28,40,52,64,76,88,100,112,124,136,148)
Largura Média Onda R	(17,29,41,53,65,77,89,101,113,125,137)
Largura Média Onda S	(18,30,42,54,66,78,90,102,114,126,138,150)
Largura Média Onda $\hat{R}$	(19,31,43,55,67,79,91,103,115,127,139,151)
Largura Média Onda $\hat{S}$	(20,32,44,56,68,80,92,104,116,128,140,152)
Número de Defleções Intrínsecas	(21,33,45,57,69,81,93,105,117,129,141,153)
Irregularidade na onda R	(22,34,46,58,70,82,94,106,118,130,142,154)
Derivação difásica da onda R	(23,35,47,59,71,83,95,107,119,131,143,155)
Irregularidade na onda P	(24,36,48,60,72,84,96,108,120,132,144,156)
Derivação difásica da onda P	(25,37,49,61,73,85,97,109,121,133,145,157)
Irregularidade na onda T	(26,38,50,62,74,86,98,110,122,134,146,158)
Derivação difásica da onda T	(27,39,51,63,75,87,99,111,123,135,147,159)
Amplitude da onda J	(160,170,180,190,200,210,220,230,240,250,260,270)
Amplitude da onda Q	(161,171,181,191,201,211,221,231,241,251,261,271)
Amplitude da onda R	(162,172,182,192,202,212,222,232,242,252,262,272)
Amplitude da onda S	(163,173,183,193,203,213,223,233,243,253,263,273)
Amplitude da onda $\hat{R}$	(164,174,184,194,204,214,224,234,244,254,264,274)
Amplitude da onda $\hat{S}$	(165,175,185,195,205,215,225,235,245,255,265,275)
Amplitude da onda P	(166,176,186,196,206,216,226,236,246,256,266,276)
Amplitude da onda T	(167,177,187,197,207,217,227,237,247,257,267,277)
Soma das áreas dos segmentos QRS divididos por 10 ( <i>QRSA</i> )	(168,178,188,198,208,218,228,238,248,258,268,278)
$QRSA + (0,5 \cdot largura da onda T) \cdot (0,1 \cdot altura da onda T)$	(169,179,189,199,209,219,229,239,249,259,269,279)

Tabela 3.3: Atributos específicos por derivação

## 3.2 Pré-processamento

Antes de utilizar a base para o treinamento do modelo proposto, foram necessários alguns procedimentos. Primeiramente foi constatado que 17 atributos da base não variavam entre as

amostras. Como estes não influenciariam o resultado da classificação, eles foram retirados do conjunto de treinamento. Em seguida foi observado a ausência de valores de atributos para algumas amostras, a esses valores foi atribuída a média dos atributos restantes. Por último foi feita a normalização dos atributos, já que esses possuem intervalos de variações diferentes. A normalização utilizada foi a linear (3.1) com valores de  $a=1$  e  $b=0$  (VALENÇA, 2010).

$$y = \frac{((a - b)(x_i - x_{max}))}{(x_{max} - x_{min})} \quad (3.1)$$

Ao final desse processo, o conjunto de treino ficou com 262 atributos, com todos eles possuindo valores entre 0 e 1.

### 3.3 Modelagem dos dados para classificação

A classificação da rede implementada é feita como o vencedor leva tudo. Nessa tipo de classificação, o neurônio com maior valor da sua função de ativação assume o valor 1 e os demais recebem 0. Para atender a essa forma de classificação, as classes foram modeladas como mostrado na Tabela 3.4.

Classe	Representação												
Ausência de arritmia	1	0	0	0	0	0	0	0	0	0	0	0	0
Alterações isquêmicas	0	1	0	0	0	0	0	0	0	0	0	0	0
Infarto do Miocárdio Anterior	0	0	1	0	0	0	0	0	0	0	0	0	0
Infarto do Miocárdio Inferior	0	0	0	1	0	0	0	0	0	0	0	0	0
Taquicardia Sinusal	0	0	0	0	1	0	0	0	0	0	0	0	0
Braquicardia Sinusal	0	0	0	0	0	1	0	0	0	0	0	0	0
Contração Ventricular Prematura	0	0	0	0	0	0	1	0	0	0	0	0	0
Contração Supraventricular Prematura	0	0	0	0	0	0	0	1	0	0	0	0	0
Bloqueio do Ramo Esquerdo	0	0	0	0	0	0	0	0	1	0	0	0	0
Bloqueio do Ramo Direito	0	0	0	0	0	0	0	0	0	1	0	0	0
Hipertrofia do Ventrículo Esquerdo	0	0	0	0	0	0	0	0	0	0	1	0	0
Fibrilação Arterial	0	0	0	0	0	0	0	0	0	0	0	1	0
Outros	0	0	0	0	0	0	0	0	0	0	0	0	1

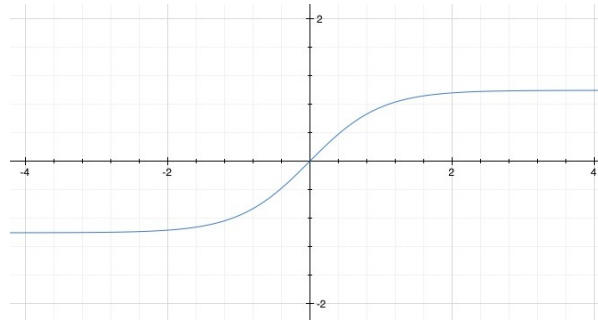
Tabela 3.4: Modelagem das Classes

### 3.4 Seleção de Parâmetro

A SLFN utilizada nesse trabalho tem 262 neurônios na camada de entrada e 13 na camada de saída. Os pesos ligando a camada de entrada a camada escondida são gerados aleatoriamente

