

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO

Alexandre Corrêa Grassi Bissacot

**Estudo Comparativo entre Regressão Logística
Binária e Redes Neurais Artificiais na Avaliação
dos Resultados Clássicos de Hosmer, Lemeshow
e Sturdivant**

Itajubá, 2015

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO

Alexandre Corrêa Grassi Bissacot

**Estudo Comparativo entre Regressão Logística
Binária e Redes Neurais Artificiais na Avaliação
dos Resultados Clássicos de Hosmer, Lemeshow
e Sturdivant**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção como parte dos requisitos para obtenção do título de **Mestre em Ciências em Engenharia de Produção**.

Área de Concentração: Qualidade e Produtos

Orientador: Prof. Dr. Pedro Paulo Balestrassi

Coorientador: Prof. Dr. Rogério Santana Peruchi

Itajubá, 2015

DEDICATÓRIA

Aos meus avós, fonte de carinho inesgotáveis.

AGRADECIMENTOS

A Deus, por estar sempre comigo.

Aos meus pais, Fernando e Gorete, pelo incentivo, investimento e apoio incondicional e aos meus irmãos, Fernando e Leonardo, pelos ótimos exemplos.

Aos meus primos, tios, avós, sogros e cunhados pelo apoio.

À Larissa, minha namorada e companheira, pela dedicação, paciência e amor em todos os momentos.

Aos professores que guiaram meu aprendizado até aqui, em especial, aos professores Pedro Paulo Balestrassi pela confiança, ideias, incentivo e discussões e Rogério Santana Peruchi pelo acompanhamento e ajuda constantes.

Aos meus amigos, especialmente aos amigos Ladenga e aos colegas do GEPE de Qualidade, indispensáveis para a motivação e evolução do trabalho.

À CNPq e à UNIFEI pelo apoio financeiro e estrutural à pesquisa brasileira viabilizando a realização deste trabalho.

“Se você não sabe para qual porto está navegando, nenhum vento é favorável”

Sêneca

RESUMO

Problemas de classificação estão presentes em diversas áreas do conhecimento. Para explorá-los e avaliá-los, diferentes metodologias podem ser utilizadas. Este trabalho busca comparar a habilidade de classificação de duas destas técnicas: regressão logística e redes neurais artificiais treinadas por planejamento ou delineamento de experimentos (*Design of Experiments*, ou DOE). A primeira pode ser considerada como técnica padrão e já é consagrada como ferramenta de auxílio na tomada de decisões. A segunda tem sido utilizada em profusão para problemas similares e apresenta-se como alternativa à regressão logística, pois assume menos restrições de uso, apesar de exigir uma simulação computacional mais complexa. Este estudo tem como objetivo realizar uma análise comparativa de oito problemas bem sedimentados, descritos no livro *Applied Logistic Regression* de Hosmer, Lemeshow e Sturdivant. Para tal análise, foi calculada a área sob a curva ROC (*Receiver Operating Characteristic*) de cada metodologia para todos os casos. A hipótese de que as redes neurais superariam os resultados encontrados pela regressão logística foi confirmada, sendo que a média para a área sob a curva dos modelos de redes neurais foram estatisticamente superiores à média dos casos modelados por regressão logística.

Palavras-Chaves: Redes Neurais Artificiais, Regressão Logística, Problemas de classificação, Curvas ROC, Planejamento de Experimentos

ABSTRACT

Classification problems are present in many different fields. Several methodologies can be used to explore and assess them. This paper compares two techniques – logistic regression and artificial neural networks trained by DOE (design of experiments) – on their classification performances. The first one can be considered as the standard statistical technique and it is well-consolidated on the literature. The second one has been used in profusion in the recent literature for similar problems and presents itself as an alternative to the logistic regression since it assumes fewer restrictions of use, in spite of requiring more computational efforts. This study aims to perform a comparative analysis of 8 well-established problems, described on the book “Applied Logistic Regression” by Hosmer, Lemeshow and Sturdivant. This analysis was carried out using the area under the ROC (Receiver Operating Characteristic) curves. The hypothesis that ANNs models would outperform LR ones was confirmed, wherein the mean for the areas under the curve modeled by neural networks were statistically superior to those modeled by logistic regression.

Keywords: Artificial Neural Networks, Logistic Regression, Classification Problems, ROC Curves, Design of Experiments

LISTA DE FIGURAS

Figura 1 - Publicações (a) e citações (b) no <i>ISI Web of Knowledge</i> para o termo " <i>logistic regression</i> "	17
Figura 2 - Publicações (a) e citações (b) no <i>ISI Web of Knowledge</i> para o termo " <i>artificial neural networks</i> "	17
Figura 3 - Publicações e citações no <i>ISI Web of Knowledge</i> para os termos " <i>Logistic Regression</i> " e " <i>Neural Networks</i> "	18
Figura 4 - Fluxograma das etapas da pesquisa	23
Figura 5 - Processo de análise de dados	26
Figura 6 - Escolha de uma técnica multivariada Fonte:	28
Figura 7 - Rede Multilayer Perceptron (MLP)	38
Figura 8- Gráfico de Pareto dos resultados expostos nos artigos revisados	47
Figura 9 - Matriz de confusão.....	48
Figura 10 - Gráfico ROC básico com 5 classificadores discretos	49
Figura 11 - Área sob a curva de 2 classificadores probabilísticos.....	50
Figura 12 - Curva ROC para o caso 1	59
Figura 13 - Efeitos principais para média e amplitude da diferença	69

LISTA DE QUADROS

Quadro 1 - Tipos de pesquisas baseados em modelos.....	22
Quadro 2 - Critérios para uso e documentação da regressão logística	32
Quadro 3 - Arranjo Plackett-Burman para doze ensaios	42
Quadro 4 - Vantagens e Desvantagens das Redes Neurais	44
Quadro 5 - Critérios de análise para ambas as metodologias	45
Quadro 6 - Função de Ligação e Informações da Resposta (Caso 1).....	57
Quadro 7 - Tabela de Regressão Logística (Caso 1)	57
Quadro 8 - Testes de aderência (Caso 1).....	57
Quadro 9 - Tabela das frequências observadas e esperadas (Caso 1)	58
Quadro 10 - Medidas de Associação (Caso 1)	58
Quadro 11 - Parâmetros de treinamento da Rede Neural para Problemas de Classificação	61

LISTA DE TABELAS

Tabela 1 - Comparação do poder discriminatório	45
Tabela 2 - Lista de artigos que comparam as técnicas	46
Tabela 3 - Descrição dos casos abordados no livro	51
Tabela 4 - Descrição das variáveis do caso 1 (Baixo Peso ao Nascimento)	52
Tabela 5 - Caso 1 (Baixo Peso ao Nascimento)	53
Tabela 6 - Resultados da Regressão Logística Univariada.....	54
Tabela 7 - Resultados da Regressão Logística Multivariada.....	54
Tabela 8 - Resultado Final da Regressão Logística Multivariada	55
Tabela 9 - Matriz de Confusão do caso 1	59
Tabela 10 - Área sob a curva ROC dos 8 casos modelados por RL.....	60
Tabela 11 - Arranjo de Plackett-Burman para o caso 1.....	67
Tabela 12 - Média e amplitude da diferença para cada rodada experimental	68
Tabela 13 - Rede escolhida para o caso 1	69
Tabela 14 - Configurações das redes para cada caso investigado	70
Tabela 15 - AUC das redes para cada caso investigado	70
Tabela 16 - AUC de ambos os classificadores	70
Tabela 17 - Teste t para dados emparelhados.....	71

LISTA DE ABREVIATURAS E SIGLAS

AD	Axiomática Descritiva
AN	Axiomática Normativa
ANN	Artificial Neural Network
AUC	Area Under the Curve
BFGS	Algoritmo Broyden-Fletcher-Goldfarb-Shanno
DCBD	Descoberta de Conhecimento em Bases de Dados
DOE	Design of Experiments
ED	Empírica Descritiva
EN	Empírica Normativa
IC	Intervalo de Confiança
LR	Logistic Regression
MLP	Multilayer Perceptron
RBF	Radial Basis Function
RNA	Rede Neural Artificial
ROC	Receiver Operating Characteristic
RL	Regressão Logística
SSM	Soft System Methodology
SVM	Support Vector Machine
UTI	Unidade de Tratamento Intensivo

SUMÁRIO

1. INTRODUÇÃO	14
1.1 Contexto.....	14
1.2 Relevância.....	16
1.3 Justificativa	18
1.4 Problema de pesquisa.....	18
1.5 Objetivos.....	19
1.5.1 Objetivo geral	19
1.5.2 Objetivos específicos	19
1.6 Delimitações	19
1.7 Método de Pesquisa	20
1.7.1 Classificação do método de pesquisa	20
1.7.2 Modelagem e simulação	20
1.7.3 Fluxograma das etapas.....	22
1.8 Estrutura do trabalho.....	23
2. FUNDAMENTAÇÃO TEÓRICA.....	25
2.1 Considerações iniciais.....	25
2.2 Estatística multivariada.....	25
2.2.1 Regressão Logística.....	29
2.2.2 Aplicações da Regressão Logística	31
2.3 Mineração de dados	35
2.3.1 Redes Neurais Artificiais.....	35
2.3.2 Aplicações das Redes Neurais Artificiais.....	42
2.4 Regressão Logística x Redes Neurais Artificiais.....	43
2.5 Análise do poder discriminatório.....	47
2.5.1 Curvas ROC (<i>Receiver Operating Characteristic</i>)	48
3. DESENVOLVIMENTO DA PESQUISA.....	51
3.1 Considerações iniciais.....	51
3.2 Modelagem através de regressão logística.....	53
3.2.1 Seleção das variáveis	53
3.2.2 Resultados da regressão logística	56

3.3	Modelagem através de Redes Neurais treinadas por DOE	60
3.3.1	Planejamento pré-experimental	60
3.3.2	Arranjo Plackett-Burman.....	66
3.4	Comparação dos resultados	70
4.	CONCLUSÕES.....	72
4.1	Considerações finais	72
4.2	Trabalhos futuros	73
	APÊNDICE A - Banco de dados de cada caso abordado.....	74
	APÊNDICE B - Resultados da Regressão Logística	81
	APÊNDICE C - Planejamento experimental das RNAs	88
	REFERÊNCIAS BIBLIOGRÁFICAS	95

1. INTRODUÇÃO

1.1 Contexto

Dados são coletados e acumulados em um ritmo acelerado. Surge assim, uma necessidade da nova geração de ferramentas e teorias computacionais auxiliar humanos a extrair informação útil (conhecimento) dos cada vez maiores volumes de dados digitais (FAYYAD, PIATETSKY-SHAPIO e SMYTH, 1996). Tais informações úteis ficaram conhecidas como a Descoberta de Conhecimento em Bases de Dados (DCBD) e se preocupam com o desenvolvimento de métodos e técnicas para dar sentido aos dados, sendo que a aplicação de métodos específicos de mineração de dados para descoberta e extração de padrões é de suma importância. Segundo Aggarwal *et al.* (2012), o progresso da pesquisa em mineração de dados tornou possível a exploração de enormes bases de dados. Apesar dessa ser uma importante contribuição, eles salientam que o objetivo final da técnica é o de possibilitar aos usuários a construção de modelos advindos de tais bases, utilizando-os para fins preditivos e analíticos, além de compartilhá-los com outras aplicações.

Segundo Witten, Frank e Hall (2011) existem quatro estilos diferentes de aprendizagem na aplicação da mineração de dados. Na aprendizagem por classificação, o esquema de aprendizagem é apresentado através de um conjunto de exemplos já classificados, a partir do qual, espera-se obter uma maneira de classificar novos exemplos. Os outros estilos são conhecidos como aprendizagem por associação, *clustering* e predição numérica.

Entretanto, antes do crescimento e popularização das técnicas de mineração de dados, os métodos estatísticos clássicos, através de métodos da estatística multivariada, já eram amplamente utilizados. Razi e Athappilly (2005) afirmam que tais métodos são utilizados na indústria há muitos anos e que eles são atrativos devido à sua metodologia solidamente estabelecida, histórico vasto de aplicação, disponibilidade de softwares e profunda aceitação. Paliwal e Kumar (2009), dizem que algumas das técnicas estatísticas comumente utilizadas para classificação e previsão são a regressão múltipla, a análise discriminante e a regressão logística.

Os problemas de classificação podem ser encontrados em uma variedade de áreas e possuem diversas aplicações. Junior (2003), Premachandra, Bahbra e Sueyochi (2009) e Kim (2011), por exemplo, verificam a possibilidade de falência em empresas. Yilmaz (2009b) e Das *et al.* (2010) utilizam técnicas de classificação para avaliar a susceptibilidade a deslizamentos de terra. Vallejos e McKinnon (2013) as utilizam para classificação de registros sísmicos. Em hidrologia, Dixon (2009) se utiliza das técnicas para identificar poços

contaminados. Em criminologia, Wouters *et al.* (2012) avaliam a influência da presença de *coffee shops* no consumo de maconha e outras drogas. Diversos outros trabalhos fazem uso da classificação para melhor identificar, modelar e prever situações e parâmetros de interesse, inclusive na indústria, onde a aplicação de classificadores se faz necessária em muitas situações.

Em pesquisas recentes, Nguyen *et al.* (2014) utilizam classificadores para fazer uma triagem no diagnóstico de gripe. Madadi *et al.* (2014) analisam os fatores que levam as mulheres a realizarem testes de mamografia. Lee *et al.* (2014) estudam-na para pesquisar a espessura do corpo caloso no início da doença de Alzheimer. Em pesquisas consagradas, Valenzuela *et al.* (1997) desenvolvem um modelo preditivo de sobrevivência em casos de parada cardíaca por fibrilação ventricular. Timmerman *et al.* (2005) visam diferenciar casos de tumores malignos e benignos em massas anexiais. Langer *et al.* (2009) fazem uso da classificação nos casos de câncer de próstata. Enfim, há uma infinidade de casos em que o uso de classificadores se faz necessário.

Para a classificação de dados existem diversas técnicas que podem ser aplicadas, derivadas da estatística clássica ou da mineração de dados, tais como: classificação bayesiana, árvores de decisão, análise discriminante, regra dos k-vizinhos mais próximos, classificadores de *kernel*, máquinas de vetores de suporte, regressão logística, redes neurais artificiais, dentre outros. Entretanto, segundo Dreiseitl e Ohno-Machado (2002), os modelos de classificação mais populares são a regressão logística seguida pelas redes neurais, responsáveis por 72% e 21,5% das publicações na base MEDLINE, na época do estudo.

