

UNIVERSIDADE FEDERAL DE VIÇOSA
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

REGINALDO RODRIGUES DO CARMO JUNIOR

**REDES NEURAIS ARTIFICIAIS NO AUXÍLIO DO DIAGNÓSTICO DE
CARDIOPATIAS**

VIÇOSA
2011

REGINALDO RODRIGUES DO CARMO JUNIOR

**REDES NEURAIS ARTIFICIAIS NO AUXÍLIO DO DIAGNÓSTICO DE
CARDIOPATIAS**

Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 490 – Monografia e Seminário e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Leonardo Bonato Felix

VIÇOSA
2011

REGINALDO RODRIGUES DO CARMO JUNIOR

**REDES NEURAIS ARTIFICIAIS NO AUXÍLIO DO DIAGNÓSTICO DE
CARDIOPATIAS**

Monografia apresentada ao Departamento de Engenharia Elétrica do Centro de Ciências Exatas e Tecnológicas da Universidade Federal de Viçosa, para a obtenção dos créditos da disciplina ELT 490 – Monografia e Seminário e cumprimento do requisito parcial para obtenção do grau de Bacharel em Engenharia Elétrica.

Aprovada em 14 de junho de 2011.

COMISSÃO EXAMINADORA

Prof. Dr. Leonardo Bonato Felix - Orientador
Universidade Federal de Viçosa

Prof. M.Sc. Alexandre Santos Brandão - Membro
Universidade Federal de Viçosa

Prof. Dr. Denílson Eduardo Rodrigues - Membro
Universidade Federal de Viçosa

"...E nunca considerem seu estudo como uma obrigação, mas sim como uma oportunidade invejável de aprender, sobre a influência libertadora da beleza no domínio do espírito, para seu prazer pessoal e para o proveito da comunidade à qual pertencerá o seu trabalho futuro."

(Albert Einstein)

Aos meus pais pelo apoio incondicional.

Agradecimentos

Agradeço a todos aqueles que me viram crescer, amadurecer e mudar dentro da universidade e permaneceram ao meu lado apesar de tudo. Inicialmente aqueles a quem devo a minha vida, meus pais, que sempre me orientaram e estiveram ao meu lado independentemente de qualquer coisa. Ao meu falecido pai que me apoiou e lutou ao meu lado durante todo o período da minha graduação e me ensinou valores que irei carregá-los comigo e a minha mãe que sempre me deu suporte mesmo quando era impossível. Aos meus irmãos por serem ouvidos quando eu queria falar e ombros quando eu precisei de apoio. Aos meus amigos e colegas de faculdade que me acompanharam ao longo do curso e dedicaram comigo longas horas de estudo para que este momento fosse possível. Ao meu orientador Leonardo Bonato por ter me indicado o caminho das pedras por tantas vezes e me fazer enxergar que eu era capaz.

Resumo

Este trabalho visa à utilização de ferramentas computacionais para auxílio na descoberta de indivíduos portadores de cardiopatias através de classificação de padrões. Há uma previsão de que doenças cardíacas em 2030, quase 23,6 milhões de pessoas morrerão de doenças cardiovasculares, principalmente por doença cardíaca e derrame sendo a principal causa de morte no planeta. A metodologia proposta foi baseada em um banco de dados chamado de “*Heart Disease Database*” (Banco de dados de doenças cardíacas). Os dados analisados foram obtidos em Cleveland nos Estados Unidos, por Robert Detrano. Como ferramenta utilizada para classificação deste padrão de doença cardíaca foi escolhido as Redes Neurais Artificiais que são sistemas paralelos distribuídos compostos de unidades de processamento simples tendo um criterioso processo de seleção das variáveis analisadas que é fundamental para um bom desempenho de uma rede neural. Outro ponto importante a ser considerado em uma rede neural é quando ocorre um número excessivo de neurônios na camada oculta, que implica na perda de capacidade de aproximar-se do diagnóstico de cardiopatias. Essa rede neural então terá como entrada as avaliações dos exames feitos e retirados do banco de dados e sua saída esperada será a avaliação do paciente sendo que ele será classificado em sadio ou doente. Os resultados obtidos nesse trabalho aponta um nível de acerto de cerca de 96% enquanto estudos correlatos tiveram índices de acerto abaixo de 91% o que mostra um progresso nesta linha de raciocínio e também abre outro leque de pesquisa no sentido de uma busca contínua pelo índice máximo de acerto. Assim, conclui-se que as RNAs são ferramentas úteis na classificação de cardiopatias e podem auxiliar no diagnóstico médico precoce.

Palavras-chave: Redes Neurais, Reconhecimento de Padrão, Cardiopatia Isquêmica.

Abstract

This work aims to use computational tools to aid in the discovery of individuals with heart disease via pattern classification. It is predicted that diseases trophy in 2030, nearly 23.6 million people will die of cardiovascular disease, especially heart disease and stroke are the leading cause of death on the planet. The proposed methodology was based on a database called "Heart Disease Database" (Database of heart disease). Data were obtained from Cleveland in the United States, by Robert Detrano. As a tool used to classify this pattern of cardiac disease was chosen artificial neural networks are parallel distributed systems composed of simple processing units with a judicious selection of process variables is essential for good performance of a neural network. Another important point to be considered in a neural network when there is an excessive number of neurons in the hidden layer, which implies the loss of ability to approach the diagnosis of heart disease. This neural network will then input the evaluation of tests done and removed from the database and its expected output will be the evaluation of the patient and it will be classified as healthy or sick. The results of this study highlights a level of accuracy of about 96%, while related studies have hit rates below 91% which shows progress in this line of reasoning and also opens up another set of research towards a continuous search for content maximum accuracy. Thus, we conclude that the RNAs are useful tools in the classification of heart disease and can aid in medical diagnosis early.

Keywords: Neural Networks, Pattern Recognition, Ischemic Heart Disease.

Sumário

1	Introdução.....	14
1.1	Doenças Cardiovasculares.....	14
1.2	Redes Neurais Artificiais.....	15
1.2.2	Neurônios Artificiais.....	16
1.2.3	Funções de Ativação.....	17
1.2.4	Arquitetura de uma Rede Neural Artificial.....	19
1.2.5	Métodos de Estimação.....	20
1.2.6	Redes com múltiplas camadas.....	21
1.3	Objetivos Gerais.....	22
2	Materiais e Métodos.....	23
2.1	Base de Dados.....	23
2.2	Tratamento do banco de dados.....	25
2.2.1	Informação Mútua Conjunta.....	26
2.3	Classificadores de Padrão.....	27
2.3.1	Redes Neurais Artificiais.....	27
3	Resultados e Discussões.....	29
4	Conclusões.....	36
	Referências Bibliográficas.....	37