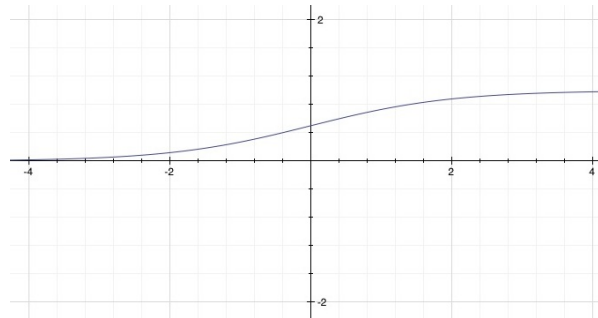
e são utilizadas duas funções de ativação para os neurônios da camada escondida: a Tangente Hiperbólica (3.2) e a Sigmoide Logística (3.3).

$$net(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.2)$$



**Figura 3.1: Tangente Hiperbólica.**

$$net(x) = \frac{1}{1 + e^{-x}} \quad (3.3)$$



**Figura 3.2: Sigmoide Logística.**

O algoritmo utilizado nesse trabalho é constituído por uma busca por uma quantidade  $n_{best}$  de neurônios na camada escondida que retorne um menor erro de classificação.

A busca é feita atribuindo-se uma quantidade inicial de neurônios da camada escondida  $n_0$ , calculando-se um erro de classificação para essa configuração  $e_{min}$  e incrementando esse tamanho a um passo  $p$  a cada ciclo da busca. Em cada ciclo é calculado o erro de classificação  $e_{atual}$  e caso esse erro seja menor que o valor de  $e_{min}$ , o valor de  $n_{best}$  assume o valor de  $n_{atual}$  e a busca persiste por mais  $m$  ciclos. Caso não seja encontrado nenhum valor melhor nos  $m$  ciclos seguintes, a busca se encerra.

**Algoritmo 2:** Algoritmo Utilizado

---

```

1 Selecionar tamanho inicial da camada escondida:  $n_0$ 
2 Selecionar incremento do tamanho da camada escondida:  $p$ 
3 Selecionar um número máximos de ciclos de busca após ser encontrado um valor de
   $e_{min}$ :  $m$ 
4 Criar um contador de ciclos:  $i$ 
5 enquanto  $i < m$  faça
6   | criar camada escondida com  $n$  neurônios
7   | gerar pesos aleatórios entre a camada de entrada e a camada escondida
8   | calcular pesos entre a camada escondida e a camada de saída
9   | calcular  $e_{atual}$  para a arquitetura
10  | se  $e_{atual} < e_{min}$  então
11  |   |  $e_{min} = e_{atual}$ 
12  |   |  $n_{otimo} = n$ 
13  |   |  $i = 0$ 
14  | senão
15  |   |  $i = i + 1$ 
16  | incrementar  $n$  para o próximo ciclo

```

---

## 3.5 Avaliação de Desempenho

### 3.5.1 Validação Cruzada

A validação cruzada é um método estatístico para avaliação e comparação de algoritmos de aprendizado de máquina que divide os dados em duas partes dois conjuntos distintos: um para o treinamento e outro para a validação do modelo (REFAEILZADEH; TANG; LIU, 2009).

O método mais comum para validação cruzada é o denominado *k-fold*, onde os dados são divididos em  $k$  segmentos (*folds*). Em seguida são feitas  $k$  iterações de treino, utilizando  $k - 1$  *folds* para o treinamento e o *fold* restante para o teste do modelo.

A Figura 3.3 mostra um exemplo com 5 *folds*.

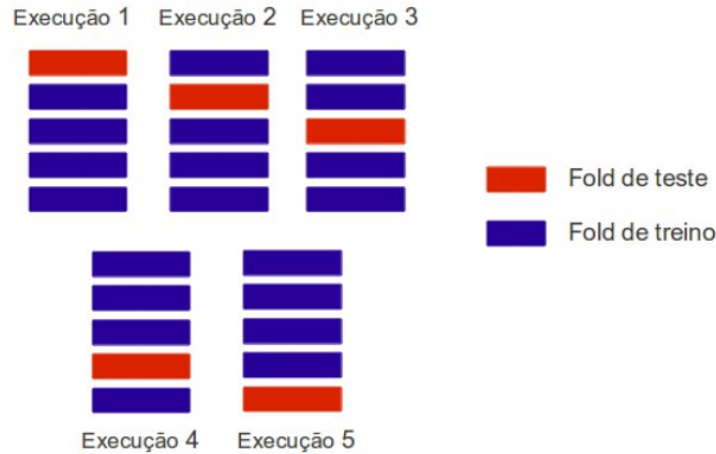
Para esse trabalho foi decidido utilizar 10 *folds*, por ser uma quantidade que fornece uma boa estimativa e é comumente utilizada para validação de algoritmos de aprendizado de máquina (REFAEILZADEH; TANG; LIU, 2009).

### 3.5.2 Métricas

A métrica escolhida para mensurar o desempenho da rede é o erro de classificação ( $e_c$ ). Essa métrica é calculada somando-se o total de amostras erroneamente classificadas  $n_e$  e dividindo

esse valor pelo total de amostras  $n_t$ .

$$e_c = \frac{n_e}{n_t} \quad (3.4)$$



**Figura 3.3: Validação Cruzada 5-fold.**

### 3.5.3 Testes de Hipóteses

Um vez realizados os experimentos com o modelo proposto, o próximo passo é fazer comparações entre esses experimentos, para saber qual o melhor, e em seguida comparar com o desempenho de outras técnicas. Para tal, serão repetidos experimentos para então coletar uma amostra das métricas para os modelos em questão e compará-las.

