



Pós-Graduação em Ciência da Computação

*“Rede Neural Artificial para Análise de Fatores
Relacionados a Transtornos Mentais Comuns”*

Por

CLAUDIA RIBEIRO SANTOS LOPES

Dissertação de Mestrado



Universidade Federal de Pernambuco

posgraduacao@cin.ufpe.br

www.cin.ufpe.br/~posgraduacao

RECIFE, JANEIRO/2003



UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE INFORMÁTICA

PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CLAUDIA RIBEIRO SANTOS LOPES

“Rede Neural Artificial para Análise de Fatores Relacionados a Transtornos Mentais Comuns”

Este trabalho foi apresentado à pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientadora: Prof.^a Dr.^a Teresa Bernarda Ludermir
Co-Orientador: Prof. Dr. Marcílio Carlos P. de Souto

Recife, Janeiro / 2003

DEDICATÓRIA

Aos dois grandes amores de minha vida, João e Thais.

À minha querida mãe Carlinda (in memorian) que, mesmo não estando mais conosco, sempre esteve ao meu lado, e através de seu exemplo de força, luta e garra me ajudou a chegar até o final desta caminhada, sendo, sem sombra de dúvidas, parte desta conquista.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer à Deus, que sempre esteve presente em minha vida, e que sem ele, nada seria possível.

À minha querida Mãe *Carlinda (in memorian)*, pelo exemplo de vida que foi para mim e meus irmãos, o qual me deu forças para chegar ao final desta caminhada.

Ao meu querido companheiro de todos os momentos de minha vida, meu marido *João*, e minha querida filha *Thais*, pela compreensão, apoio e confiança em mim depositados, que fizeram com que eu tivesse condições para alcançar essa vitória.

À minha orientadora Prof^a Dr^a *Teresa Bernarda Ludermir*, por sua paciência, dedicação, competência e todo o seu apoio a mim dispensados que contribuíram de forma significativa para o meu crescimento intelectual e profissional, além da conclusão deste trabalho.

Ao meu co-orientador Prof. Dr. *Marcílio Carlos P. de Souto* pela sua colaboração para a conclusão deste trabalho.

Ao meu amigo *Akio Yamazaki* pela sua importante contribuição, além da amizade, que contribuiu de forma definitiva para a conclusão deste trabalho.

À Prof^a. Dr^a *Ana Bernarda Ludermir* por ter disponibilizado a base de dados que utilizamos em nossos experimentos, e por estar sempre disponível em todos os momentos que precisamos durante a realização deste trabalho.

Aos meus queridos irmãos, *Sônia, Sonilda, Vera, Carlúcia, Milton e Eduardo* por terem sempre acreditado em mim, ajudando-me em todos os momentos.

À todos os meus familiares e amigos que me deram apoio nos momentos mais difíceis que atravessei durante o curso.

À minhas amigas e companheiras *Valéria, Maísa e Isledna*, que compartilharam comigo tantos momentos alegres e difíceis vivenciados durante esse curso.

A todos os colegas do Mestrado, e em especial à *Cátia, José Carlos, Silvia e Francisco Carvalho*.

Ao Prof. Dr. *Benedito Acioly* e *Corina*, pelo seu importante apoio durante a realização do curso.

Aos colegas do Departamento de Química e Exatas da Universidade Estadual do Sudoeste da Bahia - UESB, e em especial à *Cândido, Telessom e Derval* pelo importante apoio e colaboração a mim dispensados.

Ao Prof. Dr. Alúizio F. R. Araújo e Prof. Dr. Alan Kardeck de Barros pela valiosa contribuição prestada.

À todos que, direta ou indiretamente, acreditaram na minha capacidade de realizar este estudo e contribuíram para a sua conclusão, **MUITO OBRIGADA!**

RESUMO

Os Transtornos Mentais Comuns (TMC's), e dentre eles **ansiedade e depressão**, têm sido causas comuns de morbidade tanto em países desenvolvidos como em desenvolvimento. Esses transtornos levam a um considerável uso dos serviços de saúde e destacam-se como um importante fator que leva a uma perda de dias de trabalho. Além disso, causam um sério impacto tanto para os serviços de saúde como à sociedade de um modo geral. O uso de técnicas que possam vir a conduzir à identificação dos fatores que apresentam maior probabilidade de estarem relacionados aos TMC's é de grande relevância para auxiliar no processo de tomada de decisões acerca do planejamento e intervenção de saúde pública nessa área.

As Redes Neurais Artificiais (RNA's) são conhecidas porque geralmente obtêm um bom desempenho na precisão dos resultados, e tem sido amplamente utilizadas para prognóstico e diagnóstico de doenças. A aplicação de RNA's para análise dos fatores relacionados aos TMC's foi motivada principalmente pelo fato que em RNA's não há necessidade de independência entre as diversas variáveis estudadas, pela sua habilidade em detectar todas as possíveis interações entre as diversas variáveis explicativas, bem como por sua capacidade de aprendizado a partir do ambiente. Essas características têm sido destacadas como um atrativo para a aplicação de RNA's em estudos epidemiológicos.

Técnicas clássicas de estatística, a exemplo de regressão logística, geralmente são mais utilizadas em estudos epidemiológicos, porém, RNA's vêm surgindo como uma alternativa interessante, principalmente em situações onde são exibidas variáveis dependentes e independentes complexas com relações não lineares. Esse trabalho teve como objetivo principal utilizar RNA para análise dos fatores relacionados a TMC's, comparando os resultados obtidos nos experimentos realizados com os alcançados com o modelo estatístico de regressão logística. Para o desenvolvimento dos experimentos foi utilizada uma rede Multilayer Perceptron com uma camada escondida, treinada com o algoritmo Backpropagation e com o método de otimização Simulated Annealing.

Palavras-chave: Transtornos Mentais Comuns, Estudos Epidemiológicos, Redes Neurais Artificiais, Multilayer Perceptron, Backpropagation e Simulated Annealing.

ABSTRACT

The Common Mental Disorders (CMDs), like anxiety and depression, are the common causes of morbidity in the developed countries as well as in the developing countries. The mental disorders demand a considerable use of the health services and they stand out as an important factor associate to loss of days of work. Beside this, they cause a serious impact for the health services as well as for the society in a general way. The use of techniques for identification of the most important factors related to CMDs is of great relevance for decision making process on planning and intervention of the public health service in this area.

The Artificial Neural Networks (ANNs) are known by their good performance in the precision of the results, and they have been used for the prognostic and diagnosis of diseases. The application of ANNs for analysis of the factors related to CMDs was mainly motivated by the fact that ANNs do not require independence of the studied variables, and by their ability to detect all the possible interactions of the explanatory variables and by the ANN ability to learn from environment. These characteristics have been outstanding as an attractive point for the application of ANNs in epidemiologic studies.

Classic techniques of statistics, like as logistic regression, are usually used in epidemiologic studies. However, ANNs are appearing as an interesting alternative, mainly in situations where dependent and independent variables are exhibited with nonlinear relationships. This work had as main objective the use of ANN for analysis of the factors related to CMDs, comparing the results obtained in the ANN experiments with that obtained by using the statistical model of logistic regression. For the development of the experiments, it was used a Multilayer Perceptron with a hidden layer, trained with the Backpropagation algorithm and the Simulated Annealing method.

Key words: Common Mental Disorders, Epidemiologic Studies, Artificial Neural Networks, Multilayer Perceptron, Backpropagation and Simulated Annealing.

CONTEÚDO

RESUMO.....	vi
ABSTRACT.....	vii
ÍNDICE DE FIGURAS.....	x
ÍNDICE DE TABELAS.....	xiv
CAPÍTULO 1 - INTRODUÇÃO.....	1
1.1 MOTIVAÇÃO.....	2
1.2 OBJETIVOS.....	3
1.3 ORGANIZAÇÃO DA DISSERTAÇÃO.....	4
CAPÍTULO 2 - REVISÃO DA LITERATURA.....	5
2.1 INTRODUÇÃO.....	6
2.2 REDES NEURAS ARTIFICIAIS.....	6
2.3 SIMULATED ANNEALING PARA O PROJETO DE REDES NEURAS.....	9
2.4 REGRESSÃO LOGÍSTICA.....	12
2.5 ALGUMAS SIMILARIDADES ENTRE REDES NEURAS ARTIFICIAIS E REGRESSÃO LOGÍSTICA.....	14
2.6 ALGUMAS VANTAGENS NO USO DE REDES NEURAS ARTIFICIAIS.....	15
2.7 ALGUMAS DESVANTAGENS NO USO DE REDES NEURAS ARTIFICIAIS.....	17
2.8 CONSIDERAÇÕES FINAIS.....	17
CAPÍTULO 3 - MATERIAIS E MÉTODOS.....	19
3.1 INTRODUÇÃO.....	20
3.2 DESCRIÇÃO DO PROBLEMA E BASE DE DADOS.....	21
3.3 EXPERIMENTOS COM MULTILAYER PERCEPTRON.....	23
3.3.1 <i>Divisão do Conjunto de Padrões em Treinamento, Validação e Teste</i>	23
3.3.2 <i>Pré-processamento dos Dados</i>	23
3.3.3 <i>Topologias das Redes Neurais</i>	24
3.3.4 <i>Metodologia de Treinamento</i>	25
3.3.4.1 <i>Metodologia dos Treinamentos com o Algoritmo Backpropagation</i>	25
3.3.4.2 <i>Metodologia dos Treinamentos com o Algoritmo Simulated Annealing</i>	26
3.3.5 <i>Aspectos Observados</i>	27
3.4 O TESTE DE HIPÓTESE <i>T-STUDENT</i>	28
3.5 CONSIDERAÇÕES FINAIS.....	29
CAPÍTULO 4 - EXPERIMENTOS COM MULTILAYER PERCEPTRON E O ALGORITMO BACKPROPAGATION.....	30
4.1 INTRODUÇÃO.....	31
4.2 ANÁLISE INDIVIDUAL DE CADA PARTIÇÃO.....	31
4.2.1 <i>Experimentos realizados com a partição 01</i>	32
4.2.1.1 <i>Utilizando todas as variáveis do conjunto de dados</i>	32
4.2.1.2 <i>Retirando variáveis do conjunto de dados</i>	33
4.2.2 <i>Experimentos realizados com a partição 02</i>	45
4.2.2.1 <i>Utilizando todas as variáveis do conjunto de dados</i>	46
4.2.2.2 <i>Retirando variáveis do conjunto de dados</i>	46
4.3 ANÁLISE OBSERVANDO A MÉDIA OBTIDA COM AS DUAS PARTIÇÕES.....	59
4.3.1 <i>Utilizando todas as variáveis do conjunto de dados</i>	60
4.3.2 <i>Retirando variáveis do conjunto de dados</i>	60
4.4 CONCLUSÕES.....	72

CAPÍTULO 5 - EXPERIMENTOS COM MULTILAYER PERCEPTRON E O ALGORITMO SIMULATED ANNEALING	77
5.1 INTRODUÇÃO	78
5.2 ANÁLISE INDIVIDUAL DE CADA PARTIÇÃO.....	79
5.2.1 <i>Experimentos realizados com a partição 01</i>	79
5.2.1.1 Utilizando todas as variáveis do conjunto de dados.....	79
5.2.1.2 Utilizando as variáveis que mais se destacaram	81
5.2.2 <i>Experimentos realizados com a partição 02</i>	82
5.2.2.1 Utilizando todas as variáveis do conjunto de dados.....	83
5.2.2.2 Utilizando as variáveis que mais se destacaram	84
5.3 ANÁLISE OBSERVANDO A MÉDIA OBTIDA COM AS DUAS PARTIÇÕES	85
5.3.1 <i>Utilizando todas as variáveis do conjunto de dados</i>	86
5.3.2 <i>Utilizando as variáveis que mais se destacaram</i>	87
5.4 CONCLUSÕES.....	88
CAPÍTULO 6 - CONCLUSÕES	90
6.1 INTRODUÇÃO	91
6.2 COMPARAÇÃO DE RESULTADOS OBTIDOS	91
6.3 CONSIDERAÇÕES FINAIS	93
6.4 PRINCIPAIS CONTRIBUIÇÕES.....	94
6.5 TRABALHOS FUTUROS	95
REFERÊNCIAS BIBLIOGRÁFICAS	97

ÍNDICE DE FIGURAS

<i>Figura 2-1 Estrutura básica do algoritmo Simulated Annesling, adaptado de [Yamazaki et al, 2002]</i>	11
<i>Figura 4-1 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Idade (Partição 01)</i>	34
<i>Figura 4-2 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Alfabetização (Partição 01)</i>	35
<i>Figura 4-3 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Migração (Partição 01)</i>	36
<i>Figura 4-4 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Escolaridade (Partição 01)</i>	37
<i>Figura 4-5 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Propriedade da Casa (Partição 01)</i>	38
<i>Figura 4-6 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Inserção no Processo Produtivo (Partição 01)</i>	39
<i>Figura 4-7 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Condições de Moradia (Partição 01)</i>	40
<i>Figura 4-8 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Sexo (Partição 01)</i>	42
<i>Figura 4-9 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Situação Conjugal (Partição 01)</i>	43
<i>Figura 4-10 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Renda (Partição 01)</i>	44

<i>Figura 4-11 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Posse de Bens Duráveis (Partição 01)</i>	45
<i>Figura 4-12 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Idade (Partição 02)</i>	47
<i>Figura 4-13 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Alfabetização (Partição 02)</i>	48
<i>Figura 4-14 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Migração (Partição 02)</i>	49
<i>Figura 4-15 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Escolaridade (Partição 02)</i>	50
<i>Figura 4-16 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Propriedade da Casa (Partição 02)</i>	51
<i>Figura 4-17 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Inserção no Processo Produtivo (Partição 02)</i>	53
<i>Figura 4-18 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Condições de Moradia (Partição 02)</i>	54
<i>Figura 4-19 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Sexo (Partição 02)</i>	55
<i>Figura 4-20 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Situação Conjugal (Partição 02)</i>	56
<i>Figura 4-21 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Renda (Partição 02)</i>	57
<i>Figura 4-22 Comparação de resultados dos experimentos com MLP realizados utilizando</i>	

<i>todas as variáveis do conjunto e experimentos retirando a variável Posse de Bens Duráveis (Partição 02)</i>	58
<i>Figura 4-23 Comparação de resultados dos experimentos realizados com MLP e todas as variáveis do conjunto e experimentos retirando a variável Idade. Resultados observando a média das partições.</i>	61
<i>Figura 4-24 Comparação de resultados dos experimentos realizados com MLP e todas as variáveis do conjunto e experimentos retirando a variável Alfabetização. Resultados observando a média das partições.</i>	62
<i>Figura 4-25 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Migração. Resultados observando a média das partições.</i>	63
<i>Figura 4-26 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Escolaridade. Resultados observando a média das partições.</i>	64
<i>Figura 4-27 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Propriedade da Casa. Resultados observando a média das partições.</i>	65
<i>Figura 4-28 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Inserção no Processo Produtivo. Resultados observando a média das partições.</i>	66
<i>Figura 4-29 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Condições de Moradia. Resultados observando a média das partições.</i>	67
<i>Figura 4-30 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Sexo. Resultados observando a média das partições.</i>	68
<i>Figura 4-31 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Situação Conjugal. Resultados observando a média das partições.</i>	69
<i>Figura 4-32 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Renda. Resultados</i>	

<i>observando a média das partições.</i>	70
<i>Figura 4-33 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Posse de Bens Duráveis. Resultados observando a média das partições.</i>	71
<i>Figura 4-34 Curva do percentual de classificação correta dos casos com diagnósticos positivo e negativo nos experimentos utilizando todas as variáveis do conjunto e retirando variáveis (observando a médias entre as duas partições)</i>	74
<i>Figura 4-35 Curva do erro de classificação da rede nos experimentos utilizando todas as variáveis do conjunto e retirando variáveis (observando a médias entre as duas partições)</i>	74
<i>Figura 5-1 Percentual de utilização de cada variável do conjunto. Experimentos com todas as variáveis utilizando o simulated annealing (Partição 01)</i>	80
<i>Figura 5-2 Percentual de utilização de cada variável do conjunto. Experimentos com as sete variáveis principais, utilizando o simulated annealing (Partição 01)</i>	81
<i>Figura 5-3 Comparação de resultados dos experimentos realizados com o algoritmo simulated annealing utilizando todas as variáveis do conjunto e com as sete variáveis principais (Partição 01)</i>	82
<i>Figura 5-4 Percentual de utilização de cada variável do conjunto. Experimentos com todas as variáveis utilizando o simulated annealing (Partição 01)</i>	83
<i>Figura 5-5 Percentual de utilização de cada variável do conjunto. Experimentos com as sete variáveis principais, utilizando o simulated annealing (Partição 02)</i>	84
<i>Figura 5-6 Comparação de resultados dos experimentos realizados com o algoritmo simulated annealing utilizando todas as variáveis do conjunto e com as sete variáveis principais (Partição 02)</i>	85
<i>Figura 5-7 Percentual de utilização de cada variável do conjunto. Experimentos com todas as variáveis utilizando o simulated annealing, observando a médias entre as duas partições.</i>	86
<i>Figura 5-8 Percentual de utilização das variáveis que se destacaram nos experimentos com as duas partições (utilizando as sete principais como entrada em cada partição)</i>	87
<i>Figura 6-1 Resultados obtidos com redes neurais quanto ao percentual de classificação correta dos casos com diagnósticos positivo e negativo para os TMC's.</i>	92

ÍNDICE DE TABELAS

<i>Tabela 2-1 Termos similares entre Redes Neurais Artificiais e Estatística.....</i>	15
<i>Tabela 2-2 Algumas vantagens e desvantagens em utilizar uma rede neural artificial para a predição de resultados</i>	16
<i>Tabela 3-1 Variáveis que compõem o conjunto de dados.....</i>	22
<i>Tabela 3-2 Codificação das variáveis do conjunto de dados.....</i>	24
<i>Tabela 4-1 Resultados com MLP e todas as variáveis do conjunto (partição 01).....</i>	32
<i>Tabela 4-2 Resultados retirando a variável Idade (Partição 01)</i>	34
<i>Tabela 4-3 Resultados retirando a variável Alfabetização (Partição 01).....</i>	35
<i>Tabela 4-4 Resultados retirando a variável Migração (Partição 01).....</i>	36
<i>Tabela 4-5 Resultados retirando a variável Escolaridade (Partição 01).....</i>	37
<i>Tabela 4-6 Resultados retirando a variável Propriedade da Casa (Partição 01).....</i>	38
<i>Tabela 4-7 Resultados retirando a variável Inserção no Processo Produtivo (Partição 01)..</i>	39
<i>Tabela 4-8 Resultados retirando a variável Condições de Moradia (Partição 01).....</i>	40
<i>Tabela 4-9 Resultados retirando a variável Sexo (Partição 01).....</i>	41
<i>Tabela 4-10 Resultados retirando a variável Situação Conjugal (Partição 01).....</i>	42
<i>Tabela 4-11 Resultados retirando a variável Renda (Partição 01)</i>	44
<i>Tabela 4-12 Resultados retirando a variável Posse de Bens Duráveis (Partição 01)</i>	45
<i>Tabela 4-13 Resultados com uma MLP e todas as variáveis do conjunto (partição 02)</i>	46
<i>Tabela 4-14 Resultados retirando a variável Idade (Partição 02).....</i>	47
<i>Tabela 4-15 Resultados retirando a variável Alfabetização (Partição 02).....</i>	48
<i>Tabela 4-16 Resultados retirando a variável Migração (Partição 02).....</i>	49
<i>Tabela 4-17 Resultados retirando a variável Escolaridade (Partição 02).....</i>	50
<i>Tabela 4-18 Resultados retirando a variável Propriedade da Casa (Partição 02).....</i>	52
<i>Tabela 4-19 Resultados retirando a variável Inserção no Processo Produtivo (Partição 02)</i>	52

<i>Tabela 4-20 Resultados retirando a variável Condições de Moradia (Partição 02)</i>	54
<i>Tabela 4-21 Resultados retirando a variável Sexo (Partição 02)</i>	55
<i>Tabela 4-22 Resultados retirando a variável Situação Conjugal (Partição 02)</i>	56
<i>Tabela 4-23 Resultados retirando a variável Renda (Partição 02)</i>	57
<i>Tabela 4-24 Resultados retirando a variável Posse de Bens Duráveis (Partição 02)</i>	58
<i>Tabela 4-25 Quadro comparativo das diferenças obtidas nos experimentos realizados retirando variáveis (partições 01 e 02), com seus respectivos valores de p. Os resultados são equivalentes à comparação feita com experimentos utilizando todas as variáveis do conjunto. O sinal “+” significa que houve um aumento no percentual e o sinal “-” que houve uma redução.</i>	59
<i>Tabela 4-26 Resultados com MLP utilizando todas as variáveis do conjunto (média das duas partições)</i>	60
<i>Tabela 4-27 Resultados retirando a variável Idade do conjunto de dados, observando as duas partições conjuntamente</i>	61
<i>Tabela 4-28 Resultados retirando a variável Alfabetização do conjunto de dados, observando a média entre as duas partições</i>	62
<i>Tabela 4-29 Resultados retirando a variável Migração do conjunto de dados, observando a média entre as duas partições</i>	63
<i>Tabela 4-30 Resultados retirando a variável Escolaridade do conjunto de dados, observando a média entre as duas partições</i>	64
<i>Tabela 4-31 Resultados retirando a variável Propriedade da Casa do conjunto de dados, observando a média entre as duas partições</i>	65
<i>Tabela 4-32 Resultados retirando a variável Inserção no Processo Produtivo do conjunto de dados, observando a média entre as duas partições</i>	66
<i>Tabela 4-33 Resultados retirando a variável Condições de Moradia do conjunto de dados, observando a média entre as duas partições</i>	67
<i>Tabela 4-34 Resultados retirando a variável Sexo do conjunto de dados, observando a média entre as duas partições</i>	68
<i>Tabela 4-35 Resultados retirando a variável Situação Conjugal do conjunto de dados,</i>	

<i>observando a média entre as duas partições.....</i>	<i>69</i>
<i>Tabela 4-36 Resultados retirando a variável Renda do conjunto de dados, observando a média entre as duas partições.</i>	<i>70</i>
<i>Tabela 4-37 Resultados retirando a variável Renda do conjunto de dados, observando a média entre as duas partições.</i>	<i>71</i>
<i>Tabela 4-38 Quadro comparativo das diferenças apresentadas nos experimentos retirando variáveis do conjunto, observando a média entre as duas partições, e seus respectivos valores de p. O sinal “+” significa que houve um aumento no percentual e o sinal “-” que houve uma redução.</i>	<i>72</i>
<i>Tabela 4-39 Resumo dos resultados obtidos (diferenças) quanto ao erro de classificação, percentual de classificação correta dos casos positivos e percentual de classificação correta dos casos negativos nos experimentos retirando variáveis do conjunto de dados. O sinal “+” significa que houve um aumento do percentual com relação aos resultados obtidos com todas as variáveis, e o sinal “-” significa que houve uma diminuição do percentual.....</i>	<i>75</i>
<i>Tabela 5-1 Valor médio dos resultados obtidos nos experimentos com simulated annealing (partição 01, com onze variáveis).....</i>	<i>80</i>
<i>Tabela 5-2 Valor médio dos resultados obtidos nos experimentos com simulated annealing (Partição 01, com sete variáveis no conjunto de entrada).....</i>	<i>81</i>
<i>Tabela 5-3 Valor médio dos resultados obtidos nos experimentos com simulated annealing (Partição 02 com onze variáveis)</i>	<i>83</i>
<i>Tabela 5-4 Valor médio dos resultados obtidos nos experimentos com simulated annealing (Partição 02 com sete variáveis no conjunto de entrada).....</i>	<i>84</i>
<i>Tabela 5-5 Resultados obtidos nos experimentos com simulated annealing , observando a média das duas partições (com todas as variáveis no conjunto de entrada).....</i>	<i>87</i>
<i>Tabela 5-6 Resultados obtidos nos experimentos com simulated annealing , observando a média das duas partições (com as sete variáveis que se destacaram)</i>	<i>87</i>
<i>Tabela 6-1 Classificação correta de casos com diagnósticos positivo e negativo para os TMC´s obtidos no modelos neural.....</i>	<i>92</i>

Capítulo 1 - Introdução

1.1 Motivação

Segundo [Ludermir, 2000], Transtornos Mentais Comuns (TMC's) é uma expressão criada por Goldenberg & Huxley [1992] para descrever os "*transtornos comumente encontrados na população e que sinalizam uma interrupção do funcionamento normal*". Esses transtornos consistem em sintomas como insônia, fadiga, irritabilidade, esquecimento, dificuldade de concentração e queixas somáticas".

Os Transtornos Mentais Comuns (TMC's), e dentre eles **ansiedade** e **depressão**, são causas comuns de morbidade tanto em países desenvolvidos quanto naqueles em desenvolvimento como o Brasil. Esses transtornos representam um alto custo social e econômico pois são incapacitantes, constituem causa importante de dias perdidos de trabalho, além de elevarem a demanda nos serviços de saúde. [Ludermir, 1998]

O uso de técnicas que possam vir a conduzir a uma identificação dos fatores que apresentam maior probabilidade de estarem relacionados aos TMC's é de grande relevância para auxiliar no processo de tomada de decisões acerca do planejamento e intervenção de saúde pública preventiva e curativa nessa área.

Utilizando o modelo estatístico de regressão logística, Ludermir em 1998 realizou um estudo para identificar as possíveis relações entre condições de vida e trabalho com os TMC's.

As Redes Neurais Artificiais (RNA's), entretanto, vêm surgindo como uma alternativa interessante ao uso de técnicas clássicas de estatística, a exemplo de regressão logística, principalmente em situações onde são exibidas variáveis dependentes e independentes complexas, com relações não lineares [Leung and Tran, 1999]. Porém, ainda existem dúvidas se as redes neurais podem realmente trazer algum progresso nesse campo [Schwarzer et al, 2000].

O método neural é conhecido por, geralmente, obter um bom desempenho na precisão dos resultados em experimentos diversos, tendo sido amplamente utilizado para o prognóstico e diagnóstico de doenças [Chtioui et al, 1999][Legoff et al, 2000] [Leung et al, 2000][Marble, 1999][Naguib et al, 1998] [Nair, 1999] [Pendharkar et al, 1999] [Schwarzer et al,2000][Wyman, 1996]. Em estudos epidemiológicos, as redes neurais artificiais também têm tido aceitação [Lamers et al, 1997] [Nair et al, 1999]

[dos Santos et al, 2001].

[Lamers et al, 1997], por exemplo, com base no efeito epidemiológico de heterogeneidade em respostas, propôs uma nova estratégia de modelagem utilizando RNA's, a qual visou a redução de correlações entre covariáveis para resultar em uma melhor modelagem.

Uma característica que, certamente, tem tornado o uso de redes neurais muito popular, quando comparado à técnica estatística de regressão logística, é que não existe necessidade de independência entre as diversas variáveis estudadas. No modelo logístico, se a suposição da independência entre as variáveis estudadas for violada, pode ocorrer distorções no modelo [Tu, 1996] [Leung, 2000].

Além disso, a rede neural possui uma capacidade de aprendizado a partir do ambiente, ou seja, é capaz de aprender as informações a partir de um conjunto de dados de entrada (conjunto de treinamento) e de generalizar a informação aprendida, reconhecendo padrões que não estavam presentes no conjunto de treinamento.

Outra característica interessante, a qual também nos motivou para a realização desse trabalho e vem sendo apontada como uma das vantagens para o uso de RNA's, é exatamente a sua habilidade em detectar todas as possíveis interações entre as diversas variáveis explicativas. [Tu, 1996]

Considerando que os TMC's têm sido identificados como um significativo problema de saúde pública, a realização de experimentos com redes RNA's para a análise dos fatores relacionados aos TMC's apresenta-se como algo bastante interessante.

1.2 Objetivos

O objetivo principal deste trabalho é analisar os fatores relacionados a TMC's com o uso de RNA's. Em seguida, os resultados obtidos nos experimentos com RNA's serão comparados com aqueles apresentados por [Ludermir, 1998] a partir da análise feita com regressão logística. Com isso, para o alcance desse objetivo, iremos:

- Fazer um estudo sobre as aplicações das RNA's, enfatizando a área de saúde;

- Realizar experimentos práticos utilizando RNA's com a base de dados gerada por [Ludermir, 1998];
- Verificar a viabilidade prática do uso das RNA's para a análise dos fatores relacionados ao problema estudado, a partir dos resultados apresentados;
- Apresentar algumas similaridades conceituais entre RNA's e o modelo estatístico de regressão logística;
- Comparar os resultados obtidos nos experimentos utilizando RNA's com aqueles apresentados por [Ludermir, 1998] utilizando o modelo de regressão logística e a mesma base de dados.

1.3 Organização da Dissertação

Neste Capítulo, são apresentados as motivações e objetivos que levaram à realização deste trabalho, além da organização desta dissertação. No Capítulo 2 é apresentada uma revisão da literatura, com algumas considerações a respeito de RNA's, mostrando algumas aplicações na saúde, além de uma visão geral sobre o uso do *simulated annealing* [Yamazaki et al, 2002] para o projeto de RNA's.

No Capítulo 2 também é dada uma visão geral sobre o modelo de regressão logística, apresentando algumas similaridades conceituais entre o método neural e o estatístico. Além disso, são destacadas algumas vantagens e desvantagens no uso das duas técnicas.

Os materiais e métodos utilizados para a realização dos experimentos, com uma descrição da base de dados, são apresentados no Capítulo 3. No Capítulo 4 são apresentados os resultados obtidos nos experimentos realizados com uma rede MLP e o algoritmo *backpropagation* [Braga et al, 2000]

Os resultados obtidos com os experimentos realizados com uma rede MLP treinada com o algoritmo *simulated annealing* serão apresentados no Capítulo 5. E, finalmente no Capítulo 6, são apresentadas as considerações finais acerca do trabalho desenvolvido, além de uma comparação dos resultados obtidos com os modelos neural e estatístico. Nesse Capítulo são apresentadas também as principais contribuições e perspectivas de trabalhos futuros.

Capítulo 2 - Revisão da Literatura

2.1 Introdução

Neste Capítulo será apresentada uma revisão da literatura, com algumas considerações a respeito de RNA's, mostrando algumas aplicações na saúde, além de uma visão geral sobre o uso do *simulated annealing* para o projeto de RNA's.

Também é apresentado uma visão geral sobre o modelo de regressão logística, apresentando algumas similaridades conceituais entre o método neural e o estatístico. Além disso, são destacadas algumas vantagens e desvantagens no uso das duas técnicas.

2.2 Redes Neurais Artificiais

Rede Neural Artificial é um modelo de computação não algorítmica, que tem como característica básica sistemas paralelamente distribuídos, compostos por unidades de processamento bastante simples, os quais são chamados de *nodos* [Braga et al, 2000][Haykin, 2001].

As redes neurais artificiais foram inspiradas na estrutura do cérebro humano, no qual o processamento das informações é altamente complexo, paralelo e não-linear. Essa semelhança entre a rede neural e o cérebro humano é destacado por [Haykin, 2001] sobre dois aspectos: a partir do ambiente através de um processo de aprendizagem, onde a rede adquire o conhecimento; o conhecimento adquirido é armazenado pela rede por meio da força de conexão entre os neurônios, conhecidos como pesos sinápticos.

Aplicações de redes neurais podem ser vistas em campos diversos, tais como predição de doenças em culturas de camarões [Leung and Tran, 1999], dados epidemiológicos de hepatite A [dos Santos et al, 2001], doenças de plantas [Chtioui et al, 1999], reconhecimento de imagens [Karkanis et al, 2000][Silipo and Marchesi, 1998][Wyman et al, 1996], classificação cromossômica de camundongos [Musavi et al, 1998], entre outras.

Um dos atrativos da rede neural que leva a essa diversidade de aplicações é, sem dúvidas, a sua capacidade de *aprender* através de exemplos que lhe são apresentados como dados de entrada, e de *generalizar* a informação aprendida.

Em geral, isso significa que a rede é capaz de *aprender* uma informação com um número reduzido de exemplos (*treinamento*) e posteriormente fazer o reconhecimento de padrões desconhecidos (*teste*) [Braga et al, 2000][Haykin, 2001].

Para a realização do processo de aprendizagem faz-se necessário à utilização de um algoritmo denominado “*algoritmo de aprendizagem*”. Esse algoritmo possui a função de modificar os pesos sinápticos da rede de forma ordenada, visando o alcance do objetivo pretendido com o uso da rede neural.

Nesta fase de aprendizagem é apresentado à rede um conjunto de exemplos (*padrões*) do qual ela extrai as características necessárias para a representação da informação que foi fornecida. Outra característica muito importante é o paralelismo inerente à arquitetura de uma rede neural, possibilitando uma maior rapidez no processamento dos dados.

Os nodos de uma rede são ativados por uma função (*função de ativação*) que serve para produzir o valor de saída de um neurônio, podendo essa saída ser contínua ou discreta. Dentre as funções que são utilizadas para ativação em uma rede neural, as que merecem maior destaque são: *sigmoïdal* e *tangente hiperbólica*. Maiores detalhes sobre Redes Neurais Artificiais podem ser encontrados facilmente na literatura [Bishop, 1996] [Braga et al, 2000] [Haykin, 2001] [Kovacs, 1996] [Pham and Karaboga, 2000].

O *Multilayer Perceptron* (MLP), como modelo mais tradicional de rede neural, tem sido amplamente utilizado em problemas diversos na área de saúde, a exemplo de diagnóstico médico e dados epidemiológicos. Além disso, a rede MLP também tem sido utilizada em vários trabalhos com o objetivo de comparar o método neural com métodos estatísticos tradicionais [Tu, 1996].

Como exemplo de aplicação de uma MLP no diagnóstico de doenças podemos citar [Tang et al, 1997] que utilizou uma rede neural MLP para o diagnóstico de doenças nas artérias coronárias com base nos dados pessoais dos pacientes, tais como idade e sexo, na descrição subjetiva do paciente e nos resultados de vários exames (pressão sanguínea e eletrocardiograma).

Em [Tang et al, 1997] chegou-se à conclusão que, com relação ao problema então estudado, a rede neural artificial apresentou-se como uma alternativa potencial para predição quando comparado a técnicas estatísticas.

[Ronco, 1999], por sua vez, utilizou uma rede neural MLP para melhorar o diagnóstico ultrasonográfico de câncer de próstata, onde obteve, com o modelo neural, uma sensibilidade de 81,82% e especificidade de 96,95%. Esses resultados foram melhores dos que os obtidos com o modelo logístico, onde fora observado uma sensibilidade de 67,18% e especificidade de 90,97%.

A sensibilidade de um modelo nos dirá o quanto o modelo está classificando corretamente os indivíduos com diagnóstico positivo para o problema estudado, enquanto que a especificidade refere-se à classificação correta dos indivíduos que apresentam diagnóstico negativo [dos Santos, 2001].