Devido à sua importância na modelagem para a classificação de dados, o uso adequado da regressão logística passou a ser discutido em publicações como em: Steyerberg *et al.* (2000), Bagley *et al.* (2001) e Ottenbacher *et al.* (2004). O livro de Hosmer, Lemeshow e Sturdivant (2013), cuja primeira edição é de 1989, surgiu como referência na área, pois foi exclusivamente escrito para tratar da aplicação da técnica em diversos casos. É importante salientar que, apesar dos conjuntos de dados abordados no livro pertencerem a casos médicos, os procedimentos adotados se adequam a qualquer tipo de abordagem, sendo muito utilizados, também na engenharia. Zhang, Patuwo e Hu (1998), por sua vez, estudam o estado da arte na previsão através de redes neurais e atentam para o seu ótimo desempenho no reconhecimento de padrões. Assim, naturalmente, surgiram trabalhos a fim de comparar o desempenho das duas técnicas supracitadas. Tang *et al.* (2013), por exemplo, comparam as técnicas no contexto da disfunção autônoma cardíaca. Shi *et al.* (2013), por sua vez visam prever a mortalidade dentro de hospitais depois de cirurgia de traumatismos cranianos.

Como mencionado anteriormente, o clássico livro de Hosmer, Lemeshow e Sturdivant (2013), se utiliza da regressão logística para a modelagem de oito problemas, explicando minuciosamente a ferramenta. O presente trabalho visa desenvolver os modelos de regressão logística abordados no livro e criar modelos de redes neurais artificiais para os mesmos casos através de procedimentos estruturados, bem como comparar o desempenho atingido por cada um dos métodos através das áreas sob a curva da Característica de Operação do Receptor ou, *Receiver Operating Characteristic (ROC) curve*, em inglês.

1.2 Relevância

Já existem, na literatura, diversos estudos que remetem às técnicas de regressão logística e redes neurais artificiais, separadamente, para tratar de problemas de classificação e previsão. No entanto, quanto ao uso de ambas as técnicas e posterior comparação entre o desempenho das mesmas, o número de trabalhos é menor, principalmente aqueles que trazem uma metodologia estruturada na construção de ambos os modelos e principalmente na comparação dos mesmos. Além disso, oito diferentes casos serão discutidos e comparados enquanto que na maioria dos trabalhos, apenas um caso é analisado.

No trabalho de Dreiseitl e Ohno-Machado (2002), em que foram revisados trabalhos que comparam as duas técnicas, observou-se que mais de 90% dos trabalhos publicados se referem ao uso da regressão logística ou das redes neurais artificiais, o que reitera a importância dessas técnicas para as aplicações em questão, visto sua superioridade em comparação às outras técnicas. De forma geral, as redes neurais obtiveram desempenho similar ou superior àquele encontrado pela regressão logística na revisão realizada pelos autores. Da mesma forma, nos artigos revisados para a elaboração deste trabalho, constatou-se uma considerável superioridade das redes neurais artificiais (47% dos trabalhos consideraram-nas como a melhor técnica) em relação à regressão logística (8% dos trabalhos consideraram-na como a melhor técnica). Essa comparação é melhor detalhada na seção 2.4. De posse dessas informações, surgiu o interesse de se reavaliar os resultados do livro de Hosmer, Lemeshow e Sturdivant (2013), modelados através de regressão logística, utilizando as redes neurais artificiais.

Com relação ao interesse da comunidade científica no tema, uma busca no banco de dados *ISI Web of Knowledge* no dia 13 de janeiro de 2015 pelo termo “*logistic regression*” mostra um acréscimo nas publicações e nas citações de trabalhos a este respeito nos últimos anos, conforme ilustrado pela Figura 1.

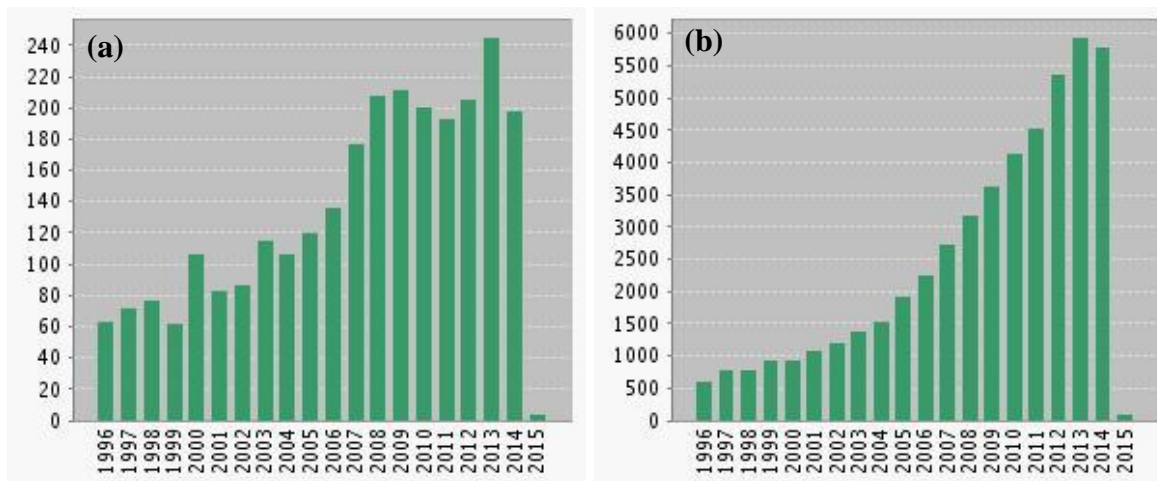


Figura 1 - Publicações (a) e citações (b) no *ISI Web of Knowledge* para o termo "logistic regression"

O mesmo ocorre ao buscar pelo termo "artificial neural networks" e pode ser observado na Figura 2.

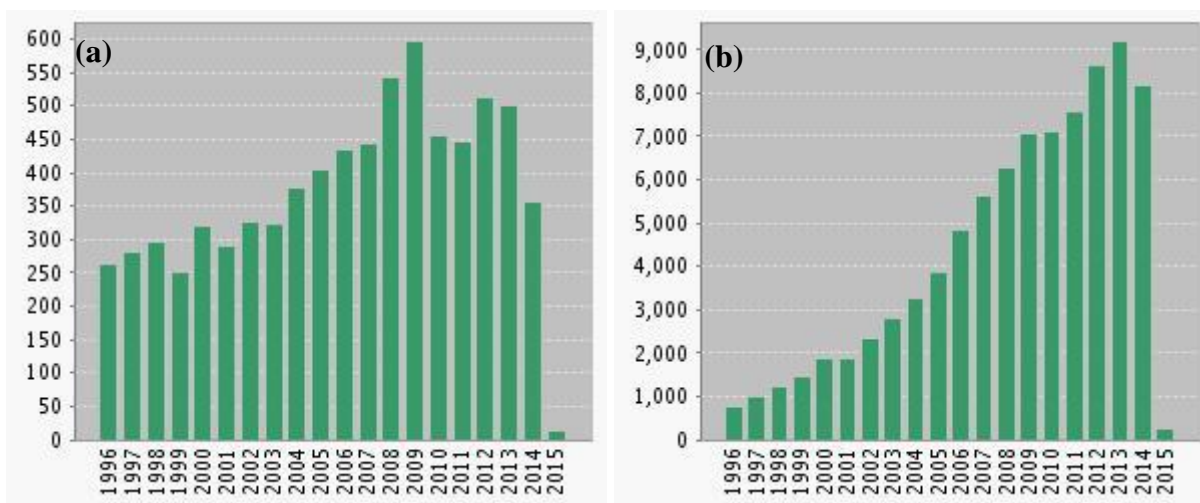


Figura 2 - Publicações (a) e citações (b) no *ISI Web of Knowledge* para o termo "artificial neural networks"

A busca conjunta pelos dois termos é apresentada na Figura 3. Pode-se constatar que o número de publicações é naturalmente inferior aos termos separados mas o número de citações continua expressivo nos últimos anos.

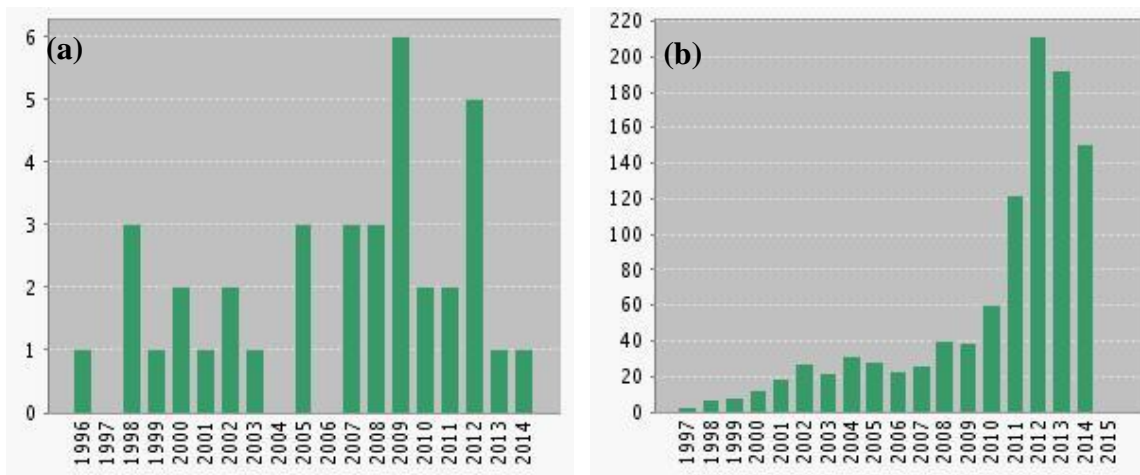


Figura 3 - Publicações e citações no *ISI Web of Knowledge* para os termos “*Logistic Regression*” e “*Neural Networks*”

1.3 Justificativa

A classificação de dados é um problema de suma importância pois, devido à sua natureza, pode estar presente em inúmeras situações. Para a obtenção de classificadores, diversos métodos e modelos foram desenvolvidos, mas nenhum deles ficou tão popular quanto a regressão logística nesse âmbito. Com o surgimento e aprimoramento das redes neurais e através de sua habilidade no reconhecimento de padrões, uma alternativa eficaz foi possivelmente encontrada. Para verificar o desempenho das redes neurais nada melhor que compará-lo, utilizando testes estatísticos, com o desempenho alcançado pela regressão logística através dos exemplos abordados no clássico livro de ensino dessa importante metodologia.

1.4 Problema de pesquisa

A metodologia de regressão logística é a mais popular e fortemente consolidada na abordagem, classificação e previsão de saídas dicotômicas. Paralelamente, o estudo e aplicação de redes neurais em diversos domínios aumentou nos últimos anos, inclusive no reconhecimento de padrões para tarefas de classificação e previsão.

O problema de pesquisa é:

Como comparar o desempenho das duas técnicas mais utilizadas em problemas de classificação? E, principalmente, será que a mais recente abordagem através de redes neurais artificiais supera os resultados conseguidos através da regressão logística?

1.5 Objetivos

1.5.1 Objetivo geral

O objetivo deste trabalho é comparar o desempenho das redes neurais artificiais treinadas por DOE com o obtido pela regressão logística binária, através das áreas sob as curvas ROC (*Receiver Operating Characteristic*), dos problemas de classificação abordados no livro de ensino de regressão logística de Hosmer, Lemeshow e Sturdivant (2013).

1.5.2 Objetivos específicos

De modo a cumprir com o objetivo geral desse trabalho, surgem os seguintes objetivos específicos:

- Criar um modelo de regressão logística para cada um dos casos analisados.
- Criar um modelo de rede neural artificial para cada um dos casos analisados.
- Obter a área sob a curva ROC para cada um dos casos, utilizando ambos os métodos.
- Utilizar testes de hipótese para verificar se há diferença estatística nos desempenhos encontrados.

1.6 Delimitações

Esta pesquisa não vislumbra esgotar inteiramente todas as questões sobre o assunto devido ao tempo escasso para tanto. Dessa forma, delimitou-se o tema da seguinte maneira:

- Utilizar bases de dados já abordadas e disponíveis na literatura.
- Comparar apenas duas das ferramentas (regressão logística e redes neurais) mais utilizadas na abordagem de problemas de classificação.
- Não desenvolver softwares específicos para a modelagem, utilizando os pacotes comerciais disponíveis no mercado.
- Utilizar apenas os efeitos principais na construção dos modelos de regressão logística e dos parâmetros das redes neurais artificiais, não considerando interações entre os mesmos.
- No planejamento ou delineamento de experimentos (DOE), para estimar os parâmetros das redes, utilizar somente o delineamento Plackett-Burman sem testar outro delineamento de aprimoramento.
- Para desenvolvimento das redes neurais, utilizar somente a arquitetura perceptron multicamadas (*multilayer perceptron* ou MLP).
- Para a comparação dos resultados, fazer uso das áreas sob a curva ROC, e não

utilizar outras ferramentas para comparação.

1.7 Método de Pesquisa

1.7.1 Classificação do método de pesquisa

Com relação aos métodos de pesquisa em Engenharia de Produção, Martins, Mello e Turrioni (2013) afirmam que a pesquisa pode ser feita através dos seguintes métodos: experimento, modelagem e simulação, *survey*, estudo de caso, pesquisa-ação e *soft system methodology*.

- O experimento é usado como método de pesquisa quando se tem um objeto de estudo, selecionam-se variáveis de influência e definem-se formas de controle e observação dos efeitos.

- A *survey* ou pesquisa levantamento é empregada quando envolver o questionamento direto das pessoas cujo comportamento se deseja conhecer e analisar.

- A modelagem e simulação serve para experimentar um sistema real através de um modelo, como neste trabalho, além de determinar como este sistema responderia às modificações propostas.

- O estudo de caso envolve a pesquisa profunda e exaustiva de um ou poucos objetos a fim de permitir o seu amplo e detalhado conhecimento.

- A pesquisa-ação é concebida e realizada para a resolução de um problema. Neste método, os pesquisadores e participantes da situação devem estar envolvidos de modo cooperativo ou participativo.

- O *soft system methodology* (SSM) auxilia a formulação e estruturação do pensamento sobre problemas em situações complexas. Seu princípio está na construção de modelos conceituais e na comparação destes com o estado real.

1.7.2 Modelagem e simulação

Segundo Chung (2004), a modelagem e simulação é o processo de criar e experimentar um sistema físico através de um modelo matemático computadorizado. Seus propósitos são:

- Aprender sobre a operação de um sistema.

Alguns sistemas são tão complexos que é difícil entender as suas operações e suas interações sem dispor de um modelo dinâmico. Em outras palavras, pode ser impossível estudar o sistema interrompendo-o e examinando cada componente isoladamente. Um típico exemplo seria tentar entender como gargalos de processo de fabricação ocorrem.

- Desenvolver políticas operacionais e de recursos para melhorar o desempenho.

Pode-se ter um sistema que seja compreensível mas que necessita de melhorias. Duas maneiras fundamentais de implementá-las são através da alteração nas políticas operacionais ou de recursos. Na primeira, pode-se incluir prioridades diferentes no agendamento das ordens de produção. Na segunda, pode-se alterar o nível de pessoal ou as pausas no trabalho.

- Testar novos conceitos e/ou sistemas antes da implantação.