Lista de Figuras

Figura 1 - Causa de mortes com pessoas com mais de 60 anos (2004).....	15
Figura 2 - Constituintes da célula do sistema nervoso central.	16
Figura 3 - Modelo McCulloch e Pitts de uma célula do sistema nervoso central.	17
Figura 4 – Exemplo de função linear.....	17
Figura 5 – Exemplo de função linear.....	18
Figura 6 – Exemplo de função Degrau	18
Figura 7 – Exemplo de função Sigmoidal.	18
Figura 8 - Melhor rede neural utilizando algoritmo de treinamento “Levenberg-Marquardt”.30	
Figura 9 - Melhor rede neural utilizando algoritmo de treinamento Gradiente Descendente. .31	
Figura 10 - Rede neural com a configuração 3 – 3 – 1 utilizando 13 entradas.	33
Figura 11 - neural com a configuração 3 – 1 – 1 utilizando 8 entradas.....	34

Lista de Tabelas

Tabela 1 – Algoritmo de treinamento “Levenberg-Marquardt”.....	29
Tabela 2 - Algoritmo de treinamento Gradiente Descendente	29
Tabela 3 – Classificação de variáveis através de Informação Mútua Conjunta.	32
Tabela 4 – Percentagem de acerto na classificação de acordo com a inserção de variáveis	33

1 Introdução

1.1 Doenças Cardiovasculares

A doença arterial coronariana mantém-se no século XXI, a despeito dos importantes avanços conquistados, como primeira causa de morte nos países desenvolvidos e em grande parte das nações em desenvolvimento. A cardiopatia isquêmica, que é resultado de um desequilíbrio entre o suprimento e a demanda do coração por sangue oxigenado bem como uma baixa disponibilidade de insumos básicos para atividade cardíaca [12], também se enquadra neste retrospecto.

Em 1988, no Brasil, as DCV foram responsáveis pela maior proporção de óbitos no país: 31% das mortes em homens e 39% nas mulheres, sendo a principal causa de morte a partir dos 40 anos de idade e contabilizando 33% dos óbitos na faixa de 40 a 49 anos de idade. Na figura 1 pode-se observar que em 2004 as doenças cardiovasculares mataram cerca de 220 mil homens e mulheres, sendo a principal causa de morte no Brasil.

De acordo com o primeiro relatório global sobre doenças não-transmissíveis, que foi publicado pela Organização Mundial de Saúde (OMS), elas foram responsáveis por 36 milhões de mortes em 2008 – 63% do total mundial.

Assim pode-se concluir que mesmo com um avanço do nível de tecnologia nos últimos anos o mesmo não aconteceu na detecção de doenças cardiovasculares.

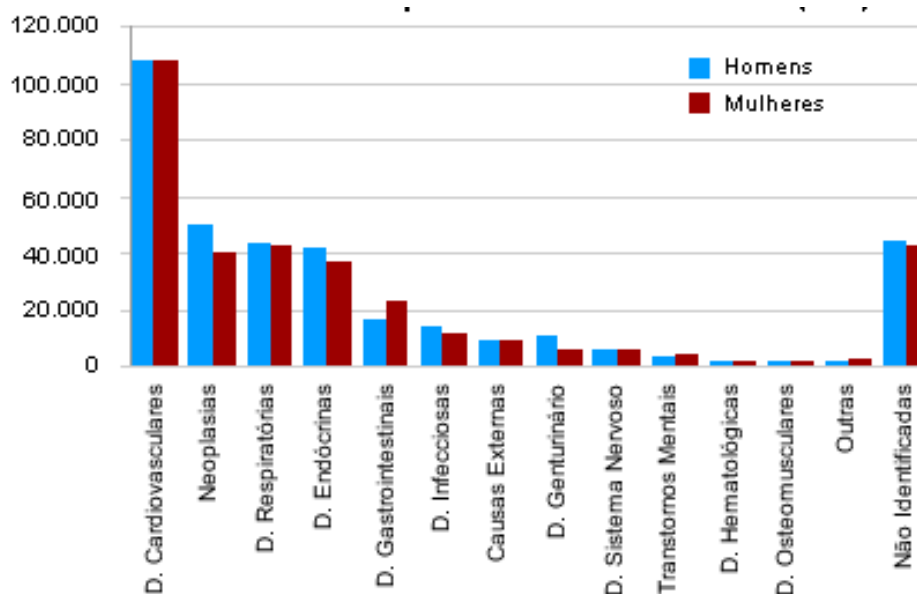


Figura 1 - Número de pessoas mortas (com mais de 60 anos) em função de várias *causa mortis* no ano de 2004.

Adaptado de [<http://pessoas.hsw.uol.com.br/indices-mortalidade-brasil6.htm>].

Há uma previsão de que doenças cardíacas em 2030, quase 23,6 milhões de pessoas morrerão de doenças cardiovasculares, principalmente por doença cardíaca e derrame sendo a principal causa de morte no planeta.

As doenças cardíacas podem ser separadas em diversas vertentes, sendo as mais citadas em literaturas que serão definidas de acordo com a Sociedade Brasileira de Cardiologia, entre elas têm-se as anginas, arritmias cardíacas, doenças valvares, endocardite infecciosa, hipertensão arterial e infarto agudo do miocárdio.

1.2 Redes Neurais Artificiais

O método escolhido para detecção destes padrões de cardiopatias foi as Redes Neurais Artificiais [3] [4] [5] que vem ganhando espaço no ramo de modelagem matemática, provando ser sistemas capazes de resolver problemas em situações onde é difícil criar modelos explícitos. Redes Neurais Artificiais são sistemas paralelos distribuídos compostos por unidades de processamento simples que calculam determinadas funções matemáticas. Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais.

Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede [6]. As redes neurais simulam o comportamento de neurônios humanos, os quais são capazes de realizar tarefas complexas. Vislumbra-se mostrar que a eficiência do computador simulando este neurônio é superior ao cérebro humano realizando a mesma tarefa, sendo assim, capaz de superar o nível de acerto no diagnóstico de doença cardíaca [13], pois o nível de acerto do cérebro humano está condicionado às condições externas, como o estresse, e além do fato do computador analisar todas as variáveis sempre que lhe for dada a entrada.

1.2.2 Neurônios Artificiais

O modelo que foi criado na década de 40 por McCulloch e Pitts é uma aproximação do modelo biológico, visto da figura 2, que é composto de um corpo principal (Soma) que processa o sinal trazido pelos dendritos e envia um sinal de saída através do axônio. Sua descrição matemática resultou em um modelo com n terminais de entrada x_1, x_2, \dots, x_n (que representam os dendritos e apenas um terminal de saída y (representando o axônio).

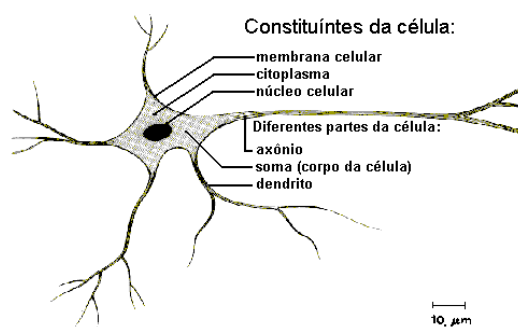


Figura 2 - Constituintes da célula do sistema nervoso central. Adaptado de
[<http://www.icmc.usp.br/~andre/research/neural/>]

Para que este comportamento lógico se comporte como o das sinapses, as entradas do neurônio têm pesos (pesos sinápticos) w_1, w_2, \dots, w_n que podem ser positivos ou negativos dependendo se as sinapses são excitatórias ou inibitórias. Os pesos determinam a prioridade no qual o neurônio deve considerar sinais de disparo que ocorrem naquela conexão.