Segundo [dos Santos, 2001], quanto mais sensível for o modelo, quer dizer, quanto maior for a probabilidade do modelo classificar corretamente os casos positivos (indivíduos portadores da doença), maior será a segurança do especialista de que um paciente, com resultado negativo, não tenha a doença. Entretanto, quanto mais específico for o modelo, maior será a segurança do especialista de que um resultado positivo confirma ou mantém o diagnóstico procurado.

Uma rede MLP também foi utilizada por [Ronco, 1999] para modelar associações existentes entre os fatores discriminantes de câncer de mama com a presença ou ausência da doença. A rede, nesse caso, obteve uma sensibilidade de 94,04% e especificidade de 97,60%.

Os resultados alcançados, em [Ronco, 1999], foram comparados com uma análise feita, utilizando a mesma base de dados, através do modelo estatístico de regressão logística, o qual apresentou uma sensibilidade de 77,46% e especificidade de 68,81%. Ou seja, a rede neural nos experimentos realizados com o problema estudado, destacou-se como uma ferramenta interessante para identificar mulheres que apresentavam alto risco para câncer de mama.

[Nair, 1999] realizou experimentos com uma rede neural artificial MLP para mapear as relações de sintomas associados à depressão, definindo o grau de importância do sintoma para o diagnóstico apresentado.

Em um estudo realizado por [Reid et. al. 1996] sobre desordem de pânico, também foi aplicado uma rede MLP para análise dos fatores mais relacionados com o problema. [dos Santos et al, 2001] aplicou uma rede MLP para dados epidemiológicos de Hepatite A.

[Tu, 1996], comparou o modelo neural com o modelo estatístico de regressão logística para predição de resultados médicos, apontando as vantagens e desvantagens no uso dos dois métodos. [Ottenbacher et al, 2001], entretanto, comparou os dois modelos, o neural e o logístico, em experimentos realizados para predição de rehospitalização de pacientes. Outros pesquisadores também utilizaram uma rede MLP aplicada à área médica, como por exemplo, para práticas urológicas [Wei et al. 1998], entre outros.

2.3 Simulated Annealing para o projeto de Redes Neurais

A definição da arquitetura de uma Rede Neural Artificial é de grande importância para o sucesso das aplicações propostas, podendo causar um impacto significativo na capacidade de processamento da rede.

Com isso, dada uma tarefa de aprendizagem, uma rede neural com poucas conexões pode ter uma capacidade de aprendizagem limitada. Mas, por outro lado, um número grande de conexões pode provocar *overfitting*, causando falha na capacidade de generalização [Yamazaki et al, 2002].

Esse dilema da definição da arquitetura “ótima” a ser aplicada em experimentos com redes neurais, buscando uma rede com complexidade mínima e que tenha uma boa capacidade de generalização, pode ser amenizado com a aplicação de algoritmos de otimização no seu treinamento.

Um exemplo disso pode ser visto em [Yamazaki et al, 2002] que fez uso do algoritmo *simulated annealing* para a otimização simultânea da arquitetura (número de nodos e conexões) além dos pesos de uma rede neural aplicada ao reconhecimento de odores em um nariz artificial, onde obteve uma rede com complexidade mínima (quantidade pequena de nodos ocultos e conexões) e com um erro de classificação inferior aos resultados obtidos nos experimentos realizados com a mesma base de dados, com a aplicação de outros métodos. Contudo, segundo [Yamazaki et al, 2002] a utilização de algoritmos que vise a otimização simultânea da arquitetura e pesos de uma RNA não tem sido muito popular.

O método de otimização *simulated annealing* é uma técnica de busca não-determinística. Esse método gera uma seqüência de soluções candidatas para o problema, tendo como objetivo a convergência probabilística à candidatos (vizinhos),

os quais otimizam a função objetivo.

O *simulated annealing* é uma ferramenta poderosa para problemas de otimização, usada em casos de grande complexidade [Cuenca et al, 2001]. Esse método procura minimizar a incidência de mínimos locais, podendo aceitar vizinhos piores (maior custo) com uma dada probabilidade. Essa probabilidade é diminuída ao longo das iterações, e o parâmetro utilizado para o controle dessa probabilidade é chamado de temperatura.

O algoritmo *Simulated Annealing* foi proposto por Kirkpatrick et al. (1983) [Pham and Karaboga, 2000] e consiste na seqüência de iterações, e, cada iteração por sua vez, da mudança aleatória da solução corrente para criar uma nova solução na sua vizinhança.

A vizinhança é definida com a escolha do mecanismo de geração. Uma vez que a nova solução é criada, a mudança correspondente na função de custo é computada para decidir se a solução recentemente produzida pode ser aceita como a solução atual. Se a mudança na função de custo for negativa, então a solução recentemente produzida é diretamente levada como solução corrente [Pham and Karaboga, 2000].

De acordo com o critério Metrópolis, citado por [Pham and Karaboga, 2000], se a diferença entre os valores da função de custo da solução atual e a solução recentemente produzida for igual ou maior que 0, então um número aleatório δ entre [0,1] de uma distribuição uniforme, é gerado. Se

$$\delta \leq e^{(-\Delta E/T)} \quad (2.1)$$

então a solução produzida recentemente é aceita como a solução atual. Caso contrário, a solução corrente fica inalterada. [Phan and Karaboga, 2000] Na equação (2.1), ΔE é a diferença entre os valores da função de custo das duas soluções e T é a temperatura atual.

Para implementação do algoritmo *Simulated Annealing* faz-se necessário a escolha de quatro elementos principais: representação das possíveis soluções, definição da função de custo, definição do mecanismo para geração da vizinhança e um parâmetro de controle, a temperatura [Pham and Karaboga, 2000][Yamazaki et al, 2002].

Para a definição da temperatura, quatro elementos devem ser especificados no algoritmo *Simulated Annealing*: uma temperatura inicial, a regra para atualização da temperatura, o número de iterações que deve ser executado cada passo da temperatura, e, finalmente, o critério de parada para a busca. [Pham and Karaboga, 2000]. Na Figura 2.1 é apresentada a estrutura básica do algoritmo *simulated annealing*.

Figura 2-1 Estrutura básica do algoritmo *Simulated Annesling*, adaptado de [Yamazaki et al, 2002]

```

 $S_0 \leftarrow$  Solução Inicial em  $S$ 
Para cada  $i = 0$  até  $I - 1$ 
    geração da solução  $s'$  da vizinhança
    Se  $f(s') \leq f(s_i)$ 
         $s_{i+1} \leftarrow s'$ 
    Senão
         $s_{i+1} \leftarrow s'$  com probabilidade  $e^{-[f(s') - f(s_i)]/T_{i+1}}$ 
        caso contrário  $s_{i+1} \leftarrow s_i$ 
Retorna  $s_i$ 

```

Na Figura 2.1, a variável I significa o total de iterações. Para implementação do algoritmo *simulated annealing* deve-se seguir os seguintes passos:

1. Escolha aleatória de uma solução;
2. Geração de uma nova solução (*vizinhança*) a partir da solução atual;
3. Se o custo (*solução nova*) for < custo (*solução atual*),
 - a. Então a solução nova é aceita;
 - b. Caso contrário

A solução nova é aceita com uma certa probabilidade

$$p = \exp[-(\text{custo}(\text{solução nova}) - \text{custo}(\text{solução atual})) / \text{Temperatura}]$$

4. Os passos 2 e 3 são repetidos até terminarem as iterações permitidas.

O parâmetro *temperatura* deve ser diminuído a cada N iterações. Se a nova solução gerada pelo algoritmo tiver um custo maior que a solução atual, a probabilidade dela ser aceita diminui quando a *temperatura* diminui e o aumento no custo é mantido fixo, ou se for mantido fixo o parâmetro *temperatura* e a variação no custo aumentar.

2.4 Regressão Logística

Os métodos estatísticos são, tradicionalmente, utilizados na área de saúde em testes de avaliação diagnóstica, para determinar valores normais, monitorar pacientes, entre outros. A inferência estatística tem sido a técnica mais popularmente utilizada em pesquisas nessa área [Bland, 1996][Walczak and Scharf, 2000].

No que se refere à análise de estudos epidemiológicos, tradicionalmente são utilizadas técnicas estatísticas muito simples, e o modelo de regressão logística também tem sido o mais popularmente utilizado [Plate et al, 1997][Kleinbaum, 1992].

Regressão é um método estatístico que serve para estimar relações numéricas entre variáveis. [Valentini, 2001], destaca que regressão é a técnica mais adequada quando se deseja estudar o comportamento de uma variável dependente (*variável resposta*) em relação a outras variáveis independentes, que são responsáveis pela sua formação (*variáveis explicativas*).

Conforme [Kleinbaum, 1992], o modelo logístico é baseado na função logística $f(z)$, a qual é dada por

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2.2)$$

quando z assume um valor $-\infty$, a função logística $f(z)$ tende a 0, e quando z assume um valor $+\infty$, então $f(z)$ tende a 1.

Esse modelo é usado para descrever a relação entre as diversas variáveis independentes e uma variável dependente, a qual é usualmente dicotômica (por exemplo “sim” ou “não”). Isto quer dizer que a variável dependente pode receber o valor 1 com uma probabilidade θ , ou o valor 0 com probabilidade $1 - \theta$ [Bland, 1996][Hosmer and Lemeshow, 1989][Ferreira et al, 2001]. O modelo prediz a população de indivíduos que tem determinada característica ou calcula a probabilidade de que um indivíduo venha a tê-la.

Quanto às variáveis independentes, essas podem tomar qualquer forma. Assim sendo, a regressão logística não faz suposição acerca da distribuição das variáveis independentes. Essas variáveis não têm que ser distribuídas normalmente, linearmente relacionadas ou de variâncias iguais em cada grupo.

A relação entre as variáveis de predição e variável resposta não é uma função linear, e a função usada é a seguinte [Bland, 1996]:

$$\theta = \frac{e(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + e(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)} \quad (2.3)$$

onde α é a constante da equação e β é o coeficiente das variáveis.

Uma forma alternativa de equação da regressão logística é

$$\text{Logit}[\theta(x)] = \log\left[\frac{\theta(x)}{1 - \theta(x)}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (2.4)$$

Na regressão logística a probabilidade de resultados dicotômicos é relatada em um conjunto potencial de variáveis explanatórias [Leung and Tran, 2000]

$$\text{Log}[p/(1-p)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (2.5)$$

onde p é a probabilidade do resultado, β_1, \dots, β_n são os coeficientes associados com cada uma variável explanatória x_1, \dots, x_n . O método usado para estimar os coeficientes β_1, \dots, β_n na regressão logística é a máxima verossimilhança. O termo β_0 é denominado de intercepto.

Esse método retorna, para um dado conjunto de observações, valores para os parâmetros desconhecidos, de forma a maximizar a probabilidade de que os dados tenham sido originados da população correspondente.

A interpretação dos coeficientes é dada através do *odds ratio* (razão das chances), que significa o quão mais provável a variável dependente tende a assumir o valor positivo $y=1$ [Ferreira et al, 2001]. Maiores detalhes podem ser vistos na literatura [Hosmer and Lemeshow, 1989][Kleinbaum, 1992].

O modelo de regressão logística é comumente usado para predição de doenças, destacando-se como uma técnica de seleção em medição de resultados médicos dicotômicos, sendo citada como a técnica mais apropriada [Bland, 1996], calculando a probabilidade de um certo evento acontecer.

[Plate et al, 1997] destaca como uma característica importante do modelo logístico a simplicidade em que as relações identificadas nas variáveis podem ser interpretadas e explicadas.

2.5 Algumas Similaridades entre Redes Neurais Artificiais e Regressão Logística

Redes Neurais Artificiais têm sido destacadas como uma alternativa nova a regressão logística, e, entre essas duas técnicas existem muitas similaridades conceituais.

O treinamento de uma rede neural é similar, no método estatístico, a aprender um modelo de regressão [Ottenbacher et al, 2001], ou à estimação de parâmetros [Schwarzer et al, 1997].

Segundo [De Souza, 1999], “(...) a aprendizagem estatística consiste em definir os parâmetros de um modelo de predição para uma amostra de casos”. Isso é similar à aprendizagem em redes neurais, pois, a rede ajusta os pesos baseado em um conjunto de dados que são apresentados à rede, visando a generalização da informação acerca do conjunto de dados.

Em uma rede neural cada nodo de entrada da rede pode ser visto, na estatística, como variáveis independentes, explanatórias, preditivas ou covariáveis. Existe uma similaridade entre os pesos utilizados nas camadas adjacentes da rede neural para o cálculo da saída com os chamados coeficientes de regressão em estatística. [Schwarzer et al, 1997].

Quanto ao *bias*, que nas redes neurais tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo respectivamente [HAYKIN, 2001], em estatística é conhecido como intercepto.

O erro em uma rede neural supervisionada é calculado através da diferença entre a saída real com a saída desejada da rede. Essa diferença (erro) é similar ao conceito de minimização de resíduos na regressão estatística. O erro total (E) de uma rede neural é usualmente determinado utilizando todo o conjunto de dados e pode ser calculado da seguinte forma

$$E = \sum_p (t_p - o_p)^2 \quad (2.6)$$

onde E é o erro total da rede, t_p é a saída desejada, e o_p é a saída real calculada pela rede [Tu, 1996].

No modelo de regressão logística o processo converge quando a função de

probabilidade é maximizada, já em redes neurais, usualmente, a função de erro dos mínimos quadrados é minimizada.

[Tu, 1996], destaca ainda em seu artigo que, o poder de uma rede neural sobre o modelo de regressão logística está na sua camada de neurônios escondidos. Esta camada atua como detectores de características dos dados e torna possível que a rede “aprenda” implicitamente quaisquer relações não lineares complexas entre as variáveis dependentes e independentes. Um quadro resumo com algumas similaridades conceituais existentes entre redes neurais e o modelo estatístico é apresentado na Tabela 2.1.

Tabela 2-1 Termos similares entre Redes Neurais Artificiais e Estatística

Termos em Redes Neurais	Termos equivalentes em Estatística
Treinamento, aprendizagem	Estimação de parâmetros
Unidades de Entrada da rede	Variáveis independentes, explanatórias, preditivas, covariáveis.
Camada de saída da rede	Variável dependente ou resposta, valores previstos
Pesos nas conexões da rede	Coefficientes de regressão
Bias	Parâmetro de intercepto
Erro	Resíduo
Casos de treinamento, padrões	Observação

2.6 Algumas vantagens no uso de Redes Neurais Artificiais

[Tu, 1996], destaca como uma vantagem importante quanto ao uso de Redes Neurais Artificiais na realização de experimentos, o fato dos usuários precisarem de pouco conhecimento formal, além disso, qualquer dos conjuntos de dados que seja analisado com uso do modelo de regressão logística também pode ser utilizado para desenvolver uma rede neural artificial [Tu, 1996]. Outra questão importante é que as redes neurais podem ser treinadas utilizando tanto variáveis contínuas quanto categóricas nas entradas e saídas da rede.

Porém, quanto ao conhecimento formal, o mesmo não podemos falar sobre regressão logística. Neste caso, para a utilização do modelo logístico é necessário que se tenha um conhecimento formal em estatística, a exemplo de *odds ratios*, *p* valor, regressão *stepwise*, entre outros. E, para isso, pode-se levar um tempo maior para aprender a técnica de modelagem de regressão logística [Tu, 1996].

A capacidade que a rede neural tem de detectar relações não lineares complexas existentes entre variáveis dependentes e independentes, bem como possíveis interações ou inter-relações existentes entre elas, ajustando seus pesos e conexões para que a rede possa modelar os dados, pode ser destacada como uma vantagem importante com relação à regressão logística. Isso é possível através da camada escondida da rede. Nesse caso, geralmente, as redes neurais podem apresentar resultados melhores que a técnica estatística.

Tabela 2-2 Algumas vantagens e desvantagens em utilizar uma rede neural artificial para a predição de resultados

Vantagens	Desvantagens
<ol style="list-style-type: none"> 1. O uso de redes neurais para a realização de experimentos exige do usuário comum pouco conhecimento formal. 2. Modelos de redes neurais artificiais podem detectar implicitamente relações complexas não lineares entre variáveis dependentes e independentes, bem como as possíveis interações ou inter-relações existentes entre elas. 3. Em redes neurais não existem muitas suposições que precisam ser verificadas antes dos modelos serem construídos, além disso, um número grande de variáveis podem ser incluídas na análise com RNA. 4. Vários algoritmos de aprendizagem podem ser utilizados no treinamento das redes neurais artificiais. 	<ol style="list-style-type: none"> 1. É definida como uma “caixa preta”, tendo uma capacidade limitada para explicitamente identificar possíveis relações causais. 2. O uso de redes neurais requer maiores recursos computacionais nos experimentos. 3. Os modelos de Redes Neurais Artificiais são propensos à ocorrência de <i>overfitting</i>. 4. O processo de treinamento e testes de uma rede neural pode vir a ser, muitas vezes, lento em relação ao modelo de regressão. 5. Não existe nenhum critério bem estabelecido sobre a interpretação da matriz de pesos das conexões em uma rede, tornando difícil a sua interpretação.

[Ottenbacher et al, 2001] destaca também como vantagem em utilizar redes neurais o fato de não existirem muitas suposições que precisam ser verificadas antes dos modelos serem construídos, além disso, um número grande de variáveis podem ser incluídas na análise com RNA.

Outra vantagem em usar redes neurais, citada por [Tu, 1996], é o fato das redes poderem ser treinadas por diferentes algoritmos de aprendizagem. O mais popularmente utilizado dentre eles, conforme descrito no Apêndice A, é o *backpropagation*. O quadro resumo, apresentado na Tabela 2.2, destaca algumas vantagens e desvantagens no uso de RNA em relação a métodos estatísticos.

2.7 Algumas desvantagens no uso de Redes Neurais Artificiais

Redes neurais artificiais são definidas como uma “caixa preta” e têm uma habilidade limitada para identificar explicitamente possíveis relações causais. O modelo de regressão logística tem sido superior às redes neurais quando o objetivo é identificar possíveis relações causais. Regressão logística pode identificar quais variáveis preditivas são fortemente relacionadas com o resultados através dos coeficientes β e *odds ratios* (razão das chances) associados [Tu, 1996].

Um outro aspecto importante, que muitas vezes é destacado como desvantagem para a utilização das redes neurais, é que não existe nenhum critério bem estabelecido sobre a interpretação da matriz de pesos em uma rede, tornando difícil a sua interpretação. A dificuldade de interpretação no nível de variáveis individuais (preditivas), também é citado na literatura como uma desvantagem para o uso de redes neurais artificiais [Ottenbacher et al, 2001].

[Tu, 1996] também destaca que, muitas vezes, as redes são treinadas com um conjunto de variáveis de entrada que não são importantes para o problema. Logo, isso pode causar prejuízos para a rede generalizar a informação sobre o conjunto de dados.

Uma dessas técnicas, apresentadas por [Tu, 1996], consiste em treinar a rede removendo cada nodo de entrada por vez, observando o efeito da retirada de tal variável na performance da rede. Outros utilizam técnicas de regressão para investigar as conexões dos pesos das variáveis de entrada, determinando quais variáveis podem ser removidas sem afetar a performance da rede. Porém, nenhuma dessas técnicas tem alcançado sucesso na comunidade médica, pois não oferecem a clareza do modelo de regressão logística com o *odds ratios* associado.

Com isso, o desenvolvimento de técnicas utilizando redes neurais artificiais que possam identificar possíveis relações entre variáveis independentes e dependentes, ou analisar o efeito das variáveis independentes com relação ao resultado, é um desafio.

2.8 Considerações Finais

Neste capítulo foi apresentada uma revisão da literatura, com algumas

considerações acerca de redes neurais artificiais, mostrando algumas aplicações na área da saúde. Também foi apresentada uma visão geral sobre o algoritmo simulated annealing bem como a sua utilização para o projeto de RNA's.

Além disso, é apresentada uma visão geral sobre o modelo de regressão logística, apresentando algumas similaridades conceituais entre o método neural e o estatístico, destacando algumas vantagens e desvantagens no uso das duas técnicas.

Observando os resultados obtidos nos exemplos apresentados neste Capítulo, foi possível perceber que redes neurais artificiais também podem ser utilizadas com sucesso em problemas epidemiológicos para análise de fatores relacionados ao diagnóstico de doenças. Fato este que motivou a realização deste trabalho.

Um dos problemas apontados por pesquisadores sobre o uso de redes neurais, conforme descrito anteriormente, é que muitas vezes o treinamento das redes é realizado com um conjunto de variáveis de entrada que não são importantes para o problema.

Essas dificuldades motivaram em nosso trabalho, a realização de experimentos que pudessem resultar em uma técnica para estabelecer o grau de relação de variáveis explicativas com o problema estudado.

Dos trabalhos citados neste Capítulo, os mais relacionados com o desenvolvimento desta pesquisa são [Tu, 1996] e [Yamazaki et al, 2002], dos quais foram adaptadas as técnicas utilizadas na realização dos experimentos. No Capítulo 3 serão apresentados os materiais e métodos utilizados para a realização dos experimentos, com uma descrição da base de dados.

Capítulo 3 - Materiais e Métodos

3.1 Introdução

Um dos motivos que tem tornado o uso de redes neurais artificiais muito popular, certamente, é o fato de não existir necessidade de independência entre as diversas variáveis estudadas, além da sua capacidade de aprendizado a partir do ambiente. É importante destacar que, a aplicação de Redes Neurais Artificiais a problemas epidemiológicos tem sido apresentada como uma alternativa interessante [Lamers et al, 1997] [Nair et al, 1999][dos Santos, 2001].

Além disso, o fato das redes neurais apresentarem uma habilidade de detectar todas as possíveis interações ou inter-relações entre as diversas variáveis explicativas e variáveis dependentes também motivou a realização dos experimentos neste trabalho, cujo objetivo principal foi analisar os fatores relacionados aos Transtornos Mentais Comuns (TMC's) a partir de uma rede neural artificial.

Para isso foram realizados experimentos com a base de dados gerada por [Ludermir, 1998], utilizando duas técnicas:

- 1) O uso de uma rede neural *Multi-Layer Perceptron*, com uma camada escondida, treinada com o algoritmo *Backpropagation* e utilizando todas as variáveis do conjunto de dados, num total de onze. Em seguida, foi retirado uma variável do conjunto de entrada da rede por vez, observando a sua capacidade de generalização e o percentual de classificação correta dos casos com diagnósticos positivo e negativo para os TMC's. Esse processo foi repetido com todas as variáveis do conjunto. Essa técnica foi adaptada de [Tu, 1996]. Para verificar se as variações ocorridas quanto ao erro de classificação da rede e o percentual de classificação correta dos casos com diagnósticos positivo e negativo para os TMC's foram estatisticamente significativos, foi aplicado o teste de hipótese *t-student* [Arango, 2001], através do software estatístico **Statistica** da StatSoft, versão 5.5;
- 2) Com o uso de uma rede neural *Multilayer Perceptron* treinada com o algoritmo *simulated annealing* para otimização simultânea da arquitetura (nodos e conexões) e pesos da rede, observando as variáveis que eram mais utilizadas para obtenção do resultado a cada execução do algoritmo e, com isso, identificar aquelas que apresentavam maior probabilidade de estarem

relacionadas com o problema estudado. Técnica adaptada de [Yamazaki et al, 2002].

Os resultados obtidos nesses experimentos foram comparados com aqueles apresentados por [Ludermir, 1998], aplicando o modelo estatístico de regressão logística. Nos capítulos 4 e 5 serão apresentados os resultados dos experimentos realizados com as duas técnicas citadas nesta Seção, e no Capítulo 6, será apresentada uma comparação dos resultados obtidos com redes neurais e regressão logística.

3.2 Descrição do Problema e Base de Dados

A base de dados utilizada em nossos experimentos foi gerada por [Ludermir, 1998] a partir de uma pesquisa realizada na cidade de Olinda, situada na região metropolitana do Recife (PE), a qual determinou a prevalência de Transtornos Mentais Comuns (TMC's) nessa área, e analisou a sua associação com condições de vida e trabalho.

O estudo foi desenvolvido envolvendo 621 indivíduos, maiores de quinze anos, de uma amostra domiciliar aleatória. Para isso, foi utilizado como instrumento de pesquisa uma versão do *Self Reporting Questionnaire (SRQ-20)* e um questionário socioeconômico. A análise dos dados foi realizada com o uso do modelo estatístico de regressão logística.

As condições de vida foram mensuradas a partir das variáveis **alfabetização, escolaridade, propriedade da casa, condições de moradia e posse de bens duráveis**. Quanto às condições de trabalho, as variáveis observadas foram operacionalizadas mediante a **inserção no processo produtivo e renda familiar per capita** mensal.

O conjunto de padrões é composto de onze variáveis de entrada, as quais correspondem às mesmas utilizadas originalmente por [Ludermir, 1998]: **idade, condições de moradia, sexo, alfabetização, escolaridade, renda, situação conjugal, posse de bens duráveis, propriedade da casa, inserção no processo produtivo e migração**. Para a saída da rede foram definidos dois nodos cujo valor **1 0** representou os casos positivos e **0 1** os casos negativos para TMC.

Todas as variáveis do estudo são ordinal/categóricas, e, para a realização dos experimentos com RNA's, foram codificadas com números discretos compreendidos entre 0 e 1. Na Tabela 3.1 é feita uma descrição das variáveis que compõem o conjunto de dados.

A prevalência total dos transtornos mentais comuns na amostra estudada foi aproximadamente de 35% dos casos, perfazendo um total de 216 padrões. Os demais casos (405) apresentavam diagnóstico negativo para TMC's. Dos casos positivos 74% (160) eram do sexo feminino e apenas 26% (56) do sexo masculino.

Tabela 3-1 Variáveis que compõem o conjunto de dados

Variável	Descrição
Idade	Utilizada para identificar os indivíduos quanto à idade, sendo agrupada em três categorias: <i>15 a 24 anos</i> , <i>25 a 54 anos</i> e <i>≥ 55 anos</i> .
Condições de Moradia	Variável utilizada para medir as condições de moradia dos indivíduos constantes na amostra. Essa variável foi definida a partir de três categorias: <i>boa</i> , <i>média</i> e <i>precária</i> .
Sexo	Utilizada para identificar os indivíduos quanto ao sexo: <i>masculino</i> ou <i>feminino</i> .
Alfabetização	Utilizada para classificar o indivíduo quanto à alfabetização: <i>alfabetizado</i> ou <i>analfabeto</i> .
Escolaridade	Expressa em anos de escolarização, foi dividida em quatro categorias: <i>0 a 4</i> , <i>5 a 7</i> , <i>8 a 10</i> e <i>≥ 11</i> .
Renda	Essa variável foi utilizada para identificar os indivíduos no conjunto de dados com relação à sua renda familiar per capita mensal. Ela foi expressa em salários mínimos, e dividida nas seguintes categorias: <i>0 à ¼</i> , <i>¼ a 1</i> e <i>>1</i> .
Situação Conjugal	Utilizada para identificar os indivíduos do conjunto de dados em três categorias: <i>solteiro</i> , <i>casado</i> e <i>separado/viúvo</i> .
Posse de Bens Duráveis	Com a variável <i>posse de bens duráveis</i> , os indivíduos da amostra foram identificados segundo a quantidade de utensílios domésticos que possuía (geladeira, televisão, máquina de lavar, etc) sendo agrupado em três categorias: <i>boa</i> , <i>média</i> e <i>precária</i> .
Propriedade da Casa	Utilizada para classificar os indivíduos que moravam em casa <i>própria</i> ou <i>alugada</i> .
Inserção no Processo Produtivo	Variável dividida em três grupos equivalentes à inserção do trabalhador no processo produtivo, expressos da seguinte forma: <i>formal</i> , <i>informal</i> , <i>desempregado</i> e <i>inativo</i> .
Migração	A partir de duas categorias distintas, foi utilizada para agrupar indivíduos <i>nativos</i> de Olinda ou <i>migrantes</i> .

Como apenas 216 padrões apresentavam diagnóstico positivo, e, para garantir o balanceamento dos dados por indivíduos portadores e não portadores de TMC's, foram selecionados aleatoriamente 216 casos dos 405 com diagnóstico negativo.

Para a retirada dos padrões negativos excedentes foi obedecido o seguinte procedimento: 1) Retirada de padrões com informações ausentes; 2) Após o

embaralhamento dos dados restantes, foram selecionados aleatoriamente os 216 padrões para a composição da base de dados.

Duas partições foram geradas para a realização dos experimentos com RNA's a partir do conjunto de dados original. O processo de seleção dos padrões para a composição das partições foi o mesmo, entretanto, a ordem de apresentação das variáveis nos conjuntos (treinamento, validação e teste) foi alterada para a formação da segunda partição.

3.3 Experimentos com Multilayer Perceptron

Os experimentos foram realizados com uma rede neural *Multi-Layer Perceptron* treinada com o algoritmo *backpropagation* e com o método de otimização *simulated annealing*, conforme descrito anteriormente.

Nos experimentos com o algoritmo *simulated annealing*, o objetivo foi otimizar simultaneamente a arquitetura (nodos de entrada e intermediários além das conexões) e pesos da rede neural, visando com isso à identificação das variáveis que apresentavam maior probabilidade de estarem relacionadas com os TMC's.

3.3.1 Divisão do Conjunto de Padrões em Treinamento, Validação e Teste

Foram geradas duas partições com a base de dados, cada uma delas composta por 432 padrões, os quais foram embaralhados e em seguida divididos em conjuntos segundo o que sugere o relatório *Proben1* [Prechelt, 1994]: 50% dos padrões para o conjunto de treinamento (216 casos); 25% dos padrões para o conjunto de validação (108 casos); e 25% dos padrões para o conjunto de teste (108 casos).

Para todos os conjuntos foi garantido o balanceamento dos dados, com 50% de casos positivos e 50% de casos negativos para TMC's, sendo a escolha dos padrões em cada partição feita aleatoriamente.

3.3.2 Pré-processamento dos Dados

Como todas as variáveis do conjunto de dados são ordinal/categóricas, para a

realização dos experimentos com redes neurais foram codificadas com números discretos entre 0 e 1. Na Tabela 3.2 são exibidas as variáveis do conjunto com sua respectiva codificação.

Tabela 3-2 Codificação das variáveis do conjunto de dados

Variável	Codificação
Idade	15 a 24 anos = 0.0 25 a 54 anos = 0.50 ≥ 55 anos = 1.0
Alfabetização	Alfabetizado = 0.0 Analfabeto = 1.0
Migração	Nativo = 0.0 Migrante = 1.0
Escolaridade	0 a 4 anos = 0.25 5 a 7 anos = 0.50 8 a 10 anos = 0.75 ≥ 11 anos = 1.0
Propriedade da Casa	Própria = 0.0 Alugada = 1.0
Inserção no Processo Produtivo	Formal = 0.25 Informal = 0.50 Desempregado = 0.75 Inativo = 1.0
Condições de Moradia	Boa = 0.0 Média = 0.50 Precária = 1.0
Sexo	Masculino = 0.0 Feminino = 1.0
Situação Conjugal	Solteiro = 0.0 Casado = 0.50 Separado/Viúvo = 1.0
Renda	0 a ¼ = 0.0 ¼ a 1 = 0.50 >1 = 1.0
Posse de Bens Duráveis	Boa = 0.0 Média = 0.50 Precária = 1.0

3.3.3 Topologias das Redes Neurais

Os experimentos foram realizados com uma rede neural MLP contendo apenas uma camada intermediária. Foram definidas onze unidades de entrada, equivalente ao número de variáveis presentes em cada padrão do conjunto de dados. A camada de saída foi composta por dois nodos.

Nos experimentos realizados com o algoritmo *backpropagation* a rede continha todas as possíveis conexões *feedforward* entre as camadas adjacentes. Foram treinadas cinco topologias diferentes com 4, 8, 12, 16 e 20 nodos na camada intermediária, com onze unidades de entrada (todas as variáveis presentes no conjunto de dados) e com dez unidades de entrada (nos experimentos retirando uma variável por vez do conjunto).

Já os experimentos com o algoritmo *simulated annealing* resultaram em uma rede com arquitetura bem simples. A topologia inicial da rede foi então definida com quatro nodos na camada intermediária e onze nodos nas unidades de entrada. Posteriormente foram realizados novos experimentos com sete nodos de entrada, valores esses resultantes dos experimentos iniciais.

3.3.4 Metodologia de Treinamento

Conforme descrito na Seção 3.1, foram realizados dois tipos de experimentos: 1) com o algoritmo *backpropagation* com todo o conjunto de dados e retirando variáveis do conjunto; 2) com o algoritmo *simulated annealing* para otimização simultânea da arquitetura (nodos e conexões) e pesos da rede. Nas Seções 3.3.4.1 e 3.3.4.2 é descrita a metodologia aplicada nos experimentos realizados em cada uma das técnicas utilizadas.

A função de ativação utilizada nos experimentos com *Backpropagation* foi a sigmóide logística [Tu, 1996] e naqueles realizados com o *simulated annealing* a tangente hiperbólica [Yamazaki et al, 2002]. Todos os experimentos foram realizados utilizando o software Matlab, versão 6 – R12.

O critério de parada adotado para o treinamento da rede neural foi o GL_5 do *Proben1* [Prechelt, 1994]. Esse critério é baseado na soma do erro quadrado no conjunto de validação, buscando minimizar o risco de *overfitting*. O erro foi medido com base no erro de classificação no conjunto de validação. O número máximo de iterações definidas também foi utilizado para a parada do treinamento, caso o critério GL_5 não fosse satisfeito.

3.3.4.1 Metodologia dos Treinamentos com o Algoritmo Backpropagation

Nesses experimentos foi utilizado o algoritmo *backpropagation* com termo *momentum* [Haykin, 2001] [Braga et al, 2000]. Para cada topologia, nos experimentos com todas as variáveis do conjunto, foram realizadas dez execuções com inicializações diferentes e aleatórias de pesos entre -1.0 e $+1.0$ e, em todos os casos o número máximo de iterações permitidas foi 10000. As taxas de aprendizagem utilizadas foram 0,001 e 0,0001.

Nos experimentos realizados retirando variáveis do conjunto de entrada da rede, foi retirada uma variável por vez, sendo realizados experimentos para cada topologia testada (4, 8, 12, 16 e 20), também com dez execuções distintas e inicializações aleatórias de pesos.

3.3.4.2 Metodologia dos Treinamentos com o Algoritmo Simulated Annealing

Nos experimentos realizados com método *simulated annealing* foi definida a topologia inicial da rede com uma camada intermediária contendo quatro nodos, tendo todas as possíveis conexões *feedforward* entre as camadas adjacentes, e todas as variáveis do conjunto de dados.

A partir dos resultados obtidos, novos experimentos foram realizados com o número de variáveis de entrada resultantes dos experimentos iniciais (sete). Dessa forma, foi possível estabelecer as variáveis que apresentavam maior probabilidade de estarem relacionadas com os TMC's.