O teste de hipóteses é uma ferramenta estatística bastante utilizada na comunidade científica para se comparar o desempenho de algoritmos. Essa técnica é um método que fornece meios para determinar se uma dada hipótese é real ou acontece devido ao acaso, baseando-se nos dados utilizados no teste (SIEGEL, 2012). Para a comparação de desempenho de algoritmos, as hipóteses formuladas buscam investigar se a medida das métricas de um modelo é melhor do que outro.

Um teste de hipótese fundamenta-se em um par de hipóteses por vez, hipótese nula ( $H_0$ ) e hipótese alternativa ( $H_1$ ) e a partir dos dados uma delas será escolhida como válida e a outra será descartada.  $H_0$  representa a conclusão que será aceita apenas caso não se consiga provar o contrário e  $H_1$  é a que se deseja provar como certa.

O procedimento do teste de hipótese segue as seguintes etapas: construção das hipóteses;

escolha de um critério de decisão (nível de significância); cálculo de um determinado teste estatístico; comparação do valor obtido nesse teste com o nível de significância e decisão por uma das hipóteses (SIEGEL, 2012).

O teste estatístico consiste em determinar um valor, chamado *p-value*, que indica a probabilidade dos dados estarem de acordo com a possibilidade nula. Se o *p-value* for menor que o nível de significância, decide-se pela hipótese alternativa e rejeita-se a nula.

Nesse trabalho serão aplicados até quatro testes estatísticos, sendo eles Teste de Shapiro-Wilk, Teste F, Teste t-Student e o Teste de Soma de Postos de Wilcoxon.

### **Teste de Shapiro-Wilk**

O teste de Shapiro-Wilk é utilizado para testar se as amostras provêm de uma população normalmente distribuída.

As hipóteses formuladas para esse teste são:

- $H_0$ : A amostra provem de uma população normal
- $H_1$ : A amostra não provem de uma população normal

A hipótese nula sendo escolhida, aceita-se que as amostras provêm de uma população normalmente distribuída e em seguida aplicamos o teste F. Caso a hipótese nula seja rejeitada, aplicamos o teste da Soma dos Postos de Wilcoxon.

### **Teste F**

O teste F visa avaliar se duas amostras provindas de populações normais independentes possuem a mesma variância. As hipóteses desse teste são:

- $H_0: s_x = s_y$
- $H_1: s_x \neq s_y$

Caso a hipótese nula seja aceita, isso mostra que as amostras possuem uma mesma variância e assim pode-se aplicar o teste t-Student para amostras independentes

### Teste t-Student

Esse teste é um método muito utilizado para se avaliar médias de amostras independentes, contanto que essas amostras sejam normalmente distribuídas as variâncias não sejam significativamente diferentes.

Existem três pares de hipóteses que podem ser formuladas para esse teste:

- Teste Bicaudal: 
$$\begin{cases} H_0 : \bar{x} = \bar{y} \\ H_1 : \bar{x} \neq \bar{y} \end{cases}$$
- Teste Unicaudal a direita: 
$$\begin{cases} H_0 : \bar{x} = \bar{y} \\ H_1 : \bar{x} \geq \bar{y} \end{cases}$$
- Teste Unicaudal a esquerda: 
$$\begin{cases} H_0 : \bar{x} = \bar{y} \\ H_1 : \bar{x} \leq \bar{y} \end{cases}$$

### Soma dos Postos de Wilcoxon

Caso a amostra não passe no teste de Shapiro-Wilk ou no teste F, será aplicado o teste da Soma dos Postos de Wilcoxon. A Soma dos Postos de Wilcoxon é um teste não-paramétrico que usa postos de dados amostrais de duas populações independentes. Esse teste é utilizado para testar se as duas amostras independentes provêm de populações com medianas iguais. As hipóteses para esse teste são:

- $H_0$  : As amostras provêm de populações com medianas iguais
- $H_1$  : As amostras provêm de populações com medianas diferentes



## 4 Resultados

Durante os experimentos, foi observado que os limites dentre os quais eram gerados os pesos da camada de entrada para a camada escondida exerciam uma influência no desempenho do modelo. Foram feitos experimentos utilizando valores -1, -0,5 e 0 para os valores mínimos e 0,5 e 1 para os valores máximos.

Outro parâmetro que foi alterado entre os experimentos foi a função de ativação dos neurônios da camada escondida. As funções utilizadas foram a Tangente Hiperbólica (3.2) e a Sigmoide Logística (3.3).

A Tabela 4.1 mostra as configurações utilizadas em cada experimento.

Experimento	Peso Máximo	Peso Mínimo	Função de ativação
ELM1	1	0	Tangente Hiperbólica
ELM2	1	0	Sigmoide Logística
ELM3	0,5	-0,5	Tangente Hiperbólica
ELM4	0,5	-0,5	Sigmoide Logística
ELM5	1	-1	Tangente Hiperbólica
ELM6	1	-1	Sigmoide Logística

**Tabela 4.1: Experimentos Realizados**

Para todos os experimentos foi executados o algoritmo apresentado na seção 3.4, com 5 neurônios inicialmente na camada escondida, incremento de 5 neurônios a cada ciclo e um valor máximo de 50 ciclos de busca caso não encontre um valor melhor.