Se um sistema ainda não existe ou se a compra de novos sistemas está sendo considerada, um modelo de simulação pode ajudar a prever quão bem o sistema proposto irá trabalhar. O custo para modelar um novo sistema pode ser muito pequeno em comparação com o capital investido para instalar um processo de manufatura. Os efeitos dos gastos com equipamentos também podem ser avaliados. Além disso, o uso da simulação antes da implantação do modelo pode ajudar a refinar a configuração do(s) equipamento(s) escolhidos.

- Obter informação sem alterar o sistema atual.

Modelos de simulação são possivelmente o único método disponível para experimentação com sistemas que não podem ser interrompidos. Alguns sistemas são tão críticos ou sensíveis que não é possível fazer nenhum tipo de mudança de operação ou em recursos. O clássico exemplo disso pode ser o posto de controle de segurança de aeroportos. Conduzir experimentações trariam sérios impactos na capacidade operacional ou efetividade de segurança do sistema.

Ainda segundo o autor, a modelagem e simulação trazem benefícios específicos que incluem:

- Experimentar em tempo menor;
- Requisitos analíticos reduzidos;
- Modelos mais facilmente demonstrados.

Entretanto o mesmo cita algumas desvantagens da simulação:

- Não pode prover resultados precisos quando os dados de entrada são imprecisos.
- Não pode fornecer respostas simples para problemas complexos.
- Não pode resolver problemas por si só.

Segundo Bertrand e Fransoo (2002), pesquisas baseadas em modelos podem ser

classificadas em quatro tipos: empírica descritiva, empírica normativa, axiomática descritiva e axiomática normativa. Esses tipos de pesquisa são apresentados no Quadro 1.

Quadro 1 - Tipos de pesquisas baseados em modelos

	Descritiva	Normativa
Empírica	ED	EN
Axiomática	AD	AN

Fonte: Bertrand e Fransoo (2002)

Segundo os mesmos autores, a principal preocupação do pesquisador em uma pesquisa axiomática é obter soluções dentro do modelo, de maneira que as mesmas auxiliem na percepção da estrutura do problema. Esse tipo de pesquisa produz conhecimento sobre o comportamento das variáveis, permitindo sua manipulação de acordo com o resultado desejado. Tipicamente a pesquisa axiomática é normativa e está primariamente interessada em desenvolver políticas, estratégias e ações para melhorar os resultados disponíveis na literatura existente, para encontrar uma solução ótima de um novo problema ou para comparar diferentes estratégias ao se abordar um problema específico.

Essa última definição retrata justamente aquilo que será desenvolvido no presente trabalho, de forma que essa pesquisa deve ser classificada como axiomática normativa.

1.7.3 Fluxograma das etapas

A Figura 4 apresenta o fluxograma com o passo a passo que será adotado nesta pesquisa a fim de atender os objetivos propostos. Para cada um dos oito conjuntos de dados utilizados na elaboração da pesquisa, utilizou-se a sequência ilustrada pelo fluxograma.

Primeiramente, a análise dos dados fornecidos será feita para verificar a presença de informações em branco ou relacionadas a erro de digitação. Em seguida, o conjunto de dados será modelado através da regressão logística binária e a respectiva área sob a curva ROC será calculada. Depois, será feita a parametrização da rede neural para que, em seguida, ocorra a modelagem através dela bem como o cálculo da respectiva área sob a curva ROC. Finalmente, será verificada a significância estatística da diferença encontrada nos resultados (referentes à área sob a curva calculada para cada um dos dois métodos) através de testes de hipóteses específicos.

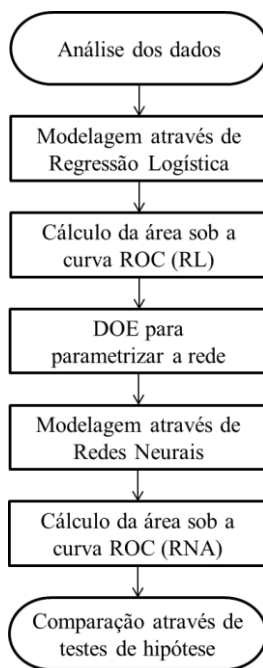


Figura 4 - Fluxograma das etapas da pesquisa

1.8 Estrutura do trabalho

Para melhor organização e compreensão, este trabalho foi dividido em 4 capítulos. O capítulo 1, introdutório, apresenta o contexto do trabalho, a relevância do tema e suas justificativas, o problema de pesquisa, os objetivos gerais e específicos a serem alcançados, a delimitação do tema, o método de pesquisa e, por fim, a explicação da estrutura adotada para elaboração do mesmo.

A fundamentação teórica é abordada no capítulo 2, trazendo os conceitos e origens da regressão logística binária e das redes neurais artificiais, com uma rápida explicação de DOE, e, como as duas técnicas têm sido aplicadas atualmente. A seguir, a comparação entre as duas técnicas é abordada com o uso de artigos que já fizeram o estudo nesse sentido e, por fim, o conceito das curvas ROC é apresentado e pormenorizado, e a utilização das mesmas para comparar os resultados obtidos é devidamente justificada.

O desenvolvimento da pesquisa propriamente dito é mostrado no capítulo 3, com algumas breves considerações iniciais e detalhamentos de cada etapa de modelagem de ambas as metodologias, regressão logística e redes neurais artificiais. A fim de cumprir com os objetivos propostos, os resultados são evidenciados e devidamente comparados.

No capítulo 4, é feita, inicialmente, a conclusão a respeito dos resultados obtidos. Em seguida, são apresentadas as considerações finais e as sugestões para trabalhos futuros. Como apêndices, foram anexados os bancos de dados dos outros 7 casos (Apêndice A), os resultados

da regressão logística para cada caso (Apêndice B) e os resultados das redes neurais artificiais para cada caso (Apêndice C).

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Considerações iniciais

As duas técnicas utilizadas para a modelagem e classificação dos dados nesse trabalho são normalmente consideradas como advindas de diferentes campos do conhecimento. A regressão logística normalmente é descrita como originária do campo de técnicas estatísticas clássicas, no ensino de técnicas de estatística multivariada (embora, obviamente, possa ser utilizada para problemas univariados). Já as redes neurais, embora possuam inúmeras aplicações, são abordadas em materiais referentes à mineração de dados, mais precisamente naqueles que tratam do aprendizado por máquinas. Primeiramente, esse capítulo explora a explicação das técnicas de estatística multivariada, enfatizando a técnica de regressão logística bem como suas aplicações. A seguir, apresenta os conceitos de mineração de dados, de onde surgem as redes neurais artificiais, que é devidamente detalhada e exemplificada. Nessa seção são apresentados também os conceitos de planejamento de experimento e do arranjo Plackett-Burman. Em seguida, a revisão de artigos que comparam as duas técnicas é devidamente realizada. Por fim, as métricas encontradas para a análise do poder discriminatório de classificadores são devidamente exploradas.

2.2 Estatística multivariada

Para Mellinger (1987), o processo de análise de dados pode ser explicado utilizando três conceitos: fatos, dados e informação, cujas relações são ilustradas na Figura 5. A realidade não pode ser acessada diretamente ou completamente e medições devem ser feitas para transformar fatos em dados. A seguir, os dados devem ser investigados antes da informação ser retirada deles: esse é o processo da análise de dados propriamente dita. Finalmente, decisões são feitas com base na informação obtida. Para o autor, a análise de dados é feita com o auxílio de vários métodos e técnicas que podem ter uma ampla ou estreita aplicabilidade e estar baseada em poucas ou numerosas suposições sobre as propriedades matemáticas dos dados em estudo. Existem, na visão do autor, diversas áreas em que se lida com um complexo sistema multidimensional que deve ser estudado como um todo, e seus componentes não fazem sentido quando são isolados uns dos outros. Nesse caso, a extração multivariada da informação se faz necessária.

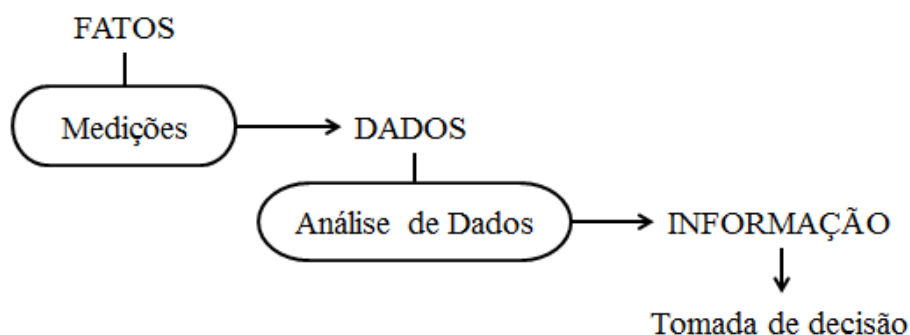


Figura 5 - Processo de análise de dados

Segundo Mingoti (2005), a estatística multivariada consiste em um conjunto de métodos estatísticos utilizados em situações nas quais várias variáveis são medidas simultaneamente, em cada elemento amostral. Em geral, as variáveis são correlacionadas entre si e quanto maior o número de variáveis, mais complexa torna-se a análise por métodos comuns de estatística univariada. A expansão na aplicação dessas técnicas somente foi possível graças ao avanço da tecnologia computacional e ao grande número de softwares estatísticos com módulos de análise multivariada implementados.

Segundo a autora, a Estatística Multivariada se divide em dois grupos: um primeiro, consistindo em técnicas exploratórias de sintetização (ou simplificação) da estrutura de variabilidade dos dados, e um segundo, consistindo em técnicas de inferência estatística. Em linhas gerais, os métodos de estatística multivariada são utilizados com o propósito de simplificar ou facilitar a interpretação do fenômeno que está sendo estudado através da construção de índices ou variáveis alternativas que sintetizem a informação original dos dados; construir grupos de elementos amostrais que apresentem similaridade entre si, possibilitando a segmentação do conjunto de dados original; investigar as relações de dependência entre as variáveis respostas associadas ao fenômeno e outros fatores (variáveis explicativas), muitas vezes, com objetivos de predição; comparar populações ou validar suposições através de testes de hipóteses.

Para auxiliar na explicação das técnicas multivariadas, Hair *et al.* (2006) apresentam uma classificação das mesmas (Figura 6). Tal classificação é baseada em três julgamentos que o pesquisador deve fazer sobre o objetivo da pesquisa e a natureza dos dados:

- As variáveis podem ser classificadas em independentes e dependentes?
- Se elas puderem, quantas variáveis são tratadas como dependentes em uma análise?

- Como ambos os conjuntos de variáveis, dependentes e independentes, são medidos?

Quando considerada a aplicação de técnicas estatísticas multivariadas, a resposta para a primeira questão indica se uma técnica de dependência ou interdependência deve ser utilizada. Uma técnica de dependência pode ser definida como uma em que uma variável ou um conjunto de variáveis é identificado como variável dependente para ser prevista ou explicada por outras variáveis conhecidas como variáveis independentes. Na técnica interdependente as variáveis não podem ser classificadas como dependentes ou independentes.

As diferentes técnicas de dependência, mostrada do lado esquerdo da Figura 6, podem ser categorizadas por duas características: o número de variáveis dependentes e o tipo da escala de medição empregada pelas variáveis. Por exemplo, se há apenas uma variável dependente e a mesma é categórica, então as técnicas apropriadas são a análise discriminante múltipla e modelos lineares de probabilidade (regressão logística binária). Para distinguir, os modelos de regressão logística diferem da análise discriminante porque eles suportam todos os tipos de variáveis independentes (métricas e não métricas) e não requerem suposição de normalidade multivariada. Em contrapartida, há também as técnicas de interdependência que são mostradas no lado direito da Figura 6. Nessas técnicas, todas as variáveis são analisadas simultaneamente num esforço para encontrar a estrutura básica para o conjunto inteiro de variáveis.

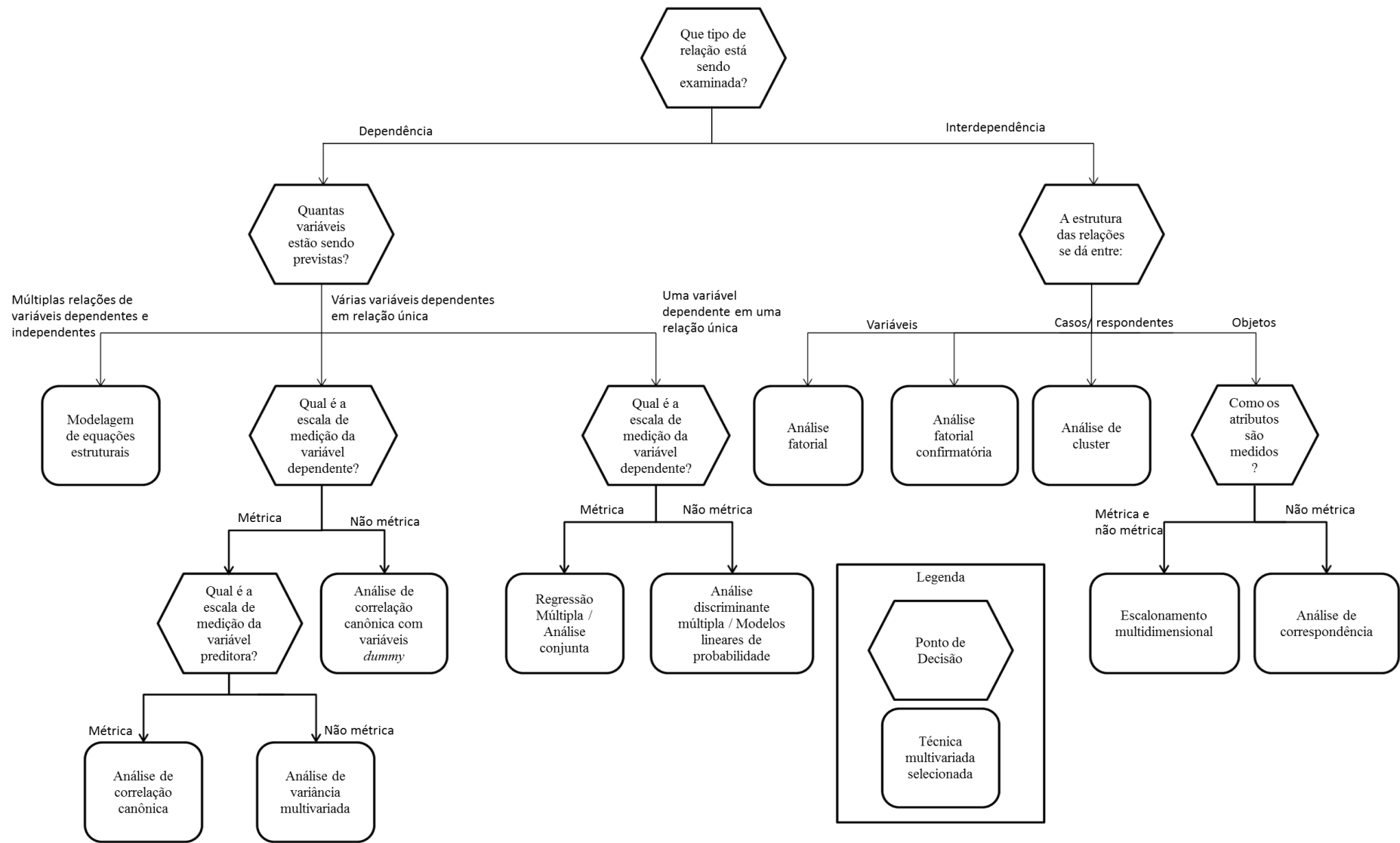


Figura 6 - Escolha de uma técnica multivariada

Fonte: Hair *et al.* (2006)

2.2.1 Regressão Logística

Segundo Montgomery e Runger (2011), a coleção de ferramentas estatísticas que são usadas para modelar e explorar relações entre variáveis que estão relacionadas de maneira não determinística é chamada de análise de regressão, uma das técnicas estatísticas mais utilizadas. Muitas aplicações da análise de regressão envolvem situações em que há mais de um regressor, para as quais o modelo de regressão múltipla deve ser desenvolvido. Para os autores, a regressão linear, técnica vastamente estudada e aplicada, funciona muito bem quando a variável de resposta é quantitativa. Já na situação em que a variável de resposta é qualitativa, como no caso de que um dispositivo obtenha “sucesso” ou “falha”, o modelo de regressão utilizado é a regressão logística, cujos parâmetros são geralmente estimados pelo método de máxima verossimilhança.