Para a realização dos experimentos, conforme descrito no Capítulo 2, Seção 2.4, foi necessário a escolha de quatro elementos principais: representação das soluções, definição da função de custo, definição do mecanismo para geração da vizinhança e temperatura [Yamazaki et al, 2002].

Na representação das soluções, nesse trabalho, cada MLP foi especificada com um vetor de conexões, e cada conexão sendo definida por dois parâmetros: a) um bit de conectividade, com o valor 1 indicando a existência da conexão, e o valor 0 indicando o contrário; b) os pesos da conexão (um número real entre 0 e 1). Se o bit de conectividade tivesse o valor igual a 0, então o peso associado não era considerado, e a conexão eliminada.

A função de custo usada para cada solução foi a média de dois importantes parâmetros: o erro de classificação no conjunto de treinamento e o percentual de conexões utilizadas pela rede. O algoritmo, entretanto, procura minimizar tanto a performance quanto a complexidade da rede.

Com isso, foi preciso definir um parâmetro para estabelecer se uma rede gerada era válida ou não. Foram consideradas como redes válidas somente aquelas que usavam, no mínimo, um nodo na camada escondida. Por isso, se uma solução

criada não fosse uma rede válida, uma nova solução era gerada na vizinhança.

Com relação ao mecanismo de geração da vizinhança, primeiro os bits de conectividade para a solução atual eram mudados obedecendo a uma dada probabilidade, a qual neste trabalho foi ajustada em 20%.

Essa operação, portanto, exclui algumas conexões e cria novas. Um número aleatório entre -1.0 e $+1.0$ é adicionado aos pesos da conexão. Esses passos podem mudar tanto a topologia quanto os pesos da conexão para produzir uma nova solução na vizinhança.

A última definição para a realização dos experimentos refere-se à temperatura. Em nosso trabalho a temperatura inicial foi definida com o valor 1, sendo ajustada a cada dez iterações, com um número máximo de 5000 iterações. O fator de ajuste da temperatura foi definido como 0.95.

3.3.5 Aspectos Observados

A medida de erro utilizada para análise dos resultados foi o *erro de classificação*, o qual corresponde à razão entre a quantidade de padrões classificados incorretamente e o número total de padrões. Assim, os aspectos observados ao final desses experimentos foram:

1. Erro de classificação no conjunto de validação (parâmetro utilizado para escolha da rede);
2. Erro de classificação no conjunto de teste;
3. O percentual de classificação correta de indivíduos com diagnósticos positivo e negativo para os TMC's.;
4. Nos experimentos realizados com o algoritmo *backpropagation* retirando variáveis do conjunto de dados, o erro de classificação no conjunto de teste foi observado para análise do comportamento da rede com relação à retirada de cada variável;

Nos experimentos realizados com o algoritmo *simulated annealing* além do erro de classificação da rede e do percentual de classificação correta dos casos com diagnósticos positivo e negativo com relação ao conjunto de teste, também foi

observado:

1. O valor médio do erro de classificação obtido nos experimentos com o *backpropagation* e *simulated annealing*. A execução do *simulated annealing* que apresentava um erro de classificação superior a essa média, não era considerado para análise das variáveis;
2. A média de variáveis utilizadas nos experimentos;
3. A partir da definição da média de variáveis, foram observadas quais as variáveis que foram mais utilizadas nos experimentos para obtenção dos resultados.

A definição do critério para a escolha das execuções do *simulated annealing* que deveriam ser consideradas para a análise das variáveis, mostrou-se de fundamental importância. [Yamazaki et al, 2002] em seus experimentos com safras de vinho utilizando essa mesma técnica, optou por excluir os dez melhores e dez piores resultados para análise dos dados.

Inicialmente em nossos experimentos, conforme resultados publicados no VII Simpósio Brasileiro de Redes Neurais – SBRN [Lopes et al, 2002], utilizamos como parâmetro para a escolha das execuções do *simulated annealing* o percentual médio do erro de classificação obtido nos experimentos com uma rede MLP treinada com o algoritmo *backpropagation* (27%).

Porém, após a realização de novos experimentos, inclusive com uma nova partição do conjunto de dados (partição 02), podemos observar que a melhor opção, em nosso caso, seria a utilização da média entre o erro de classificação obtido nos experimentos com o *backpropagation* e o *simulated annealing*. Pois, dessa forma, foi possível evitar um super ajustamento do modelo e a amostra analisada foi maior, sendo excluídas menos execuções do algoritmo.

É importante destacar que, apesar da mudança do método de exclusão das execuções do *simulated annealing*, com relação ao método utilizado nos experimentos iniciais, não houve uma alteração nos resultados finais.

3.4 O teste de hipótese *t-student*

Uma hipótese é uma pressuposição a respeito de um determinado problema, a qual,

uma vez formulada estará sujeita a uma comprovação. O mecanismo utilizado para esta comprovação é o teste de hipótese [Arango, 2001]. Quando se realiza um teste de hipótese, é verificado se o pressuposto é verdadeiro ou não.

O teste *t-student* é um teste paramétrico mais comumente utilizado para se avaliarem as diferenças entre as médias de dois grupos, podendo ser montada uma experiência aos pares e se efetuar o teste para o mesmo grupo em situações diferentes. Maiores detalhes sobre a teoria dos testes de hipótese e o teste *t-student* podem ser encontrados na literatura [Arango, 2001].

3.5 Considerações Finais

Neste Capítulo são relatadas algumas considerações acerca dos Transtornos Mentais Comuns, destacando o impacto desses com relação à saúde pública, bem como para a sociedade como um todo.

Informações sobre a base de dados, do processo de divisão dos conjuntos, pré-processamento, topologias, metodologia do treinamento e aspectos observados também estão presentes neste Capítulo. Os resultados obtidos serão apresentados nos Capítulos 4 e 5.

Capítulo 4 - Experimentos com Multilayer Perceptron e o Algoritmo Backpropagation

4.1 Introdução

Neste Capítulo são apresentados os experimentos realizados com uma rede *Multilayer Perceptron* (MLP), treinada com o algoritmo *Backpropagation*, os quais foram realizados em duas etapas distintas.

Primeiramente utilizamos todas as variáveis do conjunto de dados, num total de onze. Esses experimentos foram realizados para que pudéssemos compará-los com aqueles retirando variáveis do conjunto. Para, com isso, observar se a performance da rede sofria alguma alteração quanto ao erro de classificação e o percentual de classificação correta dos indivíduos com diagnósticos positivo e negativo para os TMC's, após a retirada de cada variável do conjunto.

Na etapa seguinte, foram realizados experimentos retirando variáveis do conjunto de entrada. O objetivo da realização desses experimentos foi analisar quais variáveis apresentavam maior possibilidade de estar relacionada aos Transtornos Mentais Comuns (TMC's), a partir da retirada de cada variável do conjunto de dados.

A análise dos resultados foi desenvolvida de duas formas distintas: 1) observando cada partição individualmente, num total de duas; 2) observando a média obtida entre as duas partições.

Foi utilizado o software Statistica da StatSoft, versão 5.5, para a realização de testes de hipóteses (*t-student*) com o objetivo de verificar se a diferença observada nos experimentos a partir da retirada de cada variável do conjunto, com relação ao erro de classificação e o percentual de classificação correta dos indivíduos com diagnósticos positivo e negativo para os TMC's eram estatisticamente significativos. Para isso, foi definido como significativo o valor de $p < 0,05$.

Na Seção 4.2 são apresentados os resultados obtidos observando cada partição individualmente. Os resultados obtidos a partir da análise feita com a média entre as duas partições são apresentados na Seção 4.3. Na Seção 4.4 são apresentadas as conclusões obtidas com esses experimentos.

4.2 Análise Individual de Cada Partição

Nesta Seção são apresentados os resultados obtidos nos experimentos a partir da

análise individual de cada partição. Os resultados são apresentados contendo cada topologia testada, bem como todos os aspectos observados, desvio padrão (st. dv.) e o nível de significância do teste de hipótese (p).

Os experimentos com a partição 01 são apresentados na Seção 4.2.1 e com a partição 02 na Seção 4.2.2. Posteriormente, são feitos os comentários a respeito do desempenho da rede neural para análise dos fatores relacionados aos TMC's.

4.2.1 Experimentos realizados com a partição 01

Nesta Seção serão apresentados os experimentos realizados com a partição 01, utilizando todas as variáveis do conjunto de dados, bem como os experimentos retirando variáveis. Além disso, é apresentada uma comparação dos resultados obtidos com a retirada de cada variável do conjunto, em relação aos experimentos utilizando todas as variáveis.

4.2.1.1 Utilizando todas as variáveis do conjunto de dados

Na Tabela 4.1 são apresentados os resultados obtidos com os experimentos realizados com todas as variáveis do conjunto de dados. A topologia considerada em nossa análise, observando a média obtida no erro de validação bem como o desvio padrão, foi definida com oito nodos na camada intermediária.

Tabela 4-1 Resultados com MLP e todas as variáveis do conjunto (partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	21,52778	2,16324	31,48148	2,39073	25,55555	1,70173	100	0	48,88889	3,40347
8	22,68518	2,08190	31,66666	2,08419	23,70370	1,91258	100	0	52,59259	3,82517
12	22,31481	2,47684	32,31481	2,24483	25,83333	2,67113	100	0	48,33333	5,34227
16	21,11111	1,69331	32,87037	2,47876	27,22222	0,64741	100	0	45,55555	1,29482
20	20,18518	1,66495	32,59259	2,38275	27,50000	2,39272	100	0	45,00000	4,78544

Com esses resultados podemos observar que a rede classificou corretamente 100% dos casos com diagnóstico positivo, porém classificou corretamente apenas 52,59% dos casos com diagnóstico negativo para os TMC's. Logo, o erro de classificação (23,70%) foi predominante nos casos com diagnóstico negativo para os TMC's.

Essa diferença existente entre a classificação dos indivíduos com diagnósticos

positivo e negativo pode indicar que a base de dados não possui informação suficiente para que a rede possa classificar de forma mais eficiente os padrões com diagnóstico negativo, ou seja, os indivíduos não portadores de TMC's.

4.2.1.2 Retirando variáveis do conjunto de dados

Nesta Seção são apresentados os resultados dos experimentos realizados, com a partição 01, retirando cada variável do conjunto de dados, bem como uma comparação com os resultados obtidos nos experimentos utilizando todas as variáveis.

Nesses experimentos é retirado uma variável por vez do conjunto de dados com o objetivo de verificar se a retirada da variável influencia nos resultados apresentados pela rede (erro de classificação e percentual de classificação correta dos indivíduos portadores e não portadores de TMC's)

Retirando a variável Idade

Os resultados obtidos nos experimentos com a partição 01, retirando a variável idade, são apresentados na Tabela 4.2. A topologia considerada foi definida com quatro nodos na camada intermediária.

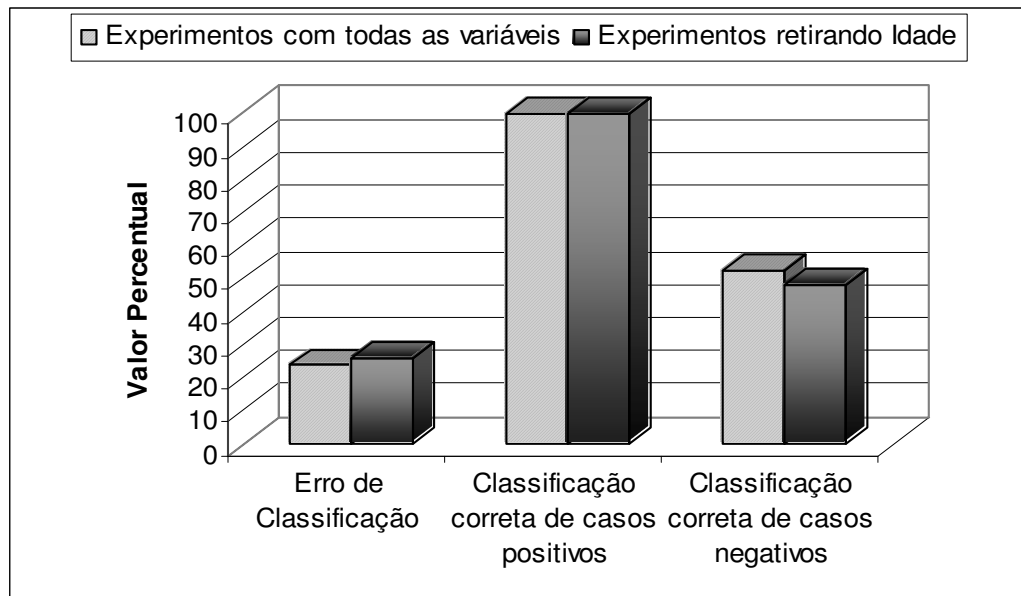
Com a retirada da variável **idade** do conjunto de dados, após a realização do teste de hipótese *t-student*, foi possível observar que a retirada dessa variável causou um crescimento estatisticamente significativo no erro de classificação (média 2,22% ___ $p = 0,0001$) e redução no percentual de classificação correta dos casos com diagnóstico negativo (média 4,44% ___ $p = 0,0001$). Uma comparação com os resultados obtidos utilizando todas as variáveis do conjunto pode ser visualizada na Figura 4.1.

É possível observar com os resultados obtidos nos experimentos que a variação ocorrida no erro de classificação da rede refletiu em uma queda na classificação correta dos casos com diagnóstico negativo, sugerindo, dessa forma, que a variável **idade** é importante para a classificação dos indivíduos com diagnóstico negativo para TMC's.

Tabela 4-2 Resultados retirando a variável Idade (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	21,94444	1,89256	31,75926	1,85442	25,92593	1,30946	100,0	0,0	48,14815	2,61891
8	23,47222	1,43195	31,85185	3,09259	24,81482	2,17368	100,0	0,0	50,37037	4,34736
12	20,87963	2,05253	33,33333	2,92804	26,66667	2,21707	99,62963	1,17121	47,03704	5,46567
16	21,71296	1,94530	33,24074	3,10028	27,03704	1,05120	100,0	0,0	45,92593	2,10239
20	20,78704	1,61042	32,77778	2,69775	27,31482	1,39744	100,0	0,0	45,37037	2,79487

Figura 4-1 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Idade (Partição 01)



Retirando a variável Alfabetização

Com a retirada da variável **alfabetização** do conjunto de dados na partição 01, foi possível observar uma queda na capacidade de generalização da rede em 0,83%. A capacidade de classificação correta dos casos com diagnóstico positivo para os TMC's, entretanto, não sofreu alteração, mas a capacidade da rede em classificar os indivíduos com diagnóstico negativo para os TMC's foi reduzida em 1,67%. Essas alterações, contudo, não apresentaram significância estatística ($p = 0,4872$). A topologia considerada foi definida com oito nodos na camada intermediária.

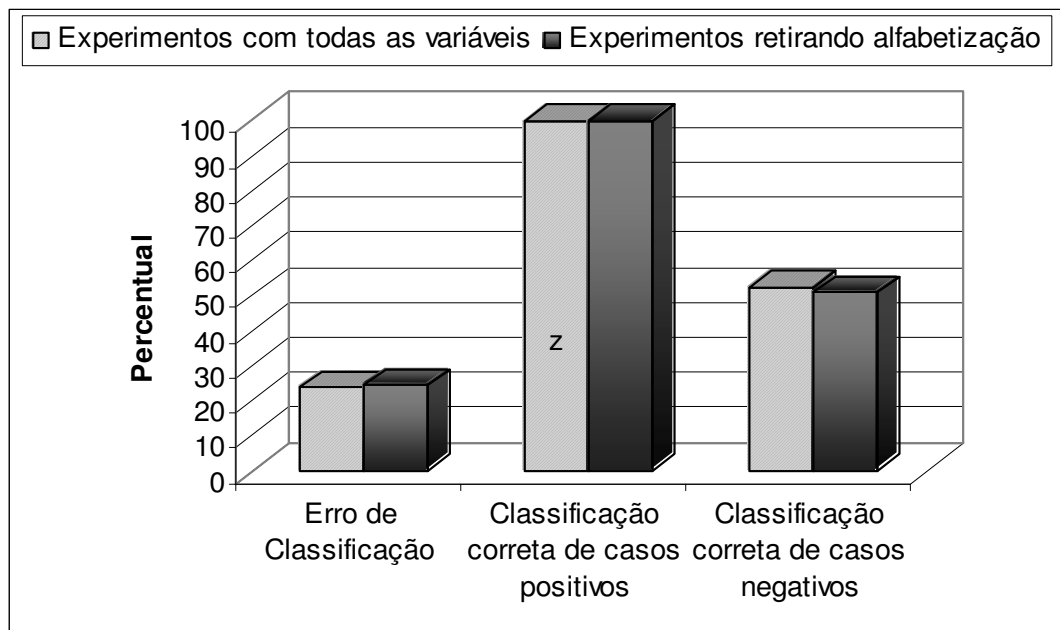
Os resultados obtidos com esses experimentos são apresentados na Tabela 4.3. e a comparação com os experimentos utilizando todas as variáveis no conjunto de dados pode ser visualizada na Figura 4.2. Esses resultados, após análise da

significância estatística, sugerem que a variável alfabetização não apresenta possibilidade de estar relacionada aos TMC's.

Tabela 4-3 Resultados retirando a variável Alfabetização (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	21,71296	1,38803	31,11111	2,27643	25,00000	2,69068	100,0	0,0	50,00000	5,38136
8	23,00926	2,69112	30,37037	2,42240	24,53704	2,55447	100,0	0,0	50,92593	5,10893
12	19,07407	1,64481	33,05555	3,11866	27,12963	1,07361	100,0	0,0	45,74074	2,14723
16	20,50926	1,63391	33,33334	2,43025	27,03704	1,43444	100,0	0,0	45,92593	2,86888
20	20,13889	1,70802	32,03704	2,58965	27,68519	1,34179	100,0	0,0	44,62963	2,68359

Figura 4-2 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Alfabetização (Partição 01)



Retirando a variável Migração

Na partição 01, a partir da retirada da variável **migração**, foi observado um crescimento no erro de classificação da rede em 1,39%. O percentual de classificação dos casos negativos foi diminuído em 2,78%, porém, o percentual de classificação dos casos positivos para TMC's não sofreu alteração. Contudo, essas alterações não apresentaram significância estatística ($p = 0,3479$). Nesses experimentos a topologia considerada também foi com oito nodos na camada intermediária.

Esses resultados sugerem que a variável **migração**, assim como **alfabetização**, não são significativos para a classificação diagnóstica dos TMC's. Os resultados obtidos com esses experimentos são apresentados na Tabela 4.4. Na Figura 4.3 é possível observar a comparação entre os experimentos realizados retirando a variável **migração** com aqueles onde foram utilizadas todas as variáveis do conjunto de dados.

Figura 4-3 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Migração (Partição 01)

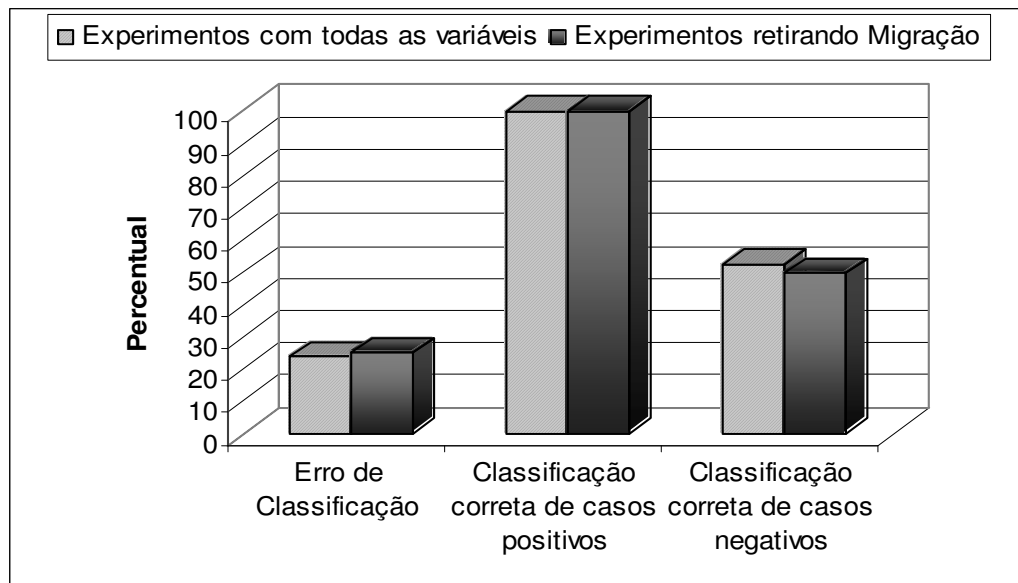


Tabela 4-4 Resultados retirando a variável Migração (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	22,17593	1,48742	30,37037	2,30140	25,64815	2,86389	100,0	0,0	48,70370	5,72778
8	22,77778	1,58583	30,27778	2,31173	25,09259	3,36547	100,0	0,0	49,81481	6,73094
12	20,97222	1,46483	31,57408	2,11369	26,94445	2,52446	99,81481	0,58561	46,29630	5,58974
16	21,38889	0,97112	31,94445	2,14944	27,22222	1,81023	100,0	0,0	45,55556	3,62046
20	20,78704	0,98693	31,75926	1,38373	27,50000	1,23842	100,0	0,0	45,00000	2,47684

Retirando a variável Escolaridade

Com a retirada da variável **escolaridade** do conjunto de dados, o erro de classificação sofreu um crescimento estatisticamente significativo de 3,70% ($p = 0,0218$) assim como a capacidade da rede para classificação correta dos casos com diagnóstico negativo foi reduzida em 7,41% ($p = 0,0218$). O percentual de

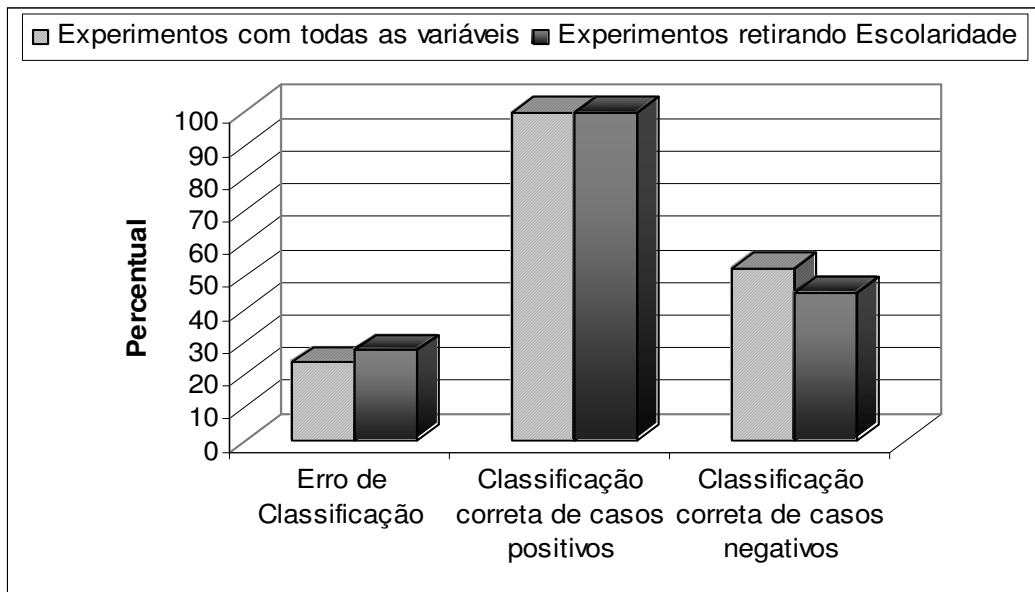
classificação correta dos casos positivos, entretanto, não sofreu alteração, mantendo-se em 100%. Esses resultados sugerem que a **escolaridade** é um fator importante, o qual pode conduzir à classificação dos indivíduos que apresentam diagnóstico negativo para TMC's.

Tabela 4-5 Resultados retirando a variável Escolaridade (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.
4	23,37963	1,60746	34,16667	2,48644	28,70370	2,39073	100,0	0,0	42,59259	4,78146
8	23,47222	1,76021	33,98148	2,22779	27,40741	3,27219	100,0	0,0	45,18519	6,54437
12	21,71296	1,71081	35,74074	2,83548	29,53704	3,47685	100,0	0,0	40,92593	6,95370
16	21,75926	1,70453	35,46297	1,80232	29,35185	1,23842	100,0	0,0	41,29630	2,47684
20	21,52778	1,23938	36,29630	1,84153	30,37037	0,85087	100,0	0,0	39,25926	1,70173

A topologia considerada foi definida com oito nodos na camada intermediária. É possível observar os resultados apresentados com esses experimentos na Tabela 4.5 e, na Figura 4.4, uma comparação com os experimentos realizados utilizando todas as variáveis do conjunto de dados.

Figura 4-4 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Escolaridade (Partição 01)



Retirando a variável Propriedade da Casa

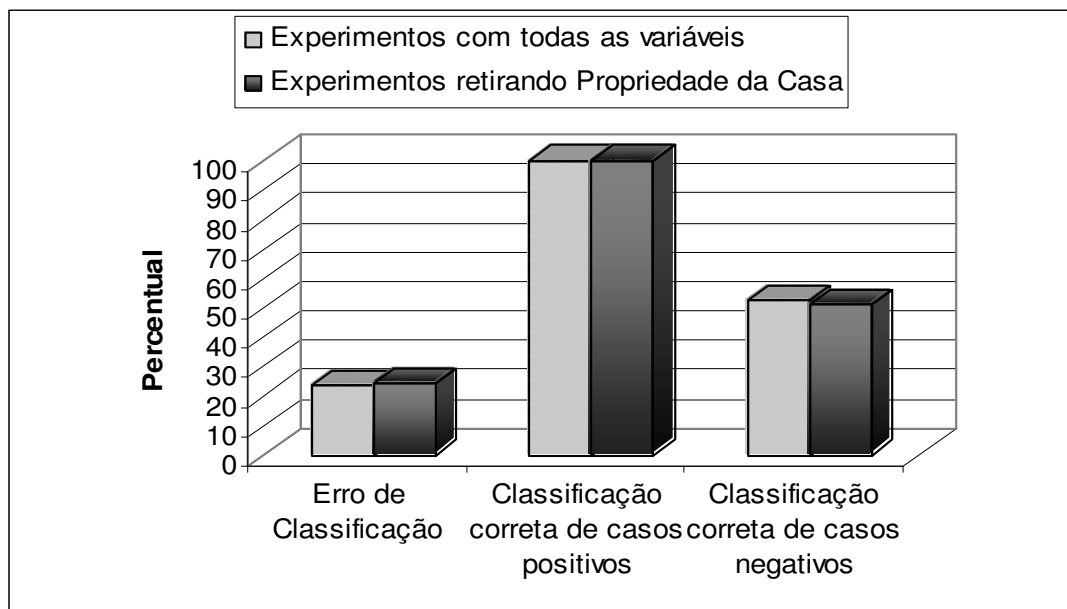
Os resultados obtidos com esses experimentos são apresentados na Tabela 4.6. A topologia considerada para a análise foi definida com oito nodos na camada

intermediária.

Tabela 4-6 Resultados retirando a variável Propriedade da Casa (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	St.dv.
4	22,63889	1,16407	31,57407	1,92500	25,83333	1,92499	100,0	0,0	48,33333	3,84999
8	22,59259	2,42829	30,92593	2,35864	24,35185	3,78134	100,0	0,0	51,29630	7,56267
12	19,95370	2,39521	33,05556	2,96200	26,85185	3,57279	99,81481	0,58561	46,48148	7,68264
16	20,46296	1,70174	33,05556	1,90510	26,94445	0,92077	100,0	0,0	46,11111	1,84154
20	20,23148	1,60449	32,12963	1,95446	27,87037	1,19137	100,0	0,0	44,25926	2,38275

Figura 4-5 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Propriedade da Casa (Partição 01)



Os resultados apresentados com esses experimentos sugerem que a variável **propriedade da casa** não é um fator importante para a classificação diagnóstica dos TMC's. Isso pode ser observado, pois as diferenças existentes entre o erro de classificação (crescimento de 0,65%) e a capacidade da rede em classificar corretamente os casos com diagnóstico negativo para TMC's (redução de 0,99%) não apresentaram significância estatística ($p = 0,6901$, para as duas situações). O percentual de classificação correta dos indivíduos que apresentavam diagnóstico positivo para TMC's foi mantida em 100%.

A comparação dos resultados obtidos nos experimentos realizados retirando a variável **propriedade da casa** com os experimentos utilizando todas as variáveis no conjunto de dados, pode ser visualizada na Figura 4.5.

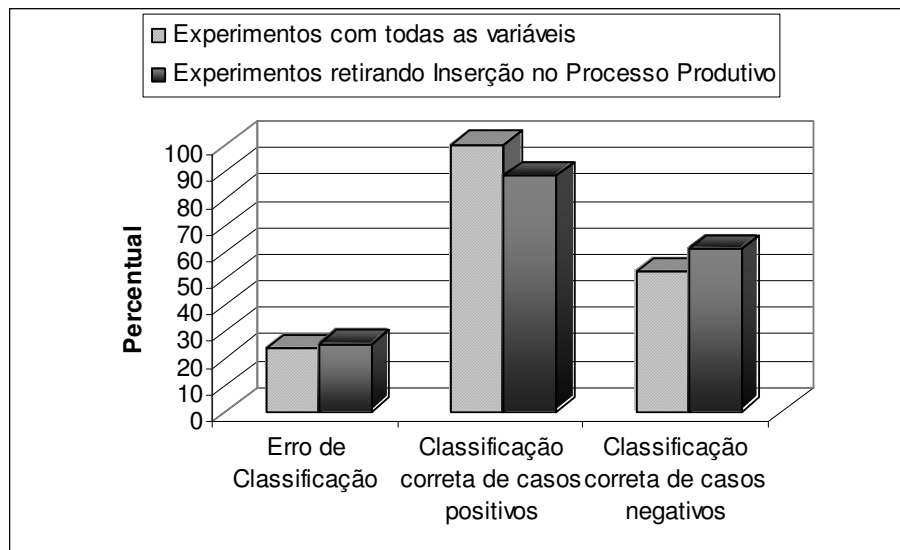
Retirando a variável Inserção no Processo Produtivo

Com a retirada da variável **inserção no processo produtivo**, foi possível observar a partir dos resultados apresentados, que essa variável é estatisticamente significativo para a classificação dos indivíduos com diagnóstico positivo para TMC's, pois a capacidade da rede em classificar corretamente os indivíduos com diagnóstico positivo foi diminuída em 11,67% ($p = 0,0001$) enquanto que a capacidade de classificação dos indivíduos com diagnóstico negativo foi melhorada em 8,33% ($p = 0,0022$).

Tabela 4-7 Resultados retirando a variável Inserção no Processo Produtivo (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	27,96296	5,03474	31,38889	2,32815	22,22222	3,67789	89,62963	2,64785	65,92593	7,51972
8	26,20371	4,04898	33,51852	3,04916	22,68519	4,23751	88,88889	2,61891	65,74074	8,87044
12	24,58333	3,52210	31,29630	3,94772	23,51852	3,60464	88,88889	3,14754	64,07407	7,20928
16	23,93519	2,90394	31,20371	2,54700	25,92593	2,96039	88,14815	2,17368	60,00000	5,60336
20	24,58334	3,05890	30,55556	2,89532	25,37037	2,86888	88,33333	1,96176	60,92593	4,48965

Figura 4-6 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Inserção no Processo Produtivo (Partição 01)



O erro de classificação, entretanto, cresceu em 1,67%, porém essa diferença não apresentou significância estatística após a realização do teste de hipótese ($p = 0,1823$). Os resultados obtidos com esses experimentos são apresentados na Tabela 4.7. A comparação desses resultados com os obtidos utilizando todas as variáveis do conjunto de dados pode ser observada na Figura 4.6.

A topologia considerada para a análise dos resultados foi definida com vinte nodos na camada intermediária. Isso mostra que houve um aumento na complexidade da rede para a classificação dos padrões, uma vez que, nos demais experimentos sempre foi utilizado uma quantidade pequena de nodos ocultos (em média oito nodos).

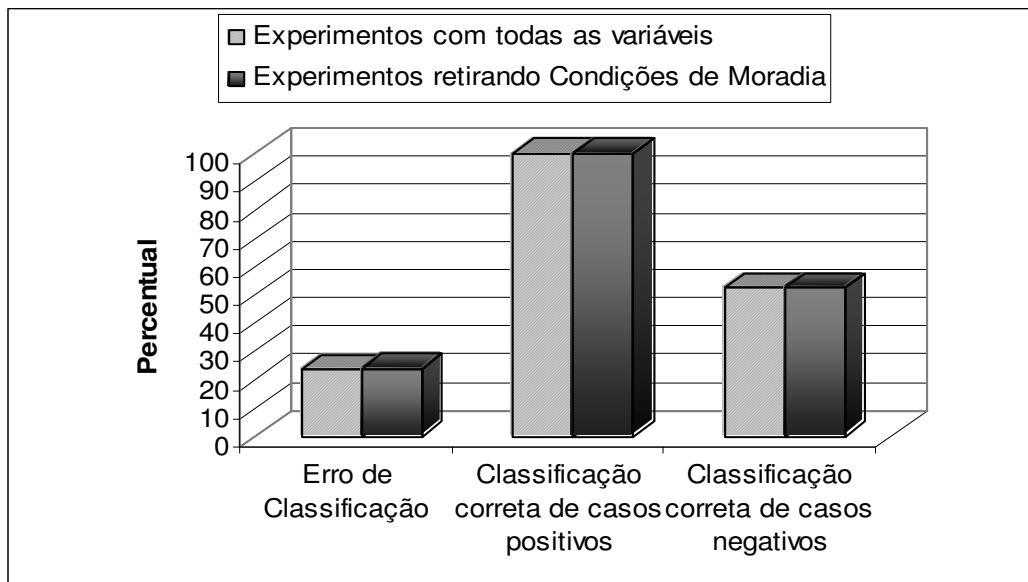
Retirando a variável Condições de Moradia

Os resultados obtidos nos experimentos realizados retirando a variável **condições de moradia**, são apresentados na Tabela 4.8 e a comparação feita com os experimentos utilizando todas as variáveis do conjunto de dados pode ser visualizada na Figura 4.7.