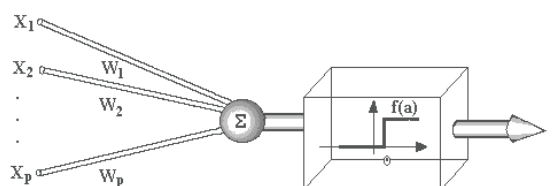


Figura 3 - Modelo McCulloch e Pitts de uma célula do sistema nervoso central. Adaptado de
[\[http://www.icmc.usp.br/~andre/research/neural/\]](http://www.icmc.usp.br/~andre/research/neural/)

As entradas têm seus “valores” multiplicados por seus respectivos pesos e esses valores finais são somados, como pode ser identificado na figura 3. Caso esta soma, que pode também ser chamado de soma ponderada, ultrapasse um limiar de excitação (*threshold*) haverá um disparo de um pulso através do axônio [7]. Quem decide se haverá esse disparo através do axônio será a função de ativação deste neurônio. Esta função é dada pela seguinte equação:

$$\sum_{i=1}^n x_i w_i \geq \Theta$$

1.2.3 Funções de Ativação

Com o modelo de McCulloch e Pitts, outros modelos com saídas que não necessariamente serão 0 ou 1 acompanhadas de outras funções de ativação surgiram e como exemplo pode-se citar:

- Função Linear

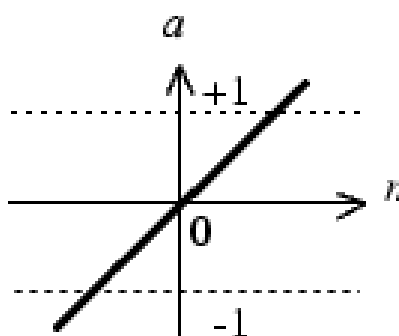


Figura 4 – Exemplo de função linear. Adaptado de [\[http://www.mathworks.com\]](http://www.mathworks.com)

- Função Rampa.

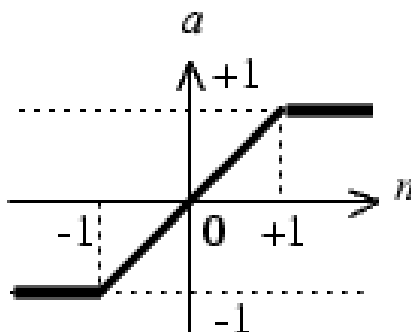


Figura 5 – Exemplo de função linear. Adaptado de [<http://www.mathworks.com>]

- Função Degrau

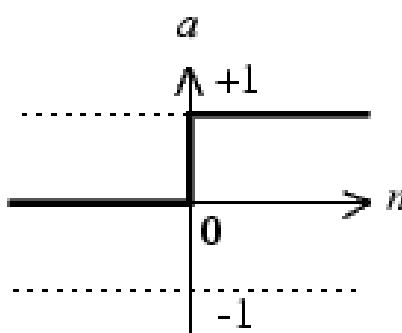


Figura 6 – Exemplo de função Degrau. Adaptado de [<http://www.mathworks.com>]

- Função Sigmoideal.

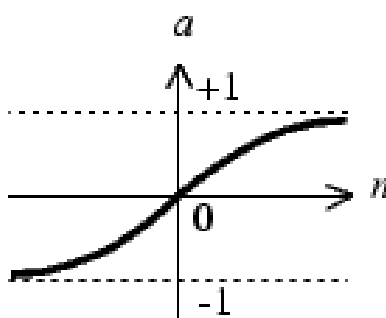


Figura 7 – Exemplo de função Sigmoideal. Adaptado de [<http://www.mathworks.com>]

A função de ativação linear é dada pela equação $y = \alpha x$, onde α é um número real e o coeficiente que define uma saída y para uma dada entrada x , como mostrado na figura 4. Caso esta função ganhe limitações, por exemplo na figura 5, em uma faixa $[-\gamma, +\gamma]$, esta função de

ativação agora será uma função rampa. O valor máximo e mínimo da saída é $+\gamma$ e $-\gamma$, respectivamente. A função passo, ilustrada na Figura 6, é similar a uma função sinal no sentido de que a função produz a saída $+\gamma$ para os valores de x maiores que zero caso contrário a função produz o valor $-\gamma$. A função degrau é definida pela equação:

$$y = \begin{cases} +\gamma & \rightarrow x > 0 \\ -\gamma & \rightarrow x < 0 \end{cases}$$

A função Sigmoidal da figura 7 é uma função semilinear, limitada e monotônica. Uma das funções sigmoidais mais importantes é a função logística definida pela equação:

$$y(x) = \frac{1}{1 + e^{\frac{-x}{T}}}$$

Onde o parâmetro T determina a suavidade da curva.

1.2.4 Arquitetura de uma Rede Neural Artificial

Com o advento da teoria do modelo de neurônio de McCulloch e Pitts problemas linearmente separáveis eram facilmente resolvidos, porém o oposto dessa vertente encontrava sérios problemas para que a solução ótima fosse encontrada [8]. Assim, com essa perspectiva foram criadas redes com camadas ocultas e com algoritmos de treinamento capazes de encontrar seus pesos ideais com maior facilidade. Então pode-se classificar esses diferentes tipos de arquiteturas em número de camadas da rede, número de nodos em cada camada, tipo de conexão entre os nodos e topologia da rede [6].

Quanto ao número de camadas, pode-se ter:

- Redes de camada única que só possuem um neurônio entre entrada e saída da rede.
- Redes de múltiplas camadas que mais de um neurônio entre entrada e saída da rede.

Os nodos podem ter conexões do tipo:

- feedforward, ou acíclica não há realimentação da informação saída de um nó com qualquer neurônio anterior.

- feedback, ou cíclica há realimentação da informação saída de um nó com qualquer neurônio anterior.

1.2.5 Métodos de Estimação

As redes neurais têm a capacidade de identificar padrões retirando característica de cada modelo que lhe é passado. Esse tipo de característica é um método de aprendizado para que se possam estimar os pesos de cada entrada em cada um de seus neurônios. Logo se faz necessário a utilização de um método bem estruturado para que esta estimação otimize o tempo e ache os melhores parâmetros. Logo essa aprendizagem acontece de forma contínua de maneira que o ambiente promove a entrada e a aprendizagem acontece de iterativa, sendo assim estimados os parâmetros da rede, no caso os pesos e bias [14]. Logo para que seja possível utilizar uma rede neural, é necessário passar por esta fase de estimativa de parâmetros onde há a extração de parâmetros e informações que são cruciais à base de dados e depois com essas informações, que nada mais é que uma relação entre entrada, saída e erro associado, passasse os pesos atualizados a cada ramo de entrada de cada neurônio.