Segundo Pardoe (2006), a análise de regressão linear múltipla é imprópria para saídas dicotômicas. Isso porque os resíduos de tal modelo não satisfariam quase nunca as quatro suposições usuais para regressão linear (média zero, variância constante, normalidade, independência).

Antes de começar um estudo de regressão logística, é importante entender que o objetivo de uma análise que utiliza esse método é o mesmo de todas as outras técnicas de construção de modelos utilizada em estatística: encontrar o modelo que melhor se adere e que seja o mais parcimonioso, e ainda, razoável para descrever a relação entre uma variável de saída (dependente ou de resposta) e um conjunto de variáveis independentes (preditoras ou explicativas). (HOSMER, LEMESHOW E STURDIVANT, 2013).

Segundo Figueira (2006), o modelo de regressão logística binária é advindo dos modelos lineares generalizados. Estes últimos são especificados por três componentes: uma componente aleatória, uma componente sistemática e uma função de ligação, que conecta os valores esperados das observações às variáveis explanatórias.

De acordo com Hosmer, Lemeshow e Sturdivant (2013), muitas funções de distribuição foram propostas para a análise de variáveis de saída dicotômicas. As razões principais para se escolher a função logística é que a mesma é extremamente flexível e facilmente utilizável e também, permite uma interpretação criteriosa. A forma específica do modelo de regressão logística que se usa é:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

Uma transformação de $\pi(x)$ que é fundamental para o estudo da regressão logística é a transformação *logit*. Definida como:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x \quad (2)$$

A importância dessa transformação é que $g(x)$ tem muitas propriedades desejáveis de um modelo de regressão linear. O *logit*, $g(x)$, é linear nos parâmetros, pode ser contínuo, e pode variar de $-\infty$ a $+\infty$, dependendo da variação de x .

O método que proverá a base para a estimação do modelo de regressão logística é o da máxima verossimilhança. De maneira geral, tal método é aquele que obtém valores para os parâmetros desconhecidos de maneira a maximizar a probabilidade de se obter o conjunto de dados observados. O completo desenvolvimento para encontrar os parâmetros de um modelo de regressão logística pode ser encontrado em Hosmer, Lemeshow e Sturdivant (2013).

Anteriormente, foi descrito o processo para modelagem no contexto univariado, mas a força de uma técnica de modelagem está na sua habilidade para modelar muitas variáveis. O modelo *logit* para o caso da regressão logística múltipla é dado pela equação:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (3)$$

nesse caso o modelo de regressão logística é:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (4)$$

A interpretação de qualquer modelo requer que se possa fazer inferências práticas dos coeficientes estimados pelo mesmo. A questão a ser feita é: o que os coeficientes estimados no modelo contam sobre as questões de pesquisa que motivaram o estudo? Os coeficientes das variáveis independentes representam a inclinação (taxa de mudança) de uma função em sua variável dependente para cada unidade de mudança nas variáveis independentes.

Um importante conceito para se interpretar os parâmetros encontrados no modelo é o de *odds ratio* ou razão de possibilidades, que é uma medida de associação com vasta aplicação pois ele informa o quão mais provável (ou improvável) se torna a presença da variável de saída de acordo com o incremento que se dá na variável de entrada. Por exemplo, se y denota a presença ou ausência de câncer de pulmão e x denota se a pessoa é fumante ou não, assim

uma razão de possibilidade = 2 estima que o câncer de pulmão é duas vezes mais provável de acontecer entre fumantes em comparação aos não fumantes.

Sempre quando se trabalha com modelos de regressão, é importante destacar os testes para adequação do modelo. Para avaliar a adequabilidade e ajuste de modelos de regressão logística, Hosmer *et al.* (1997) fazem uma revisão trazendo as principais características, vantagens e desvantagens de uma série de testes utilizados para esse fim. Três dessas medidas são utilizadas para avaliar o ajuste nesse trabalho: o teste Chi-Quadrado de Pearson, Deviance e de Hosmer-Lemeshow.

De forma conclusiva, El Sanharawi e Naudet (2013) salientam que a regressão logística é um método de análise multivariada potente que permite obter uma quantificação da associação existente entre uma saída estudada e cada um dos fatores de influência, tudo isso observando o efeito simultâneo dos outros fatores. Ela permite assim interpretar situações que seriam confusas sem o seu uso. O emprego da técnica é facilitado pela utilização de softwares estatísticos.

2.2.2 Aplicações da Regressão Logística

Uma variedade de estudos apontam para o uso da técnica de regressão logística, para modelagem, previsão e classificação de dados como em Madadi *et al.* (2014) e Lee *et al.* (2014).

Bagley, White e Golomb (2001) e Ottenbacher *et al.* (2004) examinam o uso e a documentação da regressão logística na literatura. A qualidade da análise de regressão depende fortemente do entendimento das suposições inerentes ao método e de seguir adequadamente os princípios desenvolvidos para assegurar sua boa aplicação. A modelagem de maneira explícita também é necessária para relatar os resultados de forma que outros pesquisadores possam verificá-los e replicá-los. Dessa forma, eles utilizam critérios relativos à qualidade do modelo de regressão (“analíticos”) e aqueles que abordam a maneira pela qual os resultados da regressão foram analisados (“documentação”). Tais critérios são apresentados no Quadro 2.

Quadro 2 - Critérios para uso e documentação da regressão logística

Critérios utilizados	
Critérios analíticos	Eventos suficientes por variável
	Conformidade com o gradiente linear para variáveis contínuas
	Teste de interações
	Colinearidade
	Validação
	Significância estatística
	Medições do teste de aderência
Critérios documentais	Seleção das variáveis preditoras
	Codificação das variáveis
	Procedimento de aderência

Fonte: Adaptado de Bagley, White e Golomb (2001)

Critérios analíticos:

- **Eventos suficientes por variável**

Embora seja importante que o modelo inclua todas as variáveis relevantes, é importante que o mesmo não comece com mais variáveis que seja justificável pelo número de observações. Uma regra geral seria que o número de eventos da saída que menos ocorre dividido pelo número de variáveis preditoras seja maior ou igual a 10.

- **Conformidade com o gradiente linear para variáveis contínuas**

Para melhores resultados, qualquer mudança em um preditor contínuo deve ter um efeito nas chances de uma saída que seja da mesma magnitude, independente do valor da variável preditora. Hosmer, Lemeshow e Sturdivant (2013) sugerem quatro métodos para abordar essa questão. O mais simples deles é através da representação da variável através das classes de quartis e é formado por quatro passos: obter os quartis da distribuição da variável contínua, criar uma variável categórica com quatro níveis utilizando os pontos de corte baseados nos quartis, utilizar a variável categórica ao invés da contínua para no modelo multivariado, plotar os três coeficientes estimados com relação aos pontos médios dos quartis e um coeficiente igual a zero para o ponto médio do primeiro quartil. Por fim, é necessário inspecionar visualmente a plotagem e se a mesma não for linear, escolher a forma paramétrica mais lógica para a escala da variável.

- **Teste de interações**

Interações estão presentes na forma de produtos, ou seja, o termo no modelo de regressão não é uma única variável preditora mas o produto de duas variáveis. Geralmente, com amostras de tamanho modestos a decisão de se adotar interação deve ser adotada de acordo com conhecimento prévio do assunto.

- **Colinearidade**

Duas variáveis preditoras que são altamente correlacionadas apresentam um problema para qualquer análise de regressão. Se tais variáveis forem incluídas no modelo, suas contribuições estimadas podem ser imprecisas.

- **Validação**

Mesmo se os coeficientes do modelo forem computados seguindo procedimentos padronizados, o modelo final pode não ser útil ao fazer inferências sobre a população das quais os dados foram amostrados. Portanto, é desejável validar o modelo através de técnicas como o *cross validation* ou validação cruzada.

- **Significância estatística**

Testes estatísticos de significância podem ser aplicados para cada um dos coeficientes das variáveis e também para o modelo como um todo.

- **Medições do Teste de Aderência**

Em razão do modelo não se ajustar aos dados de forma exata, alguma indicação de quão bem ele se adequa deve ser dada. Medições resumidas dos testes de aderência descrevem quão bem o modelo se adequa aos valores observados.

Crítérios documentais:

- **Seleção das variáveis preditoras**

É importante salientar o motivo pelo qual se escolheu as variáveis preditoras, seja através de pesquisa prévia, seja por meio de análises bivariadas.

- **Codificação das variáveis**

É importante detalhar a codificação utilizada para as variáveis categóricas. Esse procedimento é especialmente importante quando os termos de interação são incluídos.

- **Procedimento de aderência**

As variáveis incluídas no modelo podem ser determinadas por um procedimento automático ou podem ser especificadas *a priori*.

Hosmer, Lemeshow e Sturdivant (2013) descrevem uma metodologia de sete passos para a construção de modelos através da Regressão Logística que engloba muitos dos critérios supracitados conhecida como *Purposeful Selection of Covariate*, aqui traduzida como Seleção Intencional das Variáveis. Essa foi a metodologia utilizada para a elaboração do modelo de regressão logística.

Os sete passos seguintes descrevem esse método de seleção de variáveis. A lógica por

trás do método é que o mesmo segue passos que muitos pesquisadores empregam quando examinam um conjunto de dados para construir um modelo de regressão multivariado.

Passo 1: A seleção intencional começa com uma cuidadosa análise univariada para cada variável independente. Para variáveis categóricas o teste de chi quadrado de Pearson pode ser utilizado. Para variáveis contínuas, a melhor análise univariada envolve aplicar um modelo de regressão logística univariado. Através do uso dessas análises, os candidatos para um primeiro modelo multivariado são aqueles que apresentam um *p-value* inferior a 0,25.

Passo 2: Construir o modelo multivariado incluindo todas as variáveis identificadas como candidatas no passo 1. Variáveis que não contribuem (com níveis tradicionais de significância estatística, no nosso caso 0,05) devem ser eliminadas. O novo modelo deve ser comparado com o anterior.

Passo 3: Seguindo o modelo reduzido, compara-se os valores dos coeficientes estimados no mesmo com os respectivos valores do modelo maior. Deve-se considerar que um coeficiente que representou uma mudança significativa (superior a 20%) implica a volta da variável anteriormente excluída.

Passo 4: Acrescentar cada variável excluída no passo 1 e checar sua significância no modelo. Esse passo é vital para identificar variáveis que, por si sós, não são significativas para a resposta mas tem uma importante contribuição na presença de outras.

Passo 5: Uma vez obtido o modelo com as variáveis essenciais, examina-se de maneira mais criteriosa as variáveis. Para as contínuas, por exemplo, deve-se checar a suposição de que o *logit* aumenta/diminui linearmente em função da variável.

Passo 6: Uma vez que o modelo com os efeitos principais é obtido, checa-se a presença de interações nas variáveis do modelo. Uma interação implica que o efeito de cada variável não é constante nos níveis da outra variável. A decisão de incluir a interação no modelo deve ser baseada em considerações estatísticas e práticas.

Passo 7: Antes do modelo se tornar o modelo final, deve-se avaliar sua adequabilidade e checar o seu ajuste. É importante salientar a importância desse passo antes de utilizar o modelo para propósitos inferenciais.

Outros autores corroboram com a ideia de implantação da critérios para garantir a qualidade do modelo e da respectiva documentação, como Steyerberg *et al.* (2000), Steyerberg *et al.* (2001), Courvoisier *et al.* (2011). Este último, por exemplo, reforça a importância da relação dos eventos por variável e estudo da colinearidade entre os fatores.

2.3 Mineração de dados

Segundo Liao *et al.* (2012), as técnicas de mineração de dados formam um ramo da inteligência artificial aplicada, sendo utilizadas desde os anos 60. Inovações importantes nos sistemas computacionais levaram à introdução de novas tecnologias. Nesse contexto, a mineração de dados ganhou importância como ramo de pesquisa pois possibilita procurar informações úteis em extensos banco de dados, onde o uso da informação e conhecimento de maneira inteligente se faz necessária.

Para Witten *et al.* (2011), existem quatro estilos básicos de aprendizagem nas aplicações de mineração de dados. Dentre eles, a aprendizagem para classificação, adequada para os objetivos do presente trabalho, é aquela em que um conjunto de exemplos classificados são apresentados e espera-se classificar exemplos não vistos. Ela pode ser caracterizada como supervisionada, porque a ferramenta opera sob supervisão de modo que a saída real é fornecida para os exemplos de treinamento. O sucesso da aprendizagem para classificação pode ser julgada através da utilização dos conceitos que são aprendidos no conjunto de teste e sua posterior utilização em um conjunto independente em que se conhece as saídas reais mas as mesmas não foram utilizadas para o treinamento.

Segundo Han *et al.* (2012), a classificação de dados é um processo de dois passos, sendo eles o passo de aprendizagem (onde o modelo de classificação é construído) e o passo de classificação (onde o modelo é utilizado para prever a classe dos dados estudados). No primeiro caso, deseja-se aprender um mapeamento ou função que separa as classes de dados. No segundo, o modelo é usado para a classificação propriamente dita.

Finalmente, Dehuri e Cho (2010) enfatizam que a classificação é uma das tarefas fundamentais da mineração de dados, e que, nesse contexto, as redes neurais surgiram como uma importante ferramenta de classificação. As vastas pesquisa recentes em classificações neurais estabeleceram que as redes neurais são uma alternativa promissora quando comparadas a métodos de classificação convencionais.

2.3.1 Redes Neurais Artificiais

Segundo Adya e Collopy (1998), uma rede neural artificial (RNA) é uma estrutura computacional modelada com princípios dos processos biológicos. Exploram muitas hipóteses de maneira simultânea utilizando redes paralelas compostas de elementos computacionais relativamente não lineares com diferentes pesos. É esse conjunto de pesos interconectados que contém o conhecimento gerado pelas RNAs.