Tabela 4-8 Resultados retirando a variável Condições de Moradia (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	23,28704	2,49838	30,74074	2,78805	23,88889	3,37536	100,0	0,0	52,22222	6,75073
8	22,08333	1,63391	32,96296	2,35864	24,81482	2,85557	100,0	0,0	50,37037	5,71113
12	21,89815	1,63391	33,79630	1,86467	26,11111	1,67919	100,0	0,0	47,77778	3,35839
16	22,59259	1,52458	32,31481	2,59883	25,74074	2,08419	100,0	0,0	48,51852	4,16838
20	20,92593	0,89453	33,14815	1,89256	28,51852	2,61163	99,81481	0,58561	42,03704	6,05127

Figura 4-7 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Condições de Moradia (Partição 01)



A topologia considera para a análise dos dados foi definida com quatro nodos na camada intermediária da rede. A variação ocorrida no erro de classificação da rede

bem como no percentual de classificação dos casos negativos, não apresentaram significância estatística ($p = 0,8363$). O erro de classificação da rede aumentou em 0,18%, o percentual de classificação correta dos casos positivos não sofreu alteração e o percentual de classificação correta dos casos negativos foi reduzida em 0,37%.

Esses resultados obtidos com a partição 01 utilizada em nossos experimentos sugerem que a variável **condições de moradia** não apresenta possibilidade de estar relacionada com os TMC's.

Retirando a variável Sexo

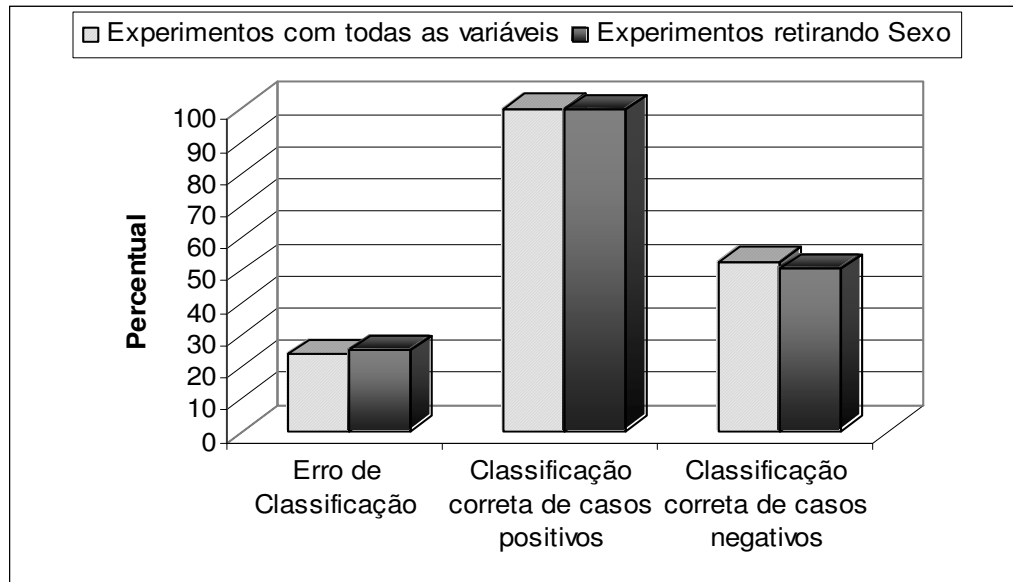
Com a retirada da variável **sexo** do conjunto de dados, os resultados obtidos nos experimentos apresentaram um aumento do erro de classificação da rede em 1,38%. A capacidade de classificação correta dos casos com diagnóstico positivo foi diminuída em 0,37% e a capacidade de classificação correta dos casos com diagnóstico negativo em 2,41%. Essas variações, entretanto, não apresentaram significância estatística ($p = 0,3566$; $p = 0,1678$; $p = 0,4447$).

Os resultados obtidos com esses experimentos sugerem que essa variável não apresenta possibilidade de estar relacionada com os TMC's. Na Tabela 4.9 podem ser visualizados os resultados obtidos com esses experimentos e a Figura 4.8 apresenta uma comparação dos resultados com os experimentos utilizando todas as variáveis.

Tabela 4-9 Resultados retirando a variável Sexo (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	23,93519	2,39123	36,20370	2,24483	25,92593	4,23188	99,81481	0,58561	48,33333	8,61769
8	23,84259	2,21761	36,01852	2,15829	25,09259	4,21949	99,62963	0,78081	50,18519	8,74933
12	22,17593	2,20900	37,31482	2,00261	28,14815	3,63097	99,81481	0,58561	43,88889	7,66278
16	22,68518	2,69068	38,24074	2,92966	27,68519	3,66362	100,0	0,0	44,62963	7,32724
20	20,37037	1,75953	39,35185	1,70453	32,03704	2,19114	100,0	0,0	35,92593	4,38228

Figura 4-8 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Sexo (Partição 01)



Retirando a variável Situação Conjugal

Nos experimentos realizados, a topologia considerada para a análise dos dados foi definida com doze nodos na camada intermediária da rede. Esses resultados são apresentados na Tabela 4.10. Na Figura 4.9 é apresentada a comparação com os resultados obtidos utilizando todas as variáveis do conjunto de dados.

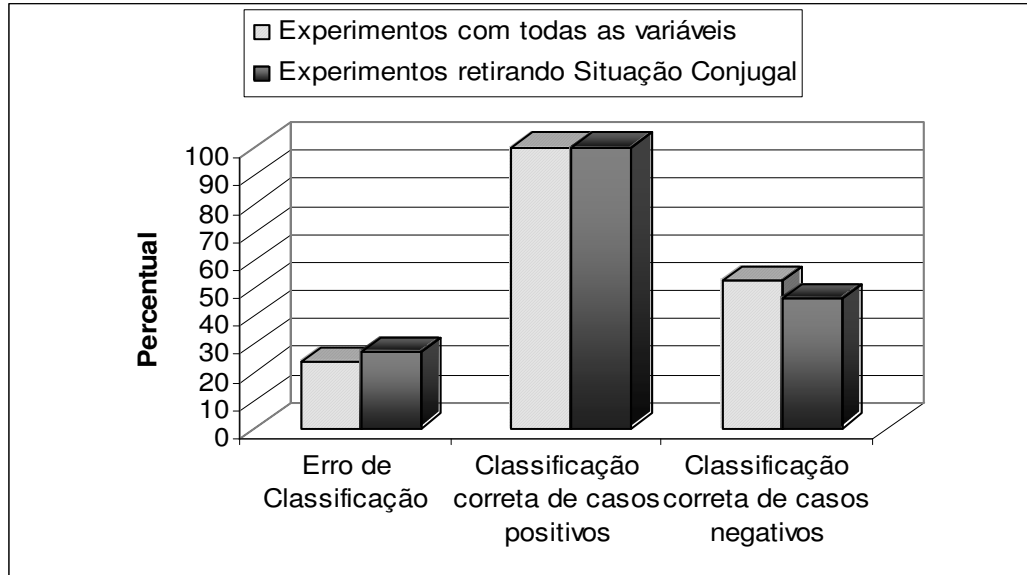
Tabela 4-10 Resultados retirando a variável Situação Conjugal (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	22,40741	2,11819	31,29630	2,21707	27,12963	2,00261	100,0	0,0	45,74074	4,00522
8	21,94444	2,15166	32,12963	2,39272	27,03704	2,53763	100,0	0,0	45,92593	5,07526
12	22,17593	1,51910	31,29630	0,73038	26,85185	1,23457	100,0	0,0	46,29630	2,46914
16	23,05556	1,68768	32,87037	1,91507	27,59259	0,95629	100,0	0,0	44,81481	1,91258
20	21,15741	1,31037	32,12963	2,27014	28,88889	2,34243	100,0	0,0	42,22222	4,68486

Com a retirada da variável **situação conjugal**, foi observado um aumento no erro de classificação da rede em 3,14%. O percentual de classificação correta dos casos negativos foi reduzido em 6,30%, porém, o percentual de classificação correta dos casos positivos não sofreu alteração. As variações observadas tanto com relação ao erro de classificação da rede ($p = 0,0020$) quanto à classificação correta dos casos negativos ($p = 0,0020$), foram estatisticamente significativos. Esses resultados

sugerem que essa variável é importante para a classificação diagnóstica para os TMC's.

Figura 4-9 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Situação Conjugal (Partição 01)



Retirando a variável Renda

A retirada da variável **renda** resultou em um aumento do erro de classificação da rede em 2,22% e uma redução do percentual de classificação correta dos casos com diagnóstico negativo em 4,44%. O percentual de classificação correta dos casos com diagnóstico positivo para TMC's, entretanto, não sofreu alteração. A topologia considerada para a análise dos resultados teve quatro nodos na camada intermediária da rede.

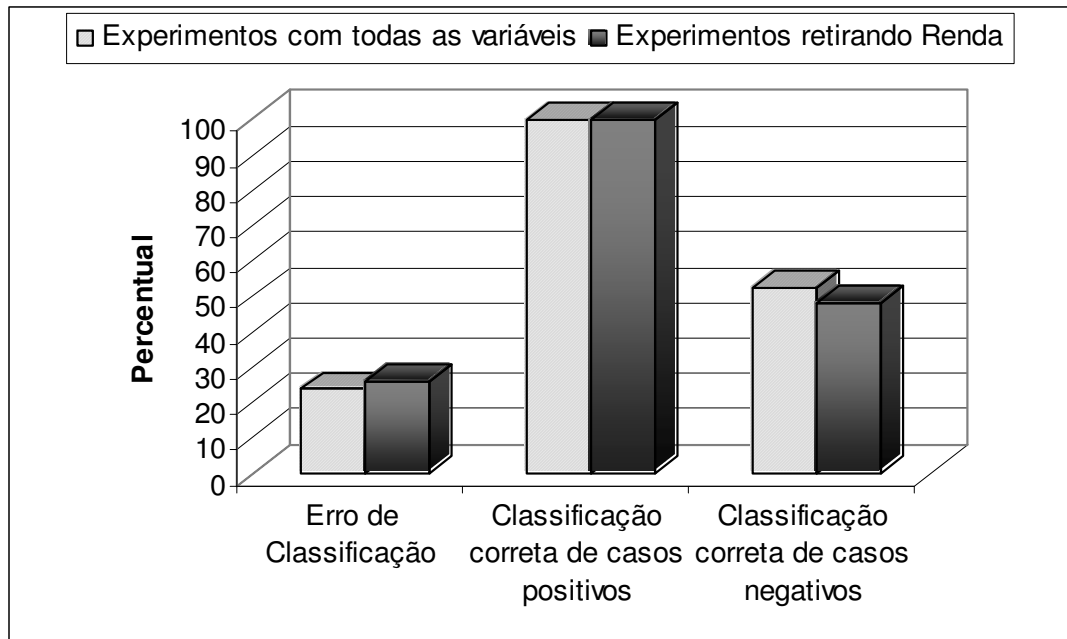
Os resultados obtidos com esses experimentos podem sugerir que a variável **renda** é importante para a classificação dos padrões com diagnóstico negativo para os TMC's, porém essas variações ocorridas no erro de classificação da rede e na capacidade de classificação correta dos casos negativos não apresentaram significância estatística ($p = 0,1076$).

Na Tabela 4.11 são apresentados os resultados obtidos nos experimentos realizados com a retirada da variável **renda**, e a Figura 4.10 uma comparação entre esses resultados e os obtidos nos experimentos utilizando todas as variáveis do conjunto de dados.

Tabela 4-11 Resultados retirando a variável Renda (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	St.dv.	Média	st.dv.
4	22,63889	2,35510	33,24074	2,84386	25,92593	3,72934	100,0	0,0	48,14815	7,45867
8	22,77778	2,12493	35,09259	3,50414	26,11111	3,59406	100,0	0,0	47,77778	7,18811
12	22,45371	1,60746	33,98148	1,95446	26,85185	3,05540	100,0	0,0	46,29630	6,11080
16	22,96296	1,74867	34,44445	2,38275	27,68519	1,47697	100,0	0,0	44,6296	2,9539
20	21,29630	0,92593	35,55556	1,70173	28,88889	2,08419	100,0	0,0	42,22222	4,16838

Figura 4-10 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Renda (Partição 01)



Retirando a variável Posse de Bens Duráveis

O erro de classificação da rede, com a retirada da variável **posse de bens duráveis**, aumentou em 0,55%. A capacidade de classificação dos casos com diagnóstico negativo, entretanto, foi reduzida em 1,11% e o percentual de classificação dos casos com diagnóstico positivo não sofreu alteração. A retirada dessa variável não apresentou significância estatística nos resultados obtidos no que se refere tanto ao erro de classificação da rede quanto ao percentual de classificação dos casos negativos ($p = 0,6577$).

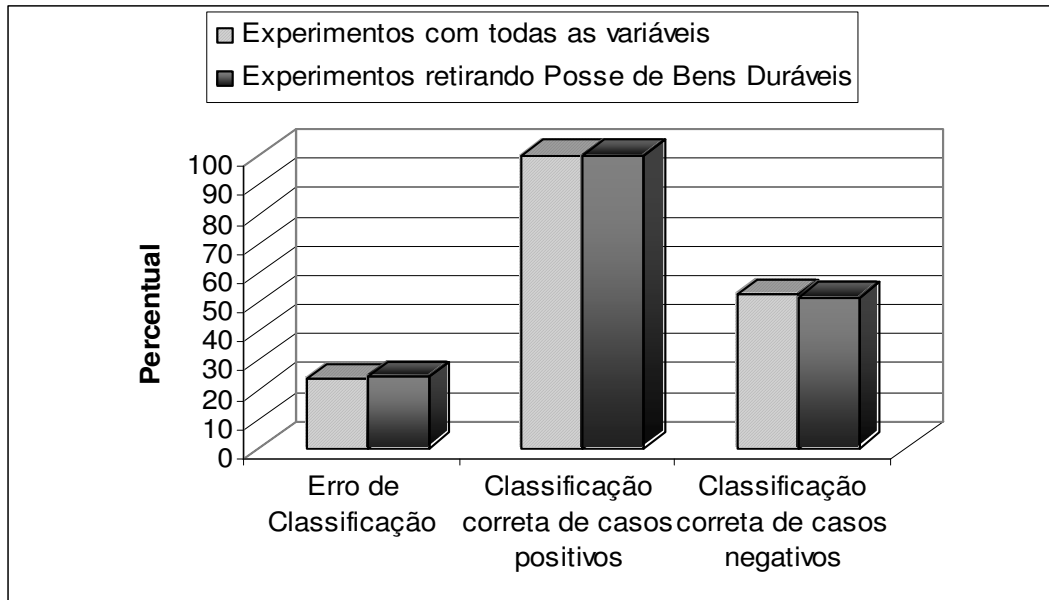
Com isso, esses resultados sugerem que a variável **posse de bens duráveis** não apresenta possibilidade de estar relacionada aos TMC's. A topologia considerada para a análise foi definida com quatro nodos na camada intermediária da rede.

Tabela 4-12 Resultados retirando a variável Posse de Bens Duráveis (Partição 01)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	22,91667	1,62221	32,31482	2,28687	24,25926	3,37536	100,0	0,0	51,48148	6,75073
8	23,24074	1,84928	32,87037	2,19331	25,55556	2,99876	100,0	0,0	48,88889	5,99751
12	21,99074	1,53159	32,87037	0,89984	26,48148	1,81023	100,0	0,0	47,03704	3,62046
16	22,91667	1,40169	33,42593	2,24483	26,38889	1,25371	100,0	0,0	47,22222	2,50742
20	21,66667	1,50888	33,14815	0,95629	27,50000	1,57680	100,0	0,0	45,00000	3,15359

Os resultados obtidos com esses experimentos podem ser observados na Tabela 4.12. Na Figura 4.11 pode ser visualizada uma comparação com os resultados obtidos utilizando todas as variáveis do conjunto de dados.

Figura 4-11 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Posse de Bens Duráveis (Partição 01)



4.2.2 Experimentos realizados com a partição 02

Os experimentos realizados com a partição 02, utilizando todas as variáveis do conjunto de dados, bem como os experimentos retirando variáveis, são apresentados nesta Seção. Também é apresentada uma comparação dos resultados obtidos com a retirada de cada variável do conjunto, em relação aos experimentos utilizando todas as variáveis.

4.2.2.1 Utilizando todas as variáveis do conjunto de dados

Os resultados obtidos com os experimentos realizados utilizando todas as variáveis do conjunto, são apresentados na Tabela 4.13. A topologia considerada em nossa análise, foi definida com quatro nodos na camada intermediária.

Nesses experimentos, mais uma vez, o erro de classificação da rede foi predominante para a classificação dos casos negativos para TMC's. O percentual de classificação dos casos positivos, nesse caso, também foi alta (95,56%). O percentual de classificação dos casos negativos, porém, foi igual à obtida na partição 01, 52,59%.

Essa diferença existente entre o percentual de classificação dos casos positivos e dos casos negativos também foi apresentada nos experimentos com a partição 01, sugerindo dessa forma, que a base de dados pode não possuir informação suficiente para uma melhor classificação dos padrões com diagnóstico negativo para TMC's.

Tabela 4-13 Resultados com uma MLP e todas as variáveis do conjunto (partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	25,83333	3,51093	39,90741	5,22782	25,92593	4,38663	95,55556	2,92152	52,59259	7,04894
8	22,82408	1,92809	43,61111	2,87717	23,88889	2,57490	98,51852	1,17121	53,70370	5,58974
12	22,82407	2,01269	43,33333	3,80019	24,72222	2,27015	98,14815	1,74594	52,40741	3,81020
16	30,55556	14,39245	45,00000	6,16048	35,55555	17,95543	82,03704	33,86690	46,85185	22,71488
20	25,13889	8,96404	44,62963	2,64785	27,50000	9,72149	92,03704	21,94703	52,96296	4,38228

4.2.2.2 Retirando variáveis do conjunto de dados

Os resultados dos experimentos realizados com a partição 02, retirando cada variável do conjunto de dados, além de uma comparação com os resultados obtidos nos experimentos utilizando todas as variáveis, são encontrados nesta Seção.

Retirando a variável Idade

Retirando a variável **idade** do conjunto, partição 02, a topologia considerada para a análise dos dados foi definida com oito nodos na camada intermediária. Os resultados obtidos são apresentados na Tabela 4.14 e uma comparação com os experimentos utilizando todas as variáveis do conjunto pode ser observada na

Figura 4.12.

Figura 4-12 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Idade (Partição 02)

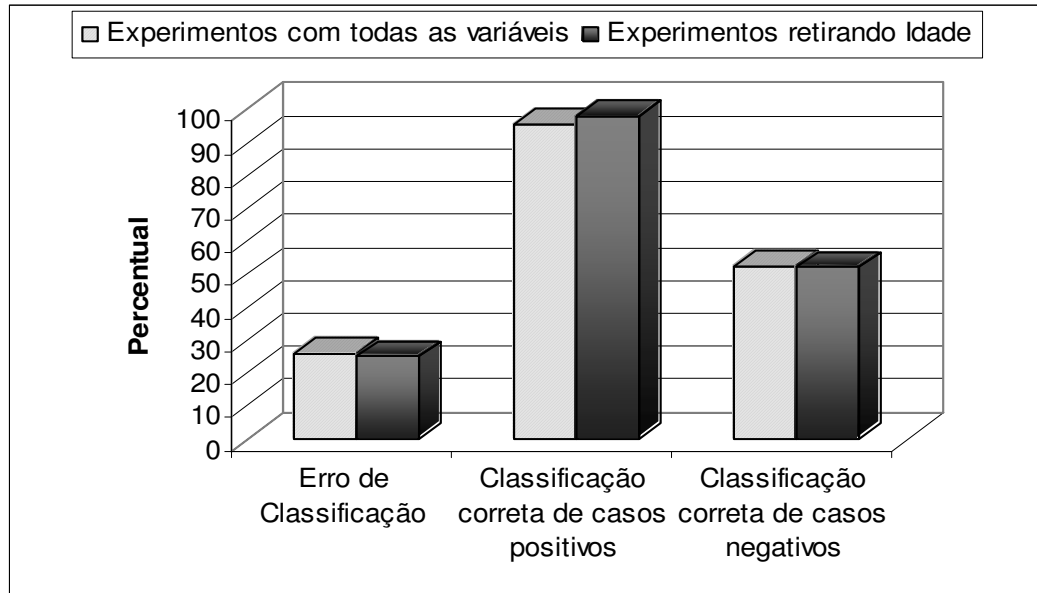


Tabela 4-14 Resultados retirando a variável Idade (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	St.dv.
4	27,68518	7,51529	41,85185	4,63580	29,53704	7,62852	86,11111	30,51968	54,81481	16,12518
8	24,58333	2,04089	41,01852	4,84184	25,00000	3,57279	97,96296	1,84154	52,03704	6,38225
12	28,70371	11,90774	41,01852	4,55809	31,11111	15,50722	89,62963	26,99444	48,14815	15,54281
16	22,12963	0,94628	44,53704	2,81017	24,53704	1,58883	99,07407	0,97601	51,85185	2,89532
20	25,37037	9,51499	46,11111	4,77747	27,22222	9,42445	95,00000	13,24070	50,55556	6,35832

A partir da retirada dessa variável foi possível observar uma redução no erro de classificação da rede em 0,93%. O percentual de classificação dos casos positivos foi melhorado em 2,41% e o percentual de classificação dos casos negativos, por sua vez, foi reduzido em 0,55%. Essas variações, entretanto, não apresentaram significância estatística ($p = 0,60537$; $p = 0,07680$; $p = 0,84798$). Com isso, os resultados sugerem que a retirada da variável **idade** do conjunto de dados não apresenta possibilidade de estar relacionada com os TMC's.

Retirando a variável Alfabetização

Nos experimentos realizados retirando a variável **alfabetização**, as variações observadas não apresentaram significância estatística. O erro de classificação da

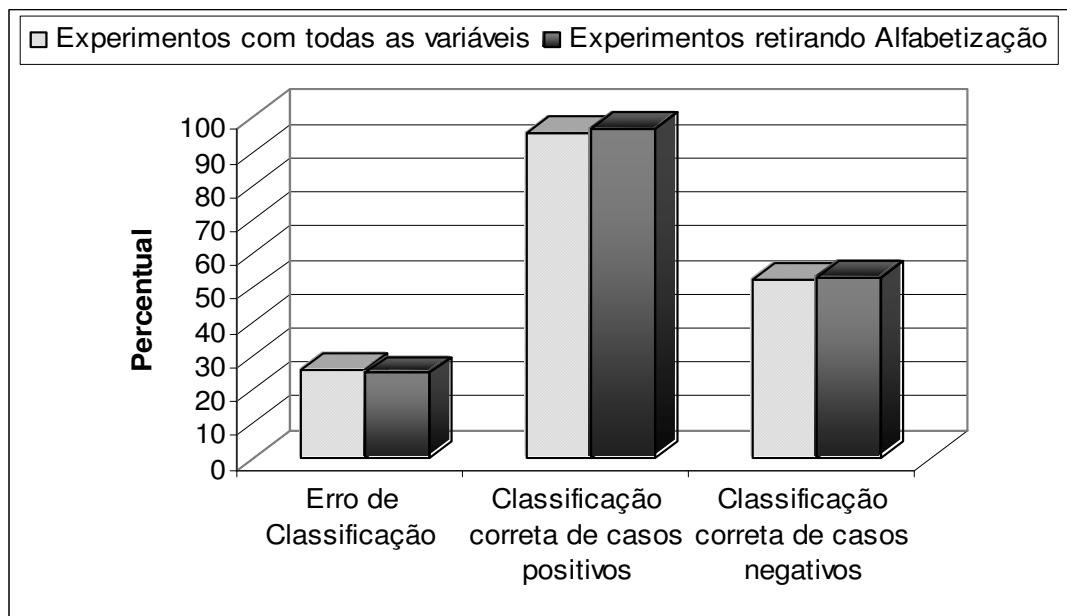
rede foi reduzido em 0,93% ($p = 0,73287$), o percentual de classificação correta dos casos positivos da rede cresceu em 1,48% ($p = 0,46643$) e o percentual de classificação correta dos casos negativos foi aumentado em 0,37% ($p = 0,92534$). A topologia considerada foi a mesma utilizada com a partição 01, com oito nodos na camada intermediária da rede.

Na Figura 4.13 é possível observar uma comparação feita entre esses resultados com os obtidos nos experimentos utilizando todas as variáveis do conjunto de dados. Na Tabela 4.15, os resultados são apresentados. Esses resultados sugerem que a variável **alfabetização** não é um fator importante para classificação diagnóstica dos TMC's.

Tabela 4-15 Resultados retirando a variável Alfabetização (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	St.dv.
4	26,11111	3,62046	41,66667	4,55704	29,62963	3,40906	92,59259	4,53609	48,14815	8,14253
8	24,16667	2,22351	40,92593	3,08024	25,00000	5,93685	97,03704	5,10520	52,96296	7,91473
12	26,94445	9,49745	43,05555	5,20590	30,83334	9,16672	92,22222	20,70985	46,11111	12,72414
16	22,08333	0,81805	44,35185	1,54012	24,53704	1,25371	99,81481	0,58561	51,11111	2,64785
20	27,87037	12,71028	46,29630	5,21961	31,20370	11,05136	92,22222	23,30686	45,37037	17,22322

Figura 4-13 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Alfabetização (Partição 02)



Retirando a variável Migração

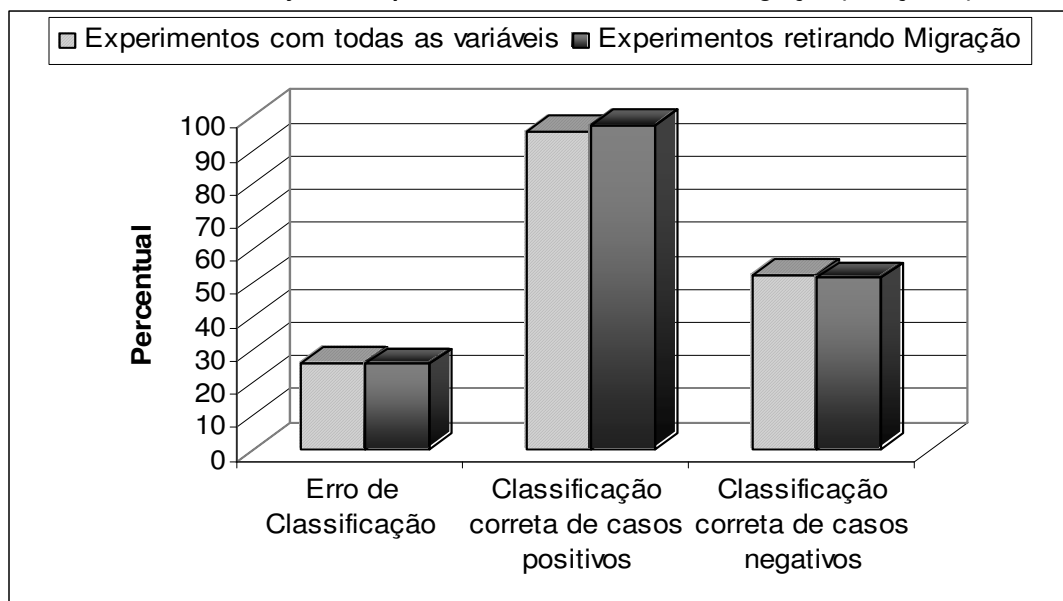
O erro de classificação da rede, nos experimentos realizados retirando a variável **migração**, foi reduzido em 0,37%. O percentual de classificação correta dos casos positivos melhorou em 1,85% e o percentual de classificação correta dos casos negativos também foi reduzida em 0,37%. Essas diferenças, entretanto, não apresentaram significância estatística ($p = 0,88914$; $p = 0,37208$; $p = 0,78807$).

Tabela 4-16 Resultados retirando a variável Migração (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	24,95370	3,26235	42,50000	4,88883	29,90741	4,84185	93,51852	4,63991	46,66667	8,40957
8	24,07407	2,41057	42,22222	3,12324	25,55556	5,25961	97,40741	5,10520	51,48148	7,44845
12	26,71296	9,91277	43,98148	4,68587	30,46297	8,73462	92,40741	20,12478	46,66667	11,77055
16	24,30556	9,07809	46,01852	2,09558	27,31482	8,08677	89,81481	31,56308	55,55556	15,85831
20	27,50000	13,02564	47,50000	4,18776	31,94445	11,18162	92,22222	23,95189	43,88889	16,72486

Os resultados obtidos são apresentados na Tabela 4.16, e uma comparação com os experimentos realizados utilizando todas as variáveis do conjunto, pode ser visualizada na Figura 4.14. Esses resultados podem sugerir que a variável **migração** não é significativa para a classificação diagnóstica dos TMC's.

Figura 4-14 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Migração (Partição 02)



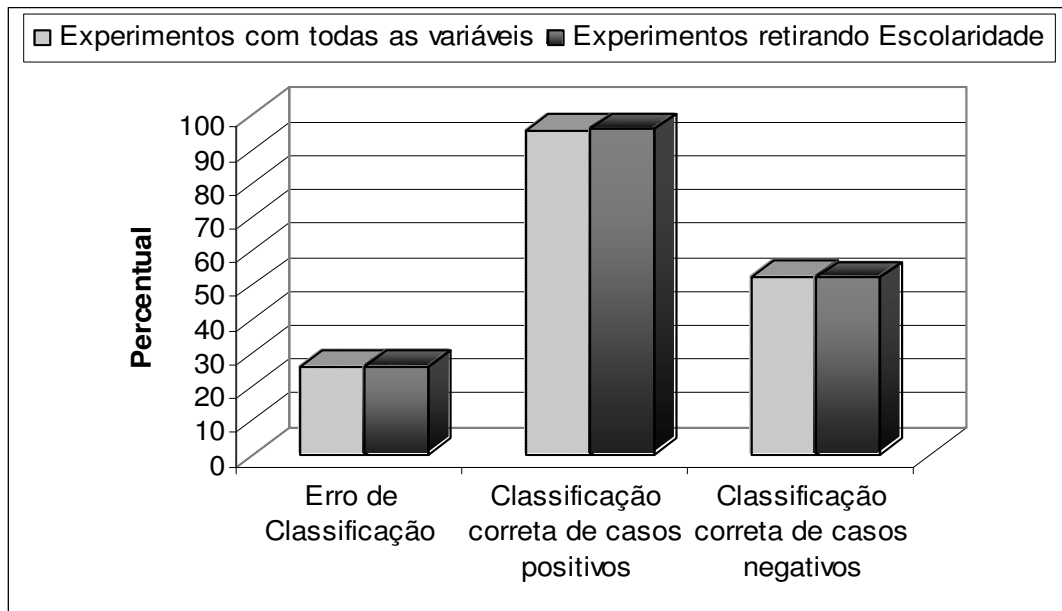
Retirando a variável Escolaridade

A retirada da variável **escolaridade** do conjunto de dados, não provocou uma alteração no erro de classificação da rede, e, além disso, a diferença existente entre o percentual de classificação correta dos casos com diagnósticos positivo e negativo foi pequena, não apresentando significância estatística. O percentual de classificação correta dos casos positivos cresceu em 0,37% ($p = 0,82266$) e o percentual de classificação correta dos casos negativos foi reduzida em 0,37% ($p = 0,90953$).

Tabela 4-17 Resultados retirando a variável Escolaridade (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.
4	26,75926	3,16866	42,68518	4,01116	29,53704	3,98734	93,14815	4,62345	47,77778	7,65034
8	25,37037	2,51501	42,31482	2,96200	25,92593	4,20932	95,92593	4,16838	52,22222	5,71113
12	27,77778	9,94622	45,18519	4,51082	30,83333	8,93517	93,14815	17,13005	45,18519	13,61388
16	23,79630	9,38495	46,75926	2,19331	28,24074	7,79894	88,51852	31,13217	55,00000	16,07429
20	33,84259	17,16567	47,03704	4,69704	36,48148	15,12786	80,55556	33,84664	46,48148	19,35261

Figura 4-15 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Escolaridade (Partição 02)



Esses resultados diferem dos apresentados com a partição 01, onde, os resultados obtidos podem indicar que a **escolaridade** é um fator estatisticamente significativo para a classificação diagnóstica dos TMC's.

Os resultados obtidos nesses experimentos são apresentados na Tabela 4.17. Na Figura 4.15 é apresentada uma comparação entre esses resultados e os obtidos nos experimentos utilizando todas as variáveis do conjunto de dados. A topologia considerada foi com oito nodos na camada intermediária da rede.

Retirando a variável Propriedade da Casa

A topologia considerada para a análise dos resultados foi com oito nodos na camada intermediária da rede. Nesses experimentos foi observada uma redução tanto no erro de classificação da rede (em 0,74%), quanto no percentual de classificação correta dos casos com diagnóstico negativo (em 1,11%), porém essas variações não foram estatisticamente significativas ($p = 0,76328$; $p = 0,79805$). A variação no percentual de classificação correta dos casos com diagnóstico positivo, entretanto, foi estatisticamente significativa ($p = 0,02946$), sendo melhorada em 2,59%. Esses resultados podem sugerir que a retirada dessa variável do conjunto de dados é significativa para a classificação diagnóstica dos TMC's.

Os resultados obtidos com esses experimentos podem ser observados na Tabela 4.18. Na Figura 4.16, é apresentada uma comparação desses resultados com aqueles obtidos nos experimentos utilizando todas as variáveis do conjunto de dados.