Existem duas estratégias de treinamento para uma rede: aprendizado supervisionado e não supervisionado. Aprendizado por reforço e o aprendizado por competição são apresentados em [3] e [6].

O aprendizado supervisionado conta com um professor que consegue analisar cada saída para um conjunto de entradas e para cada uma delas calcula um erro associado e atribui um novo peso para cada entrada de cada neurônio. O algoritmo se repete até que o peso adequado que produza uma saída ideal possa ser vista. Esse ajuste acontece passo a passo com cada entrada produzindo uma nova saída e assim sucessivamente. A tendência então é de que a cada iteração se aproxime cada vez mais da resposta desejada. O aprendizado não supervisionado não conta com um professor e mesmo com certa semelhança entre o aprendizado supervisionado e o aprendizado dos seres humanos, muitos sistemas biológicos recorrem ao aprendizado não supervisionado como visão e audição [6]. Este formato de aprendizado conta somente com dados de entrada e utiliza da harmonia entre a rede e uma regularidade estatística de entrada de dados para que seja possível criar novas classes e grupos de maneira automática, porém dados com redundância são extremamente úteis para que esse aprendizado seja efetivo.

1.2.6 Redes com múltiplas camadas

Uma rede com uma única camada pode implementar qualquer função contínua, ou seja, pode achar solução para qualquer problema linearmente separável, porém, no final da década de 60, Minsky e Pappert colocaram em xeque a capacidade das Redes Neurais ao publicar o artigo *Perceptrons: An introduction to computational geometry* pela MIT Press em Massachusetts e apresentar *hard learning problems* (problemas de difícil aprendizagem). Como exemplo problemas que apresentam paridade, conectividade ou ainda qualquer tipo de simetria. Durante a década de 70 essa abordagem conexionista ficou esquecida, uma grande parte devido a esta publicação, porém em 1982, John Hopfield publicou um artigo intitulado *Neural Networks and physical systems with emergent collective properties* mostrando a relação entre redes recorrentes auto-associativas. Logo depois surgia o algoritmo de treinamento Backpropagation o que moveu novamente os estudos nesta área. Segundo George Cybenko em um artigo publicado em 1988, intitulado *Continuous valued neural network with two hidden layers are sufficient*, a utilização de camadas ocultas torna qualquer problema solucionável.

As Redes Neurais de Múltiplas Camadas possuem quantas entradas e saídas forem necessárias para o sistema que está sendo analisado em questão, porém existem quantas camadas ocultas forem necessárias para que a complexidade deste problema seja satisfatoriamente resolvida [14]. Então tem-se que o processamento do neurônio atual depende da combinação dos processamentos da camada de neurônios anteriores conectados a ele. Essas camadas interiores captam características que são cruciais ao desenvolvimento de uma solução do sistema e por esta razão se torna interessante que o número de neurônios nesta camada oculta atenda plenamente o nível de complexidade do problema. Este é um dos problemas mais críticos dentro de uma rede neural multicamadas já que caso um número acima de neurônio (*overfitting*) iria tornar a rede demasiadamente grande, ou seja, difícil de encontrar uma solução ótima ou ainda caso o número de neurônios estiver abaixo do número ideal (*underfitting*) implicando em uma rede que não consegue convergir para uma solução satisfatória. Este número ideal fica ligado a fatores como a quantidade de dados dentro do universo de treinamento, a quantidade de dados que adicionam ruídos ao treinamento, a complexidade da função em questão a ser aprendida e a distribuição estatística dos dados.

1.3 Objetivos Gerais

Este estudo visa o auxílio no diagnóstico de cardiopatia utilizando as redes neurais artificiais, através de uma variabilidade de camadas ocultas e neurônios em cada uma destas camadas e tipos de treinamento de redes neurais para comparação quanto acerto no diagnóstico. A vantagem está no fato de que o cérebro humano está sujeito a ações de interferência externa, como o estresse, e assim pode se interferir em um diagnóstico de um paciente. Além disso, temos que um médico com o passar dos anos cria certo padrão de análise dentro dos exames podendo avaliar alguns itens com maior valor que outros. Dentro de uma Rede Neural Artificial temos uma análise sempre de maneira total dos dados de entrada, sendo assim, sempre estará sendo avaliado a situação com um todo.

Dado o objetivo geral, têm-se como objetivos específicos:

- Criação de um algoritmo capaz de identificar a melhor rede neural no reconhecimento de cardiopatias;
- Utilizar uma ferramenta de seleção de variáveis capaz de elencar as variáveis mais importantes;
- Estudar Redes Neurais Artificiais (RNA) e, posteriormente, aprofundar na área dos problemas envolvendo reconhecimento de cardiopatias;
- Comparar este estudo com os que utilizaram métodos semelhantes.

2 *Materiais e Métodos*

2.1 *Base de Dados*

A base de dados utilizada foi a “Heart Disease Database” (Base de dados de doenças cardíacas) [2]. A base de dados original possui quatro bancos de dados distintos sendo eles da Hungria, Suíça e dois dos EUA. Aqui dentro deste estudo estará sendo utilizado os dados de Cleveland já que os outros possuem muitos dados incompletos. Estes dados de Cleveland foram obtidos no V.A. Medical Center, em Long Beach, e no Cleveland Clinic Foundation, em Cleveland, ambas nos Estados Unidos, por Robert Detrano. São 303 indivíduos, destes nas quais foram retirados 6 por não possuírem dados completos e desta restando assim 297 amostras, que foram analisados em 76 atributos, porém, grande parte das pesquisas publicadas que utilizaram esta mesma base de dados utilizaram apenas 14 deles, sendo 13 os atributos e o resultado final a respeito da cardiopatia analisada os quais são listados abaixo:

1. Idade (AGE) – variando de 29 a 77 anos;
2. Sexo (SEX) – masculino ou feminino, sendo representados por 1 ou 0, respectivamente;
3. Tipo de dor no peito (CP) – quatro tipos de dor no peito:
 - 3.1. Valor 1: angina típica;
 - 3.2. Valor 2: angina atípica;
 - 3.3. Valor 3: sem dor anginal;
 - 3.4. Valor 4: assintomático.
4. Pressão arterial em repouso (TRESTBPS) – medida em mm Hg;
5. Colesterol no soro sanguíneo (CHOL) – medido em mg/dl;
6. Concentração de açúcar no sangue (FBS) > 120 mg/dl – verdadeiro (1) ou falso (0);
7. Resultado da eletrocardiografia em repouso (RESTECG):
 - 7.1. Valor 0: Normal;
 - 7.2. Valor 1: Com onda ST-T anormal;
 - 7.3. Valor 2: Mostrando provável (ou definida) hipertrofia do ventrículo esquerdo.
8. Máxima taxa de batimento cardíaco atingido (THALACH);
9. Angina induzida por exercício (EXANG):
 - 9.1. Valor 1: Sim;

- 9.2. Valor 0: Não.
- 10. Depressão ST induzida por exercício relativamente sossegado (OLDPEAK);
- 11. Inclinação da extremidade do segmento ST no exercício (SLOPE):
 - 11.1. Valor 1: Inclinado para cima;
 - 11.2. Valor 2: Plano;
 - 11.3. Valor 3: Inclinado para baixo.
- 12. Número de vasos coloridos pela fluoroscopia (CA) – valor de 0 a 3;
- 13. Talassemias (THAL):
 - 13.1. Valor 1: normal;
 - 13.2. Valor 2: defeito fixo (irreparável);
 - 13.3. Valor 3: defeito reversível (reparável).