### 4.1 Análise dos Resultados

Para os experimentos ELM1 e ELM2, foi observado que não havia diferença entre os resultados gerados por tamanhos distintos da camada escondido, como pode ser visto nas Figuras 4.1 e 4.2.

Como os dados foram normalizados entre 0 e 1, todos os pesos nesses experimentos eram

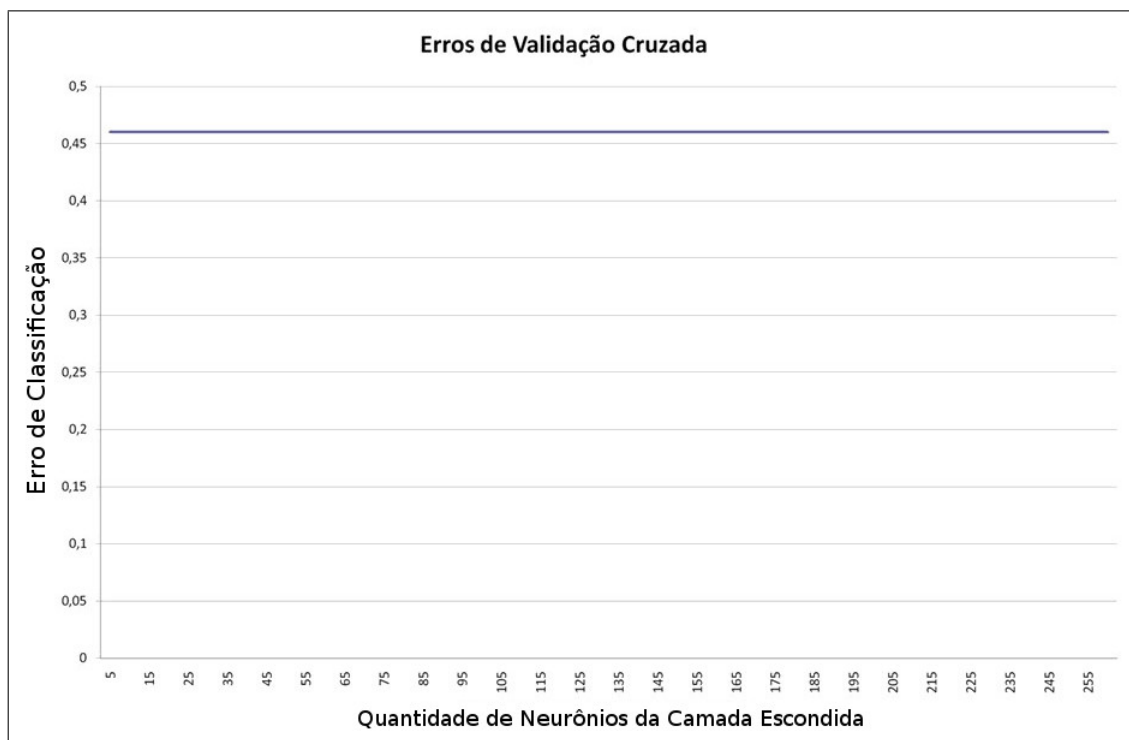


Figura 4.1: Busca ELM1

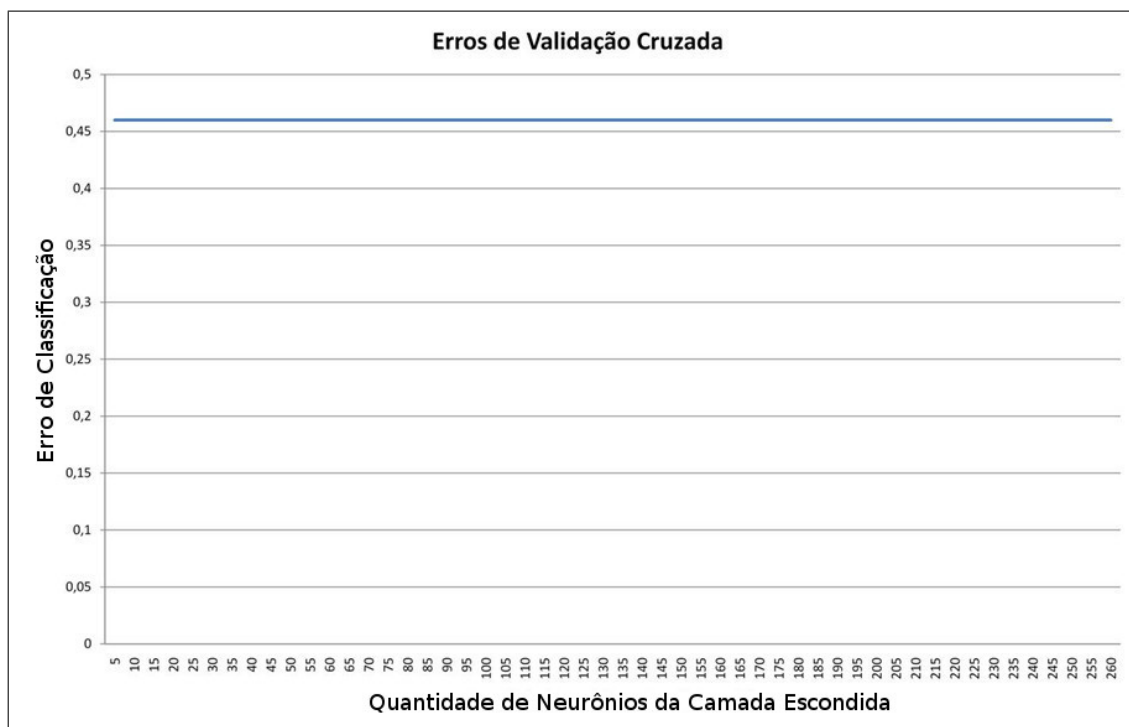
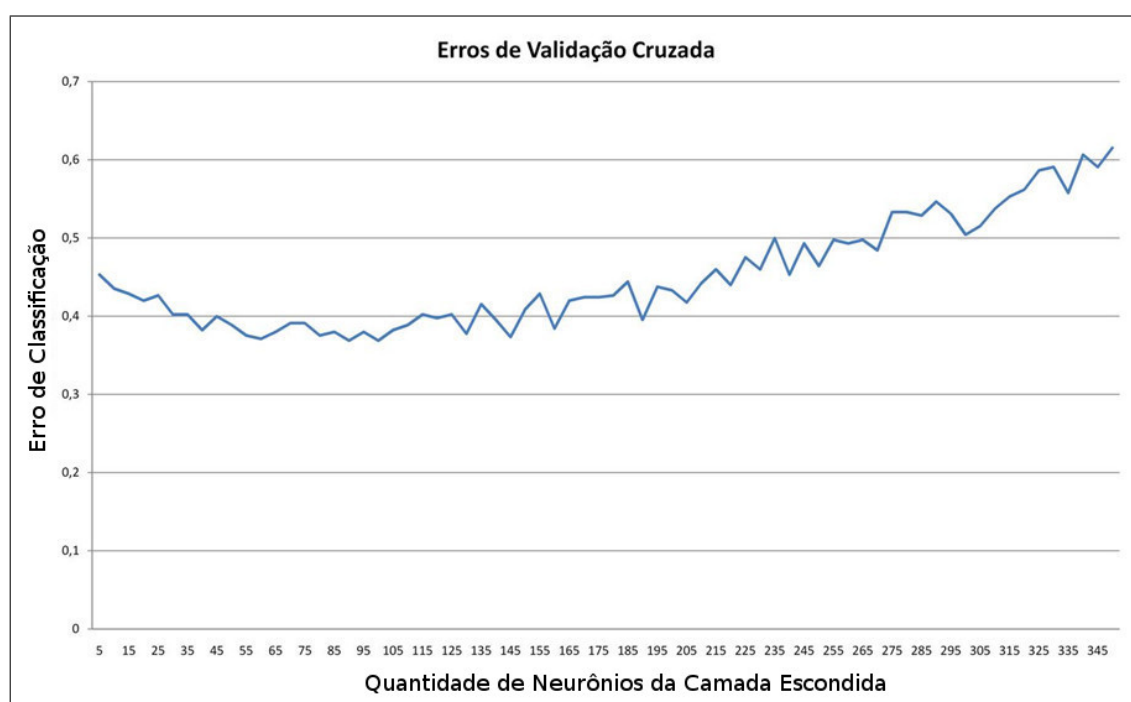


Figura 4.2: Busca ELM2

positivos e a quantidade de entradas foi muito grande, o somatório dos valores de entrada dos neurônios da camada escondida resultavam valores que tinham uma média aproximada de 40. Como esses valores eram maiores que o ponto de saturação tanto da Tangente Hiperbólica quanto da Sigmóide Logística (Figuras 3.1 e 3.2), a saída dos neurônios da camada escondida sempre foi igual a 1.

Os experimentos ELM3, ELM4, ELM5 e ELM6 apresentaram desempenho melhor que o dos apresentados pelos anteriores, como mostrado nas Figuras 4.3, 4.4, 4.5 e 4.6.