Figura 4-16 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Propriedade da Casa (Partição 02)

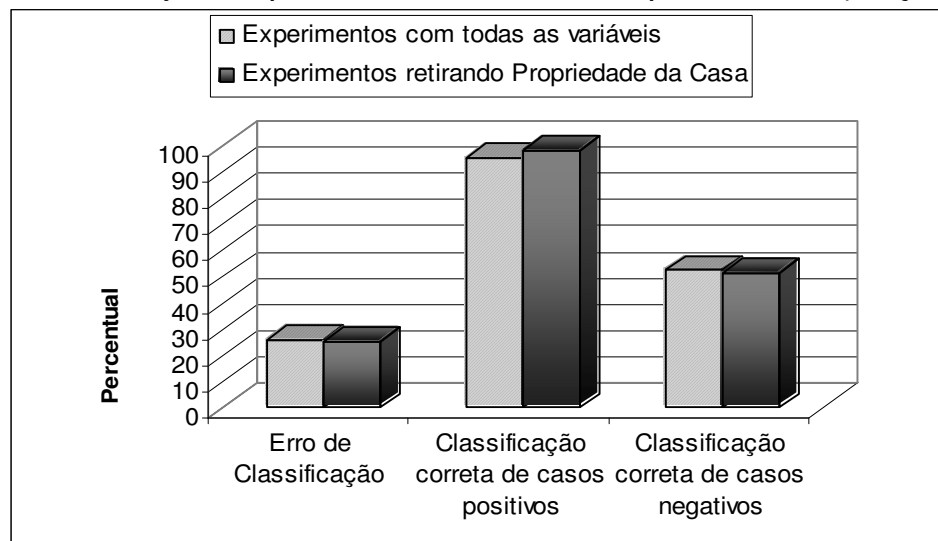


Tabela 4-18 Resultados retirando a variável Propriedade da Casa (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.
4	26,25000	3,21529	40,46296	3,73061	26,94444	3,00669	95,18519	4,20479	50,92593	6,99740
8	24,35185	2,01446	40,64815	3,71526	25,18519	4,99199	98,14815	1,23457	51,48148	9,47486
12	27,03704	9,89870	43,70370	3,56745	30,64815	10,46155	91,66667	23,11726	47,03704	16,07785
16	24,72222	8,96178	45,46296	1,97386	27,31482	8,16882	89,25926	31,37600	56,11111	15,66613
20	32,17593	17,06660	46,57407	5,59060	35,64815	15,40275	82,40741	34,93249	46,29630	20,07834

Retirando a variável Inserção no Processo Produtivo

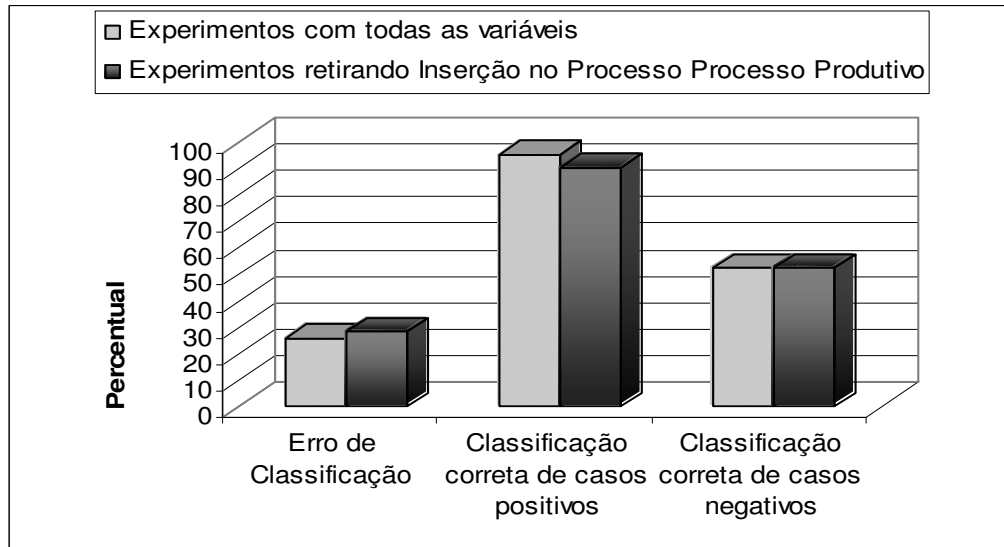
Quando realizamos os experimentos com a partição 02, foi possível observar uma diferença nos resultados obtidos com a partição 01, onde foi retirada a mesma variável, no que se refere ao percentual de classificação correta dos casos positivos e negativos para os TMC's.

Com a partição 01 houve uma variação estatisticamente significativa ($p = 0,00225$) no percentual de classificação correta dos casos negativos (aumento de 8,33%), porém o mesmo não ocorreu com a partição 02, onde o percentual de classificação correta dos casos com diagnóstico negativo para TMC's foi reduzido em 0,18%. Na partição 02, entretanto, só foi observada uma diferença estatisticamente significativa ($p = 0,01556$) com relação ao percentual de classificação correta dos casos com diagnóstico positivo, onde foi possível observar um aumento do mesmo em 5,0%. Com relação à variação ocorrida no percentual de classificação correta dos casos negativos, não foi estatisticamente significativo ($p = 0,93473$). O erro de classificação da rede, nos experimentos com a partição 02, também cresceu em 2,59%, porém esse resultado não apresentou significância estatística ($p = 0,12298$).

Tabela 4-19 Resultados retirando a variável Inserção no Processo Produtivo (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	28,28704	1,16408	28,33333	1,17121	30,83333	1,38374	89,44444	1,96176	48,88889	2,49981
8	28,79630	0,64741	27,22222	0,89453	29,25926	1,91258	90,18519	90,18519	51,29630	4,09925
12	28,51852	0,76229	27,22222	0,89453	28,51852	1,73500	90,55556	2,82201	52,40741	2,14723
16	28,51852	28,51852	27,22222	0,78081	28,70370	1,95202	89,81481	3,05540	52,77778	2,92803
20	28,75000	1,50334	27,22222	1,46076	28,79630	1,34179	89,44444	3,03035	52,96296	2,92152

Figura 4-17 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Inserção no Processo Produtivo (Partição 02)



Os resultados obtidos com esses experimentos podem indicar que a variável **inserção no processo produtivo** é um fator significativo para o diagnóstico dos TMC's. Esses resultados são apresentados na Tabela 4.19, e uma comparação com os experimentos realizados utilizando todas as variáveis pode ser analisada na Figura 4.17.

Retirando a variável Condições de Moradia

A partir dos resultados obtidos nos experimentos realizados retirando a variável **condições de moradia** do conjunto de dados, podemos observar que o erro de classificação da rede foi reduzido em 0,28% e apesar do percentual de classificação correta dos casos positivos ter sido melhorado em 1,67%, o percentual de classificação correta dos casos negativos para TMC's foi diminuído na mesma proporção. Essas variações, porém, não apresentaram significância estatística ($p = 0,88409$); $p = 0,08875$; $p = 0,60436$).

A topologia considerada para a análise foi definida com oito nodos na camada intermediária. Esses resultados são apresentados na Tabela 4.20 e na Figura 4.18 pode ser visualizada uma comparação com os resultados obtidos nos experimentos realizados utilizando todas as variáveis do conjunto.

Figura 4-18 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Condições de Moradia (Partição 02)

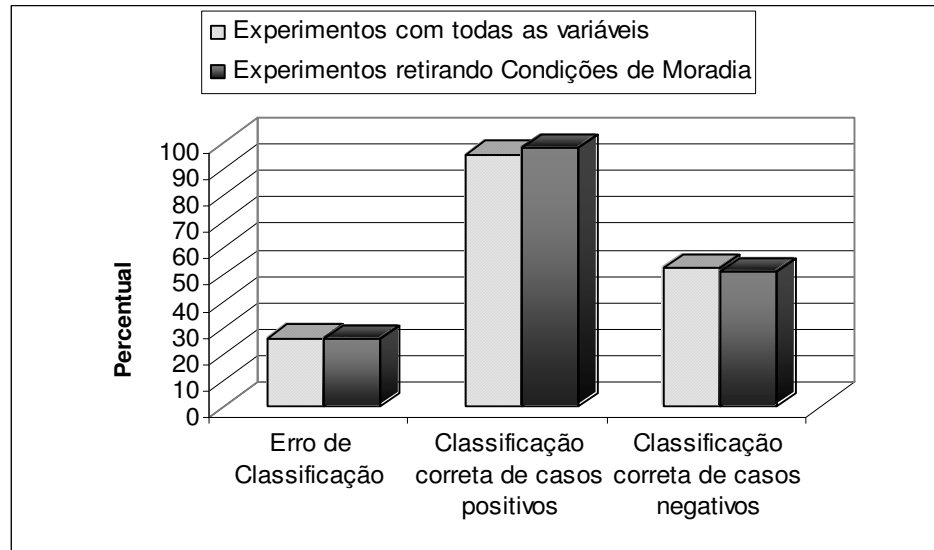


Tabela 4-20 Resultados retirando a variável Condições de Moradia (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	25,18519	2,25963	41,29630	3,83512	27,22222	3,18365	95,92593	4,34736	49,62963	3,88448
8	24,53704	2,03563	40,37037	4,34736	25,64815	3,46587	97,77778	1,91258	50,92593	6,00069
12	28,19445	11,10779	41,94445	3,85617	41,94445	10,99953	91,29630	21,75521	48,88889	13,91835
16	24,90741	8,92344	45,18519	2,92152	27,77778	7,97707	89,07407	31,31096	55,37037	15,96488
20	31,25000	15,02590	44,90741	3,38804	33,14815	14,89689	83,14815	31,94944	50,55556	16,42602

Retirando a variável Sexo

A topologia considerada para a análise com esses experimentos foi definida com oito nodos na camada intermediária. Esses resultados são apresentados na Tabela 4.21 e uma comparação feita com os experimentos realizados utilizando todas as variáveis do conjunto é apresentada na Figura 4.19.

Ao analisar os resultados obtidos com a partição 02, retirando a variável **sexo** do conjunto, não foi observada significância estatística nas variações existentes. Houve um crescimento no erro de classificação da rede em 3,42% ($p = 0,33893$), uma redução de 3,15% ($p = 0,18792$) no percentual de classificação correta dos casos com diagnóstico positivo, além de uma redução no percentual de classificação correta dos casos com diagnóstico negativo para TMC's de 3,70% ($p = 0,49088$).

Figura 4-19 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Sexo (Partição 02)

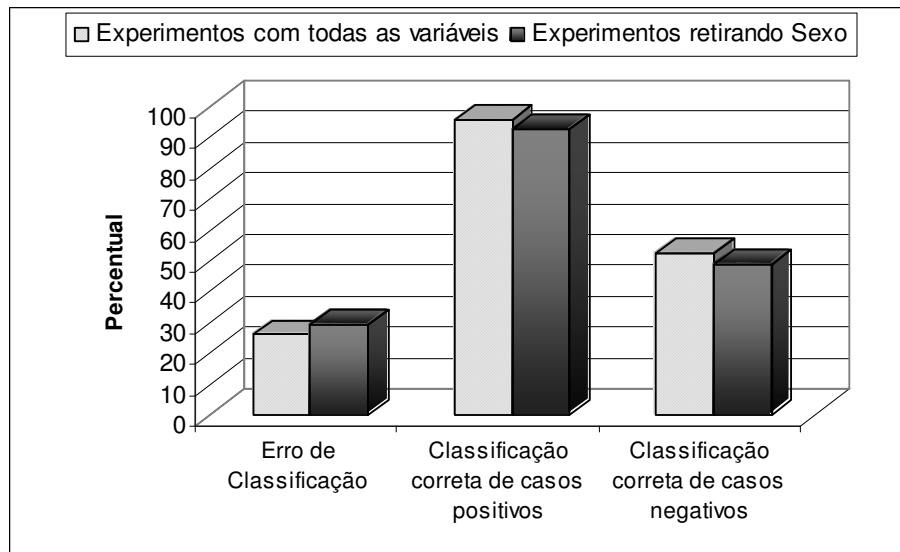


Tabela 4-21 Resultados retirando a variável Sexo (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	26,57408	7,91172	44,72222	3,41046	32,31482	8,26389	80,00000	28,76570	55,37037	17,51281
8	26,38889	8,62542	44,72222	2,54700	29,35185	10,44880	92,40741	8,06967	48,88889	14,61282
12	30,41667	13,90611	45,37037	2,86223	30,27778	14,91894	76,85185	41,31701	51,85185	22,88118
16	28,00926	11,60233	47,03704	2,08419	30,27778	10,88636	78,33333	41,31701	61,11111	21,25824
20	30,97222	16,32516	48,24074	4,19685	34,62963	13,39805	81,11111	30,67597	49,62963	15,11527

Retirando a variável Situação Conjugal

Os resultados obtidos nos experimentos retirando a variável **situação conjugal** são apresentados na Tabela 4.22. A topologia considerada para a análise foi definida com oito nodos na camada intermediária da rede.

Com esses experimentos foi possível observar uma redução no erro de classificação da rede em 4,63%, uma melhora no percentual de classificação correta dos casos negativos em 6,48% e uma redução no percentual de classificação correta dos casos positivos em 1,67%. Essas variações foram estatisticamente significativas com relação ao percentual de classificação correta dos casos negativos e o erro de classificação da rede ($p = 0,01298$; $p = 0,02557$), porém o mesmo não aconteceu com relação ao percentual de classificação dos casos positivos ($p = 0,06169$).

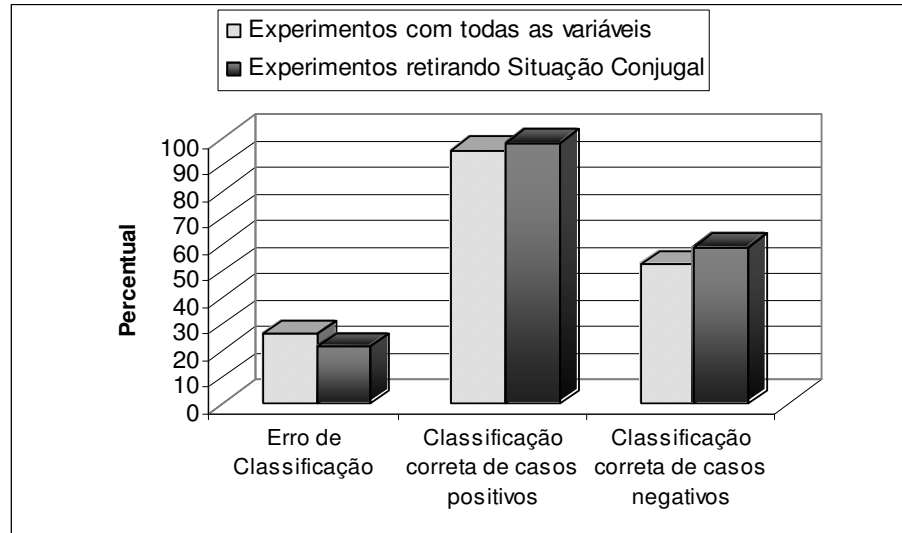
Com isso, esses resultados sugerem uma relação da variável **situação conjugal**

com a classificação diagnóstica dos TMC's. Uma comparação dos resultados é feita com os experimentos realizados utilizando todas as variáveis no conjunto de dados, e apresentada na Figura 4.20.

Tabela 4-22 Resultados retirando a variável Situação Conjugal (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	26,89815	7,84901	44,81481	5,72779	24,44444	9,89774	85,55556	30,91108	65,55556	15,37520
8	24,30556	2,72803	41,38889	3,92958	21,29630	2,86223	98,33333	1,84154	59,07407	5,34227
12	26,01852	9,47988	43,24074	3,85617	23,70370	9,86883	98,14815	2,61891	54,44444	17,61586
16	25,00000	9,08268	44,81481	3,21344	25,37037	8,85109	89,44444	31,43970	59,81481	14,47791
20	28,42593	14,85687	44,90741	2,51690	28,14815	14,43837	86,11111	31,13768	57,59259	11,66160

Figura 4-20 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Situação Conjugal (Partição 02)



Retirando a variável Renda

Os melhores resultados nos experimentos retirando a variável **renda** do conjunto de dados foram obtidos com uma topologia com oito nodos na camada intermediária da rede. Esses resultados são apresentados na Tabela 4.23 e uma comparação com os resultados obtidos nos experimentos utilizando todas as variáveis pode ser visualizada na Figura 4.21.

A retirada dessa variável resultou em uma melhora no erro de classificação da rede em 4,35% e no percentual de classificação correta dos casos com diagnóstico negativo para os TMC's em 7,78%, porém, após a realização do teste de hipótese, o resultado obtido não apresentou significância estatística

($p = 0,07831$; $p = 0,09541$) O percentual de classificação correta dos casos positivos foi reduzido em 0,92%, e essa diferença também não foi estatisticamente significativa ($p = 0,51294$). Esses resultados sugerem que a variável **renda** não apresenta possibilidade de estar relacionada aos TMC's.

Figura 4-21 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Renda (Partição 02)

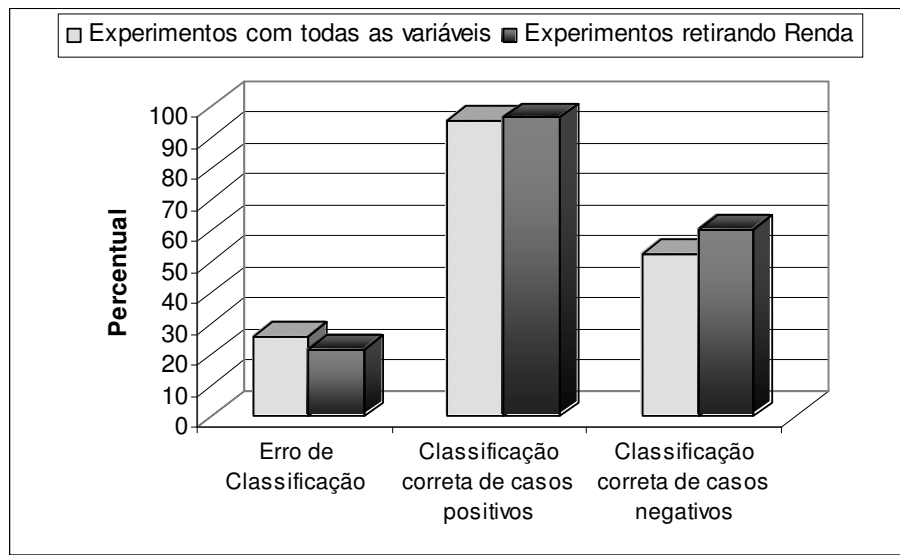


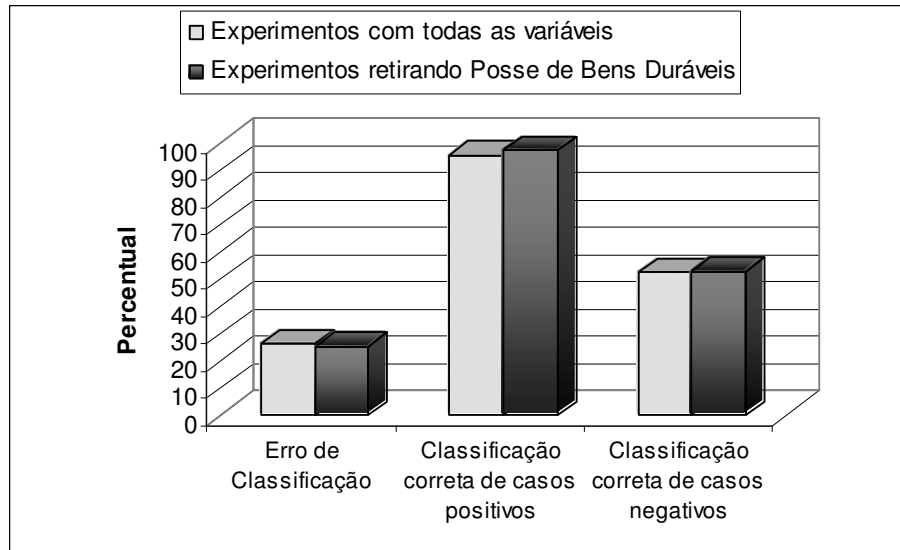
Tabela 4-23 Resultados retirando a variável Renda (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	27,91667	7,70694	41,38889	3,83139	29,35185	8,46433	85,00000	29,91695	56,29630	17,02182
8	26,15741	2,71052	40,37037	3,73444	21,57408	4,45237	96,48148	2,68359	60,37037	9,28802
12	24,95370	2,76446	42,77778	3,14149	22,96296	3,23117	97,96296	2,82201	56,11111	4,86442
16	25,32407	9,00909	44,72222	2,99399	26,75926	8,66894	88,88889	31,26895	57,59259	15,89311
20	29,12037	14,87009	46,01852	1,90510	30,55556	12,74060	85,55556	30,89875	53,33333	13,74202

Retirando a variável Posse de Bens Duráveis

Nesses experimentos, o erro de classificação da rede foi reduzido em 0,92%, e, tanto o percentual de classificação correta dos casos positivos quanto o percentual de classificação correta dos casos negativos para TMC's apresentaram uma melhora em 1,66% e 0,18%, respectivamente. Essas variações observadas não apresentaram significância estatística ($p = 0,61517$; $p = 0,24396$; $p = 0,95318$). Com isso, esses resultados podem sugerir que a variável **posse de bens duráveis** não é importante para a classificação dos TMC's.

Figura 4-22 Comparação de resultados dos experimentos com MLP realizados utilizando todas as variáveis do conjunto e experimentos retirando a variável Posse de Bens Duráveis (Partição 02)



Os resultados obtidos com esses experimentos são apresentados na Tabela 4.24, e a comparação feita com os experimentos utilizando todas as variáveis do conjunto de dados pode ser observada na Figura 4.22. A topologia considerada para a análise foi definida com oito nodos na camada intermediária da rede.

Tabela 4-24 Resultados retirando a variável Posse de Bens Duráveis (Partição 02)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	26,71296	7,81130	43,51852	3,92837	32,12963	6,35608	86,29630	29,54205	49,44444	18,47835
8	24,81482	2,50932	40,00000	3,17166	25,00000	2,65504	97,22222	2,18243	52,77778	5,32799
12	24,02778	2,07560	42,68519	3,66362	24,07408	2,13834	98,88889	1,99068	52,96296	4,29445
16	25,09259	8,86721	45,37037	3,17766	27,50000	8,04899	89,44444	31,43970	55,55556	15,85831
20	26,48148	12,11356	45,18519	2,42240	29,25926	10,94004	90,18519	27,16120	51,29630	5,72778

Um quadro comparativo das diferenças obtidas nos experimentos realizados retirando variáveis do conjunto de dados com as duas partições utilizadas, pode ser visualizado na Tabela 4.25.

Tabela 4-25 Quadro comparativo das diferenças obtidas nos experimentos realizados retirando variáveis (partições 01 e 02), com seus respectivos valores de p . Os resultados são equivalentes à comparação feita com experimentos utilizando todas as variáveis do conjunto. O sinal “+” significa que houve um aumento no percentual e o sinal “-” que houve uma redução.

Partição 01						
Experimentos Retirando Variáveis	Erro de Classificação		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Diferença	P	Diferença	p	Diferença	p
Idade	+2,22%	0,0001	0,0	0,0	-4,44%	0,0001
Alfabetização	-0,83%	0,4872	0,0	0,0	-1,67%	0,4872
Migração	+1,39%	0,3479	0,0	0,0	-2,78%	0,3479
Escolaridade	+3,70%	0,0218	0,0	0,0	-7,41%	0,0218
Propriedade da Casa	+0,65%	0,6901	0,0	0,0	-0,99%	0,6901
Inserção no Processo Produtivo	+1,67%	0,1823	-11,67%	0,0001	+8,33%	0,0022
Condições de Moradia	-0,18%	0,8363	0,0	0,0	-0,37%	0,8363
Sexo	+1,38%	0,3566	-0,37%	0,1678	-2,41%	0,4447
Situação Conjugal	+3,14%	0,0020	0,0	0,0	-6,30%	0,0020
Renda	+2,22%	0,1076	0,0	0,0	-4,44%	0,1076
Posse de Bens Duráveis	+0,55%	0,6577	0,0	0,0	-1,11%	0,6577
Partição 02						
Experimentos Retirando Variáveis	Erro de Classificação		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Diferença	P	Diferença	P	Diferença	P
Idade	-0,93%	0,60537	+2,41%	0,07680	-,055%	0,84798
Alfabetização	-0,93%	0,73287	+1,48%	0,46643	+0,37%	0,92534
Migração	-0,37%	0,88914	+1,85%	0,37208	-0,37%	0,78807
Escolaridade	0,0	0,0	+0,37%	0,82266	-0,37%	0,90953
Propriedade da Casa	-0,74%	0,76328	+2,59%	0,02946	-1,11%	0,79805
Inserção no Processo Produtivo	+2,59%	0,12298	+5,0%	0,01556	-0,18%	0,93473
Condições de Moradia	-0,28%	0,88409	+1,67%	0,08875	-1,67%	0,60436
Sexo	+3,42%	0,33893	-3,15%	0,18792	-3,70%	0,49088
Situação Conjugal	-4,63%	0,02557	-1,67%	0,06169	+6,48%	0,01298
Renda	+4,35%	0,07831	-0,92%	0,51294	+7,78%	0,09541
Posse de Bens Duráveis	-0,92%	0,61517	+1,66%	0,24396	+0,18%	0,95318

4.3 Análise observando a média obtida com as duas partições

Nesta Seção são apresentados os resultados obtidos nos experimentos a partir da análise conjunta das duas partições. Os resultados são apresentados contendo cada topologia testada, bem como todos os aspectos observados, desvio padrão e o nível de significância do teste de hipótese (p).

Na Seção 4.3.1 são apresentadas as análises feitas com todas as variáveis no conjunto de dados e na Seção 4.3.2 retirando variáveis do conjunto. Posteriormente, são feitos os comentários a respeito do desempenho da rede neural para análise

dos fatores relacionados aos TMC's.

4.3.1 Utilizando todas as variáveis do conjunto de dados

Na Tabela 4.26 são apresentados os resultados com todas as variáveis do conjunto de dados. A topologia considerada para a análise foi definida com quatro nodos na camada intermediária da rede. A escolha da topologia foi feita observando o erro no conjunto de validação.

O erro de classificação da rede apresentado, nesse caso, foi 25,74%, o percentual de classificação correta dos casos com diagnóstico positivo foi 97,78% e o percentual de classificação correta dos casos com diagnóstico negativo para TMC's foi 50,74%. Esses resultados servirão de parâmetro para a comparação com os resultados obtidos nos experimentos retirando variáveis do conjunto de dados.

Tabela 4-26 Resultados com MLP utilizando todas as variáveis do conjunto (média das duas partições)

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	23,68056	3,044484	35,69445	5,958032	25,74074	0,261898	97,77778	3,142694	50,74074	2,618911
8	22,75463	0,098217	37,63889	8,446002	23,79630	0,130949	99,25926	1,047565	53,14815	0,785673
12	22,56944	0,360101	37,82407	7,791270	25,27778	0,785673	99,07408	1,309456	50,37037	2,880810
16	25,83334	6,678235	38,93519	8,576944	31,38889	5,892554	91,01852	12,701731	46,20370	0,916623
20	22,66204	3,502802	38,61111	8,511473	27,50000	0,000000	96,01852	5,630663	48,98148	5,630663

4.3.2 Retirando variáveis do conjunto de dados

Nesta Seção são apresentados os resultados dos experimentos realizados retirando cada variável do conjunto de dados, bem como uma comparação com os resultados obtidos nos experimentos utilizando todas as variáveis. Todos os resultados são analisados observando a média entre as duas partições.

Retirando a variável Idade

Com a análise feita observando conjuntamente as duas partições, foi possível observar que topologia selecionada foi definida com oito nodos na camada intermediária da rede. O erro de classificação da rede foi reduzido em 0,83%, o percentual de classificação correta dos casos positivos e negativos para TMC's

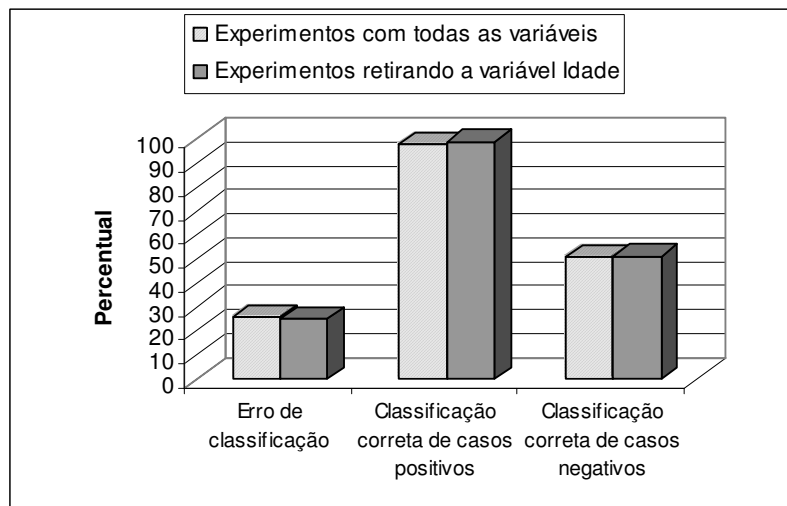
foram melhorados em 1,20% e 0,46%, respectivamente.

Essas variações, entretanto, não foram estatisticamente significativas ($p = 0,37428$; $p = 0,07886$; $p = 0,77072$). Esses resultados sugerem que a variável **idade** não apresenta possibilidade de estar relacionada aos TMC's.

Tabela 4-27 Resultados retirando a variável Idade do conjunto de dados, observando as duas partições conjuntamente

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	24,81481	4,05932	36,80556	7,13654	27,73149	2,55344	93,05556	9,82093	51,48148	4,71404
8	24,02778	0,78567	36,43519	6,48181	24,90741	0,13094	98,98148	1,44040	51,20371	1,17851
12	24,79167	5,53246	37,17593	5,43425	28,88889	3,14269	94,62963	7,07107	47,59260	0,78567
16	21,92130	0,29463	38,88889	7,98769	25,78704	1,76777	99,53704	0,65473	48,88889	4,19026
20	23,07871	3,24090	39,44445	9,42809	27,26852	0,06548	97,50000	3,53553	47,96297	3,66648

Figura 4-23 Comparação de resultados dos experimentos realizados com MLP e todas as variáveis do conjunto e experimentos retirando a variável idade. Resultados observando a média das partições.



Os resultados são apresentados na Tabela 4.27 e uma comparação com os experimentos utilizando todas as variáveis do conjunto de dados pode ser visualizada na Figura 4.23.

Retirando a variável Alfabetização

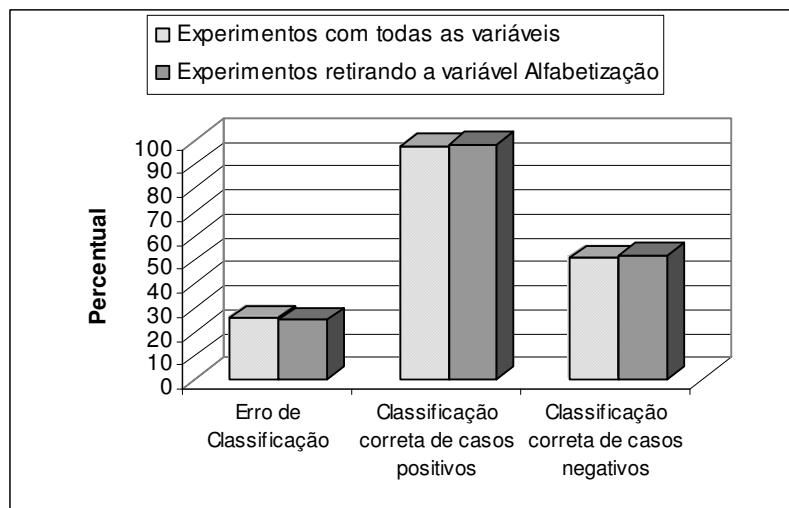
Nos experimentos realizados retirando a variável **alfabetização**, foi observada uma melhora na performance da rede. O erro de classificação foi reduzido em 0,97%, o percentual de classificação correta de casos positivos aumentou em 0,74%, assim como o percentual de classificação correta de casos negativos em 1,20%. Essas

alterações observadas nos resultados porém, não foram estatisticamente significativas ($p = 0,47942$; $p = 0,45131$; $p = 0,56601$). A topologia considerada foi definida com oito nodos na camada intermediária. Esses resultados podem sugerir que a variável **alfabetização** não apresenta possibilidade de estar relacionada aos TMC's.

Tabela 4-28 Resultados retirando a variável Alfabetização do conjunto de dados, observando a média entre as duas partições

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	23,91204	3,10996	36,38889	7,46391	27,31482	3,27364	96,29630	5,23783	49,07408	1,30946
8	23,58797	0,81841	35,64815	7,46391	24,76852	0,32736	98,51852	2,09513	51,94445	1,44040
12	23,00926	5,56520	38,05555	7,07107	28,98149	2,61892	96,11111	5,49972	45,92593	0,26189
16	21,29630	1,11304	38,84260	7,79126	25,78704	1,76777	99,90741	0,13095	48,51852	3,66648
20	24,00463	5,46698	39,16667	10,08282	29,44445	2,48796	96,11111	5,49972	45,00000	0,52378

Figura 4-24 Comparação de resultados dos experimentos realizados com MLP e todas as variáveis do conjunto e experimentos retirando a variável Alfabetização. Resultados observando a média das partições.



Na Tabela 4.28 são apresentados os resultados obtidos com esses experimentos, e uma comparação dos resultados obtidos utilizando todas as variáveis pode ser observada na Figura 4.24.

Retirando a variável Migração

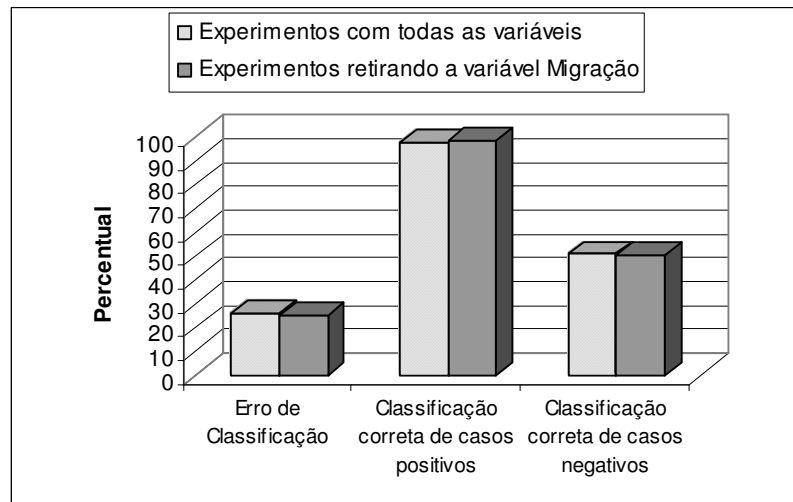
Foi possível observar nesses experimentos algumas variações nos resultados, porém estatisticamente não significativos. O erro de classificação da rede foi reduzido em 0,41% ($p = 0,76309$), o percentual de classificação dos casos com

diagnóstico positivo foi melhorado em 0,92% ($p = 0,35788$) e o percentual de classificação dos casos com diagnóstico negativo para TMC's foi reduzido em 0,09% ($p = 0,96731$). Com isso, esses resultados sugerem que essa variável não apresenta possibilidade de estar relacionada com os TMC's. A topologia considerada, mais uma vez, tinha oito nodos na camada intermediária.

Tabela 4-29 Resultados retirando a variável Migração do conjunto de dados, observando a média entre as duas partições

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	23,56482	1,96418	36,43519	8,57694	27,77778	3,01175	96,75926	4,58310	47,68519	1,44040
8	23,42593	0,91662	36,25000	8,44599	25,32408	0,32737	98,70371	1,83324	50,64815	1,17851
12	23,84259	4,05932	37,77778	8,77336	28,70371	2,48797	96,11111	5,23782	46,48149	0,26189
16	22,84723	2,06240	38,98149	9,95187	27,26852	0,06548	94,90741	7,20202	50,55556	7,07107
20	24,14352	4,74678	39,62963	11,13038	29,72223	3,14270	96,11111	5,49972	44,44445	0,78567

Figura 4-25 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Migração. Resultados observando a média das partições.



Esses resultados são apresentados na Tabela 4.29, e a comparação com os experimentos realizados com todas as variáveis do conjunto, pode ser observada na Figura 4.25.