A décima quarta variável desde banco de dados refere-se à saída, ou seja, se o paciente que foi avaliado com os outros 13 itens deste banco de dados possui ou não doença cardíaca, demonstrados através de um código binário, 0 para indivíduos saudáveis e 1 para indivíduos que possuem alguma cardiopatia.

Os resultados conseguidos por este estudo foram analisados juntamente com outras pesquisas que utilizaram métodos semelhantes ou ainda o mesmo método para análise, a fim de que pudesse ser comparado. Tem-se dos 297 indivíduos analisados, 160 indivíduos saudáveis e 137 indivíduos doentes. Originalmente, o banco de dados contém 76 atributos para cada indivíduo e sofreu essa alteração, de 75 atributos para 13 acrescidos da coluna de saída que se mostra dentro do meio médico decisivo quanto ao laudo cardiovascular do paciente. Se fosse considerado todo ele, iria ser inserida uma grande quantidade de ruído o que prejudicaria o reconhecimento dos padrões. Este ruído seria em parte dados que não irão interferir na doença cardiovascular do paciente, por exemplo, o nome. Caso a rede fosse treinada com todos os dados os pesos referentes a dados, que não fossem pertinentes ao reconhecimento deste padrão de doença cardíaca, durante o treinamento da rede tenderia para um valor nulo. Porém como temos uma capacidade de processamento limitada e a possibilidade de *overfitting* é real o corte de variáveis foi necessário.

Os treze dados mais a saída inicialmente considerados nos mostram apenas quais dados dentre todos obtidos através da pesquisa de Cleveland estão completos e adicionariam pouco ruído a esta experiência o que não garante que eles de fato possam ajudar no reconhecimento de

padrões. Então se fez necessário uma análise de quais arquiteturas de redes poderiam identificar com uma maior facilidade este padrão dentro destas 13 variáveis.

2.2 Tratamento do banco de dados

O banco de dados começou a ser manipulado através da separação entre as duas classes, sadios e doentes, de maneira que pudessem randomizar a cada iteração todo o banco de dados. Este artifício é usado para que a rede neural nunca repita o mesmo banco de dados em suas iterações. Uma vez randomizados as duas classes separadas foi retirada da classe de sadios 70% para que fosse composta a classe de treinamento, 15% para que fosse composta a classe de teste e os 15% restantes compuseram a parcela de validação da rede neural. O mesmo procedimento foi repetido com o grupo dos indivíduos sadios. Este procedimento é necessário, pois a rede utiliza dos dados de treinamento para poder chegar ao pesos ideais que formarão através da soma ponderada das entradas a saída desejada. Ao ser treinada a rede tende a decorar o padrão apresentado e para isso os dados de validação são utilizados para que se monitores o nível de generalização da rede. O grupo de teste mostra o nível de acerto para indivíduos jamais apresentados a rede.

O primeiro passo após a randomização do banco de dados e separação proporcional deste, foi a remoção de linhas da entrada que possuíam valores que poderiam ser constantes, os quais não adicionariam informações suficientes à rede neural e os valores de cada linha foram normalizados, para que a rede receba valores que possam ter seus padrões mais facilmente identificados.

Uma grande base de dados é um desafio contínuo para uma rede neural, sendo que dentro deste banco de dados haverá muitos dados serão irrelevantes ao reconhecimento de determinado padrão, ou seja, não adicionarão nenhuma informação relevante implicando em um *overfitting* de informação e possivelmente fazendo nossa rede atingir soluções de máximo local [15]. Seria então extremamente útil a seleção de variáveis que possam mostrar reais características dentro do banco de dados. Neste momento do trabalho então será focado em escolher algumas variáveis que podem ser retiradas do banco de dados utilizando Informação Mútua que é uma relação entre uma entrada X com sua respectiva saída Y de maneira que se uma determinada entrada. Por exemplo, no resultado de um exame particular como o nível de glicose no sangue. Caso faça sua saída variar, diminuindo assim a incerteza com relação à

observação que está sendo feita então ter uma informação mútua grande. Derivada da informação mútua a técnica de Informação mútua conjunta será utilizada neste estudo para que as variáveis que tragam mais informações para o reconhecimento de padrões possam ser escolhidas.

2.2.1 Informação Mútua Conjunta

Uma grandeza fundamental para a seleção de variáveis é a informação mútua entre duas variáveis aleatórias X e Y [16], que aqui neste estudo em especial será tratada como a uma variável de entrada e sua respectiva saída. Essa grandeza, de certa forma, indica a redução no nível de incerteza associado a uma variável graças à informação trazida pela outra variável. Em outras palavras, a informação mútua associada a um par de variáveis X e Y é dada por:

$$I(x, y) = H(x) - H(x/y) = H(y) - H(y/x)$$

Onde H é a entropia de uma variável qualquer medida em função da sua probabilidade e relação a sua ocorrência. Um exemplo que pode ser citado é que seja feita a leitura de dois bilhetes: o primeiro diz “Seu amigo, o Sr. Y, acabou de saber que ganhou na loteria.” e o segundo diz “Sua amiga, a Sra. Z, descobriu que o bebê que está esperando é uma menina.”. É razoável esperar que a primeira notícia seja bastante surpreendente, enquanto a segunda, provavelmente, não causará tanto espanto. Este impacto causado pode ser interpretado como a quantidade de entropia da informação.

Proposto por Yang e Moody [17] é uma informação entre a saída desejada para cada caso conjunto com uma variável qualquer definida pelo emparelhamento entre um candidato X_n com cada uma das variáveis. Essa relação pode ser simplificada pela equação,

$$I_{JMI} = I(x_n; y) - \frac{1}{n-1} \sum_{k=1}^{n-1} [I(x_n; x_k) - I(x_n; x_k / y)]$$

A informação Mútua Conjunta captura completamente um termo em redundância, mas toma apenas seu valor médio.

2.3 Classificadores de Padrão

Dentro de um universo de classificadores de padrão que aqui podem ser citados alguns exemplos como Lógica Fuzzy, Máquina de Vetor Suporte e as Redes Neurais Artificiais, aqui neste estudo será utilizado apenas esta última técnica.