Para Zhang, Patuwo e Hu (1998), as redes neurais artificiais, originalmente desenvolvidas para imitar os sistemas neurais biológicos (particularmente o cérebro humano), são compostas de um número de elementos processadores simples interconectados chamados de neurônios ou nós. Cada um recebe um sinal de entrada que é a informação total de outros neurônios ou estímulo externo, processa-o localmente através de uma função de ativação ou de transferência e produz um sinal de saída transformado para outros nós ou saídas externas. Embora cada neurônio individual implemente sua função vagarosamente ou imperfeitamente, coletivamente uma rede pode realizar um número surpreendente de tarefas de maneira eficiente. Essa característica de processamento de informação faz das redes neurais um dispositivo computacional poderoso e capaz de aprender por exemplos e depois generalizar para exemplos nunca vistos antes.

Balestrassi *et al.* (2009) citam que as redes neurais, primariamente utilizadas no campo da ciência cognitiva e engenharia, são aproximadores de funções universais e altamente flexíveis. Zhang (2000) destaca que as redes neurais emergiram e se sedimentaram como uma ferramenta importante para a classificação. A vasta atividade em classificações neurais estabeleceu que as mesmas são uma alternativa promissora com relação aos métodos de classificação convencional.

A popularidade das RNAs, segundo Zhang (2007), se deve à sua poderosa capacidade de modelagem no reconhecimento de padrões, classificação de objetos, e futuras previsões sem que seja necessária suposições irreais feitas a priori sobre a estrutura e o processo de geração do modelo. O processo de modelagem é altamente adaptativo e o modelo é largamente determinado pelas características ou padrões que a rede aprende dos dados no processo de aprendizagem. Essa abordagem orientada por dados é altamente aplicável em qualquer situação do mundo real onde a teoria das relações entre variáveis são escassas ou difíceis de descrever mas os dados são numerosos ou fáceis de coletar.

Embora haja muitos tipos de redes neurais para propósitos de classificação, as redes alimentadas diretamente com múltiplas camadas são as mais estudadas e aplicadas em problemas de classificação. Cerca de 95% das aplicações encontradas na literatura utilizam esse tipo de modelo neural (ZHANG, 2000; ZHANG, 2007).

Pontes (2011) afirma que o projeto de uma rede neural envolve a escolha das variáveis utilizadas como entrada, a definição do número de variáveis de saída e a determinação do número de casos utilizados para treinamento. Envolve também a escolha de uma determinada arquitetura de rede. Após a seleção da arquitetura, deve-se proceder ao projeto dos parâmetros da rede, que incluem a determinação do número de unidades na camada oculta, a escolha dos

algoritmos de treinamento (que também podem ter seus próprios parâmetros de configuração), técnicas de poda e a estratégia de pré e pós processamento dos dados. A etapa final do projeto envolve o treinamento da rede.

Existem diferentes arquiteturas de redes neurais, como as redes MLP (perceptron multi-camada), redes RBF (função de base radial) e outras. Redes neurais podem apresentar uma única camada oculta ou múltiplas camadas. As redes podem ser acíclicas (ou *feedforward*), nas quais saídas de um neurônio podem ser utilizadas como entrada somente pelas camadas seguintes; ou podem ser cíclicas (ou *feedback*), nas quais saídas de um neurônio podem ser usadas como entradas por neurônios da mesma camada ou até de camadas anteriores. Redes neurais podem ser conectadas parcial ou completamente. Neste último caso, que corresponde à grande maioria das aplicações, cada neurônio é conectado a todos os neurônios da camada seguinte da rede (BRAGA; CARVALHO E LUDERMIR, 2007; HAYKIN, 2009).

Braga, Carvalho e Ludermir (2007) afirmam que a capacidade de aprender através de exemplos e de generalizar a informação aprendida é o atrativo principal da solução de problemas através de redes neurais artificiais. Por aprendizado, entende-se o processo pelo qual os parâmetros de uma rede neural são ajustados através de uma forma continuada de estímulo pelo ambiente em que a rede está operando, sendo o tipo específico de aprendizado definido pela maneira particular como ocorrem os ajustes realizados nos parâmetros (HAYKIN, 2009). Existem dois tipos de aprendizados principais: aprendizado supervisionado e aprendizado não supervisionado.

No aprendizado supervisionado, objeto deste trabalho, um conjunto de exemplos contendo valores de entradas e correspondentes saídas é reunido por um supervisor e apresentado à rede um determinado número de vezes. Os exemplos podem ter sua origem em dados históricos, ou modelos previamente existentes. Durante o treinamento, os parâmetros da rede são ajustados de acordo com algum algoritmo. O objetivo é o ajuste de pesos e limiares de forma a mapear-se o relacionamento entre as grandezas de entrada e saída (BRAGA; CARVALHO E LUDERMIR, 2007). Se o treinamento for bem realizado, a rede terá modelado a relação em suas sinapses, e poderá ser utilizada para realizar previsões para casos de entrada diferentes dos utilizados para treinamento, para os quais a saída não é conhecida de antemão pela rede. As já mencionadas redes MLP se utilizam do aprendizado supervisionado. Já no aprendizado não supervisionado, não utilizado no presente trabalho, não há um supervisor para acompanhar o processo de aprendizagem.

Sick (2002) recomenda conduzir mais de um experimento com cada configuração de rede. Com um único experimento, uma dada configuração pode apresentar um resultado

correspondente a um mínimo local da função de erro, ou ter seu desempenho influenciado pela iniciação aleatória dos pesos de suas sinapses. Resultados assim obtidos podem superestimar ou subestimar o potencial da rede.

REDES DE ARQUITETURA MLP (PERCEPTRON MULTICAMADA)

Redes com arquitetura perceptron multi-camada, ou MLP (do inglês *Multilayer Perceptron*) constituem o modelo de redes neurais mais popular na literatura. Cada unidade desse tipo de rede efetua o produto escalar do seu vetor de entrada pelo vetor de pesos associados a cada entrada. O resultado é comparado a um valor de limiar. Em caso de disparo, o resultado do produto escalar é utilizado como variável independente de uma função de ativação, cuja variável dependente constituirá a saída do neurônio. A função sigmóide é a mais comumente utilizada como função de ativação, embora outras funções também sejam utilizadas. Os neurônios são conectados em modo acíclico. Isso significa que neurônios de uma camada são ligados unicamente a neurônios da camada seguinte (HAYKIN, 2009). Na Figura 7 é ilustrado o conceito de rede MLP.

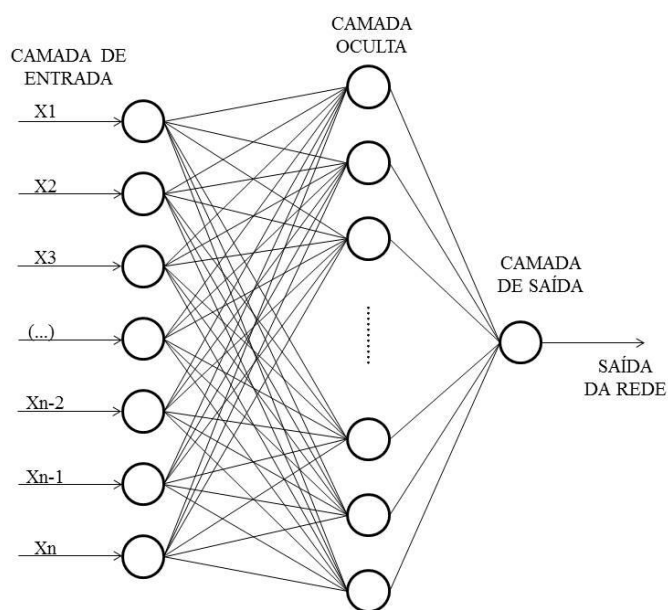


Figura 7 - Rede Multilayer Perceptron (MLP)

Fonte: Adaptado de Özel; Karpap (2005)

Segundo Bishop (2006), tal modelo de rede é capaz de modelar funções complexas. O número de camadas ocultas e o número de neurônios em cada uma destas camadas

determinam a complexidade da função que a rede é capaz de modelar. Os parâmetros livres do modelo da rede são os pesos das sinapses e o valor do limiar de disparo dos neurônios.

TREINAMENTO DE REDES MLP

Segundo Zhang, Patuwo e Hu (1998), dentre os algoritmos utilizados para treinamento de redes MLP, o mais conhecido e amplamente utilizado é o algoritmo de retro-propagação, que é essencialmente o método do gradiente descendente. O algoritmo procura minimizar o erro obtido ajustando pesos e limiares de ativação em busca dos pontos de mínimo da superfície de erro.

O treinamento através do algoritmo de retro-propagação repete-se por um certo número de vezes. As repetições são chamadas de épocas de treinamento. O algoritmo de retro-propagação pode ser executado baseado em exemplos ou em bateladas. No modo baseado em exemplos, o erro é retro-propagado, e os parâmetros livres são ajustados após a apresentação de cada exemplo. No modo baseado em bateladas, os parâmetros são ajustados somente após um ciclo completo de apresentação de exemplos, ciclo este que recebe o nome de época de treinamento (HAYKIN, 2009).

Os casos de treinamento são submetidos um de cada vez à rede. A saída da rede para cada caso é comparada ao valor de saída real do caso correspondente. Ao final de cada ciclo (após um exemplo ou após uma época, dependendo do modo de treinamento), os erros de predição são agrupados e uma função de erro é calculada.

Este valor da função erro, juntamente com o valor calculado do gradiente da superfície de erro, é utilizado para ajustar os parâmetros livres da rede. Os valores dos pesos e limiares são ajustados no sentido do maior decremento do erro. Uma sequência de ajustes nesta direção leva uma redução do erro para um mínimo (que pode ser um mínimo local ou absoluto). Ajustes mais amplos podem levar a uma convergência mais rápida, mas podem levar também à ultrapassagem do ponto de mínimo. Ajustes muito pequenos, por sua vez, requerem maior número de ciclos para a obtenção de um mínimo. Na prática, os ajustes são proporcionais à derivada da função de erro, e também a um parâmetro denominado taxa de aprendizado. O algoritmo também prevê um termo denominado de momento, através do qual o ajuste em um sentido da função de erro é incrementado quando sucessivos ajustes ocorrem naquela direção (HAYKIN, 2009).

O treinamento ocorre, portanto, em duas etapas: a etapa direta (ou *forward*), na qual se define a saída da rede para um determinado padrão de entrada ou conjunto de padrões de entrada; e a etapa reversa (ou *backward*), que utiliza a diferença entre a saída de rede e a saída

desejada, o erro da rede, para atualizar os parâmetros livres. O número de vezes em que o processo é repetido pode ser definido de diferentes maneiras. O projetista da rede pode escolher por realizar um número arbitrário de épocas, pode realizar o treinamento até que o erro seja reduzido a um valor requerido ou pode realizar o treinamento até que o erro comece a crescer após atingir um mínimo.

Segundo Haykin (2009), na aprendizagem por retropropagação, o início é realizado através de uma amostra de treinamento e o algoritmo é utilizado para calcular os pesos sinápticos de um perceptron de múltiplas camadas carregando (codificando) tantos exemplos de treinamento quanto possível para dentro da rede. Espera-se que a rede neural assim projetada seja capaz de generalizar, ou seja, deseja-se que o mapeamento de entra-saída computado pela rede seja correto para dados não utilizados. Entretanto, há a possibilidade da rede aprender um número excessivo de exemplos e acabar memorizando os dados de treinamento. Ela pode fazer isso devido à presença de características nos dados de treinamento que não estão presentes na função a ser modelada. Tal fenômeno é conhecido como excesso de ajuste (*overfitting*).

Existem outros algoritmos que procuram aumentar a eficiência na busca pelo mínimo global da função de erro, bem como acelerar o processo de treinamento das redes. O método do gradiente conjugado é um método avançado para treinar MLPs. Normalmente, sua performance é significativamente melhor que a retropropagação e pode ser usado nas mesmas situações que a mesma. É a técnica recomendada para qualquer rede com um vasto número de pesos e múltiplas unidades de saída. Para redes menores, o método Quasi-Newton pode ser melhor. Ambos os algoritmos são atualizados em lote: enquanto a retropropagação ajusta os pesos da rede após cada caso, o gradiente conjugado e o Quasi-Newton BFGS trabalham no gradiente médio da superfície de erro de todos os casos antes de atualizar os pesos no fim da época (BISHOP, 2006; SHEPHERD, 1997).

TREINAMENTO DE REDES MLP ATRAVÉS DE DOE – PLACKETT-BURMAN

Montgomery (2009) define DOE (do Inglês, *Design of Experiments*) ou planejamento de experimentos como uma ferramenta utilizada para descobrir relacionamentos de causa e efeito em um determinado processo ou sistema. O autor acrescenta que se trata de um teste ou de uma série deles, no qual as mudanças intencionais são realizadas nas variáveis de entrada do processo ou sistema, refletindo em alterações que podem ser identificadas na variável de resposta do mesmo.

Coleman e Montgomery (1993) apresentam algumas diretrizes para planejar um experimento que apesar de ter um enfoque na área industrial pode ser adaptado para a simulação computacional. São elas:

- a. **Reconhecimento e definição do problema.** Esse pode parecer um ponto óbvio mas em experimentos industriais – e também em simulações computacionais – não é simples visualizar todos os detalhes de um problema. Uma correta definição do problema normalmente fornece maior foco para alcançar o objetivo proposto.
- b. **Escolha dos fatores e níveis.** Em experimentos do mundo real, apenas um pequeno número de fatores são normalmente variados pois é praticamente impossível de controlar um número grande de fatores. Muitos experimentos lidam com cinco ou menos fatores. Entretanto, uma variedade de fatores existem para modelos de simulação usados na prática.
- c. **Seleção da variável de resposta.** Na simulação computacional, a resposta usualmente é uma variável estimada que pode ser repetida sem erro ao se estabelecer as mesmas condições iniciais.
- d. **Escolha do arranjo experimental.** Esse termo denota a matriz onde as colunas representam os fatores de entrada e cada linha representa uma combinação dos níveis de cada fator. Envolve o tamanho da amostra, a ordem dos ensaios (o que é irrelevante para simulação computacional), além de algumas restrições.
- e. **Condução do experimento.** Conduzir os experimentos em simulação computacional é normalmente mais fácil que em experimentos industriais.
- f. **Análise estatística dos dados.** Métodos estatísticos não precisam ser muito elaborados se as demais diretrizes tiverem sido seguidas.
- g. **Conclusões e recomendações.** Se combinadas com o conhecimento do processo e senso comum, a análise estatística pode levar à conclusões e recomendações contundentes.

Existem diversas metodologias para se planejar um experimento. Balestrassi *et al.* (2009) se utilizaram da metodologia de Taguchi e do fatorial fracionado para estimar os melhores parâmetros de redes neurais visando a previsão de séries temporais. Magallanes e Olivieri (2010), por sua vez, atentam para o fato de que a metodologia Plackett-Burman é extremamente econômica do ponto de vista de ensaios realizados. Doze ensaios Plackett-Burman, por exemplo, possibilitam o estudo de onze fatores. Em comparação, um fatorial completo de dois níveis requer 32 experimentos para cinco fatores.