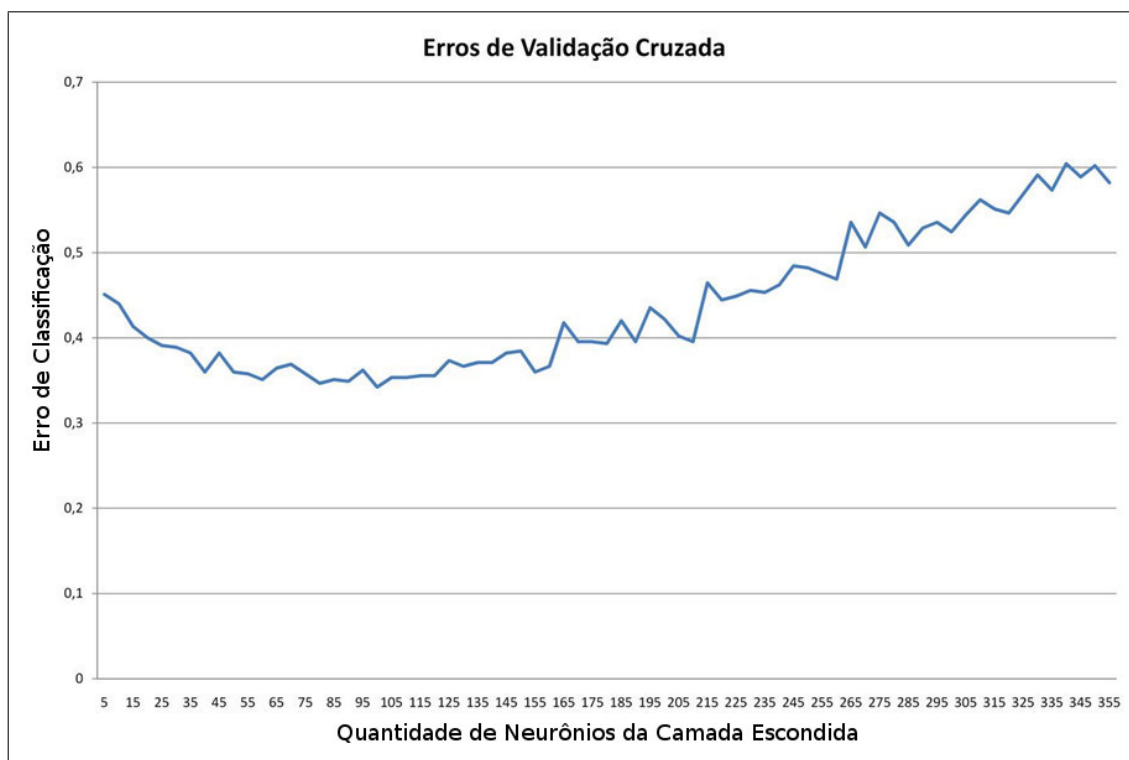
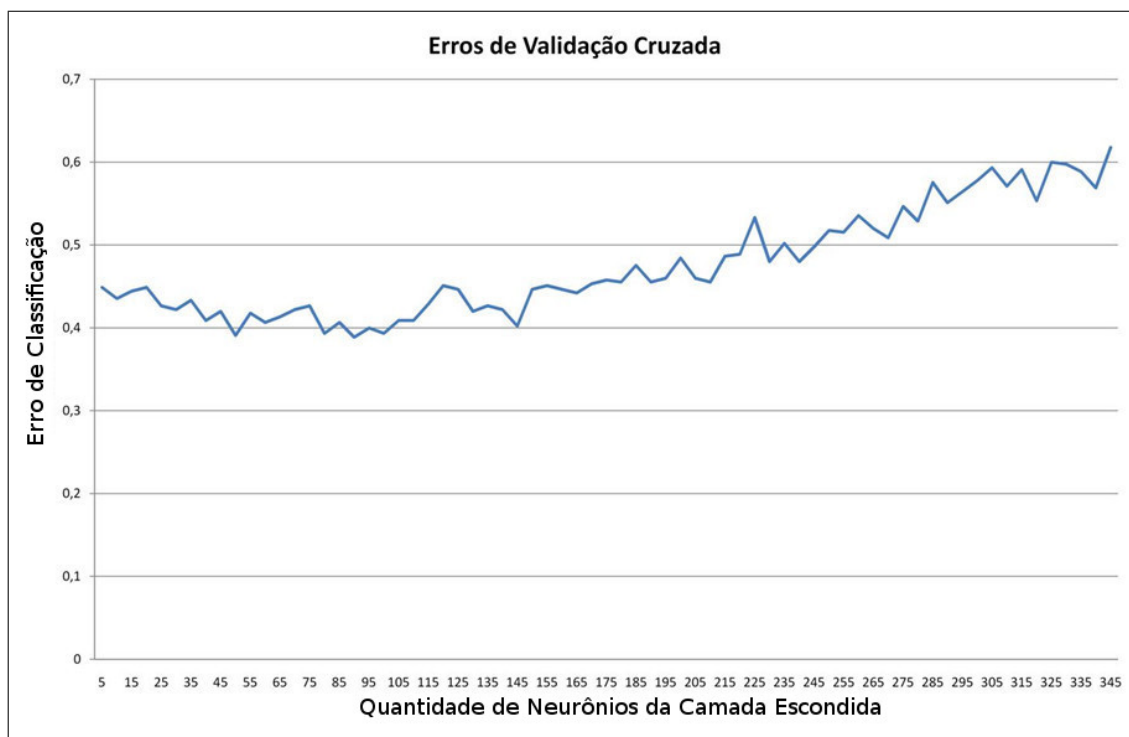
**Figura 4.3: Busca ELM3**

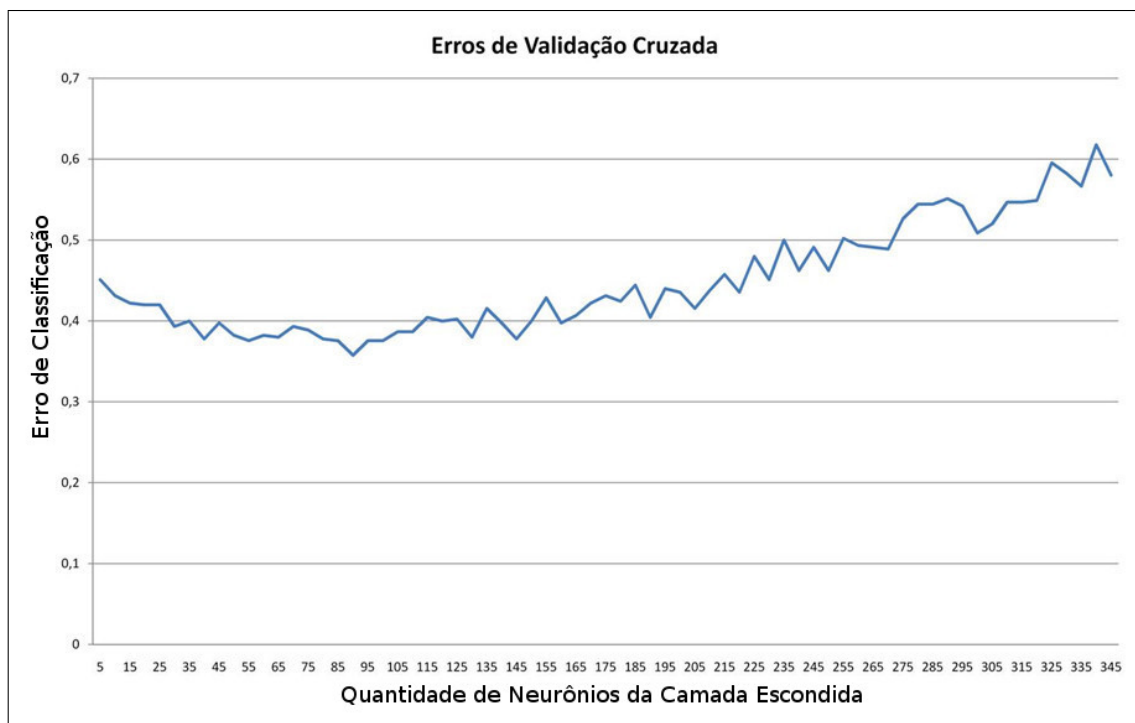
Quando encontrado o número de neurônios que retornava o menor erro de classificação, foram realizadas 30 simulações com essa mesma quantidade de neurônios para serem feitas as comparações. A Tabela 4.2 mostra os resultados encontrados para cada experimento.