2.3.1 Redes Neurais Artificiais

Buscou-se a melhor rede neural que seria capaz de reconhecer um maior número de indivíduos dentro do universo do banco de dados. Para isso foi criado um algoritmo que é capaz de fazer iterações com o número de neurônios nas camadas ocultas de uma rede neural, de maneira que a cada iteração é testada por até 100 vezes esta mesma configuração para que se chegue aos melhores pesos e bias de uma rede. Os testes seguiram o seguinte padrão:

1 – Foram escolhidos os algoritmos de treinamento “Levenberg-Marquardt” e Gradiente descendente para solução dos pesos e bias da Rede Neural que possuem características distintas. O primeiro possui característica de rápida convergência e o segundo possui a característica de alterar o passo de adaptação dos pesos e bias para que caso esteja convergindo para um ponto de sela ou mínimo local ele consiga ir à busca de algum outro ponto de solução do problema.

2 – A função sigmoideal foi escolhida para que fosse a função de ativação desta rede neural em sua camada oculta, pois sua saturação nos valores 0 e 1 já era conhecida e como a saída do sistema analisado é binário esta função se encaixa perfeitamente.

3 – Depois de ter dado início ao treinamento, a rede simula 100 redes neurais para cada caso e escolhe a melhor dentro todas para solução do problema. Esta escolha se baseia na performance da rede que é uma relação entre a saída desejada e a saída real que a rede através do erro quadrático médio e que está produzindo para os pesos escolhidos naquela determinada iteração.

4 – Uma vez que a rede foi treinada com a mesma configuração de neurônios e camadas ocultas a iteração volta ao início, onde é escolhida uma nova configuração de neurônios e camadas ocultas e novos treinamentos são feitos até que novamente se escolha pesos e bias ideais para esta outra configuração e assim sucessivamente. Essas alterações de neurônios e

camadas ocultas se dá de maneira limitada, pois aqui é desejável apenas mostrar que o algoritmo é capaz de sugerir redes que possam ser a solução do problema, sendo que, foi sugerido as camadas variando de 1 a 3 com neurônios em cada uma delas variando de 1 a 3.

3 Resultados e Discussões

Foram analisadas ao todo 27 configurações distintas de redes neurais todas utilizando o mesmo algoritmo iterativo para que seu cálculo de acerto/performance fosse atingido o que diferencia os resultados analisados é que foram testados dois algoritmos de treinamento, “Levenberg-Marquardt” e Gradiente Descendente, respectivamente mostrados nas duas tabelas abaixo os seus 10 melhores resultados. Estas tabelas estão separadas em colunas sendo CAM1, CAM2 e CAM3 as colunas que mostram a quantidade de neurônios em cada camada oculta e as colunas Treino, Validação, Teste e Médio mostram taxas de acerto em cada um desses grupos.

Tabela 1 – Algoritmo de treinamento “Levenberg-Marquardt”.

CAM1	CAM2	CAM3	Treino	Validação	Teste
2	1	2	83,30%	72,70%	88,60%
2	3	3	87,10%	84,10%	86,40%
1	3	2	87,60%	77,30%	84,10%
2	2	1	82,80%	81,80%	84,10%
2	2	2	85,20%	88,60%	84,10%
2	3	2	90,90%	77,30%	81,80%
1	1	1	84,70%	90,90%	81,20%
1	2	3	88,00%	75,00%	79,50%
3	1	2	88,00%	75,00%	79,50%
3	1	3	86,10%	77,30%	79,50%

Tabela 2 - Algoritmo de treinamento Gradiente Descendente

CAM1	CAM2	CAM3	Treino	Validação	Teste
2	1	2	83,30%	72,70%	88,60%
2	3	3	87,10%	84,10%	86,40%
1	3	2	87,60%	77,30%	84,10%
2	2	1	82,80%	81,80%	84,10%
2	2	2	85,20%	88,60%	84,10%
2	3	2	90,90%	77,30%	81,80%
1	1	1	84,70%	90,90%	81,20%
1	2	3	88,00%	75,00%	79,50%
3	1	2	88,00%	75,00%	79,50%
3	1	3	86,10%	77,30%	79,50%

Sua matriz confusão para o melhor caso das redes neurais está mostrada nas duas figuras abaixo sendo a primeira utilizada o algoritmo de treinamento “Levenberg-Marquardt” e a segunda o algoritmo de treinamento Gradiente Descendente. Esta matriz mostra os dados de maneira que o eixo vertical mostra a saída real da rede e o eixo horizontal a saída desejada.

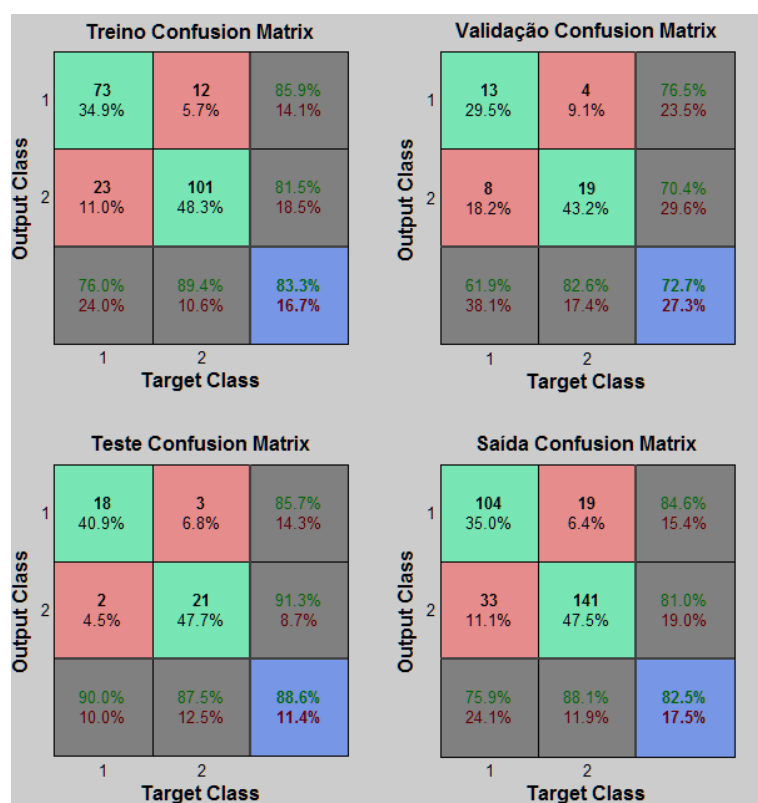


Figura 8 - Melhor rede neural utilizando algoritmo de treinamento “Levenberg-Marquardt”.