Retirando a variável Escolaridade

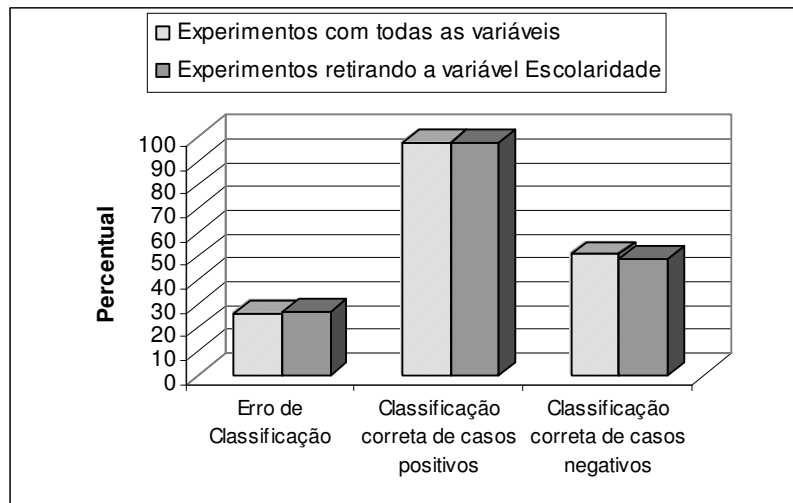
A retirada dessa variável causou um crescimento no erro de classificação da rede em 0,92%, uma queda no percentual de classificação correta dos casos negativos para TMC's em 2,03% e um crescimento no percentual de classificação correta dos

casos positivos em apenas 0,18%. Mas, foi observado após a realização do teste *t-student* que essas variações não foram estatisticamente significativas ($p = 0,45252$; $p = 0,81539$; $p = 0,33293$). Esses resultados podem sugerir que a variável **escolaridade** não apresenta possibilidade de estar relacionada com os TMC's.

Tabela 4-30 Resultados retirando a variável Escolaridade do conjunto de dados, observando a média entre as duas partições

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	25,06945	2,38976	38,42593	6,02350	29,12037	0,58926	96,57408	4,84499	45,18519	3,66648
8	24,42130	1,34219	38,14815	5,89256	26,66667	1,04756	97,96297	2,88080	48,70371	4,97593
12	24,74537	4,28848	40,46297	6,67823	30,18519	0,91662	96,57408	4,84499	43,05556	3,01175
16	22,77778	1,44040	41,11112	7,98768	28,79630	0,78567	94,25926	8,11863	48,14815	9,68998
20	27,68519	8,70789	41,66667	7,59485	33,42593	4,32121	90,27778	13,74930	42,87037	5,10688

Figura 4-26 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Escolaridade. Resultados observando a média das partições.



Nesse caso, a topologia considerada foi definida com oito nodos na camada intermediária. Esses resultados são apresentados na Tabela 4.30, e, a comparação feita com os experimentos realizados utilizando todas as variáveis do conjunto pode ser vista na Figura 4.26.

Retirando a variável Propriedade da Casa

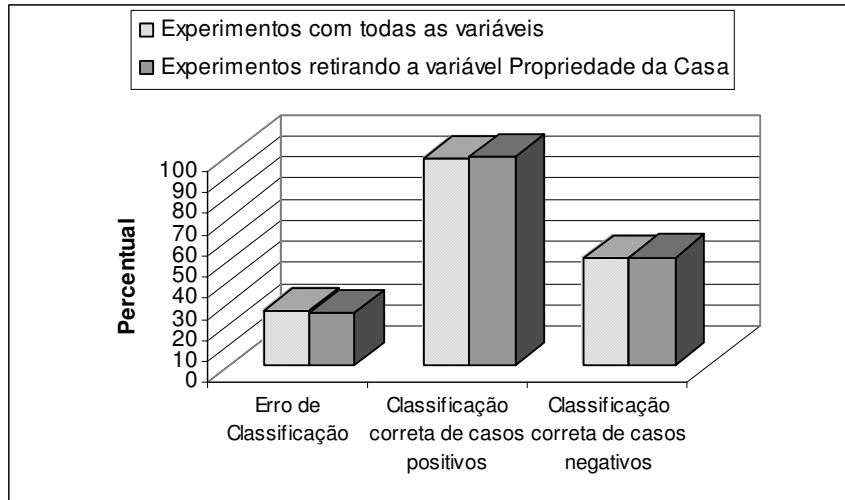
A topologia considerada para análise com esses experimentos foi definida com oito nodos. Os resultados obtidos podem ser encontrados na Tabela 4.31 e na Figura

4.27 pode ser visualizada uma comparação com os experimentos realizados utilizando todas as variáveis do conjunto.

Tabela 4-31 Resultados retirando a variável Propriedade da Casa do conjunto de dados, observando a média entre as duas partições

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	24,44445	2,55344	36,018515	6,28539	26,38889	0,78567	97,59260	3,40458	49,62963	1,83325
8	23,47222	1,24398	35,78704	6,87465	24,76852	0,58926	99,07408	1,30946	51,38889	0,13094
12	23,49537	5,00868	38,37963	7,52937	28,75000	2,68439	95,74074	5,76161	46,75926	0,39284
16	22,59259	3,01175	39,25926	8,77336	27,12964	0,26189	94,62963	7,59485	51,11111	7,07107
20	26,20371	8,44600	39,35185	10,21376	31,75926	5,49972	91,20371	12,43984	45,27778	1,44040

Figura 4-27 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Propriedade da Casa. Resultados observando a média das partições.



Com a retirada da variável **propriedade da casa** foi observada uma variação estatisticamente significativa apenas com relação à capacidade da rede em classificar corretamente os casos com diagnóstico positivo para os TMC's. O erro de classificação da rede foi reduzido em 0,97% ($p = 0,47115$), o percentual de classificação correta dos casos positivos aumentou em 1,30% ($p = 0,03518$) e o percentual de classificação dos casos negativos foi melhorado em 0,65% ($p = 0,79351$).

Esses resultados podem sugerir que a permanência da variável **propriedade da casa** no conjunto de dados pode influenciar de forma negativa a classificação dos padrões com diagnóstico positivo para os TMC.

Retirando a variável Inserção no Processo Produtivo

Com a retirada da variável **inserção no processo produtivo** o erro de classificação da rede aumentou em 1,34%. Foi possível observar também, uma redução de 8,88% no percentual de classificação correta dos casos positivos e uma melhora de 6,20% no percentual de classificação dos casos negativos para TMC's. Essas variações observadas tanto no percentual de classificação dos casos positivos quanto no percentual de classificação dos casos negativos foram estatisticamente significativos ($p = 0,00001$; $p = 0,00621$), porém com relação ao erro de classificação da rede não foi significativo ($p = 0,19412$). Com isso, esses resultados sugerem que essa variável é um importante fator para a predição dos casos com diagnóstico positivo para os TMC's.

Figura 4-28 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Inserção no Processo Produtivo. Resultados observando a média das partições.

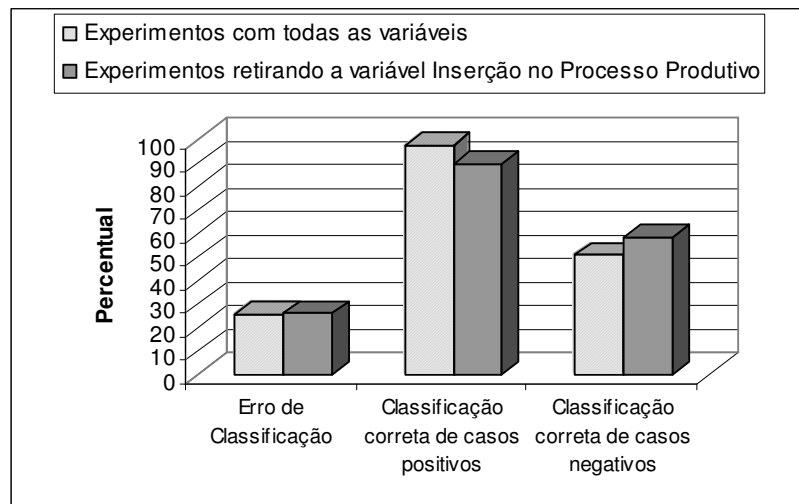


Tabela 4-32 Resultados retirando a variável Inserção no Processo Produtivo do conjunto de dados, observando a média entre as duas partições

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	St.dv.
4	28,12500	0,22916	29,86111	2,16061	26,52778	6,08897	89,53704	0,13095	57,40741	12,04701
8	27,50001	1,83324	30,37037	4,45216	25,97223	4,64857	89,53704	0,91662	58,51852	10,21376
12	26,55093	2,78260	29,25926	2,88081	26,01852	3,53553	89,72223	1,17851	58,24074	8,24957
16	26,22686	3,24090	29,21297	2,81534	27,31482	1,96418	88,98148	1,17851	56,38889	5,10688
20	26,66667	2,94627	28,88889	2,35703	27,08334	2,42250	88,88889	0,78567	56,94445	5,63067

A topologia considerada para a análise dos resultados foi definida com vinte nodos na camada intermediária da rede. Na Tabela 4.32 são apresentados os resultados

obtidos com esses experimentos e na Figura 4.28 uma comparação desses resultados com os obtidos nos experimentos realizados utilizando todas as variáveis do conjunto.

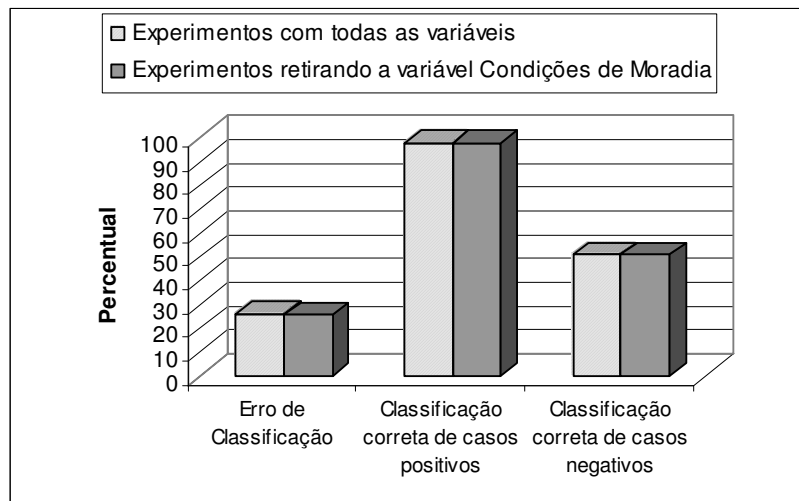
Retirando a variável Condições de Moradia

Analisando a média obtida com as duas partições, foi possível observar que com a retirada da variável **condições de moradia**, tanto o erro de classificação da rede quanto do percentual de classificação correta dos casos positivos e negativos para os TMC's foram reduzidos em um valor pouco significativo, 0,18%. Essas variações não foram estatisticamente significativas ($p = 0,84621$; $p = 0,78942$; $p = 0,91538$), sugerindo dessa forma que essa variável não é um fator importante para a classificação diagnóstica dos TMC's.

Tabela 4-33 Resultados retirando a variável Condições de Moradia do conjunto de dados, observando a média entre as duas partições.

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	24,23612	1,34219	36,01852	7,46391	25,55556	2,35702	97,96297	2,88080	50,92593	1,83324
8	23,31019	1,73503	36,66667	5,23783	25,23149	0,58925	98,88889	1,57135	50,64815	0,39284
12	25,04630	4,45216	37,87038	5,76161	34,02778	11,19586	95,64815	6,15445	48,33334	0,78567
16	23,75000	1,63682	38,75000	9,10073	26,75926	1,44040	94,53704	7,72580	51,94445	4,84499
20	26,08797	7,30022	39,02778	8,31505	30,83334	3,27364	91,48148	11,78511	46,29630	6,02350

Figura 4-29 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Condições de Moradia. Resultados observando a média das partições.



A topologia que apresentou os melhores resultados foi definida com apenas quatro

nodos na camada intermediária. A Tabela 4.33 apresenta os resultados obtidos e, na Figura 4.29 é possível visualizar a comparação desses resultados com os experimentos utilizando todas as variáveis do conjunto.

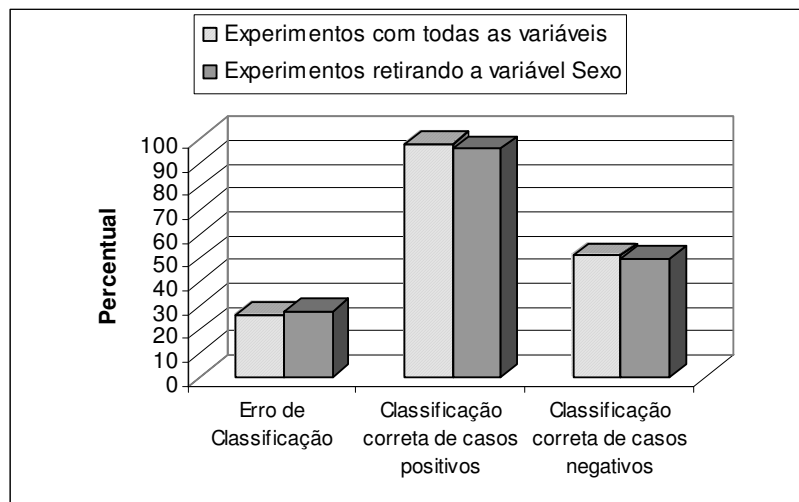
Retirando a variável Sexo

Comparando os resultados com os experimentos realizados com todas as variáveis presentes no conjunto de dados, foi possível observar que a retirada da variável **sexo** influenciou negativamente a performance da rede, ou seja, o erro de classificação cresceu em 1,48%. Além disso, tanto o percentual de classificação correta dos casos positivos quanto o percentual de classificação correta dos casos negativos para os TMC's foram diminuídos em 1,76% e 1,20%, respectivamente. Entretanto, essas variações não foram estatisticamente significativas ($p = 0,43841$; $p = 0,13529$; $p = 0,69406$).

Tabela 4-34 Resultados retirando a variável Sexo do conjunto de dados, observando a média entre as duas partições.

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	25,25464	1,86598	40,46296	6,02350	29,12038	4,51763	89,90741	14,01119	51,85185	4,97594
8	25,11574	1,80051	40,37037	6,15445	27,22222	3,01175	96,01852	5,10688	49,53704	0,91662
12	26,29630	5,82708	41,34260	5,69613	29,21297	1,50588	88,33333	16,23726	47,87037	5,63066
16	25,34722	3,76469	42,63889	6,21992	28,98149	1,83324	89,16667	15,32065	52,87037	11,65417
20	25,67130	7,49664	43,79630	6,28539	33,33334	1,83324	90,55556	13,35646	42,77778	9,68998

Figura 4-30 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Sexo. Resultados observando a média das partições.



A topologia selecionada por apresentar os melhores resultados foi definida com oito nodos na camada intermediária. Esses resultados podem ser vistos na Tabela 4.34 e Figura 4.30.

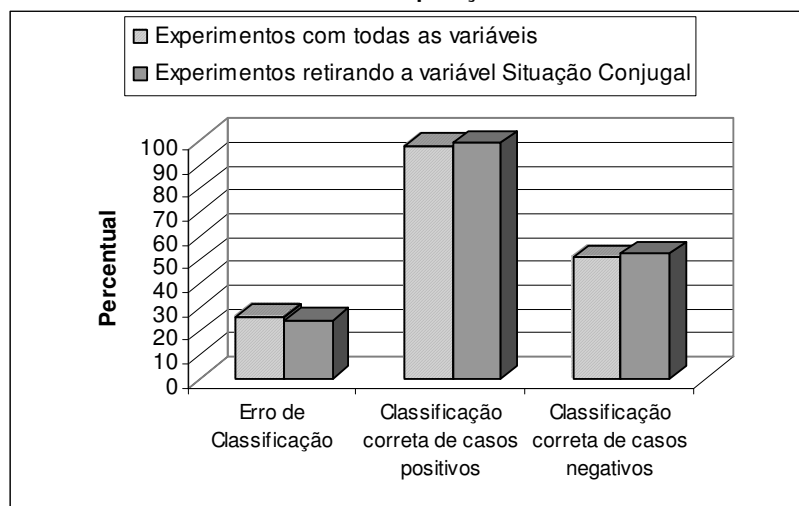
Retirando a variável Situação Conjugal

Observamos com esses experimentos que a retirada da variável **situação conjugal** resultou em uma redução do erro de classificação da rede em 1,57%. Tanto o percentual de classificação correta dos casos positivos quanto o percentual de classificação correta dos casos negativos para os TMC's aumentaram em 1,39% e 1,76%, respectivamente. Porém, essas variações não foram estatisticamente significativas ($p = 0,18848$; $p = 0,06506$; $p = 0,37803$).

Tabela 4-35 Resultados retirando a variável Situação Conjugal do conjunto de dados, observando a média entre as duas partições.

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	24,65278	3,17543	38,05556	9,55903	25,78704	1,89872	92,77778	10,21376	55,64815	14,01119
8	23,12500	1,66956	36,75926	6,54729	24,16667	4,05932	99,16667	1,17851	52,50000	9,29714
12	24,09723	2,71712	37,26852	8,44599	25,27778	2,22608	99,07408	1,30946	50,37037	5,76161
16	24,02778	1,37493	38,84259	8,44599	26,48148	1,57135	94,72222	7,46391	52,31481	10,60660
20	24,79167	5,13962	38,51852	9,03525	28,51852	0,52378	93,05556	9,82093	49,90741	10,86849

Figura 4-31 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Situação Conjugal. Resultados observando a média das partições.



Os resultados obtidos sugerem que a variável **situação conjugal** não representa um fator importante para a classificação de indivíduos portadores de TMC's. A topologia

que apresentou os melhores resultados foi definida com oito nodos na camada intermediária da rede. Esses resultados podem ser observados na Tabela 4.35 e Figura 4.31.

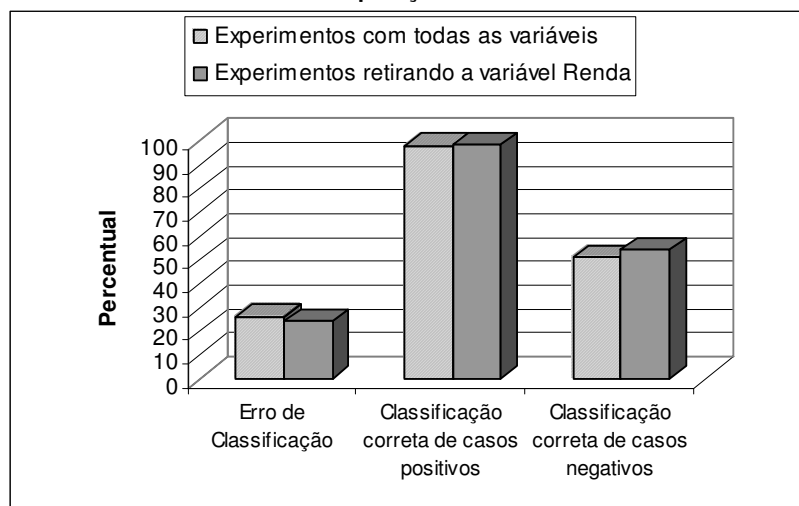
Retirando a variável Renda

Na análise dos experimentos realizados retirando a variável **renda**, observamos uma redução na capacidade de generalização da rede, com um crescimento do erro de classificação em 1,89%. Também foi observado uma redução no percentual de classificação dos casos positivos em 5,27% e aumento no percentual de classificação dos casos negativos em 1,48%. As variações, entretanto, não foram estatisticamente significativas ($p = 0,31274$; $p = 0,29528$; $p = 0,65184$).

Tabela 4-36 Resultados retirando a variável Renda do conjunto de dados, observando a média entre as duas partições.

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4	25,27778	3,73195	37,31482	5,76161	27,63889	2,42249	92,50000	10,60660	52,22223	5,76161
8	24,46760	2,38976	37,73148	3,73195	23,84260	3,20816	98,24074	2,48797	54,07408	8,90431
12	23,70371	1,76776	38,37963	6,21992	24,90741	2,74986	98,98148	1,44040	51,20371	6,94012
16	24,14352	1,66956	39,58334	7,26748	27,22223	0,65473	94,44445	7,85674	51,11110	9,16622
20	25,20834	5,53245	40,78704	7,39843	29,72223	1,17851	92,77778	10,21376	47,77778	7,85674

Figura 4-32 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Renda. Resultados observando a média das partições.



Após análise da significância estatística das variações que foram observadas com esses experimentos, os resultados obtidos podem sugerir que a variável **renda** não

é importante para a classificação dos TMC's. Os resultados são apresentados na Tabela 4.36 e Figura 4.32. A topologia considerada para análise foi definida com quatro nodos na camada intermediária da rede.

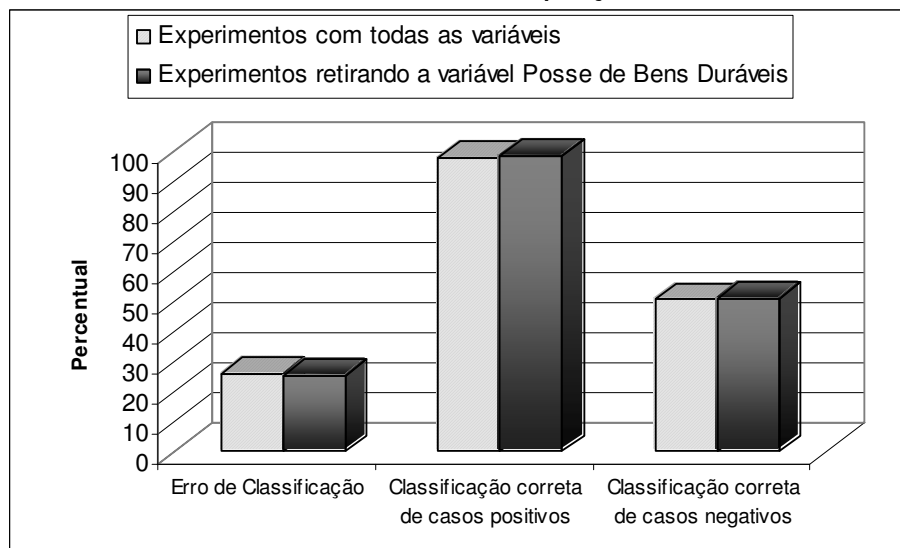
Retirando a variável Posse de Bens Duráveis

Com os resultados obtidos, podemos observar que o erro de classificação da rede diminuiu 0,46% e a capacidade de classificação correta dos casos com diagnóstico positivo para os TMC's aumentou em 0,83%. O percentual de classificação correta dos casos negativos também aumentou, porem em um percentual pouco significativo, 0,09%. As variações observadas, contudo, não foram estatisticamente significativas ($p = 0,65910$; $p = 0,23409$; $p = 0,96079$). Esses resultados sugerem que essa variável, com relação à base estudada, não foi um importante fator para a classificação dos TMC's.

Tabela 4-37 Resultados retirando a variável Renda do conjunto de dados, observando a média entre as duas partições.

Nodos	Erro Treinamento		Erro Validação		Erro Teste		% de classificação dos casos positivos		% de classificação dos casos negativos	
	Média	st.dv.	Média	st.dv.	Média	st.dv.	Média	St.dv.	Média	st.dv.
4	24.81482	2.68438	37.91667	7.92221	28.19445	5.56519	93.14815	9.68998	50.46296	1.44040
8	24.02778	1.11304	36.43519	5.04141	25.27778	0.39284	98.61111	1.96419	50.83334	2.74986
12	23.00926	1.44040	37.77778	6.94013	25.27778	1.70229	99.44445	0.78567	50.00000	4.19026
16	24.00463	1.53861	39.39815	8.44599	26.94445	0.78567	94.72222	7.46391	51.38889	5.89256
20	24.07408	3.40458	39.16667	8.51147	28.37963	1.24398	95.09260	6.94012	48.14815	4.45216

Figura 4-33 Comparação de resultados dos experimentos realizados com MLP utilizando todas as variáveis do conjunto e experimentos retirando a variável Posse de Bens Duráveis. Resultados observando a média das partições.



Os resultados obtidos, e uma comparação feita com os experimentos realizados utilizando todas as variáveis do conjunto, podem ser observadas na Tabela 4.37 e Figura 4.33. A topologia considerada na análise teve oito nodos na camada intermediária.

Tabela 4-38 Quadro comparativo das diferenças apresentadas nos experimentos retirando variáveis do conjunto, observando a média entre as duas partições, e seus respectivos valores de p . O sinal “+” significa que houve um aumento no percentual e o sinal “-” que houve uma redução.

Experimentos Retirando Variáveis	Média entre as duas Partições					
	Erro de Classificação		classificação correta dos casos positivos		Classificação correta dos casos negativos	
	Diferença	P	Diferença	p	Diferença	P
Idade	-0,83%	0,37428	+1,20%	0,07886	+0,46%	0,77072
Alfabetização	-0,97%	0,47942	+0,74%	0,45132	+1,20%	0,56601
Migração	-0,41%	0,76309	+0,92%	0,35789	-0,09%	0,96731
Escolaridade	+0,92%	0,45252	+0,18%	0,81539	-2,03%	0,33293
Propriedade da Casa	-0,97%	0,47115	+1,29%	0,03518	+0,65%	0,79351
Inserção no Processo Produtivo	+1,34%	0,19412	-8,88%	0,00001	+6,20%	0,00621
Condições de Moradia	-0,18%	0,84621	+0,18%	0,78942	+0,18%	0,91538
Sexo	+1,48%	0,43841	-1,75%	0,13529	-1,20%	0,69406
Situação Conjugal	-1,57%	0,18848	+1,38%	0,06506	+1,76%	0,37803
Renda	-1,89%	0,31274	-5,27%	0,29528	+1,48%	0,65184
Posse de Bens Duráveis	-0,46%	0,65910	+0,83%	0,23409	+0,09%	0,96079

Na Tabela 4.38 é apresentado um resumo das diferenças ocorridas nos experimentos observando a média entre as duas partições, além dos respectivos valores de p . Essas diferenças foram obtidas comparando os resultados dos experimentos realizados utilizando todas as variáveis do conjunto com aqueles retirando uma variável por vez.

4.4 Conclusões

É possível observar que os resultados apresentados em nossos experimentos utilizando o algoritmo *Backpropagation* retirando uma variável do conjunto por vez, não nos ofereceu uma visão clara de quais variáveis (fatores) estavam relacionadas aos TMC's. Foram obtidos resultados diferentes em cada análise feita observando individualmente cada partição ou a média entre elas.

Com a partição 01, as variáveis que ao serem retiradas do conjunto de dados causou uma variação estatisticamente significativa quanto ao erro de classificação da rede e do percentual de classificação correta dos casos com diagnósticos positivo

e negativo foram **idade**, **escolaridade**, **inserção no processo produtivo** e **situação conjugal**.

Quanto ao erro de classificação, as variáveis que provocaram um crescimento estatisticamente significativo foram **idade**, **escolaridade** e **situação conjugal** ($p = 0,0001$; $p = 0,0218$; $p = 0,0020$). Com relação à classificação correta dos casos com diagnóstico positivo, apenas a retirada da variável **inserção no processo produtivo** provocou uma redução significativa ($p = 0,0001$). No que se refere ao percentual de classificação correta dos casos negativos, a retirada das variáveis **idade**, **escolaridade** e **situação conjugal** influenciou na sua redução ($p = 0,0001$; $p = 0,0218$; $p = 0,0020$).

Nos experimentos com a partição 02 a variável que ao ser retirada do conjunto de dados provocou uma redução significativa no erro de classificação foi **situação conjugal** ($p = 0,02557$). Quanto ao percentual de classificação correta dos casos positivos, foi observado uma variação significativa com a retirada das variáveis **propriedade da casa** ($p = 0,02946$) e **inserção no processo produtivo** ($p = 0,01556$), aumentando o percentual de classificação correta dos casos positivos. Apenas a retirada da variável **situação conjugal** apresentou uma variação significativa quanto ao percentual de classificação correta dos indivíduos com diagnóstico negativo para os TMC's ($p = 0,01298$), a qual foi aumentada.

Observando a média entre as duas partições, a retirada das variáveis **propriedade da casa** e **inserção no processo produtivo** foram significativas quanto à variação ocorrida no percentual de classificação correta dos casos com diagnóstico positivo ($p = 0,03518$; $p = 0,00001$), sendo que a primeira aumentou e a segunda reduziu esse percentual. Quanto a classificação correta dos casos com diagnóstico negativo, apenas a retirada da variável **inserção no processo produtivo** provocou uma variação estatisticamente significativa ($p = 0,00621$), com um aumento do seu percentual.

Com relação ao erro de classificação da rede, observando a média entre as duas partições, as variações ocorridas nos experimentos retirando variáveis não foi estatisticamente significativo. Um resumo das diferenças observadas em todos os experimentos pode ser visualizado nas Tabelas 4.25 e 4.38.

Essa diferença existente nos resultados obtidos pode ter sido influenciada pelo fato do balanceamento dos dados, para a composição dos conjuntos nas duas partições, só ter sido feito para o diagnóstico apresentado, ou seja, quanto aos indivíduos portadores e não portadores de TMC's. A utilização, entretanto, da técnica de *crossvalidation* poderia ter amenizado tal problemática, porém, por uma questão de tempo, não foi possível a utilização dessa técnica nesta pesquisa.

Figura 4-34 Curva do percentual de classificação correta dos casos com diagnósticos positivo e negativo nos experimentos utilizando todas as variáveis do conjunto e retirando variáveis (observando a médias entre as duas partições)

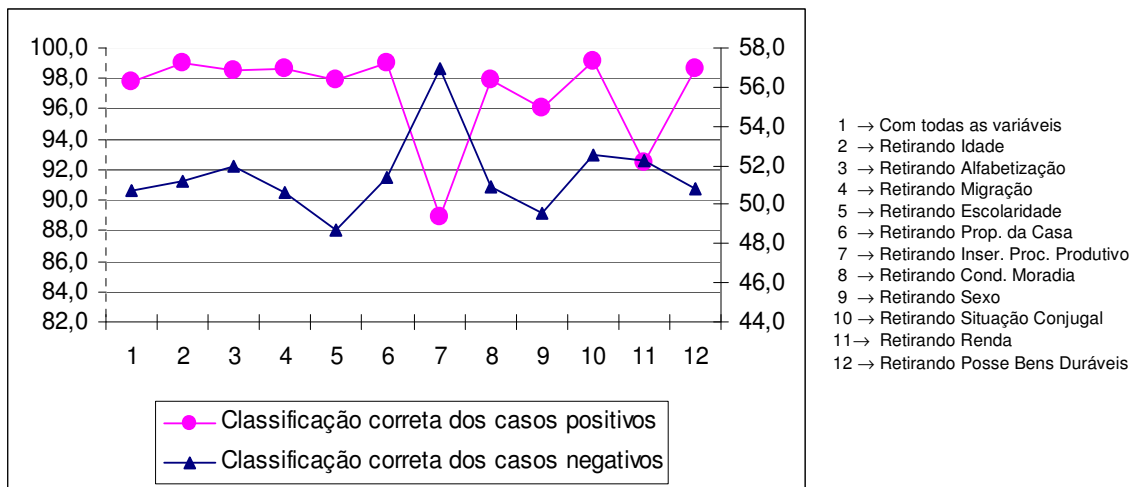


Figura 4-35 Curva do erro de classificação da rede nos experimentos utilizando todas as variáveis do conjunto e retirando variáveis (observando a médias entre as duas partições)



Outra questão importante que merece destaque, é que quando retiramos apenas uma variável do conjunto de dados por vez não está sendo levado em consideração

que possa existir uma inter-relação desta com outras variáveis do conjunto. Isso pode causar distorções nos resultados apresentados pela rede neural.

Tabela 4-39 Resumo dos resultados obtidos (diferenças) quanto ao erro de classificação, percentual de classificação correta dos casos positivos e percentual de classificação correta dos casos negativos nos experimentos retirando variáveis do conjunto de dados. O sinal “+” significa que houve um aumento do percentual com relação aos resultados obtidos com todas as variáveis, e o sinal “-” significa que houve uma diminuição do percentual.

Partição 01						
Variável Retirada	Erro Médio no Conjunto de Teste		Média do percentual de classificação correta dos casos positivos		Média do percentual de classificação correta dos casos negativos	
	Erro Obtido	Diferença	% médio	Diferença	% médio	Diferença
Idade	25,92593	+2,22223	100,0	0,0	48,14815	-4,44444
Alfabetização	24,53704	-0,83334	100,0	0,0	50,92593	-1,66666
Migração	25,09259	+1,38889	100,0	0,0	49,81481	-2,77778
Escolaridade	27,40741	+3,70371	100,0	0,0	45,18519	-7,40740
Propriedade da Casa	24,35185	+0,64815	100,0	0,0	51,29630	-0,99629
Inserção no Processo Produtivo	25,37037	+1,66667	88,33333	-11,66667	60,92593	+8,33334
Condições de Moradia	23,88889	+0,18519	100,0	0,0	52,22222	-0,37037
Sexo	25,09259	+1,38489	99,62963	-0,37037	50,18519	-2,40740
Situação Conjugal	26,85185	+3,14415	100,0	0,0	46,29630	-6,29629
Renda	25,92593	+2,22223	100,0	0,0	48,14815	-4,44444
Posse de Bens Duráveis	24,25926	+0,55156	100,0	0,0	51,48148	-1,11111
Partição 02						
Variável Retirada	Erro Médio no Conjunto de Teste		Média do percentual de classificação correta dos casos positivos		Média do percentual de classificação correta dos casos negativos	
	Erro Obtido	Diferença	% médio	Diferença	% médio	Diferença
Idade	25,0	-0,92593	97,96296	+2,40740	52,03704	-0,55555
Alfabetização	25,0	-0,92593	97,03704	+1,48148	52,96296	+0,37037
Migração	25,55556	-0,37037	97,40741	+1,85185	51,48148	-1,11111
Escolaridade	25,92593	0,0	95,92593	+0,37037	52,22222	-0,37037
Propriedade da Casa	25,18519	-0,74074	98,14815	+2,59259	51,48148	-1,11111
Inserção no Processo Produtivo	28,51852	+2,59259	90,55556	-5,0	52,40741	-0,18518
Condições de Moradia	25,64815	-0,27778	97,77778	+1,66666	50,92593	-1,66666
Sexo	29,35185	+3,42592	92,40741	-3,14815	48,88889	-3,70730
Situação Conjugal	21,29630	-4,62963	98,33333	-1,66667	59,07407	+6,48148
Renda	21,57408	-4,35185	96,48148	-0,92592	60,37037	-7,77778
Posse de Bens Duráveis	25,0	-0,92593	97,22222	+1,66666	52,77778	+0,18519

É importante destacar que, apesar de termos utilizados uma quantidade variada de nodos escondidos nos experimentos realizados com as duas partições (4, 8, 12, 16 e 20), as topologias que foram consideradas para a análise, por terem obtido os melhores resultados, tinham um número de nodos escondidos bastante pequeno, em média oito nodos.