A respeito desse delineamento, Montgomery (2009) observa que o mesmo é classificado como delineamento fatorial fracionado de dois níveis desenvolvido por Plackett e Burman (PLACKETT e BURMAN, 1946). Os arranjos experimentais são criados em tamanho múltiplos de quatro e são compostos de sinais positivos para os níveis superiores, e sinais negativos para os níveis inferiores. Além disso, os projetos seguem a regra $k = (N-1)$, onde k é o número de parâmetros estudados e N é o total de ensaios realizados. Como ilustração, o **Erro! Fonte de referência não encontrada.** apresenta uma matriz com doze ensaios, $N=12$ e $k=11$, onde a primeira linha é o padrão para delineamentos Plackett-Burman com doze ensaios. A segunda linha é formada a partir da primeira, copiando-se a mesma com um deslocamento da coluna para a direita e assim sucessivamente até a última linha que é formada somente por sinais negativos.

Quadro 3 - Arranjo Plackett-Burman para doze ensaios

Fonte: Couto (2012)

+	-	+	-	-	-	+	+	+	-	+
+	+	-	+	-	-	-	+	+	+	-
-	+	+	-	+	-	-	-	+	+	+
+	-	+	+	-	+	-	-	-	+	+
+	+	-	+	+	-	+	-	-	-	+
+	+	+	-	+	+	-	+	-	-	-
-	+	+	+	-	+	+	-	+	-	-
-	-	+	+	+	-	+	+	-	+	-
-	-	-	+	+	+	-	+	+	-	+
+	-	-	-	+	+	+	-	+	+	-
-	+	-	-	-	+	+	+	-	+	+
-	-	-	-	-	-	-	-	-	-	-

2.3.2 Aplicações das Redes Neurais Artificiais

Segundo Ahmad *et al.* (2012), uma infinidade de aplicações das redes neurais artificiais podem ser encontradas na literatura. Em casos médicos, similares àqueles encontrados no livro sob investigação, elas são especialmente utilizadas para diagnósticos, análise de imagens e previsão de doenças. São usadas em diversas áreas tais quais patologia, radiologia, cardiologia, pneumologia, gastroenterologia. Redes neurais tem sido amplamente empregadas na classificação e diagnóstico de diferente doenças.

Lisboa (2002) faz uma revisão para avaliar os benefícios traduzidos pela aplicação de redes neurais artificiais nas funções de diagnóstico, prognóstico, análise de sobrevivência, em oncologia, cuidados críticos e medicina cardiovascular. Lisboa e Taktak (2006), por sua vez,

fazem uma revisão sistemática para acessar o benefício das redes neurais artificiais como ferramentas de tomadas de decisão em doenças cancerígenas. Kuruvilla e Gunavathi (2014) utilizam a ferramenta para ajudar na classificação de câncer de pulmão na tomografia computadorizada.

Itchhaporia *et al.* (1996) analisa a situação do uso das redes neurais com relação à aplicação na medicina cardiovascular. O autor cita a possibilidade de se utilizar a ferramenta para problemas de diagnóstico e tratamento da doença da artéria coronária, infarto do miocárdio, interpretação de eletrocardiogramas, detecção de arritmias e análise de imagens em radiografias e sonografias. Jiang, Trundle e Ren (2010) fornecem uma revisão da aplicação das redes neurais artificiais para interpretação e detecção de bordas em imagens médicas. Hayashi e Setiono (2002) combinam redes neurais artificiais para previsões de doenças hepatobiliares.

Al-shayea (2011) avalia o uso das redes neurais artificiais em diagnósticos de doenças em dois casos. O primeiro caso é a nefrite aguda com os dados sendo os sintomas da doença e o segundo é a doença cardíaca com dados das imagens da tomografia computadorizada. Os resultados atingidos para a classificação corretas dos pacientes foram de 99% para o primeiro caso e 95% para o segundo, ratificando a significativa contribuição da metodologia para identificar uma pessoa infectada.

2.4 Regressão Logística x Redes Neurais Artificiais

Pepe (2005) afirma que a classificação e a previsão são componentes fundamentais. E, apesar de sua importância, recebem atenção inadequada dos estatísticos médicos. Desenvolvimentos tecnológicos recentes, no entanto, estimularam um aumento de interesse na avaliação de classificadores e previsores. O autor reitera que, infelizmente, os classificadores são raramente perfeitos e erros de classificação podem levar a consequências sérias. Um indivíduo com uma doença que for erroneamente classificado como não doente, por exemplo, pode não receber o tratamento vital para cura. Assim, a avaliação da performance dos classificadores é extremamente desejável.

Segundo Dreiseitl e Ohno-Machado (2002), os métodos mais populares para construir modelos de reconhecimento de padrões a partir de dados reais que sirvam de treinamento são a regressão logística e as redes neurais artificiais. Eles comprovam tal afirmativa explorando as publicações no acervo Medline, computando 28500 estudos para regressão logística e 8500 para redes neurais, o que representa mais de 90% das publicações totais.

O clássico estudo de Tu (1996) foi considerado de grande impacto para as publicações na área. Nele, foram desenvolvidas as principais vantagens e desvantagem no uso das redes neurais artificiais em comparação com a regressão logística. O esquema comparativo pode ser observado no Quadro 4.

Dreiseitl e Ohno-Machado (2002) analisaram 72 artigos que usavam ambas a metodologias de acordo com os seguintes critérios: se detalhes sobre a construção do modelos foram fornecidos (seleção de variáveis para regressão logística, seleção de parâmetros e evitar sobreajuste nas redes neurais), se estimativas não enviesadas do erro de generalização foram incluídas, se medidas do poder discriminatório (tópico que será abordado na próxima sessão) foram dadas e se informações relativas à validação foram feitas. O Quadro 5 descreve cada um dos critérios de análise adotados para ambas as metodologias e o que é analisado em cada um deles.

Quadro 4 - Vantagens e Desvantagens das Redes Neurais

Vantagens	Desvantagens
1- Modelos de redes neurais requerem menos treinamento estatístico	1- Redes neurais são "caixas pretas" e têm habilidade limitada para identificar explicitamente relações causais
2- Modelos de redes neurais podem detectar complexas relações não-lineares entre as variáveis	2- Modelos de redes neurais podem ser mais difíceis de serem aplicados em campo
3- Modelos de redes neurais têm a habilidade de detectar todas as possíveis interações entre as variáveis	3- A modelagem de redes neurais requer melhores recursos computacionais
4- Redes neurais podem ser desenvolvidas utilizando diferentes algoritmos de treinamento	4- Modelos de redes neurais são propensas ao sobreajuste
	5- O desenvolvimento do modelo é empírico e problemas metodológicos ainda devem ser resolvidos

Fonte: Adaptado de Tu (1996)

Quadro 5 - Critérios de análise para ambas as metodologias

Critério	Exemplo
1- Inclusão de detalhes do processo de construção do modelo de Regressão Logística	Como as variáveis foram selecionadas?
2- Inclusão de detalhes do processo de construção do modelo de Redes Neurais	Como os parâmetros foram selecionados e o que foi feito para evitar o sobreajuste?
3- Presença de estimativas não enviesadas do erro de generalização	Foram utilizadas técnicas para evitar o erro de generalização (conjunto de teste, validação cruzada ou <i>bootstrapping</i>)
4- Fornecimento de medidas do poder discriminatório	Quão bem as duas classes da variável de saída foram separadas?
5- Inclusão de informações sobre a calibração	Quão perto as previsões do modelo estão do modelo real?

Fonte: Adaptado de Dreiseitl e Ohno-Machado (2002)

Com relação aos resultados encontrados comparando-se o poder discriminatório dos modelos de regressão logística e de redes neurais artificiais, pôde-se constatar que ambos os modelos tem desempenho similar na maioria dos casos, com as redes neurais superando a regressão logística nos casos restantes, conforme ilustrado pela Tabela 1.

Tabela 1 - Comparação do poder discriminatório

	<i>RNA - melhor (%)</i>	<i>RL - melhor (%)</i>	<i>Sem diferença (%)</i>
Com testes estatísticos	18	1	42
Sem testes estatísticos	33	6	0

Fonte: Adaptado de Dreiseitl e Ohno-Machado (2002)

No decorrer dos anos, tem havido um número crescente de artigos que exploram o uso das RNAs como uma metodologia promissora quando comparada à Regressão Logística (RL), técnica mais bem consolidada. Uma revisão de 38 artigos desde 1996 até 2013 foi realizada e é resumida na Tabela 2. A primeira coluna apresenta o artigo revisado. A segunda coluna mostra a natureza do artigo, isto é, se os autores priorizaram uma abordagem conceitual, uma revisão ou uma aplicação. A terceira apresenta os objetivos do artigo analisado e a quarta expõe qual metodologia (RNAs ou RL) obteve um melhor desempenho.

Tabela 2 - Lista de artigos que comparam as técnicas

Artigo	Natureza	Objetivo	Performance (RNA x RL)
Tu (1996)	Conceitual	Prever resultados médicos	Inconclusiva
Schumacher <i>et al.</i> (1996)	Aplicação	Comparação geral de ambos os métodos	Inconclusiva
Vach <i>et al.</i> (1996)	Conceitual	Comparação geral de ambos os métodos	Inconclusiva
Freeman <i>et al.</i> (2000)	Aplicação	Prever morte hospitalar depois de angioplastia	Similar
Leung e Tran (2000)	Aplicação	Prever surto de doenças em camarões	RNA
Borque <i>et al.</i> (2001)	Aplicação	Prever o estágio patológico (pacientes sem próstata)	Similar
Chun <i>et al.</i> (2007)	Aplicação	Prever a probabilidade de câncer de próstata	RL
Kawakami <i>et al.</i> (2008)	Aplicação	Prever a probabilidade de câncer de próstata	RL
Ottobacher <i>et al.</i> (2001)	Aplicação	Prever rehospitalização de pacientes com enfarto	Similar
Nguyen <i>et al.</i> (2002)	Aplicação	Prever morte ou amputação na doença meningocócica	Similar
DiRusso <i>et al.</i> (2002)	Aplicação	Analisar sobrevivência em traumas pediátricos	RNA
Dreiseitl e Ohno-Machado (2002)	Revisão	Comparação geral de ambos os métodos	RNA
Hajmeer e Basheer (2003)	Aplicação	Classificar crescimento de bactérias	RNA
Ottobacher <i>et al.</i> (2004)	Aplicação	Prever cenário de vida depois de fratura no quadril	Similar
Lin <i>et al.</i> (2010)	Aplicação	Prever mortalidade em idosos com fratura no quadril	RNA
Ergün <i>et al.</i> (2004)	Aplicação	Classificar a estenose da carótida em pacientes com diabetes	RNA
Yesilnacar e Topal (2005)	Aplicação	Analisar susceptibilidade de deslizamento de terras	RNA
Yilmaz (2009)	Aplicação	Analisar susceptibilidade de deslizamento de terras	RNA
Pradhan e Lee (2010)	Aplicação	Analisar susceptibilidade de deslizamento de terras	RNA
Choi <i>et al.</i> (2012)	Aplicação	Analisar susceptibilidade de deslizamento de terras	RL
Song <i>et al.</i> (2005)	Aplicação	Diferenciar entre tumores benignos e malignos no peito	Similar
McLaren <i>et al.</i> (2009)	Aplicação	Detecção e diagnóstico de lesões no peito	Similar
Green <i>et al.</i> (2006)	Aplicação	Prever a síndrome coronária aguda	RNA
Chiang <i>et al.</i> (2006)	Aplicação	Diferenciar entre lojas online e lojas tradicionais	RNA
Liew <i>et al.</i> (2007)	Aplicação	Prever doença em pacientes após cirurgia bariátrica	RNA
Gutiérrez <i>et al.</i> (2008)	Aplicação	Mapear a infestação de <i>Ridolfia Segetum</i> (praga persistente)	Inconclusiva
Kurt <i>et al.</i> (2008)	Aplicação	Prever a doença cardíaca coronária	RNA
Al Housseini <i>et al.</i> (2009)	Aplicação	Prever o risco de cesariana em nulíparas	RNA
Caocci <i>et al.</i> (2010)	Aplicação	Prever a ocorrência da doença do enxerto contra o hospedeiro	RNA
Pavlekovic <i>et al.</i> (2010)	Aplicação	Reconhecer crianças talentosas em matemática	Similar
Trtica-Majnaric <i>et al.</i> (2010)	Aplicação	Prever a efetividade da vacinação contra gripe	RNA
Chen <i>et al.</i> (2012)	Aplicação	Diferenciar nódulos malignos e benignos no pulmão	RNA
Larasati <i>et al.</i> (2012)	Aplicação	Pesquisa psicológica	RNA
Pourshahriar (2012)	Aplicação	Pesquisa psicológica	Similar
Swiderski <i>et al.</i> (2012)	Aplicação	Avaliar a condição financeira de uma empresa	Inconclusiva
Askın e Gokalp (2013)	Aplicação	Avaliar os resultados de estudantes em matemática	Similar
Morteza <i>et al.</i> (2013)	Aplicação	Prever o nível de albuminúria em diabetes tipo 2	Inconclusiva
Vallejos e McKinnon (2013)	Aplicação	Classificar registros sísmicos	Similar

A Figura 8 resume as características observadas na última coluna, informando, através de um gráfico de Pareto, a porcentagem de trabalhos em que as RNAs superaram a RL e vice versa. Performances similares ou com resultados inconclusivos também são apresentadas. Pelo resultado, infere-se que dentre os trabalhos analisados, as RNAs tiveram um desempenho superior à regressão logística em quase metade dos mesmos, enquanto que o contrário foi verificado em menos de 10% dos artigos estudados. Fazendo um teste de hipóteses para duas proporções a fim de avaliar se essa diferença percentual é significativa, obtém-se um $p\text{-value}=0,000$. Dessa forma nega-se a hipótese nula de proporções iguais, ou seja, a superioridade observada nessa quantidade de artigos é significativa.