Experimento	Tamanho da Camada Escondida	Média do Erro
ELM3	90	0,3758621
ELM4	100	0,3521839
ELM5	90	0,4163985
ELM6	90	0,3815236

**Tabela 4.2: Resultado das buscas**

Como os resultados encontrados em ELM1 e ELM2 não foram expressivos, os mesmos foram descartados dos testes de hipótese.

**Figura 4.4: Busca ELM4****Figura 4.5: Busca ELM5**



**Figura 4.6: Busca ELM6**

## 4.2 Comparação entre os experimentos

Para os testes de hipótese, foram feitos os testes estatísticos descritos no capítulo 3. O nível de significância utilizado para todos os teste foi de 5% (0.05).

Primeiramente foi feito o teste de Shapiro-Wilk, a Tabela 4.3 mostra os resultados.

Experimento	p-value
ELM3	0,4605
ELM4	0.2407
ELM5	0.01482
ELM6	0.33

**Tabela 4.3: Shapiro-Wilk para os experimentos com ELM**

Com esses resultados pode-se concluir que apenas o ELM5 provém de uma população não normalmente distribuída. As amostras provindas desse experimento não foram utilizada para os testes F e o t-Student.

Foram feitos 3 testes F, um para cada par de conjunto de amostras normalmente distribuídas. Os resultados se encontram na Tabela 4.4.

Como todas os conjuntos de amostras possuem a mesma variância, é possível se aplicar

Experimento	p-value
ELM3,ELM4	0.3848
ELM3,ELM6	0.8952
ELM4,ELM6	0.4605

**Tabela 4.4: Teste F entre os experimentos com ELM**

o teste t-Student para amostras independentes. Primeiramente foi feito o teste bicaudal para avaliar se as amostras apresentavam médias diferentes.

Experimento	p-value
ELM3,ELM4	$1,021 \cdot 10^{-8}$
ELM3,ELM6	0.122
ELM4,ELM6	$5,212 \cdot 10^{-11}$

**Tabela 4.5: t-Student entre os experimentos com ELM**

Os resultados apresentados na Tabela 4.5 atestam que os experimentos ELM3 e ELM6 possuem a mesma média, mas o experimento ELM4 não. Foi feito um teste unicaudal para averiguar se a média de ELM4 é menor que as médias de ELM3 e ELM6.

Experimento	p-value
ELM4,ELM3	$5,14 \cdot 10^{-9}$
ELM4,ELM6	$2,606 \cdot 10^{-11}$

**Tabela 4.6: t-Student unicaudal entre os experimentos com ELM**

Por fim, foi realizado o teste da Soma de Postos de Wilcoxon com os experimentos ELM4 e ELM5, para avaliar se as amostras possuíam a mesma mediana. O *p-value* para esse teste foi  $2,934 \cdot 10^{-11}$ , comprovando que as amostras possuíam medianas diferentes;

Com os testes realizados, pode-se afirmar que o melhor desempenho foi o apresentado pelo experimento ELM4.

### 4.3 Comparação com outras técnicas

Feitos as comparações entre os experimentos realizados com o algoritmo utilizado neste trabalho, o próximo passo é a comparação com resultados obtidos com outras técnicas de aprendizado de máquina. Foram escolhidas para comparação uma MLP treinada com o *Back-propagation* com 20 neurônios na camada escondida (MLP-BP) e o algoritmo *Naive Bayes* (RISH, 2001).

Foram coletadas 30 amostras de erro para cada técnica, e realizados os mesmos testes estatísticos aplicados na comparação entre os experimentos. A Tabela 4.7 mostra as médias de erro de classificação para essas técnicas.

Técnica	Média de erro da amostra
MLP-BP	0,345614
Naive Bayes	0,3804572

**Tabela 4.7: Média de erros de MLP-BP e Naive Bayes**

Para o teste de Shapiro-Wilk, foram obtidos os seguintes resultados.

Técnica	<i>p-value</i>
MLP-BP	0.06257
Naive Bayes	0.7639

**Tabela 4.8: Teste Shapiro-Wilk para MLP-BP e Naive Bayes**

Como ambas as amostras provêm de uma população normalmente distribuída, em seguida foi feito o teste F de cada uma delas com as amostras de ELM4.

Técnicas	<i>p-value</i>
MLP-BP, ELM	0,009467
Naive Bayes, ELM	$7,365 \cdot 10^{-5}$

**Tabela 4.9: Teste F entre ELM e MLP-BP e Naive Bayes**

Os resultados do teste F mostram que nenhuma das amostras avaliadas possuem a mesma variância. Com isso só foi possível realizar o teste das Somas dos Postos de Wilcoxon.