A Tabela 4.39 apresenta um resumo com as diferenças existentes entre os resultados obtidos (média) quanto ao erro de classificação da rede e o percentual de classificação correta dos casos com diagnósticos positivo e negativo nos experimentos realizados com a retirada de variáveis, bem como os experimentos realizados com todas as variáveis presentes no conjunto de dados.

Nas Figuras 4.34 e 4.35 é possível analisar a curva dos resultados obtidos nos experimentos realizados com todas as variáveis do conjunto e com a retirada de cada variável.

No Capítulo 5 serão apresentados os resultados obtidos com os experimentos realizados com o método *simulated annealing*, otimizando simultaneamente a arquitetura e pesos da rede, para análise dos fatores relacionados aos TMC's.

Capítulo 5 - Experimentos com Multilyer Perceptron e o Algoritmo Simulated Annealing

5.1 Introdução

Neste Capítulo serão apresentados os resultados dos experimentos realizados com uma rede neural Multilayer Perceptron (MLP), treinada com o algoritmo *simulated annealing*, otimizando simultaneamente a arquitetura e pesos da rede. A partir da otimização da arquitetura (nodos de entrada, nodos ocultos e conexões), objetivou-se analisar os fatores relacionados com os Transtornos Mentais Comuns (TMC's).

Como o algoritmo *simulated annealing* é probabilístico, e a cada execução do mesmo por inicialização de pesos pode resultar em topologias diferentes, fez-se necessário a definição de uma topologia inicial contendo todas as variáveis do conjunto (onze nos experimentos iniciais e depois com sete) e todas as conexões possíveis nas camadas adjacentes da rede.

Para definição desta topologia inicial, no que se refere ao número de nodos na camada intermediária, foram realizados inicialmente alguns experimentos. Com isso, foi possível observar que o número máximo de nodos utilizados na camada intermediária era quatro. Então, o número de nodos na camada intermediária da rede definido como topologia inicial foi quatro.

A cada execução do algoritmo *simulated annealing* por inicialização de pesos, foram observadas as variáveis utilizadas (topologia resultante do experimento) para a obtenção do resultado. Com isso, a definição de quais variáveis apresentavam maior possibilidade de estarem relacionadas aos TMC's, foi feita a partir da observação das variáveis que eram mais utilizadas pelo algoritmo (percentual) a cada execução. Na partição 01, foram observadas para análise as execuções que apresentaram um erro de classificação $\leq 28,13\%$ e com a partição 02 $\leq 29,75\%$.

Esses experimentos foram realizados em duas etapas distintas: 1) inicialmente o conjunto de entrada foi composto com todas as variáveis da base de dados, num total de onze; 2) a partir dos resultados obtidos nos experimentos iniciais, foram realizados novos experimentos com apenas as variáveis resultantes, num total de sete variáveis.

A análise dos resultados foi desenvolvida de duas formas distintas: 1) observando individualmente cada partição utilizada; 2) observando a média obtida entre as

partições. Nas Seções 5.2 e 5.3 são apresentados os resultados obtidos com as análises desenvolvidas, e, finalmente, na Seção 5.4 são apresentadas as conclusões obtidas a partir da análise dos resultados.

5.2 Análise individual de cada partição

Nesta Seção são apresentados os resultados obtidos com a análise individual de cada partição utilizada para a realização dos experimentos, num total de duas. Os resultados são apresentados contendo todos os aspectos observados e o desvio padrão.

Os experimentos com a partição 01 são apresentados na Seção 5.2.1 e com a partição 02 na Seção 5.2.2. Posteriormente, são mostrados os comentários a respeito do desempenho da rede neural para a análise dos fatores relacionados aos TMC's.

5.2.1 Experimentos realizados com a partição 01

Nesta Seção serão apresentados os resultados obtidos nos experimentos realizados com a partição 01, utilizando inicialmente todas as variáveis para a composição do conjunto de entrada da rede, como também utilizando apenas as variáveis resultantes dos experimentos iniciais.

Conforme foi descrito no Capítulo 3, apenas as execuções do algoritmo *simulated annealing* cujo erro de classificação fosse $\leq 28,13\%$ (média obtida com o erro de classificação nos experimentos com os algoritmos *Backpropagation* e *Simulated Annealing*) foram escolhidas para verificação das variáveis que apresentaram maior probabilidade de estar relacionada aos TMC's. Com a definição desse percentual médio do erro de classificação, foi possível evitar um super-ajustamento do modelo.

5.2.1.1 Utilizando todas as variáveis do conjunto de dados

Na Tabela 5.1 são apresentados os resultados obtidos utilizando como topologia inicial da rede todas as variáveis presentes no conjunto de dados (num total de onze). O percentual de utilização de cada variável pode ser observado na Figura 5.1.

Nesses experimentos, com a partição 01, o erro médio de classificação da rede foi 23,08%, o percentual de classificação correta dos casos positivos 89,08% e o percentual de classificação correta dos casos negativos 64,75%.

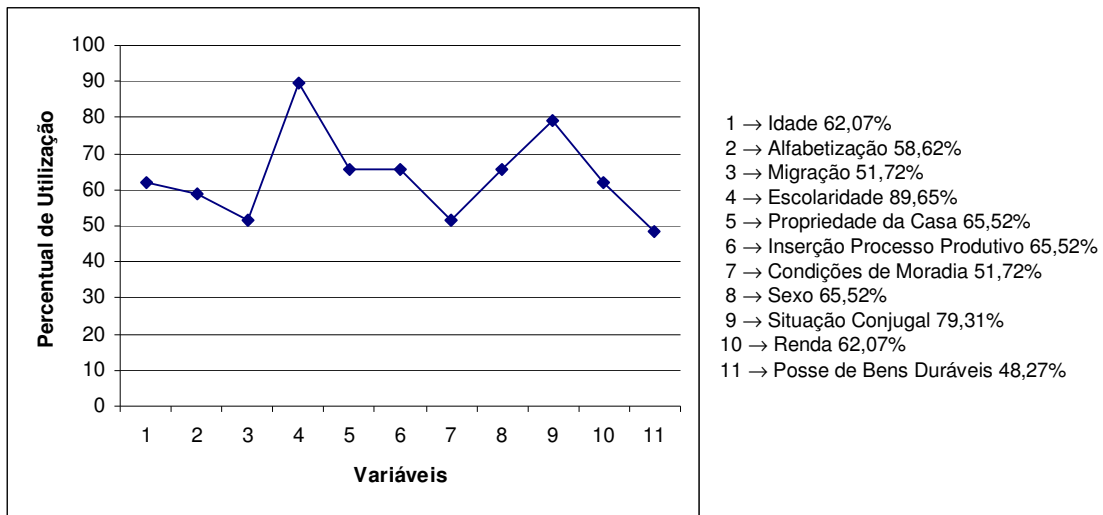
A média de variáveis resultante desses experimentos, foi sete. Ou seja, das onze variáveis utilizadas na topologia inicial da rede, apenas sete (em média) foram utilizadas pelo algoritmo nas execuções observadas.

Tabela 5-1 Valor médio dos resultados obtidos nos experimentos com simulated annealing (partição 01, com onze variáveis)

Nodos Entrada	Nodos Ocultos	Nr. Conexões	Erro Treinamento		Erro Validação		Erro Teste		% de classificação correta dos casos positivos		% de classificação correta dos casos negativos	
			Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
7,0	1,79	11,38	28,83142	6,13678	38,60153	7,70366	23,08429	4,48812	89,08046	12,80460	64,75096	15,41034

A variável que apresentou o menor percentual de utilização foi **posse de bens duráveis** com 48,27%. Sete variáveis se destacaram, apresentando um maior percentual de utilização para a obtenção dos resultados, ou seja, que foram escolhidas automaticamente em um maior número de vezes a cada execução do algoritmo *simulated annealing* por inicialização de pesos: **idade, escolaridade, propriedade da casa, inserção no processo produtivo, sexo, situação conjugal e renda.**

Figura 5-1 Percentual de utilização de cada variável do conjunto. Experimentos com todas as variáveis utilizando o simulated annealing (Partição 01)



Dentre as variáveis utilizadas para mensurar as condições de vida, **escolaridade** foi a que mais se destacou, obtendo 89,65% de utilização. Quanto às condições de

trabalho, as variáveis **inserção no processo produtivo** e **renda** obtiveram um percentual de utilização bem próximo, acima de 60% (65,52% e 62,07% respectivamente). Essa classificação acerca das variáveis utilizadas para mensurar as condições de vida e trabalho foi definida por [Ludermir, 1998].

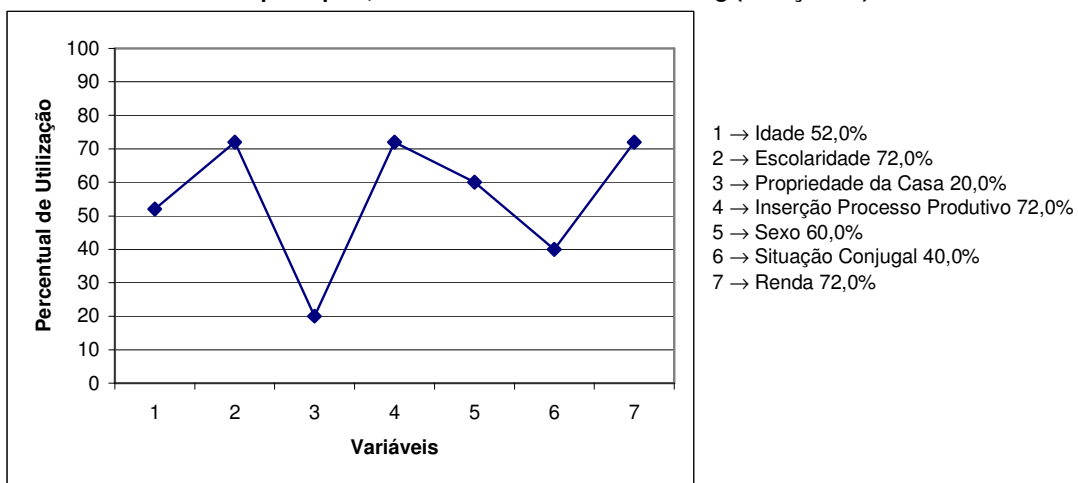
5.2.1.2 Utilizando as variáveis que mais se destacaram

Comparando os resultados apresentados utilizando apenas as variáveis que mais se destacaram nos experimentos iniciais (**idade, escolaridade, propriedade da casa, inserção no processo produtivo, sexo, situação conjugal e renda**) como topologia inicial da rede, podemos observar uma redução no erro de classificação em 1,71% . Além disso, o percentual de classificação correta dos casos positivos da rede aumentou em 1,51%, assim como o percentual de classificação correta dos casos negativos em 1,92%.

Tabela 5-2 Valor médio dos resultados obtidos nos experimentos com simulated annealing (Partição 01, com sete variáveis no conjunto de entrada)

Nodos Entrada	Nodos Ocultos	Nr. Conexões	Erro Treinamento		Erro Validação		Erro Teste		% de classificação correta de casos positivos		% de classificação correta de casos negativos	
			Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	St.dv.
3,9	1,68	6,68	27,46296	4,95702	41,51852	11,03588	21,37037	4,65978	90,59259	8,24577	66,66667	11,88171

Figura 5-2 Percentual de utilização de cada variável do conjunto. Experimentos com as sete variáveis principais, utilizando o simulated annealing (Partição 01)

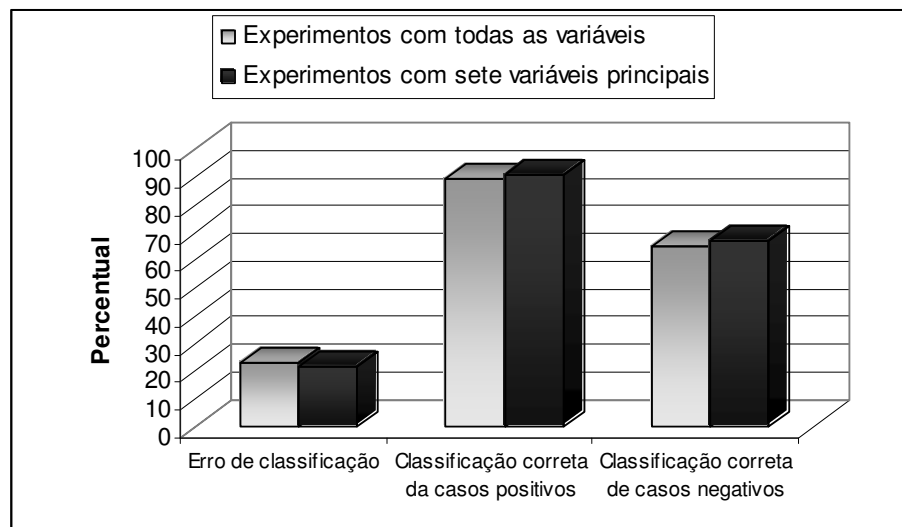


A quantidade média de nodos de entrada resultantes desses experimentos foi quatro, com isso, foi possível observar que as variáveis que eram mais utilizadas

pelo algoritmo a cada execução e inicialização de pesos foram: **escolaridade, inserção no processo produtivo, sexo e renda**. Esses resultados sugerem que essas quatro variáveis apresentam maior probabilidade de estarem relacionadas aos TMC's.

Na Tabela 5.2 e Figura 5.2 são apresentados os resultados obtidos com os experimentos realizados. Na Figura 5.3 uma comparação com os resultados obtidos utilizando todas as variáveis na entrada da rede pode ser visualizada.

Figura 5-3 Comparação de resultados dos experimentos realizados com o algoritmo *simulated annealing* utilizando todas as variáveis do conjunto e com as sete variáveis principais (Partição 01)



5.2.2 Experimentos realizados com a partição 02

Nesta Seção serão apresentados os resultados obtidos nos experimentos realizados com a partição 02, utilizando na topologia inicial da rede todas as variáveis do conjunto de entrada, e, em seguida, os experimentos realizados com apenas as variáveis resultantes dos experimentos iniciais.

Com a partição 02, para verificação das variáveis que apresentaram maior probabilidade de estar relacionada aos TMC's, foram escolhidas as execuções do algoritmo *simulated annealing* cujo erro de classificação fosse $\leq 29,75\%$. Esse percentual foi obtido com a média do erro de classificação apresentado nos experimentos com os algoritmos *Backpropagation* e *Simulated Annealing*, com essa partição.

5.2.2.1 Utilizando todas as variáveis do conjunto de dados

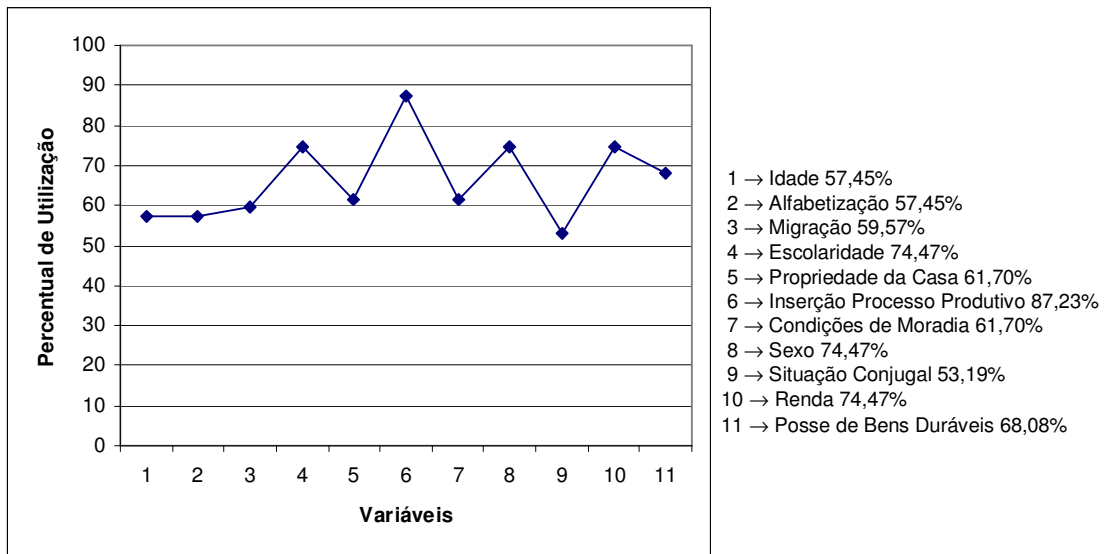
Na Tabela 5.3 são apresentados os resultados obtidos com todas as variáveis presentes no conjunto de dados, na topologia inicial da rede. O percentual de utilização de cada variável, pode ser observado na Figura 5.4.

Com esses experimentos, assim como aconteceu com a partição 01, a topologia resultante teve em média sete variáveis na unidade de entrada da rede, das onze definidas na topologia inicial. O erro de classificação foi 24,21%, o percentual de classificação correta de casos positivos 88,73% e o percentual de classificação correta de casos negativos 62,84%.

Tabela 5-3 Valor médio dos resultados obtidos nos experimentos com *simulated annealing* (Partição 02 com onze variáveis)

Nodos Entrada	Nodos Ocultos	Nr. Conexões	Erro Treinamento		Erro Validação		Erro Teste		% de classificação correta dos casos positivos		% de classificação correta dos casos negativos	
			Média	st.dv.	Média	St.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.
7,3	2,11	12,25	25,53192	6,01385	44,77935	6,10057	24,21198	6,16304	88,73128	11,88169	62,84476	17,41222

Figura 5-4 Percentual de utilização de cada variável do conjunto. Experimentos com todas as variáveis utilizando o simulated annealing (Partição 01)



Com isso, das variáveis que foram utilizadas mais vezes (em valores percentuais) pelo algoritmo foram: **escolaridade, propriedade da casa, inserção no processo produtivo, condições de moradia, sexo, renda e posse de bens duráveis**. Logo, esses resultados sugerem que essas variáveis podem estar mais relacionadas com os TMC's.

A variável **situação conjugal** (com 53,19%) apresentou o menor percentual de utilização e a que mais se destacou foi **inserção no processo produtivo**, com 87,23%.

É possível observar que, assim como aconteceu com a partição 01, a variável **escolaridade**, dentre as variáveis utilizadas para mensurar as condições de vida, foi a que apresentou o maior percentual de utilização, com 74,47%. E, quanto às condições de trabalho, a variável **inserção no processo produtivo** foi a que obteve o maior percentual de utilização.

5.2.2.2 Utilizando as variáveis que mais se destacaram

Os resultados observados com esses experimentos, quando comparados aos experimentos realizados com as onze variáveis na topologia inicial da rede, apresentaram uma redução no erro de classificação em 0,85%, o percentual de classificação correta de casos positivos foi aumentado em 5,28%, e o percentual de classificação correta de casos negativos foi reduzido em 3,58%. Esses resultados são apresentados na Tabela 5.4.

Figura 5-5 Percentual de utilização de cada variável do conjunto. Experimentos com as sete variáveis principais, utilizando o simulated annealing (Partição 02)

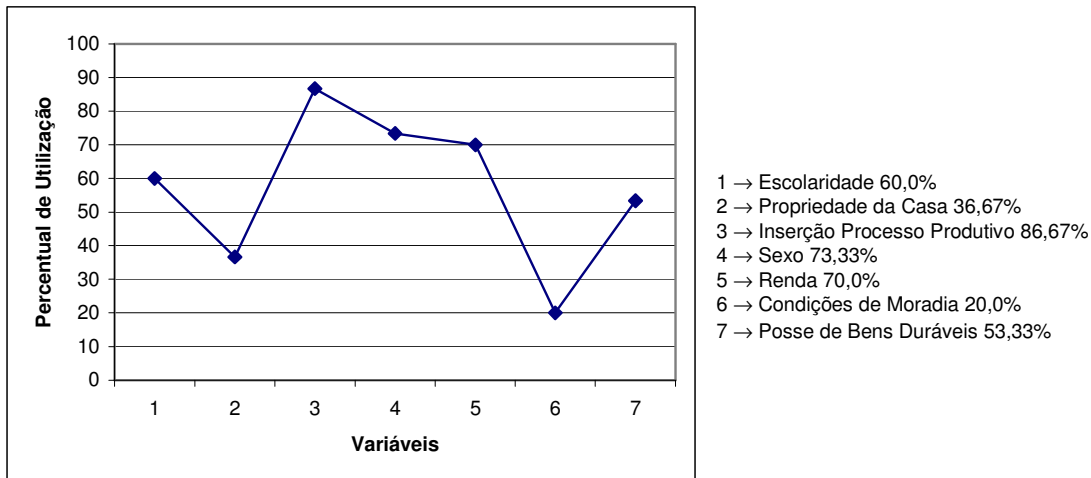
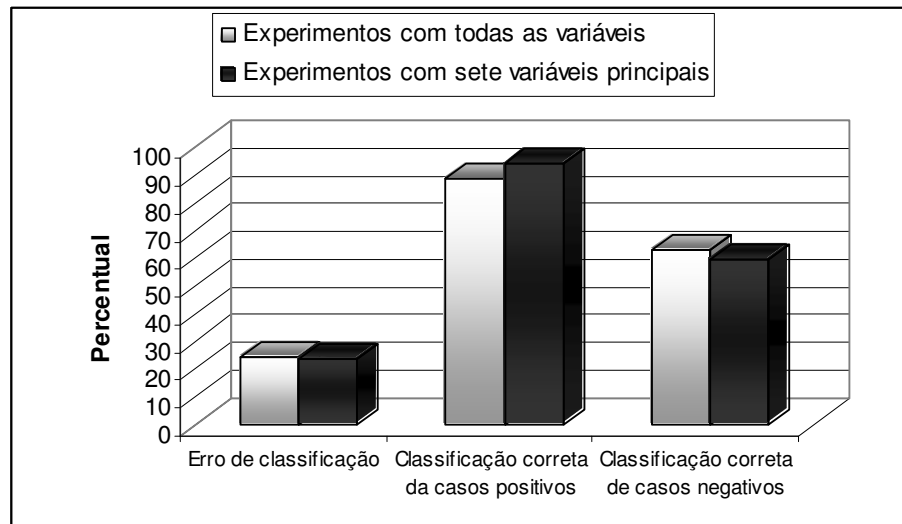


Tabela 5-4 Valor médio dos resultados obtidos nos experimentos com *simulated annealing* (Partição 02 com sete variáveis no conjunto de entrada)

Nodos Entrada	Nodos Ocultos	Nr. Conexões	Erro Treinamento		Erro Validação		Erro Teste		% de classificação correta dos casos positivos		% de classificação correta dos casos negativos	
			Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
4,0	2,1	7,83	25,37037	6,63185	42,09877	6,14061	23,36420	5,71133	94,01235	8,58577	59,25926	13,73801

A quantidade de entradas escolhidas automaticamente pelo algoritmo a partir da otimização da arquitetura da rede, nesses experimentos foi quatro. Com isso, as variáveis mais utilizadas foram **escolaridade**, **inserção no processo produtivo**, **sexo** e **renda**. Na Figura 5.5 é possível observar os percentuais de utilização de cada variável do conjunto de entrada. Esses resultados sugerem que essas variáveis apresentam maior probabilidade de estar relacionada com os TMC's.

Figura 5-6 Comparação de resultados dos experimentos realizados com o algoritmo *simulated annealing* utilizando todas as variáveis do conjunto e com as sete variáveis principais (Partição 02)



Na Figura 5.6 é possível observar uma comparação dos resultados obtidos com todas as variáveis e com os experimentos utilizando apenas as sete variáveis principais, como topologia inicial da rede.

5.3 Análise observando a média obtida com as duas partições

Nesta Seção são apresentados os resultados obtidos com a análise conjunta das duas partições (média) utilizadas nos experimentos. Os resultados são apresentados contendo todos os aspectos observados e o desvio padrão.

Os resultados obtidos com os experimentos utilizando todas as variáveis do conjunto de dados são apresentados na Seção 5.3.1 e na Seção 5.3.2 os resultados obtidos utilizando as sete variáveis principais. Posteriormente, são mostrados os comentários a respeito do desempenho da rede neural para a análise dos fatores relacionados aos TMC's.

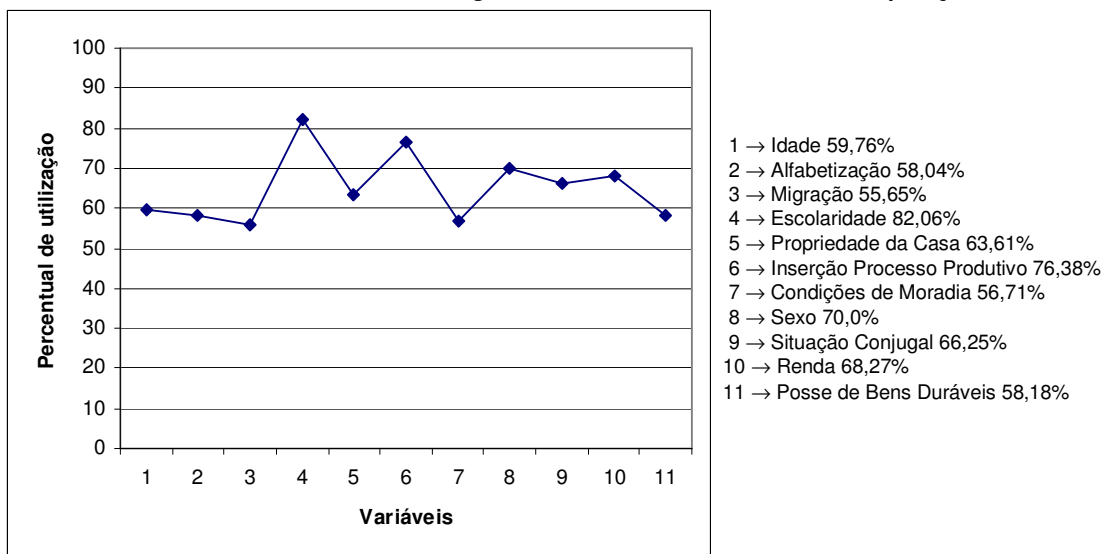
5.3.1 Utilizando todas as variáveis do conjunto de dados

Analisando conjuntamente as duas partições, foi observado um erro de classificação da rede de 23,65%, um percentual de classificação correta dos casos positivos de 88,91% e dos casos negativos de 63,80%. A média de variáveis utilizadas no conjunto de entrada permaneceu sete, como topologia resultante dos experimentos iniciais.

Essa média foi utilizada como parâmetro para escolha das variáveis, ou seja, como a topologia resultante da rede com esses experimentos teve sete nodos nas unidades de entrada, foram escolhidas as sete variáveis que apresentavam, em valores absolutos, o maior percentual.

Com isso, as sete variáveis que apresentaram maior percentual de ocorrência nos resultados foram: **idade**, **escolaridade**, **propriedade da casa**, **inserção no processo produtivo**, **sexo**, **situação conjugal** e **renda**. Na Figura 5.7 pode ser visualizada o percentual de utilização de cada variável do conjunto de dados. Os resultados são apresentados na Tabela 5.5.

Figura 5-7 Percentual de utilização de cada variável do conjunto. Experimentos com todas as variáveis utilizando o *simulated annealing*, observando a médias entre as duas partições.



Esses resultados sugerem que essas variáveis apresentam uma maior probabilidade de estarem relacionadas aos TMC's. Entre as variáveis utilizadas para mensurar as condições de vida, observamos mais uma vez que **escolaridade** foi a que mais se

destacou. Quanto às condições de trabalho, a variável **inserção no processo produtivo** mais uma vez apresenta o maior percentual de utilização.

Tabela 5-5 Resultados obtidos nos experimentos com *simulated annealing*, observando a média das duas partições (com todas as variáveis no conjunto de entrada)

Nodos Entrada	Nodos Ocultos	Nr. Conexões	Erro Treinamento		Erro Validação		Erro Teste		% de classificação correta dos casos positivos		% de classificação correta dos casos negativos	
			Média	st.dv.	Média	St.dv.	Média	st.dv.	Média	st.dv.	Média	st.dv.
7.15	1.95	11.815	27.18167	2.33310	41.69044	4.36838	23.64813	0.79740	88.90587	0.24691	63.79786	1.34789

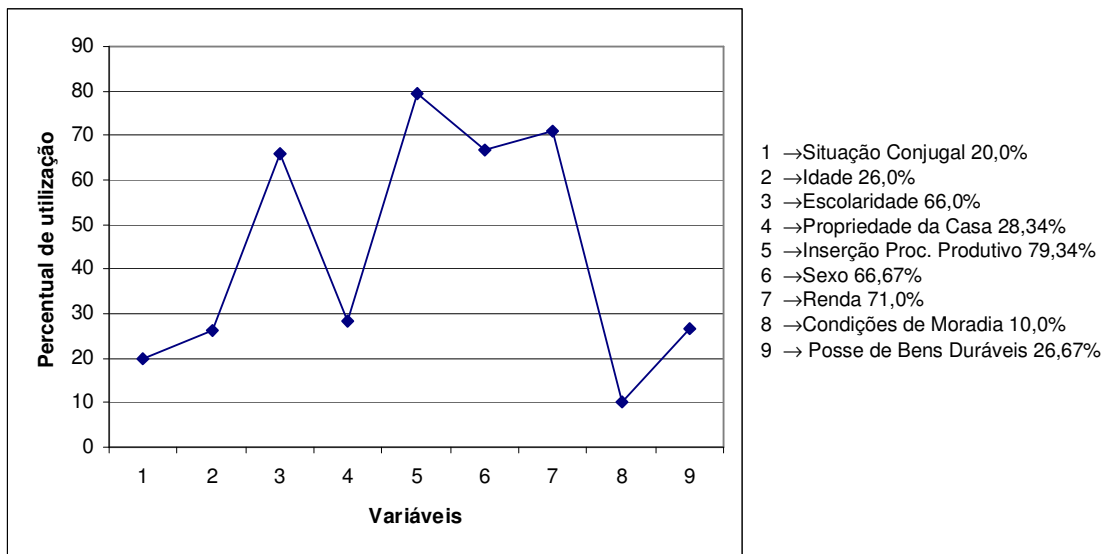
5.3.2 Utilizando as variáveis que mais se destacaram

Os resultados obtidos nos experimentos realizados utilizando como topologia inicial da rede apenas as variáveis que mais se destacaram nos experimentos iniciais, podem ser observados na Tabela 5.6. Na Figura 5.8 são apresentados os percentuais de utilização de cada variável.

Tabela 5-6 Resultados obtidos nos experimentos com *simulated annealing*, observando a média das duas partições (com as sete variáveis que se destacaram)

Nodos Entrada	Nodos Ocultos	Nr. Conexões	Erro Treinamento		Erro Validação		Erro Teste		% de classificação correta dos casos positivos		% de classificação correta dos casos negativos	
			Média	st.dv.	Média	St.dv.	Média	St.dv.	Média	st.dv.	Média	St.dv.
3,95	1,89	7,25	26,41666	1,47968	41,80864	0,41030	22,36728	1,40985	92,30247	2,41814	62,96296	5,23783

Figura 5-8 Percentual de utilização das variáveis que se destacaram nos experimentos com as duas partições (utilizando as sete principais como entrada em cada partição)



É importante destacar que entre os resultados obtidos com as duas partições, foi apresentada uma diferença quanto as variáveis mais utilizadas. Na partição 01 as variáveis **idade** e **situação conjugal** estiveram presentes entre as sete que

apresentaram um maior percentual de utilização.

Com a partição 02, entretanto, essas variáveis não estavam presentes, mas **condições de moradia** e **posse de bens duráveis**. A média dos nodos de entrada utilizados, porém, foi igual nas duas partições (quatro).

Com isso, observando a média obtida entre as duas partições, as quatro variáveis que apresentaram uma maior probabilidade de estar relacionada com os TMC's foram: **inserção no processo produtivo**, **renda**, **sexo** e **escolaridade**. Ressaltamos ainda, que essas variáveis sempre estiveram presentes em todos os resultados obtidos com as duas partições.

5.4 Conclusões

Os experimentos realizados com o algoritmo *simulated annealing*, otimizando automaticamente a arquitetura da rede (nodos de entrada, nodos ocultos e conexões) para análise dos fatores relacionados aos TMC's, foram mais simples do que os experimentos com o algoritmo *backpropagation* retirando cada variável do conjunto.

Esta simplicidade está no fato que o processo de escolha das variáveis que apresentavam maior possibilidade de estar relacionada com os TMC's foi automático, o que não ocorreu nos experimentos com o algoritmo *Backpropagation*.

Os resultados obtidos também foram mais uniformes entre as duas partições utilizadas, ou seja, nos experimentos com o algoritmo *Backpropagation* foram obtidos resultados diferentes com cada partição utilizada. Nos experimentos com o algoritmo *Simulated Annealing*, entretanto, apesar de algumas diferenças apresentadas, o resultado final com relação as variáveis que apresentavam maior probabilidade de estar relacionada com os TMC's, foi o mesmo quer seja a partir da análise individual de cada partição, ou com a média entre elas.

Foi possível observar também que todos os experimentos resultaram em arquiteturas bem simples, com complexidade mínima, além de uma média do percentual do erro de classificação ter sido inferior quando comparado com os resultados obtidos nos experimentos realizados com o algoritmo *Backpropagation*.

Os resultados obtidos quanto ao número de entradas escolhidas, nodos ocultos e

conexões entre as camadas adjacentes da rede, em todos os experimentos com as duas partições utilizadas, foram semelhantes.

Em geral, tanto o erro de classificação da rede quanto o percentual de classificação correta dos casos positivos e dos casos negativos da rede foram melhoradas nos experimentos realizados utilizando apenas as sete variáveis que mais se destacaram no conjunto de dados, como topologia inicial da rede.

Em nossos experimentos, das variáveis utilizadas para mensurar as condições de vida e trabalho, segundo [Ludermir, 1998], podemos observar que as variáveis que apresentaram um maior percentual de utilização nas execuções do algoritmo em cada inicialização aleatória dos pesos foram: **escolaridade**, com relação às condições de vida; **inserção no processo produtivo** e **renda**, quanto às condições de trabalho.

Dentre as onze variáveis inicialmente presentes no conjunto de dados (na topologia inicial da rede), as que apresentaram maior probabilidade de estarem relacionadas aos TMC's foram **escolaridade**, **renda**, **inserção no processo produtivo** e **sexo**.

Comparando os resultados obtidos nos experimentos realizados com aqueles apresentados por [Ludermir, 1998] utilizando o método estatístico de regressão logística, é possível observar que os resultados obtidos foram semelhantes à análise estatística com *odds ratios* simples.

Com isso, a partir dos resultados apresentados, podemos concluir que uma rede neural artificial treinada com o algoritmo *simulated annealing*, com otimização simultânea da arquitetura e pesos da rede, mostrou-se como uma opção interessante para a análise dos fatores relacionados aos TMC's. Isso, levando em consideração a base de dados utilizada em nossos experimentos.

A seguir, no Capítulo 6, serão apresentadas as conclusões obtidas com esse trabalho, além de uma comparação dos resultados com o método neural e o logístico. As considerações finais, principais contribuições e possibilidades de trabalhos futuros também são apresentadas.