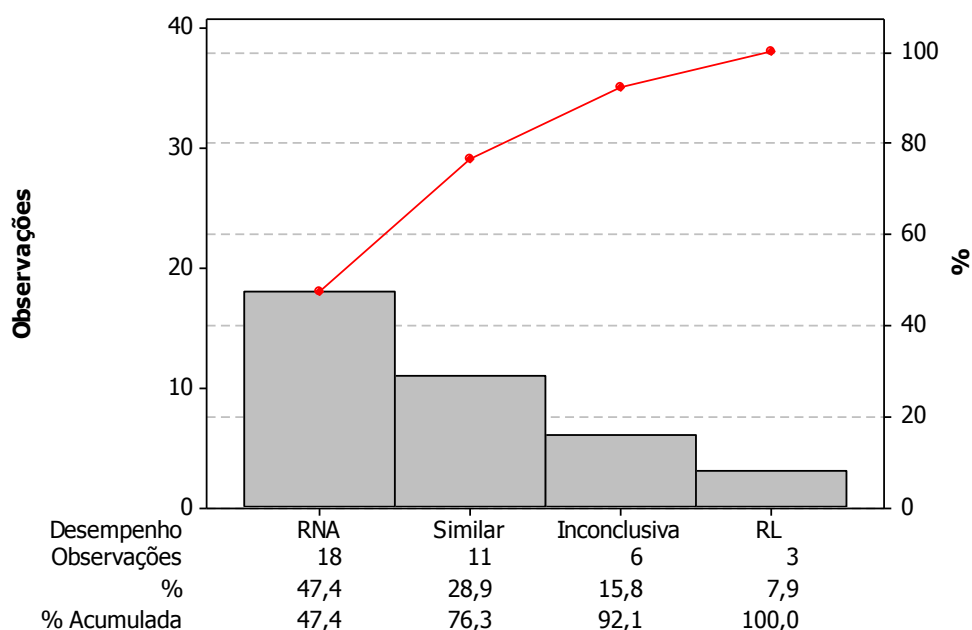


Figura 8- Gráfico de Pareto dos resultados expostos nos artigos revisados

2.5 Análise do poder discriminatório

Segundo Dreiseitl e Ohno-Machado (2002), o poder discriminatório é a análise utilizada para comparar a performance de classificadores. A discriminação é uma medida de quão bem as duas classes do conjunto de dados são separadas em consequência das características de entrada.

De acordo com Ben-David (2008), a acurácia é a primeira medida para se avaliar o desempenho de um classificador. Entretanto, é uma medida problemática porque não compensa os sucessos que ocorrem devido ao acaso. Por exemplo, em um problema de classificação binária em que as probabilidade de sucesso e fracasso sejam as mesmas, um classificador que acerta metade das classificações em metade dos casos tem o mesmo desempenho que aquele que um classificador aleatório (por exemplo que julgue sempre todas as saídas como sucesso). Ou seja, ambos os classificadores não tem utilidade. O autor atenta para o fato que existem alternativas de medidas que consideram a aleatoriedade e uma das destacadas é a curva da característica de operação do receptor (*receiver operating characteristic* ou ROC). Na grande maioria dos trabalhos revisados, a área sob a curva ROC foi a métrica utilizada para avaliar o poder discriminatório dos classificadores.

2.5.1 Curvas ROC (*Receiver Operating Characteristic*)

Segundo Fawcett (2006), Hand e Till (2001), Søreide (2009) um gráfico ROC é uma técnica para visualizar, organizar e selecionar classificadores supervisionados baseada em suas performances. Foi utilizado na teoria da detecção de sinais e visualização e análise do comportamento de sistemas de diagnósticos. Existe uma vasta literatura, nos mais diferentes campos, sobre o uso dos gráficos ROC. Conceitualmente, são simples mas há complexidades que surgem quando aplicados em pesquisa.

Fawcett (2006) afirma que dado um classificador e um exemplo, há quatro possíveis saídas. Se o exemplo é positivo e é classificado como positivo, tem-se um positivo verdadeiro; se é classificado como negativo é conhecido como falso negativo. Agora, se o exemplo é negativo e é classificado como negativo, tem-se um negativo verdadeiro e se é classificado como positivo, tem-se um falso positivo. A Figura 9 apresenta uma matriz de confusão.

		Valor Verdadeiro	
		Positivos	Negativos
Valor Previsto	Positivos	VP Verdadeiro Positivo	FP Falso Positivo
	Negativos	FN Falso Negativo	VN Verdadeiro Negativo

Figura 9 - Matriz de confusão

Os gráficos ROC são gráficos bidimensionais nos quais a sensibilidade (ou a taxa de verdadeiros positivos) é plotada no eixo Y e a taxa de falsos positivos (ou 1 – especificidade) é plotada no eixo X. A Figura 10 mostra um gráfico ROC com cinco classificadores discretos nomeados de A até E. Diversos pontos num gráfico ROC são importantes. O ponto inferior esquerdo (0,0) representa a estratégia de nunca classificar um caso como positivo. Esse classificador não comete erros FP (falsos positivos) mas também não faz acertos VP (verdadeiros positivos). A estratégia oposta, de somente conceder classificações positivas é representada pelo ponto superior direito (1,1). O ponto (0,1) representa classificação perfeita. O desempenho de D é perfeito como se vê. Um ponto é melhor que outro quanto mais a noroeste ele está no gráfico (taxa de verdadeiros positivos maior e taxa de falsos positivos menor). A medida que o ponto vai descendo no gráfico ele diminui sua performance. A e B

tem performances similares e satisfatórios, a performance de C é exatamente igual a do acaso e a performance de E é pior do que aquela encontrada pelo acaso.

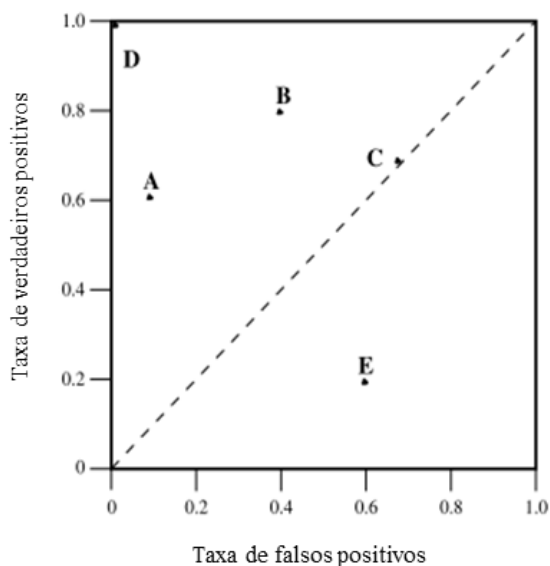


Figura 10 - Gráfico ROC básico com 5 classificadores discretos

Fonte: Fawcett (2006)

Efetivamente, para comparar classificadores, o que se faz é calcular a área abaixo da curva ROC, conhecida como AUC (*area under the curve*). Como a AUC é uma porção de uma área de 1un^2 , seu valor será sempre entre 0 e 1. Entretanto, porque adivinhações aleatórias produzem a linha diagonal entre (0,0) e (1,1), que tem uma área de 0,5, nenhum classificador realístico deve ter uma AUC menor que 0,5.

A área sob a curva de um classificador tem uma importante propriedade estatística: ela é equivalente à probabilidade do classificador ranquear uma observação positiva escolhida ao acaso com um valor superior à uma observação negativa escolhida ao acaso.

A Figura 11 apresenta as duas curvas ROC, A e B, e as áreas sob elas. O classificador B tem uma área sob a curva maior e, portanto, uma performance média superior, apesar de ser pior do que A em determinadas regiões.

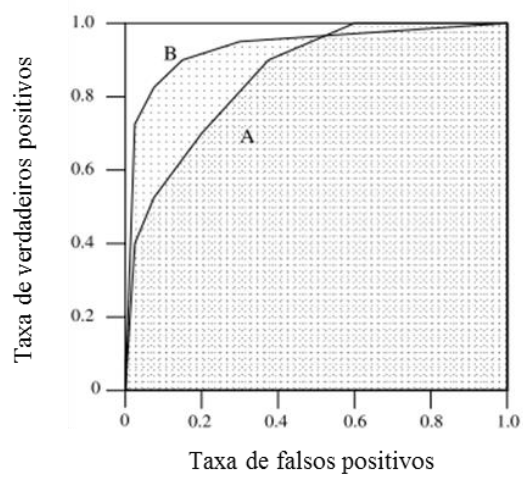


Figura 11 - Área sob a curva de 2 classificadores probabilísticos

Fonte: Fawcett (2006)

3. DESENVOLVIMENTO DA PESQUISA

3.1 Considerações iniciais

Os métodos de classificação de Regressão Logística e de Redes Neurais Artificiais serão aplicados em oito diferentes casos. Todos estes problemas foram abordados no clássico livro de Hosmer, Lemeshow e Sturdivant (2013) e os conjuntos de dados completos para cada problema podem ser encontrados no *site* da Universidade do Massachusetts. Uma descrição sucinta de cada caso analisado é apresentado na Tabela 3.

Tabela 3 - Descrição dos casos abordados no livro

Caso	Descrição
Caso 1	Estudo do baixo peso ao nascimento, definido como peso ao nascer inferior a 2500 gramas. A taxa de mortalidade infantil é maior para bebês nessa situação.
Caso 2	Estudo da sobrevivência de pacientes sujeitos a acompanhamento na unidade de tratamento intensivo (UTI).
Caso 3	Estudo de osteoporose em mulheres com mais de 55 anos. Os objetivos são elucidar os riscos de fratura no primeiro ano de acompanhamento.
Caso 4	Estudo sobre os fatores determinantes dos locais em que adolescentes hospitalizados com problemas psiquiátricos receberão cuidados após a liberação.
Caso 5	Estudo sobre hospitalizações relacionadas a queimaduras. A resposta de interesse é a sobrevivência após saída do hospital.
Caso 6	A miopia é um distúrbio visual em que um indivíduo tem dificuldade de enxergar a distância. Pessoas que tiveram 5 anos de acompanhamento e não eram míopes no início foram analisadas.
Caso 7	O estudo foi elaborado para fornecer dados sobre a saúde e a nutrição de adultos e crianças nos Estados Unidos. A variável de interesse é o índice de massa corporal (IMC).
Caso 8	Nesse estudo a variável de interesse é se o paciente está sob uso de medicamentos de três ou mais classes diferentes, o interesse é identificar fatores relacionados com essa variável.

A análise dos resultados separadamente, isto é, para cada método, será explicada apenas para um dos casos. O caso que será explicado detalhadamente é o caso 1.

Tal problema é definido como o nascimento de bebês com peso abaixo de 2500 gramas e chama a atenção de especialistas há anos. As taxas de mortalidade infantil e de bebês que nascem com deficiências são muito altas quando falamos de crianças que nascem abaixo do peso descrito. O comportamento das mães durante a gravidez (dieta, hábito de fumar, recebimento de cuidados pré-natais) pode diminuir drasticamente as chances de dar à luz um bebê com peso abaixo do normal.

O objetivo deste estudo foi o de identificar os fatores de risco associados com o baixo peso ao nascimento. Foram coletados dados de 189 mulheres das quais 59 tiveram bebês com

peso abaixo do normal e 130 tiveram bebês com peso normal. Considerou-se, no estudo, as seguintes variáveis independentes: idade, peso da mãe em seu último período menstrual, raça, ocorrência de fumar durante a gravidez, histórico de partos prematuros, histórico de hipertensão, presença de irritabilidade uterina e o número de consultas médicas durante o primeiro trimestre de gravidez. A descrição das variáveis é apresentada na Tabela 4 e parte do banco de dados utilizado é apresentado na Tabela 5.

Tabela 4 - Descrição das variáveis do caso 1 (Baixo Peso ao Nascimento)

Variável	Descrição	Valor	Nome
1	Número de identificação	1-189	ID
2	Baixo peso ao nascer	0 = >= 2500 gramas 1 = < 2500 gramas	Baixo
3	Idade da mãe	Em anos	Idade
4	Peso da mãe no último período menstrual	Em libras	PesoMae
5	Raça	1 = branco 2 = negro 3 = outro	Raça
6	Fumou durante a gravidez?	0 = não 1 = sim	Cigarro
7	Histórico de partos prematuros	0 = nenhum 1 = um 2 = dois.....	PreMAT
8	Histórico de hipertensão	0 = não 1 = sim	Hiper
9	Presença de irritabilidade uterina	0 = não 1 = sim	Irrit
10	Número de consultas durante o primeiro trimestre	0 = nenhuma 1 = uma 2 = duas....l	Cons
11	Peso do bebê	Em gramas	PesoBebe

Tabela 5 - Caso 1 (Baixo Peso ao Nascimento)

ID	Baixo	Idade	PesoMae	Raça	Cigarro	Premat	Hiper	Irrit	Cons	PesoBebe
1	1	28	120	3	1	1	0	1	0	709
2	1	29	130	1	0	0	0	1	2	1021
3	1	34	187	2	1	0	1	0	0	1135
4	1	25	105	3	0	1	1	0	0	1330
5	1	25	85	3	0	0	0	1	0	1474
6	1	27	150	3	0	0	0	0	0	1588
7	1	23	97	3	0	0	0	1	1	1588
8	1	24	128	2	0	1	0	0	1	1701
9	1	24	132	3	0	0	1	0	0	1729
10	1	21	165	1	1	0	1	0	1	1790
11	1	32	105	1	1	0	0	0	0	1818
12	1	19	91	1	1	2	0	1	0	1885
13	1	25	115	3	0	0	0	0	0	1893
14	1	16	130	3	0	0	0	0	1	1899
15	1	25	92	1	1	0	0	0	0	1928
...
180	0	20	158	1	0	0	0	0	1	3997
181	0	26	160	3	0	0	0	0	0	4054
182	0	21	115	1	0	0	0	0	1	4054
183	0	22	129	1	0	0	0	0	0	4111
184	0	25	130	1	0	0	0	0	2	4153
185	0	31	120	1	0	0	0	0	2	4167
186	0	35	170	1	0	1	0	0	1	4174
187	0	19	120	1	1	0	0	0	0	4238
188	0	24	116	1	0	0	0	0	1	4593
189	0	45	123	1	0	0	0	0	1	4990

3.2 Modelagem através de regressão logística

Toda a abordagem para a aplicação do modelo de regressão logística foi baseada nos passos para a construção do modelo de regressão logística de Hosmer, Lemeshow e Sturdivant (2013) e procurou-se abordar os critérios sugeridos por Bagley *et al.* (2001), Dreiseitl e Ohno-Machado (2002), que visam justamente garantir a correta e estruturada aplicação do método para a obtenção de resultados confiáveis e que possam ser reproduzidos. É importante salientar que como os dados já foram devidamente coletados para utilização no livro, algumas condições não são passíveis de alteração. O software utilizado para o tratamento, modelagem e interpretação dos dados foi o Minitab®.

3.2.1 Seleção das variáveis

Antes de iniciar os passos estruturados sugeridos por Hosmer, Lemeshow e Sturdivant (2013), conhecidos como *Purposeful Selection of Covariates*, o conjunto de dados foi separado em 80% para a construção do modelo de regressão logística e 20% para testar esse

modelo, avaliando o poder de generalização do modelo. A seguir, o modelo pôde ser elaborado através dos sete passos descritos na sessão 2.2.2.

Passo 1: Construir um modelo de regressão logística univariada para cada variável. Os resultados dessa análise são mostrados na Tabela 6. É importante notar que cada linha apresenta os resultados para os coeficientes de regressão estimados de um modelo que contém apenas a variável em questão.