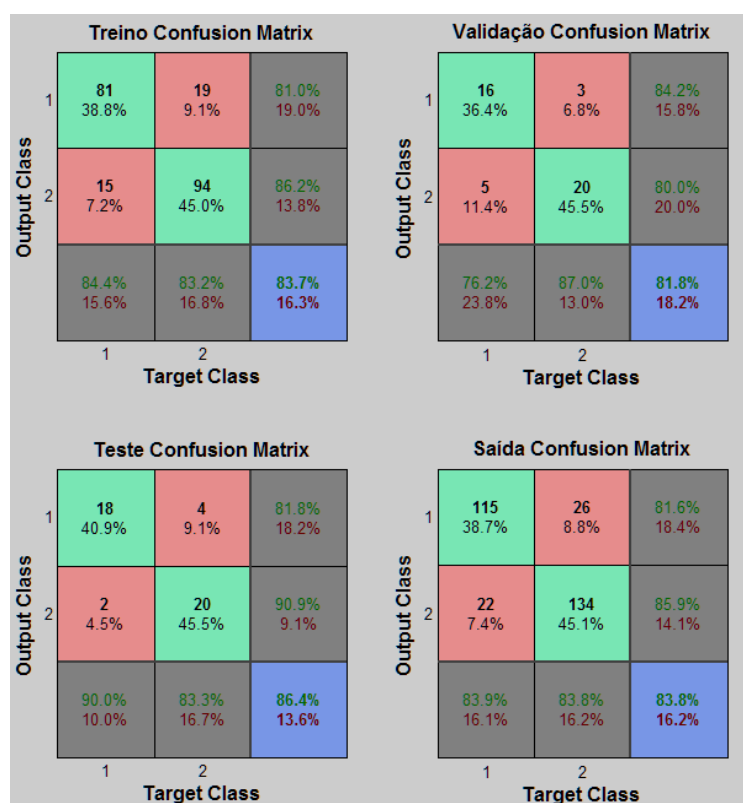


Figura 9 - Melhor rede neural utilizando algoritmo de treinamento Gradiente Descendente.

Com as tabelas acima é possível perceber que existem algumas configurações que ambos os algoritmos de treinamento afirmaram serem bons classificadores dentro desse universo apresentado. A rede que figura na primeira posição do algoritmo de treinamento “Levenberg-Marquardt” e no algoritmo de treinamento Gradiente Descendente possuem ambos o mesmo nível de acerto, porém com relação a velocidade de treinamento o algoritmo “Levenberg-Marquardt” se mostra muito mais rápido e por isso será utilizado nos testes a seguir.

Utilizando a técnica de seleção de variáveis *Informação Mútua Conjunta* foram classificadas as variáveis até que se encontrassem uma solução que fosse melhor do que a oferecida pelas treze variáveis ao mesmo tempo. A ordem seguida obtida através da Informação Mútua Conjunta foi a seguinte:

Tabela 3 – Classificação de variáveis através de Informação Mútua Conjunta.

Ordem	Parâmetro
1	CHOL
2	THALACH
3	Age
4	TRETBPS
5	Old Peak
6	CP
7	CA
8	THAL
9	Slope
10	Exang
11	Sexo
12	Restecg
13	Fbs

Assim as informações relativas a cada entrada foram sendo extraídas através deste método e assim pode-se elencar as mais importantes e que geravam um maior impacto na saída. O que foi feito foi o teste de cada modelo de rede neural, utilizando a variação da quantidade de neurônios, com a variação de números de entradas utilizando o ranking gerado através da Informação Mútua Conjunta. Assim foram testadas todas as variações possíveis e os dados relativos a estas simulações geraram uma massa de dados que foram classificadas dentro de uma matriz que esta mostrada abaixo com o melhor acerto dentro do número de entradas e quais foram utilizadas e também a configuração da rede neural que foi responsável por este acerto.

Tabela 4 – Percentagem de acerto na classificação de acordo com a inserção de variáveis

Camada 1	Camada 2	Camada 3	Numero de entradas	Treino	Validação	Teste
2	1	3	1	58,4%	54,5%	59,1%
2	2	1	2	67,5%	72,7%	84,1%
3	3	1	3	73,7%	71,1%	77,3%
3	3	1	4	79,9%	79,5%	81,8%
3	3	1	5	72,7%	72,7%	84,1%
3	3	1	6	84,7%	79,5%	88,6%
1	2	2	7	80,9%	86,4%	90,9%
3	1	1	8	93,8%	88,6%	93,2%
1	2	1	9	88,5%	88,6%	93,2%
1	1	1	10	85,6%	84,1%	93,2%
1	2	1	11	88,0%	86,4%	95,5%
3	2	3	12	87,6%	88,6%	93,2%
3	3	1	13	91,9%	86,4%	95,5%

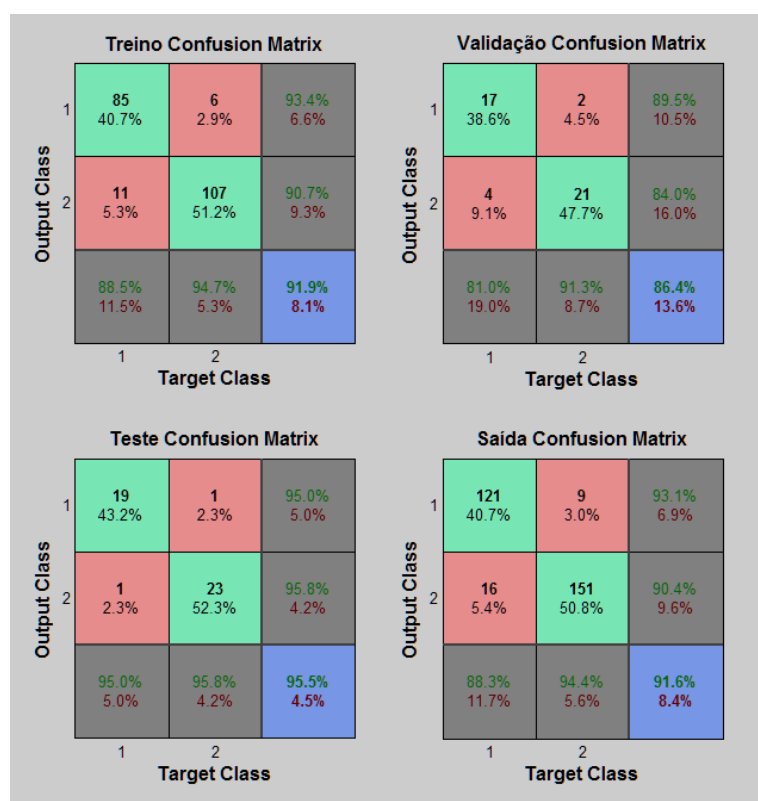


Figura 10 - Rede neural com a configuração 3 – 3 – 1 utilizando 13 entradas.