Capítulo 6 - Conclusões

6.1 Introdução

Neste Capítulo será apresentada uma comparação dos resultados obtidos nos experimentos realizados com redes neurais artificiais e o método estatístico de regressão logística. Além disso, são apresentadas as considerações finais acerca do trabalho realizado, principais contribuições e possibilidades de trabalhos futuros.

6.2 Comparação de Resultados Obtidos

Os resultados obtidos em nossos experimentos serão comparados, nesta Seção, com os apresentados por [Ludermir, 1998], que utilizou o modelo estatístico de regressão logística para investigar a associação das condições de vida e trabalho com os Transtornos Mentais Comuns (TMC's).

Para investigar essa associação, [Ludermir, 1998] estimou os *odds ratios* simples e ajustados, cuja significância estatística foi avaliada pelo teste χ^2 , considerando os intervalos de confiança a 95% e valores de p ($\leq 0,05$). O modelo estatístico de regressão logística foi utilizado para analisar a independência da associação de cada variável com os TMC's.

No modelo logístico, após o controle dos *fatores de confusão*¹ com *odds ratios* ajustados, somente **escolaridade**, entre as variáveis relativas às condições de vida, e **renda** entre as variáveis relativas às condições de trabalho, estiveram relacionadas de modo estatisticamente significativa com os TMC's. Sugerindo dessa forma, um efeito direto da **escolaridade** e da **renda** sobre a saúde mental [Ludermir, 1998].

Entretanto, quando observarmos no modelo logístico as variáveis que apresentaram o menor valor de p , com *odds-ratios* simples, é possível verificar que quatro delas foram estatisticamente significativas: **sexo** ($p < 0,0001$), **escolaridade** ($p < 0,0001$), **inserção no processo produtivo** ($p = 0,0001$) e **renda** ($p < 0,0001$) [Ludermir, 2000].

Essas variáveis (**escolaridade**, **renda**, **inserção no processo produtivo** e **sexo**) foram as mesmas encontradas nos experimentos realizados com a rede MLP

¹ Termo utilizado por epidemiologistas para descrever uma covariável que é associada com a variável resposta de interesse e a variável independente primária, ou fator de risco [Hosmer, 1989].

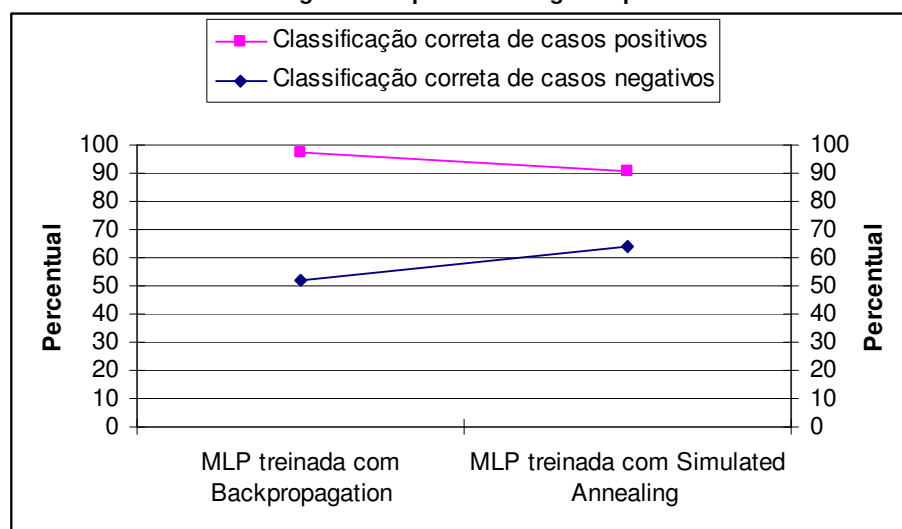
treinada com o algoritmo *simulated annealing*, as quais apresentaram maior probabilidade de estarem relacionadas aos TMC's. Ou seja, os resultados obtidos em nossos experimentos são semelhantes àqueles apresentados no modelo logístico com *odds ratios* simples.

Em geral, os experimentos com uma rede neural MLP destacou-se pela sua capacidade de classificação correta dos casos com diagnóstico positivo para TMC's, porém o mesmo não aconteceu quanto à classificação correta dos casos com diagnóstico negativo. Na Tabela 6.1 e Figura 6.1 é possível observar um resumo dos resultados obtidos com o modelo neural quanto à classificação correta dos casos com diagnóstico positivo e negativo para TMC's.

Tabela 6-1 Classificação correta de casos com diagnósticos positivo e negativo para os TMC's obtidos no modelos neural.

Experimentos	% de classificação correta de casos positivos	% de classificação correta de casos negativos
MLP treinada com <i>Backpropagation</i> (média entre as duas partições)	97,56%	51,73%
MLP treinada com <i>Simulated Annealing</i> (média entre as duas partições)	90,60%	63,88%

Figura 6-1 Resultados obtidos com redes neurais quanto ao percentual de classificação correta dos casos com diagnósticos positivo e negativo para os TMC's.



6.3 Considerações Finais

Para a realização deste trabalho, inicialmente foram realizados experimentos com outros modelos de redes neurais, a exemplo de *RBF* (Rede Neural de Funções de Base Radial) [Haykin, 2001] e *PNN* (Rede Neural Probabilística) [Leung and Tran, 1999] além de uma rede *MLP* (Multilayer Perceptron).

Entretanto, a rede MLP, com a base de dados utilizada nesse trabalho, foi a que apresentou uma maior regularidade nos resultados. Nos experimentos iniciais realizados com os outros modelos de redes neurais, foi observada uma certa inconstância nos resultados apresentados.

No caso da rede MLP, antes de optarmos pelo uso do algoritmo *backpropagation*, realizamos experimentos com o algoritmo *Levenberg Marquardt*, contudo, também foi observada uma certa inconstância nos resultados.

Apesar do modelo logístico ser a metodologia normalmente utilizada quando o objetivo é identificar os fatores de risco que possuem associação com a variável resposta [Ferreira et al, 2001], onde a interpretação dos coeficientes pode ser apresentada pelo *odds ratios*, podemos observar bons resultados com redes neurais, com relação à predição dos casos positivos para o problema estudado.

Com relação à classificação correta dos casos com diagnóstico positivo, a média obtida nos resultados dos experimentos com RNA's foram superiores à 90,0% (97,56% com o algoritmo *backpropagation* e 90,60% com o *simulated annealing*).

Quanto ao percentual de classificação correta dos casos com diagnóstico negativo para os TMC's, os resultados obtidos nos experimentos com o algoritmo *simulated annealing* (63,38%) foram superiores aos apresentados com o algoritmo *backpropagation* (51,73%).

Com os resultados apresentados neste trabalho, foi possível mostrar experimentalmente que uma rede neural MLP treinada com o algoritmo *simulated annealing*, com otimização simultânea da sua arquitetura (nodos de entrada, nodos ocultos e conexões) e pesos, apresenta-se como uma alternativa interessante para a análise dos fatores relacionados aos TMC's, quando comparado aos resultados obtidos com o modelo logístico, pois a rede neural é capaz de detectar todas as possíveis interações entre as diversas variáveis explicativas.

Além disso, os resultados apresentados nos experimentos com RNA treinada com o algoritmo *simulated annealing* com otimização automática da arquitetura da rede, apresentou-se também como uma alternativa para o processo de seleção de variáveis, eliminando variáveis do conjunto de entrada que não são importantes para o problema. A permanência de variáveis que não são importantes para a classificação dos padrões, pode provocar distorções no resultado apresentado pela rede [TU, 1996].

A identificação dos fatores (variáveis) que apresentavam possibilidade de estarem relacionados ao problema estudado, foi possível de ser feita a partir da retirada dos neurônios de entrada da rede. A análise do percentual de classificação correta dos casos positivos e negativos também foi importante em nossos experimentos, pois podemos observar o quanto cada variável influenciava na predição dos casos com diagnósticos positivo e negativo para os TMC's.

Quanto aos experimentos realizados com o algoritmo *backpropagation*, retirando manualmente cada variável do conjunto por vez, com a base de dados utilizada nesta pesquisa, não ofereceu uma visão clara de quais fatores apresentavam maior relação com os TMC's.

Nos experimentos com o *backpropagation* foram obtidos resultados diferentes em cada análise feita observando individualmente cada partição ou a média entre elas. Com isso, para que pudéssemos obter resultados mais precisos, seria interessante a realização de experimentos com a utilização de outras partições. Isso não foi possível no período de realização desta pesquisa por causa do tempo, porém apresenta-se como uma possibilidade interessante de trabalho futuro.

6.4 Principais Contribuições

A partir dos resultados obtidos neste trabalho, é possível perceber que uma rede neural pode ser utilizada como uma alternativa para a identificação dos fatores que apresentam maior probabilidade de estarem relacionados aos TMC's, podendo auxiliar no processo de tomada de decisões acerca do planejamento e intervenção de saúde pública nessa área.

Com isso, algumas contribuições importantes merecem destaque:

- A partir dos resultados obtidos, mostramos experimentalmente que o uso de uma rede MLP treinada com o algoritmo *simulated annealing*, visando a otimização simultânea da sua arquitetura (nodos e conexões) e pesos, pode ser uma alternativa interessante para a identificação de fatores que apresentam maior probabilidade de estar relacionado com os TMC's;
- Um dos problemas apontados na realização de pesquisas com redes neurais é o fato de se utilizar no conjunto de entrada, muitas vezes, variáveis que não tem importância para o problema. A permanência dessa variável no conjunto de dados pode resultar em uma generalização ruim. Em nossos experimentos podemos perceber que o uso do algoritmo *simulated annealing* para a otimização simultânea da arquitetura e pesos apresenta-se como uma alternativa para amenizar esse problema de seleção de variáveis;
- A apresentação de algumas similaridades conceituais entre o método neural e o estatístico, bem como, a apresentação de um estudo sobre aplicações de RNA's enfatizando a área de saúde, pode vir a auxiliar em pesquisas futuras, que tenham como objetivo a aplicação de RNA's na área de saúde;
- Os resultados apresentados neste trabalho podem ser úteis para futuras comparações com novos métodos propostos, cujo objetivo seja a análise de fatores relacionados ao problema estudado.

6.5 Trabalhos Futuros

Dentre as perspectivas de trabalhos futuros, podemos destacar algumas possibilidades interessantes:

- O uso de outras abordagens de RNA's para a identificação dos fatores relacionados ao problema estudado, pois, os diversos modelos de RNA tem aplicações diferentes, e diversos aspectos podem ser comparados, tais como, o erro de classificação, o percentual de classificação correta dos casos com diagnósticos positivo e negativo, entre outros;
- A realização de novos experimentos com a utilização de outras partições

geradas a partir da mesma base de dados. Seria interessante a utilização de dez partições para a realização dos experimentos, mas, em função do tempo, só duas foram utilizadas em nosso trabalho. O uso de mais partições levará a resultados mais precisos, uma vez que, a ordem dos exemplos (e os exemplos escolhidos) influenciam nos resultados obtidos pela rede neural;

- Existem várias técnicas diferentes de otimização, a utilizada em nosso trabalho foi escolhida pela simplicidade de implementação. Além disso, [Yamazaki et al, 2002] já havia utilizado a mesma técnica para a classificação de safras de vinhos, obtendo bons resultados quando comparados a outros modelos utilizados. A realização de experimentos com o uso de outras técnicas de otimização, a exemplo de *Tabu Search*, entretanto, pode ser uma alternativa interessante. *Tabu search* também é um método bem simples de ser implementado, e, guardar a lista de soluções mais recentes (tabu list), pode vir a otimizar ainda mais os resultados. Outras técnicas mais complexas, como Algoritmos Genéticos, também podem ser utilizadas.
- Finalmente, a realização de novos experimentos com o uso de outras bases de dados e aplicando a mesma metodologia utilizada neste trabalho, desenvolvendo uma análise comparativa dos resultados, apresenta-se também como uma possibilidade de trabalho futuro.

Referências Bibliográficas

ARANGO, H. G. **Bioestatística Teórica e Computacional**. Guanabara Koogan, Rio de Janeiro, 2001.

BAXT, W. G. **Use of an Artificial Neural Network for the Diagnosis of myocardial infarction**. *Annals of Internal Medicine* 115:843-848, 1991.

BEALE, R. and JACKSON, T. **Neural Computing: and introduction**. Adam Hilger, Bristol, Philadelphia and New York, 1991.

BEZERIANOS, A.; PAPADIMITRIOU, A.; ALEXOPOULOS, D. **Radial basis function neural networks for the characterization of heart rate variability dynamics**. *Artificial Intelligence I Medicine*, vol. 15 (3), pp 215-234, 1999, Elsevier Science.

BISHOP, Christopher M. **Neural Networks for pattern recognition**. Oxford University Press Inc. New York, 1996.

BLAND, Martin. **An introduction to medical statistics**. Oxford Medical Publications. 2a. ed. NY, 1996.

BRAGA, Antonio de Pádua; LUDERMIR, Teresa Bernarda; CARVALHO, André Carlos P. de L. Ferreira. **Redes Neurais Artificiais – Teoria e Aplicações**. Ed. LTC, Rio de Janeiro, 2000.

CERQUEIRA, Cezar Augusto. **Mortalidade por causas nas capitais Brasileiras e sua associação com indicadores socioeconômicos**. Dissertação de Mestrado, Universidade Federal de Pernambuco, Departamento de Estatística, 1998.

CHTIOUI, Younes; PANIGRAHI, Surujan; FRANCL, Leonard. **A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease**. *Chemometrics and Intelligent Laboratory Systems*, vol. 48, pp. 55-67, 1999, Elsevier Science.

CUENCA, Walter M. et al. **Otimização por Algoritmos Genéticos e Simulated Annealing após Tratamento de Dados por Sistemas Neuro-Fuzzy**. V Brazilian Conference on Neural Networks – V Congresso Brasileiro de Redes Neurais, pp. 175-180, 2001.

DE SOUZA, Renata Maria C. Rodrigues. **Classificação de Imagens SAR baseada em uma abordagem simbólica**. Dissertação de Mestrado, Universidade Federal de Pernambuco, Departamento de Estatística, 1999.

DOS SANTOS, Alcione M. dos. et al. **Aplicações de Redes Neurais Artificiais em dados epidemiológicos de Hepatite A**. V Brazilian Conference on Neural Networks – V Congresso Brasileiro de Redes Neurais, pp. 589-594, 2001.

DOS SANTOS, Alcione M. dos. et al. **Comparação de Redes Neurais Artificiais com regressão logística**. V Brazilian Conference on Neural Networks – V Congresso Brasileiro de Redes Neurais, pp. 595-600, 2001.

FERREIRA, C. A. et al. **Reconhecimento de padrões em estatística: uma abordagem comparativa**. V Brazilian Conference on Neural Networks – V

Congresso Brasileiro de Redes Neurais, pp. 409-414, 2001.

HAYKIN, Simon. **Redes neurais: princípios e prática**. 2^a ed., Bookman, Porto Alegre, 2001.

HOLMSTROM, Lasse et al. **Comparison of Neural and Statistical Classifiers – theory e practice**. Published in the Proceedings of ICPR'96, Vienna, Austria, vol. IV, pp. 286-290, 1996.

HOSMER, D. W. and LEMESHOW, S. **Applied logistic Regression**. John Wiley e Sons, NY, 1989.

KARKANIS, S. et al. **Image Recognition and Neuronal Networks: Intelligent Systems for the improvement of imaging information**. Minimally Invasive Therapy and Allied Technologies, vol. 9, No. 3-4, 225-230, 2000, Elsevier Science.

KLEINBAUM, David G. **Logistic Regression: a self-learning text**. Springer, 1992.

KOVACS, Z. L. **Redes Neurais Artificiais: fundamentos e aplicações**. Ed. Acadêmica, São Paulo, 1996.

LAMERS, Maarten H. et al. **Combined Neural Network Models for Epidemiological data: modelling heterogeneity and reduction of input correlations**. In Proceedings of the Third International Conference on Artificial Neural Nets and Genetic Algorithms (ICANNGA 97). Springer-Verlag, 1997.

LEGOFF, J. M.; LAVAYSSIERE, L; ROUESSE, J.; SPYRATOS, F. **Nonlinear discriminant analysis and prognostic factor classification in node-negative primary breast cancer using probabilistic neural networks**. Anticancer-Research, pp. 2213-2218, 2000, Elsevier Science.

LEUNG, PingSun; TRAN, Liem T. **Predicting shrimp disease occurrence: artificial neural networks vs. Logistic regression**. Aquaculture 187, pp. 35-49, 2000, Elsevier Science.

LOPES, Claudia R. S., LUDERMIR, Teresa B., DE SOUTO, Marcílio P., LUDERMIR, Ana B. **Rede Neural Artificial para Análise de Fatores Relacionados aos Transtornos Mentais Comuns**. VII Simpósio Brasileiro de Redes Neurais (SBRN'02), pp. 43-48, 2002.

LUDERMIR, Ana Bernarda e FILHO, Djalma A.. de Melo. **Condições de vida e estrutura ocupacional associadas a transtornos mentais comuns**. Rev. Saúde Pública, 36(2), pp. 213-221, 2002.

LUDERMIR, Ana Bernarda. **Inserção produtiva, gênero e saúde mental**. Cadernos de Saúde Pública, Rio de Janeiro, vol. 16 (3), pp. 647-659, jul-set, 2000.

LUDERMIR, Ana Bernarda. **Socioeconomic Status, Employment, Migration and Common Mental Disorders in Olinda, Northeast Brazil**. PhD Thesis, London: Department of Epidemiology and Population Sciences, London School of Hygiene and Tropical Medicine, 1998.

MARBLE, Robert P. **A neural network approach to the diagnosis of morbidity outcomes in trauma care.** Artificial Intelligence in Medicine, vol. 15 (3), pp 299-307, 1999, Elsevier Science.

MARINARO, M; SCARPETTA, S. **On-line learning in RBF Neural Networks: a stochastic approach.** Neural Network, vol. 13 (7), pp. 719-29, 2000, Elsevier Science.

MUSAVI, M. T., et al. **Mouse chromosome classification by radial basis function network with fast orthogonal search.** Neural Networks 11, pp. 769 – 777, 1998.

NAGUIB, R. N. G.; HAMDY, F. C. **A general regression neural network analysis of prognostic markers in prostate cancer.** Neurocomputing, vol. 19 (1-3), pp. 145-150, 1998, Elsevier Science.

NAIR, Jyotsna, et al. **Analysis of the symptoms of depression _ a neural network approach.** Psychiatry Research, vol. 87 (2-3), pp. 193-201, 1999, Elsevier Science.

OTTENBACHER, Kenneth J., et al. **Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke.** Journal of Clinical Epidemiology, vol. 54, 11, pp. 1159-1165, 2001, Elsevier Science.

PENDHARKAR, P. C., et al. **Association, statistical, mathematical and neural approaches for mining breast cancer patterns.** Expert Systems Applications, vol. 17 (3), pp. 223-232, 1999, Elsevier Science.

PHAM, D. T. and KARABOGA, D. "Introduction" , Intelligent Optimisation Techniques (Edited by D. T. Pham and D. Karaboga), Springer Verlag, 2000.

PLATE, Tony, et al. **A comparison between neural networks and other statistical techniques for modeling the relationship between tobacco and alcohol and cancer.** Advances in Neural Information Processing 9 (NIPS* 96) editors M. C. Mozer, M. I. Jordan and T. Petsche, MIT Press, pp. 967-973, 1997.

PRECHELT, L. **Proben 1 – A set of neural network benchmark problems and benchmarking rules.** Technical Report 21/94, Fakultat fur Informatik, Universitat Karlsruhe, Germany, September, 1994.

REID, John C. et al. **Effectiveness of stages of change and adinazolam SR in Panic Disorder: a neural network analysis.** Journal of Anxiety Disorders, vol. 10, nr. 5, pp. 331-345, 1996, Elsevier Science.

RONCO, A. L. **Use of artificial neural networks in modeling associations of discriminant factors: towards an intelligent selective breast cancer screening.** Artificial Intelligence in Medicine, vol. 16 (3), pp. 299-309, 1999, Elsevier Science.

RONCO, Alvaro L., FERNANDEZ, R. **Improving ultrasonographic diagnosis of prostate cancer with neural networks.** Ultrasound in Medicine & Biology, pp. 729-733, 1999, Elsevier Science.

SABBATINI, Renato M. E. **Um programa para o cálculo da acurácia, especificidade e sensibilidade de testes médicos.** Revista Informédica, 2 (12):pp.

19-21, 1995.

SALZBERG, Steven L. **On Comparing Classifiers: A Critique of Current Research and Methods**. Data Mining and Knowledge Discovery, 1, 1-12, Kluwer Academic Publishers, Boston. Manufactured in The Netherlands, 1999.

SARLE, Warren S. **Neural Networks and Statistical Models**. Proceedings of the Nineteenth Annual SAS User Group International Conference, Cary, NC: SAS Institute, pp. 1538-1550, 1994.

SCHWARZER, Guido, et al. **On the misuses of artificial neural networks for prognostic and diagnostic classification in Medicine**. Statistic in Medicine, vol. 19, pp. 541-561, 2000.

SCHWENKER, Friedhelm et al. **Three learning phases for radial-basis-function networks**. Neural Networks, vol. 14 (4-5) pp. 439-458., 2001, Elsevier Science.

SHUMACHER, Martin; ROBNER, Reinhard; VACH, Werner. **Neural Networks and Logistic Regression _ Part I**. Computacional Statistic and data analysis., vol. 21, pp. 661-682, 1996.

SILIPO, Rosaria; MARCHESI, Carlo. **Artificial Neural Networks for automatic ECG análisis**. IEE Trans. Signal Processing, vol. 46 (5), pp. 1417-1425, 1998.

TANG, K. Wendy; PINGLE, Girish; SRIKANT, Govardhau. **Artificial Neural Networks for the diagnosis coronary artery disease**. Journal of Intelligent Systems, vol. 7 (3-4), pp. 307-338, 1997.

TU, Jack V. **Advantages and disadvantages of using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes**. Journal Clin. Epidemiol, vol. 49 (11), pp. 1225-1231, 1996, Elsevier Science.

VALENTINE, Marilza Pereira. **Modelos lineares generalizados e algumas extensões**. Dissertação de Mestrado. Universidade Federal de Pernambuco, Departamento de Estatística, 2001.

WALCZAK, Steven; SCHARF, Jonh E. **Reducing surgical patient cost through use of an artificial neural networks to predict transfusion requirements**. Decision Suppot Systems, vol. 30, pp. 125-138, 2000, Elsevier Science.

WARNER, Brad; MISRA, Manavendra. **Understanding Neural Networks as Statistical Tools**. The American Statistician, vol. 50, pp. 284-293, 1996.

WEI, John T. et al. **Understanding artificial neural networks and exploring their potencial applications for the practicing urologist**. Urology, vol. 52, pp. 161-172, 1998, Elsevier Science.

WYMAN, Charles L. et al. **Cancer diagnostics using neural network sorting of processed images**. Proceedings of SPIE – The International Society for Optical Engineering 2760 (0), pp. 324-332, 1996.

YAMAZAKI, A.; SOUTO, M. C. P.; LUDERMIR, T. B. **Optimization of Neural**

Network Weights and Arquitetures for Odor Recognition using Simulated Annealing. 2002 International Joint Conference on Neural Networks (IJCNN' 02), May 12-17 Honolulu, Hawaii, pp. 547-552, 2002.

YANG, Zheng Rong; CHEN, Sheung. **Robust maximum likelihood training of heterocedastic probabiblistic neural networks.** Neural Networks, vol. 11 (4), pp. 739-747, 1998.

ZUPAN, A. B. C., et al. **Machine learnin for survival analysis: a case study on recurrence of prostate cancer.** Artificial Intelligence in Medicine, vol. 20 (1), pp. 59-75, 2000, Elsevier Science.

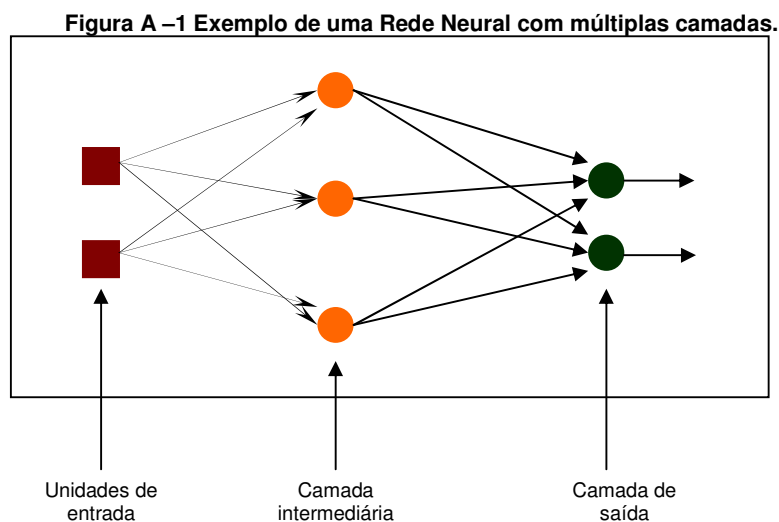
APÊNDICE A

Redes Neurais de Camadas Múltiplas

1.1 Introdução

Uma rede neural de múltiplas camadas (*Multilayer Perceptron _ MLP*) é uma rede feedforward capaz de resolver problemas não linearmente separáveis [Braga et al, 2000]. O paradigma de aprendizado utilizado é o supervisionado, ou seja, utiliza pares de entrada e saídas desejadas, ajustando os pesos da rede por meio de um mecanismo de correção de erros.

Para o treinamento de uma MLP, uma das técnicas mais utilizadas visando a minimização do erro, é o algoritmo Backpropagation ou uma de suas variações (ex.: Backpropagation com *momentum*, Rprop). A Figura A.1 mostra uma rede com múltiplas camadas.



[Haykin, 2001] destaca três características distintivas da rede Multilayer Perceptron:

1. O modelo de cada neurônio da rede inclui uma função de ativação não-linear, com uma não-linearidade suave definida normalmente pela função logística

$$y_j = \frac{1}{1 + \exp(-v_j)} \quad (\text{A.1})$$

onde v_j é a soma ponderada de todas as entradas sinápticas acrescidas do *bias* do neurônio j e y_j é a saída do neurônio.

2. A rede contém uma ou mais camadas intermediárias de neurônios, os quais capacitam a rede a aprender tarefas complexas, extraindo progressivamente as características mais significativas dos padrões de entrada.
3. A rede exibe um alto grau de conectividade, determinado pelas sinapses da rede.

A combinação destas três características e mais a habilidade que a rede possui em “aprender” a partir de um conjunto de treinamento, deriva no poder computacional do Multilayer Perceptron. Porém, essas características também são responsáveis pelas deficiências de conhecimento sobre o comportamento da rede [Haykin, 2001].

Um dos problemas enfrentados no treinamento de uma rede MLP é a definição dos seus parâmetros, sendo essa seleção um processo muitas vezes pouco compreendido. É possível que em experimentos aplicados a um determinado problema, onde fora utilizado a mesma metodologia para a sua realização, ocorra, por exemplo, tempos de treinamentos diferentes.

Outro problema que vale à pena ressaltar refere-se à definição de “quando” devemos parar de realizar os treinamentos da rede, ou seja, o critério de parada. Alguns dos critérios mais utilizados para determinar o encerramento do treinamento são:

- a) Encerramento do treinamento após um número x de iterações (épocas);
- b) Encerramento após o erro quadrático médio ficar abaixo de uma constante α ;
- c) Encerramento após a perda de generalização da rede atingir um certo limiar [Prechelt, 1994]. Esse método é conhecido como *early stopping*, o qual necessita de parte do conjunto de treinamento para a definição do conjunto de validação. O conjunto de validação não é utilizado no ajuste dos pesos, mas para medir a capacidade de generalização da rede durante a fase de treinamento.

Durante o treinamento de uma rede MLP, outro problema que pode ocorrer é o *overfitting*, ou seja, a rede tem a sua capacidade de generalização diminuída porque “memoriza” os padrões de treinamento gravando os seus ruídos.

Para evitar esse tipo de problema, *overfitting*, podemos utilizar como critério de parada o encerramento mais cedo do treinamento, quando o erro no conjunto de validação começar a subir. Outra alternativa seria utilizar *pruning*, podando os pesos da rede. [Braga et al, 2000]

1.2 Algoritmo Backpropagation

O algoritmo de aprendizagem Backpropagation é baseado na regra de aprendizagem por correção de erro. Ele utiliza pares de entradas e saídas desejadas, e, por meio de um mecanismo para correção dos erros, ajusta os pesos da rede. Para a minimização do erro obtido pela rede e o ajuste dos pesos, esse algoritmo utiliza uma generalização da regra delta de Widrow e Hoff, com aplicação do gradiente. [Braga et al, 2000]

O treinamento por retropropagação do erro consiste de duas fases: *forward* e *backward*. Cada uma dessas fases percorre a rede em um sentido através das diferentes camadas da rede.

Na fase *forward* é definido a saída da rede por um dado padrão de entrada. Nessa fase, as saídas produzidas pelos neurônios da última camada (saída real) são comparadas às saídas desejadas, e o erro para cada neurônio da camada de saída é calculado. Já na fase *backward* é utilizada a saída desejada (fornecida pelo professor externo) e a saída calculada pela rede para atualizar os pesos de suas conexões.

A partir da última camada da rede, o peso atribuído ao neurônio é ajustado visando reduzir o seu erro. O erro dos neurônios de camadas anteriores é definido pelos erros dos nodos da camada seguinte conectados a eles, ponderados pelos pesos das conexões existentes entre eles.

A fase *forward* é composta dos seguintes passos [Braga et al, 2000]:

- a) Os dados de entrada são apresentados à rede (*unidades de entrada*);
- b) Para cada camada, a partir das unidades de entrada, os nodos calculam o seu sinal de saída, e, estes sinais vão servir como entrada para a camada seguinte definir as saídas que serão produzidas nos seus nodos;

- c) As saídas que foram produzidas pelos nodos da última camada são comparadas com a saída desejada.

Na fase *backward*, por sua vez, são realizados a partir da última camada até chegar nas unidades de entrada da rede, os seguintes passos [Braga et al, 2000]:

- a) Os nodos da camada atual ajustam seus pesos de forma que reduza seus erros;
- b) Então, o erro de um nodo das camadas intermediárias é utilizado para estimar os erros dos nodos da camada seguinte que estão conectados a eles, ponderados pelos pesos das conexões existentes entre eles.

O algoritmo Backpropagation procura minimizar o erro que foi obtido pela rede com o ajuste dos pesos e limiares para que eles correspondam às coordenadas dos pontos mais baixos da superfície do erro. Com isso, ele utiliza informações do gradiente, o qual está na direção e no sentido em que a função de erro tem a taxa de variação máxima.

O algoritmo pode ficar preso em mínimos locais nas superfícies mais complexas, e visando a resolução desse problema, existem técnicas que aceleram a convergência do algoritmo, procurando dessa forma evitar a incidência de mínimos locais. Dentre essas técnicas, a que mais se destacam é a utilização de uma taxa de *aprendizado decrescente* e o uso de um termo *momentum*. Um quadro resumo do algoritmo *backpropagation*, segundo [Beale and Jackson, 1990], é exibido na Figura A.2.

A escolha do algoritmo de aprendizagem a ser utilizado no treinamento de uma rede neural artificial é influenciada pela tarefa de aprendizagem que a rede deve executar. [Haykin, 2001] identifica seis importantes tarefas que se aplicam ao uso de redes neurais, as quais são: associação de padrões, reconhecimento de padrões, aproximação de funções, controle, filtragem e formação de feixe.

Uma das dificuldades existentes na utilização do algoritmo **Backpropagation** está relacionado à sua lentidão para superfícies mais complexas, e não é raro que ele possa convergir para mínimos locais. Devido a isso, existem algumas técnicas que podem ser utilizadas para acelerar o treinamento com esse algoritmo, além de reduzir a incidência de mínimos locais. Dentre elas, destacamos:

- a) Adicionar ruídos no conjunto de dados (*jitter*), que deve ser feito quando o

conjunto de treinamento é muito pequeno;

- b) Fazer uso do termo *momentum*. Esse termo momento pode ser escrito da seguinte forma:

$$\delta_p w_{ji}(t+1) = w_{ji}(t) + \eta \delta_{pj} o_{pi} + \alpha (w_{ji}(t) - w_{ji}(t-1)) \quad (\text{A.2})$$

onde α é o fator momentum, $0 < \alpha < 1$.

- c) Adicionar nodos intermediários;
d) Utilizar uma taxa de aprendizagem decrescente.

Figura A.2: Algoritmo backpropagation adaptado de Beale e Jackson (1990).

1. Inicialização aleatória dos pesos e limiares
2. Apresentação dos exemplos de treinamento à rede (par entrada e saída desejada)

As entradas $X_p = x_0, x_1, x_2, \dots, x_{n-1}$ e as saídas $T_p = t_0, t_1, \dots, t_{m-1}$ onde n é o número de nodos de entrada e m o número de nodos de saída. Fixar w_0 para ser $-\theta$, o bias, e x_0 para ser sempre 1. X_p e T_p , para associação de padrões, representam os padrões que serão associados. E, para classificação, T_p é fixado em 0 exceto para um elemento que corresponda a classe que está em X_p que é fixada em 1.

3. Cálculo da saída atual
- Calcula cada camada

$$y_{pj} = f \left[\sum_{i=0}^{n-1} w_{ij} x_i \right]$$

e passam como entrada para a próxima camada. As saídas da camada final avaliam o_{pj} .

4. Adaptação dos pesos
- Começa da camada de saída no sentido para trás (*backwards*)

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pi}$$

$w_{ij}(t)$ representam os pesos do nodo i para o nodo j no tempo t , η é a taxa de aprendizagem, e δ_{pj} é o erro para o padrão p no nodo j .

Para as unidades de saída

$$\delta_{pj} = k o_{pj} (1 - o_{pj}) (t_{pj} - o_{pj})$$

Para unidades intermediárias

$$\delta_{pj} = k o_{pj} (1 - o_{pj}) \sum_k \delta_{pk} w_{jk}$$

onde a soma está acima do nodo k na camada sobre o nodo i .

Dentre essas técnicas, a mais frequentemente utilizada para a resolução do problema dos mínimos locais e acelerar o processo de treinamento da rede é a adição do termo *momentum* [Braga et al, 2000][Haykin, 2001].

Com o propósito de melhorar o desempenho do algoritmo *backpropagation* padrão, principalmente por causa da sua lentidão quando utilizados em problemas maiores e mais complexos, foram desenvolvidas várias alterações: *Backpropagation* com *Momentum*, *Quickprop*, *Momentum* de Segunda Ordem, *Levenberg Marquardt*, *Newton* e *Rprop* [Bishop, 1996][Braga et al, 2000][Haykin, 2001].