Técnicas	<i>p-value</i>
MLP-BP, ELM	0,9868
Naive Bayes, ELM	$1,249 \cdot 10^{-9}$

**Tabela 4.10: Soma dos Postos de Wilcoxon entre ELM e MLP-BP e Naive Bayes**

Com os resultados observados na Tabela 4.10, concluímos que as amostras provenientes de MLP-BP e ELM4 possuem a mesma mediana.

## 5 *Considerações Finais*

### 5.1 Conclusões

Esse trabalho se propôs a utilizar a técnica de *Extreme Learning Machine* para o treinamento de uma rede neural que classificasse dados provenientes de exames de Eletrocardiograma de acordo com classes de arritmia cardíaca.

Foram realizados experimentos variando-se os valores limites dos pesos gerados aleatoriamente entre a camada de entrada e a camada de saída. Também foi feita variação das funções de ativação dos neurônios da camada escondida. Dentre os experimentos, encontrou-se como melhor configuração para o problema apresentado, os pesos sendo gerados entre  $-0,5$  e  $0,5$  e a Sigmóide Logística como função de ativação.

Foi constatado que o desempenho dessa configuração é semelhante ao de outras técnicas mais tradicionais, como o a *Multilayer Perceptron* treinada com o *Backpropation* e o algoritmo *Naive Bayes*.

### 5.2 Trabalhos Futuros

Para uma melhora nos resultado obtidos pelo *Extreme Learning Machine* no diagnóstico de arritmias cardíacas baseado no Eletrocardiograma, podem ser feitos os seguintes trabalhos:

- Aplicar algoritmo para seleção de variáveis
- Utilizar um algoritmo de busca para selecionar parâmetros ótimos para o treinamento
- Utilizar uma generalização das redes ELM, chamada *Reservoir Computing*, que permite tratar sistemas dinâmicos por permitir feedback nos neurônios da camada de saída e escondida.
- Uso de outras bases de dados para validação



## *Referências*

- BARROSO, J. J.; ESPACIAIS, D. P.; CALDAS, I. L. Uso de Rede Neural Percéptron Multi-Camadas na e Classificação de Patologias Cardíacas. v. 2, n. 2, p. 255–264, 2008.
- BRAGA, A. d. P.; CARVALHO, A. P. D. L. F. D.; LUDERMIR, T. B. *Redes Neurais Artificiais - Teoria E Aplicações*. 2. ed. [S.l.]: LTC, 2007. 238 p.
- FRANK, A.; ASUNCION, A. *{UCI} Machine Learning Repository*. 2010. Disponível em: <<http://archive.ics.uci.edu/ml>>.
- GUYTON, A. C.; HALL, J. E. *Tratado de Fisiologia Médica*. 11. ed. Rio de Janeiro, RJ: Elsevier Editora Ltda, 2006.
- HAYKIN, S. *Redes Neurais*. 2. ed. [S.l.]: Bookman, 2001. 898 p.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing*, v. 70, n. 1-3, p. 489–501, dez. 2006. ISSN 09252312. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0925231206000385>>.
- JADHAV, S.; NALBALWAR, S.; GHATOL, A. Modular neural network based arrhythmia classification system using ECG signal data. *International Journal of . . .*, v. 4, n. 1, p. 205–209, 2011. Disponível em: <[http://csjournals.com/IJITKM/PDF 4-1/42.Shivajirao M. Jadhav1, Sanjay L. Nalbalwar2 & Ashok A. Gh.pdf](http://csjournals.com/IJITKM/PDF%204-1/42.Shivajirao%20M.%20Jadhav1,%20Sanjay%20L.%20Nalbalwar2%20&%20Ashok%20A.%20Gh.pdf) [http://csjournals.com/IJITKM/PDF 4-1/42.Shivajirao M. Jadhav1, Sanjay L. Nalbalwar2 %26 Ashok A. Gh.pdf](http://csjournals.com/IJITKM/PDF%204-1/42.Shivajirao%20M.%20Jadhav1,%20Sanjay%20L.%20Nalbalwar2%20%26%20Ashok%20A.%20Gh.pdf)>.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, Springer New York, v. 5, n. 4, p. 115–133, 1943. ISSN 0007-4985. Disponível em: <<http://dx.doi.org/10.1007/BF02478259>>.
- REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-Validation. In: LIU, L.; ÖZSU, M. T. (Ed.). *Encyclopedia of Database Systems*. Springer US, 2009. p. 532–538. ISBN 978-0-387-35544-3, 978-0-387-39940-9. Disponível em: <[http://dx.doi.org/10.1007/978-0-387-39940-9\\_565](http://dx.doi.org/10.1007/978-0-387-39940-9_565)>.
- RISH, I. An empirical study of the naive Bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial . . .*, 2001. Disponível em: <<http://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf>>.
- SIEGEL, A. F. Hypothesis testing: Deciding between reality and coincidence. In: *Practical Business Statistics*. [S.l.]: Elsevier, 2012.
- VALENÇA, M. J. *Fundamentos das Redes Neurais*. 2. ed. Olinda, PE: Livro Rápido, 2010. 386 p. ISBN 978-85-7716-342-7.

- 
- World Health Organization. *Cardiovascular disease*. 2012. Disponível em: <[http://www.who.int/cardiovascular\\_diseases/en/](http://www.who.int/cardiovascular_diseases/en/)>.
- ZUO, W.; LU, W. Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier. *Computers in Cardiology*, ..., p. 253–256, 2008. Disponível em: <[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4749025](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4749025)>.