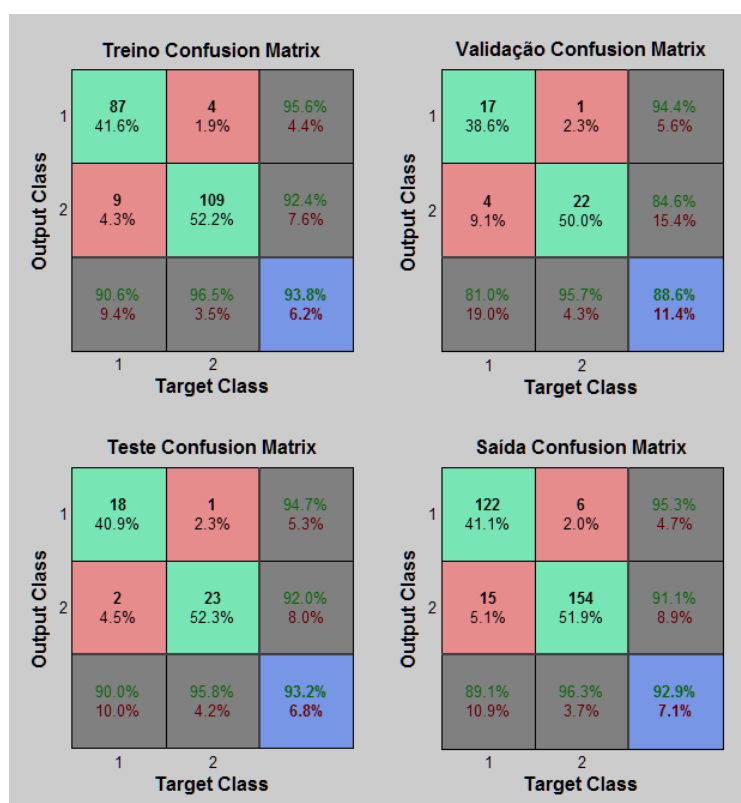


Figura 11 - neural com a configuração 3 – 1 – 1 utilizando 8 entradas.

Como a Figura 6 mostra a configuração com 3 neurônios na primeira camada, 1 neurônio na segunda e na terceira usa apenas de 8 entradas para conseguir praticamente o mesmo nível de acerto que a configuração que utiliza das 13 entradas o que valida a informação mútua conjunta como um classificador de variáveis de entrada.

Comparado a demais estudos com o mesmo banco de dados é possível perceber que Ho & Chou [9] conseguiram através a lógica Fuzzy uma taxa de acerto da ordem de 83%. Já Hu, Li, Cai & Xu [10] obtiveram uma taxa de acerto de cerca de 83,5% utilizando a “Máquina de Vetor de Suporte” todos estas taxas de acerto estão associadas ao grupo de teste. O estudo do Thiago Rodrigues [11] obteve uma taxa de acerto de cerca de 91%. Este estudo utilizando as redes neurais com as 13 características de entrada conseguiu classificar o grupo de teste com uma taxa de acerto de cerca de 96% o que é uma melhora na linha de estudos e que pode ser considerada principalmente pelo fato do algoritmo que variou as principais entradas deste sistema que são a quantidade de neurônios por camada e a quantidade de entradas sendo essas

também classificadas antes de serem modificadas. Então utilizando a técnica de *Informação Mútua Conjunta* foi observado um acerto que foi superior e se mostrou um aditivo de acerto em cada rede.

Então foi possível com este estudo criar este classificador que consegue encontrar dentro de um banco de dados apresentado a melhor configuração da rede neural além de classificar suas entradas pela ordem de importância e apresentar a rede esta ordem.

4 Conclusões

A ferramenta de seleção de variáveis se mostrou muito pertinente neste caso já que fez com que os resultados melhorassem. A melhor rede foi capaz de reconhecer até 95,5% podendo assim ser considerada como um auxílio no reconhecimento deste padrão de doença cardíaca.

Este trabalho alcançou o objetivo de contribuir para o reconhecimento de doenças cardiovasculares e estimular a implementação real do sistema para atuar em equipamentos ou como uma interface auxiliar para computadores.

Assim o trabalho segue podendo ser melhorado na busca por algum outro algoritmo que consiga identificar uma melhor combinação de variáveis para que este seja apresentado a rede. Outra melhora significativa seria o treinamento através de outros algoritmos para que a rede consiga convergir mais rápido para os pesos adequados.

Referências Bibliográficas

CASOS

- [01] Organização Mundial da Saúde. Disponível em <http://www.who.int/mediacentre/factsheets/fs317/en/>. Acesso em 22/03/2011 14
- [02] Aha, D. W., "Heart Disease Databases", link do banco de dados: <http://www.ics.uci.edu/pub/machine-learning-databases/heart-disease/heartdisease.names> (atualizado até Outubro de 2001). 25
- [03] Haykin, S. (2001). Redes Neurais, Princípios e Prática. Ed. Bookman - 2001. 17, 22
- [04] Bishop, C. M.. Neural Networks for Pattern Recognition. Oxford. Clarendon Press. 1995. 17
- [05] Principe, J. C., Euliano, N. R. and Lefebvre, W. C.. Neural and Adaptive Systems: Fundamentals Through Simulations, John Wiley. 2000. 17
- [06] Braga, A. P., Carvalho, A. C. P. L. F., Ludermir, T. B. (2000). Redes Neurais Artificiais, Teoria e Aplicações. Ed LTC - 2000. 17, 21, 22
- [07] Ludwig Jr. Costa, O. e Costa, Eduardo Montgomery M.(2007). Redes Neurais: Fundamentos e Aplicações com Programas em C. Ed.Ciência Moderna – 2007 18
- [08] Medeiros, M.C. & Pedreira, C.E.. "What are the effects of forecasting linear time series with neural networks?", Engineering Intelligent Systems, v.9, p.237- 242, 2001. 21
- [09] Ho, C.S., Chou, J.S.. Fuzzy ARTRON: A General-purpose Classifier Empowered by Fuzzy ART and Error Back-propagation Learning, Journal of Information Science and Engineering 17, 683-695, 2001 36
- [10] Hu, Z.H., Li, Y.G., Cai, Y.Z., Xu, X.M.. An Empirical Comparison of Ensemble Classification Algorithms with Support Vector Machines, Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004.. 36
- [11] Rodrigues, Thiago Baptista. Seleção de Variáveis e Classificação de Padrões por Redes Neurais como Auxílio ao Diagnóstico de Doença Cardíaca. Dezembro de 2006. 36
- [12] Cardiopatias Isquêmicas. Disponível em http://www.hc.ufpr.br/acad/clinica_medica/cardiologia/graduacao/aulas/CardiopatiasIsquemica_Fisiopatologia.ppt. Acesso em 22/03/2011. 14
- [13] Brandão, A. S., Cataldo, E. , Leta, F.R., Lucero, J.C., Usando Redes Neurais para Classificação de Padrões de voz. 2004, XXVIII CNMAC 23
- [14] J.M. Mendel and R. W. McLaren. Adaptive, Learning and Pattern Recognition Systems; Theory and Applications, Chapter Reinforcement-learning control and pattern recognition systems, pages 287-318. New York: Academic Press, 1970. 22
- [15] Gavin Brown. A New Perspective for Information Theoretic Feature Selection, School of Computer Science University of Manchester Oxford Road, UK, 2009 27
- [16] Battiti, R. Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Networks, 5(4):537–550. 1994. 28
- [17] Yang, H. and Moody, J. Data Visualization and Feature Selection: New Algorithms for Nongaussian Data. Advances in Neural Information Processing Systems, 12. 1999. 28
- [18] Cecil, Medicina Interna Básica / Editor chefe Thomas E. Andreoli – Rio de Janeiro, Elsevier 2005. 15