Tabela 6 - Resultados da Regressão Logística Univariada

	Coef	EP Coef	Z	P	Razão de pos.	95% IC	
Idade	-0,042	0,036	-1,17	0,240	0,96	0,89	1,03
PesoMae	-0,015	0,007	-2,29	0,022	0,99	0,97	1
Raça 2	0,863	0,549	1,57	0,116	2,37	0,81	6,95
Raça 3	0,366	0,377	0,97	0,333	1,44	0,69	3,02
Cigarro 1	0,942	0,357	2,64	0,008	2,57	1,28	5,16
Premat	1,024	0,407	2,51	0,012	2,78	1,25	6,18
Hiper 1	0,996	0,633	1,57	0,115	2,71	0,78	9,35
Irrit 1	0,987	0,461	2,14	0,032	2,68	1,09	6,62
Cons	-0,089	0,169	-0,53	0,597	0,91	0,66	1,27

Passo 2: Constrói-se o primeiro modelo multivariado que contenha todas as variáveis que sejam significantes em análises univariadas em níveis de até 25%. Para a construção do modelo, excluiu-se a variável “Cons” que apresentou um $p\text{-value} = 0,597$. Os resultados são mostrados na Tabela 7. Observa-se pelos resultados que a variável “Idade” obteve um $p\text{-value}$ muito superior a 0,05 (0,678) sendo candidata a se retirar do modelo. Como os resultados não se alteraram significativamente, prosseguiu-se com a retirada da variável.

Tabela 7 - Resultados da Regressão Logística Multivariada

	Coef	EP Coef	Z	P	Razão de pos.	95% IC	
Constante	0,452	1,365	0,33	0,741			
Idade	-0,017	0,042	-0,41	0,678	0,98	0,91	1,07
PesoMae	-0,016	0,007	-2,21	0,027	0,98	0,97	1
Raça 2	1,240	0,638	1,94	0,052	3,45	0,99	12,05
Raça 3	0,568	0,496	1,14	0,253	1,76	0,67	4,67
Cigarro 1	1,124	0,453	2,48	0,013	3,08	1,27	7,47
Premat	0,797	0,430	1,85	0,064	2,22	0,96	5,15
Hiper 1	1,655	0,721	2,29	0,022	5,24	1,27	21,53
Irrit 1	0,935	0,505	1,85	0,064	2,55	0,95	6,86

Passo 3: Checa-se se a variável excluída no passo 2 é necessária por alterar os coeficientes das variáveis que permaneceram no modelo. No caso em questão, a variável não alterou significativamente nenhum dos coeficientes e deve ser mantida fora do modelo. Após a retirada da variável “Idade” ainda existem outras variáveis como “Premat” e “Irrit” candidatas a serem excluídas mas constatou-se que sua retirada prejudicaria a qualidade do ajuste e optou-se por mantê-las. A Tabela 8, que nesse caso foi a final, apresenta as alterações supracitadas. O modelo *logit* final é apresentado na equação 6.

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = 0,093 - 0,017 * (PesoMae) + 1,282 * (Raça2) + 0,594 * (Raça3) + 1,135 * (Cigarro1) + 0,769 * (Premat) + 1,671 * (Hiper1) + 0,940 * (Irrit1) \quad (6)$$

Para um exemplo em que o $PesoMae = 182$; $Raça = 2$; $Cigarro = 0$; $Premat = 0$; $Hiper = 0$; $Irrit = 1$, por exemplo, tem-se os seguintes valores para o *logit* (equação 7) e para a probabilidade (equação 8):

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = -0,75758 \quad (7)$$

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} = 0,3192 \quad (8)$$

Tabela 8 - Resultado Final da Regressão Logística Multivariada

	Coef	EP Coef	Z	P	Razão de pos.	95% IC	
Constante	0,093	1,055	0,09	0,93			
PesoMae	-0,017	0,007	-2,29	0,022	0,98	0,97	1
Raça 2	1,282	0,633	2,03	0,043	3,61	1,04	12,46
Raça 3	0,594	0,490	1,21	0,225	1,81	0,69	4,73
Cigarro 1	1,135	0,451	2,52	0,012	3,11	1,28	7,53
Premat	0,769	0,424	1,81	0,07	2,16	0,94	4,95
Hiper 1	1,671	0,723	2,31	0,021	5,32	1,29	21,93
Irrit 1	0,940	0,504	1,87	0,062	2,56	0,95	6,87

Passo 4: A variável “Cons” excluída no Passo 1 é reintegrada ao modelo e seu coeficiente é verificado, como não se tornou significativo ela permanece excluída do modelo.

Foi analisada a multicolinearidade das variáveis preditoras mas não houve correlação significativa.

Passo 5: Nesse ponto, já se tem o modelo preliminar dos efeitos principais e é necessário verificar a escala do *logit* para a variável contínua *PesoMae*. Existem diferentes formas de verificar a linearidade no *logit* para variáveis quantitativas, uma das mais simples consiste na formação de classes utilizando, por exemplo, os quartis como pontos de corte. Analisou-se, para fins exemplificativos, a variável contínua peso da mãe. O primeiro, segundo e terceiro quartil para essa variável foram os valores 110, 121 e 140 respectivamente. Seguindo os demais passos explicitados na seção 2.2.1 referente ao item, pôde-se constatar aceitável linearidade da variável analisada com relação à saída. Caso a variável contínua não seja linear, uma forma paramétrica da escala da variável deve ser feita para a posterior utilização no modelo.

Passo 6: O próximo passo no procedimento de seleção é o de explorar possíveis interações entre os efeitos principais. A decisão por adotar interações compreende além de significância estatística, o conhecimento prévio do pesquisador sobre a possibilidade das interações serem coerentes no âmbito médico, o que inviabiliza a adoção desse passo.

Passo 7: Com o modelo preliminar final testes da qualidade do ajuste do modelo são realizados e analisados e serão discutidos na sessão referente à interpretação dos resultados encontrados no modelo.

3.2.2 Resultados da regressão logística

De posse do modelo final, diversos resultados são gerados pelo software. Eles são apresentados e interpretados em cada um dos quadros a seguir. O Quadro 6 apresenta informações sobre a função de ligação e sobre a resposta. O Quadro 7 traz a tabela de regressão logística. O Quadro 8 mostra os testes de aderência realizados. O Quadro 9 expõe a tabela de frequências esperadas e obtidas. Finalmente, o Quadro 10 apresenta as medidas de associação entre as variáveis de resposta e as probabilidades previstas.

Quadro 6 - Função de Ligação e Informações da Resposta (Caso 1)

Resultado do Minitab	Link Function: <i>Logit</i>																
	Response Information																
	<table border="1"> <thead> <tr> <th>Variable</th> <th>Value</th> <th>Count</th> <th></th> </tr> </thead> <tbody> <tr> <td>BAIXO</td> <td>1</td> <td>49</td> <td>(Event)</td> </tr> <tr> <td></td> <td>0</td> <td>102</td> <td></td> </tr> <tr> <td>Total</td> <td></td> <td>151</td> <td></td> </tr> </tbody> </table>	Variable	Value	Count		BAIXO	1	49	(Event)		0	102		Total		151	
Variable	Value	Count															
BAIXO	1	49	(Event)														
	0	102															
Total		151															
Interpretação	<p>Função de ligação: A função <i>logit</i> foi a escolhida entre as concorrentes (<i>Probit</i> e <i>Gompit</i>).</p> <p>Informações da resposta: O número de observações para as duas categorias da resposta foi de 49 para baixo peso ao nascimento e 102 para peso normal (somando 151, que corresponde a 80% dos casos disponíveis). O evento de referência foi “baixo peso ao nascimento”.</p>																

Quadro 7 - Tabela de Regressão Logística (Caso 1)

Resultado do Minitab	Logistic Regression Table								
	Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI		
	Constant	0,0932704	1,05542	0,09	0,930				
	PesoMae_1	-0,0168861	0,0073693	-2,29	0,022	0,98	0,97	1,00	
	Raça_1								
	2	1,28238	0,632743	2,03	0,043	3,61	1,04	12,46	
	3	0,594282	0,490135	1,21	0,225	1,81	0,69	4,73	
	Cigarro_1								
	1	1,13484	0,451134	2,52	0,012	3,11	1,28	7,53	
	Premat_1	0,769161	0,423842	1,81	0,070	2,16	0,94	4,95	
	Hiper_1								
	1	1,67092	0,722912	2,31	0,021	5,32	1,29	21,93	
	Irrit_1								
	1	0,940044	0,503812	1,87	0,062	2,56	0,95	6,87	
	Log-Likelihood = -79,892								
	Test that all slopes are zero: G = 30,541, DF = 7, P-value = 0,000								
Interpretação	<p>Tabela de Regressão logística: Os coeficientes estimados, erro padrão dos coeficientes, Z-values, p-values são representados nessa tabela. Utilizando o <i>logit</i>, a razão de possibilidades e um intervalo de confiança de 95% para a razão de possibilidades são computados. Como esse é o modelo final, os p-values dos coeficientes são todos baixos indicando sua significância no modelo. A razão de possibilidades de 3,11 para o CIGARRO quer dizer que uma mãe fumante tem uma chance 3,11 vezes maior de ter um bebê com baixo peso ao nascimento do que uma mãe não fumante. Esse valor está intimamente ligado com o valor do coeficiente do parâmetro através de uma exponencial. O valor da estatística G testa a hipótese nula de que todos os coeficientes associados com os preditores são iguais a zero versus a hipótese de que os coeficientes não são iguais a zero. No exemplo, G = 30,541, com p-value de 0,000, indicando que há evidência suficiente que ao menos um dos coeficientes é diferente de zero.</p>								

Quadro 8 - Testes de aderência (Caso 1)

Resultado do Minitab	Goodness-of-Fit Tests			
	Method	Chi-Square	DF	P
	Pearson	118,627	121	0,544
	Deviance	133,785	121	0,201
	Hosmer-Lemeshow	6,341	8	0,609
Interpretação	<p>Testes de aderência: Os p-values de todos os testes de aderência foram superiores a 0,05, assim não é possível rejeitar a hipótese nula de que o modelo é adequado.</p>			

Quadro 9 - Tabela das frequências observadas e esperadas (Caso 1)

Resultado do Minitab	Table of Observed and Expected Frequencies: (See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)											
	Value	1	2	3	4	5	Group				Total	
	1						6	7	8	9	10	
	Obs	0	2	3	2	5	4	5	10	7	11	49
	Exp	1,0	1,6	2,3	3,1	3,8	4,5	6,0	6,9	8,9	11,0	
	0											
Obs	15	13	12	13	10	11	11	5	8	4	102	
Exp	14,0	13,4	12,7	11,9	11,2	10,5	10,0	8,1	6,1	4,0		
Total	15	15	15	15	15	15	16	15	15	15	151	
Interpretação	Tabela das frequências observadas e esperadas: Essa tabela permite identificar quão bem o modelo se adequa aos dados ao comparar as frequências observadas e esperadas. Não há evidência suficiente de que o modelo não se adequa aos dados pelos dados observados.											

Quadro 10 - Medidas de Associação (Caso 1)

Resultado do Minitab	Measures of Association: (Between the Response Variable and Predicted Probabilities)			
	Pairs	Number	Percent	Summary Measures
	Concordant	3810	76,2	Somers' D 0,53
	Discordant	1165	23,3	Goodman-Kruskal Gamma 0,53
	Ties	23	0,5	Kendall's Tau-a 0,23
Total	4998	100,0		
Interpretação	Medidas de Associação (entre a Variável de Resposta e as Probabilidades Previstas): Essas estatísticas apresentam o número e a porcentagem de pares concordantes, discordantes e empatados. Esses valores medem a associação entre respostas observadas e probabilidades previstas. A tabela é calculada ao se comparar observações com respostas diferentes. No nosso exemplo há 49 casos de crianças com baixo peso e 102 com crianças com peso normal, resultando em 49*102=4998 pares com valores diferentes de resposta. Baseado no modelo, um par é concordante se uma criança com baixo peso ao nascimento tem uma probabilidade, fornecida pelo modelo, maior de ter baixo peso ao nascimento do que aquela que não tem baixo peso ao nascimento. Um par é discordante se o oposto for verdadeiro e empatado se as probabilidades forem iguais. Aqui, 76,2% dos pares são concordantes e 23,3% são discordantes. As medidas Somer' D, Goodman-Kruskal Gamma e Kendall's Tau-a se referem a capacidade preditiva do modelo, que no caso foi regular.			

Para ilustrar uma matriz de confusão gerada pelo modelo, optou-se por um limiar de 0,5, ou seja, caso a probabilidade estimada pelo modelo de Regressão Logística fosse maior que 0,5 a classificação do modelo forneceria uma resposta de valor igual a 1, ou seja, bebê com baixo peso ao nascimento. Caso contrário, a resposta seria igual a zero, ou seja, bebê com peso normal. A Tabela 9 ilustra a matriz de confusão para o caso analisado.

Tabela 9 - Matriz de Confusão do caso 1

		Real		
		0	1	Total
Classificado pela RL	0	115	33	148
	1	15	26	41
Total		130	59	189

Já a curva ROC (*receiver operating characteristic*), que plota a sensibilidade e especificidade para diferentes limiares (não apenas o de 0,5), é demonstrada na Figura 12. A plotagem da mesma para o caso 1 foi realizada com auxílio do software MedCalc®. A AUC (*area under the curve*) foi de 0,7473 que, segundo Hosmer, Lemeshow e Sturdivant (2013), é considerada como sendo uma discriminação aceitável do modelo.

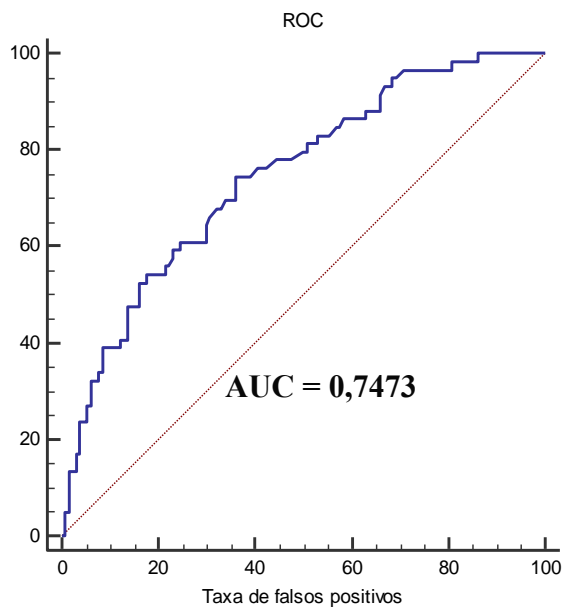


Figura 12 - Curva ROC para o caso 1

Os mesmos procedimentos foram realizados para os demais conjuntos de dados e obtiveram-se os valores apresentados na Tabela 10 com relação às áreas sob as curvas ROC. Todos os modelos de regressão logística tiveram um desempenho satisfatório, com valores superiores a 0,7. Alguns classificadores foram excelentes como o do estudo de queimaduras com uma área de 0,9654.

Tabela 10 - Área sob a curva ROC dos 8 casos modelados por RL

Caso	AUC
Caso 1	0,747
Caso 2	0,845
Caso 3	0,720
Caso 4	0,911
Caso 5	0,965
Caso 6	0,888
Caso 7	0,710
Caso 8	0,739

3.3 Modelagem através de Redes Neurais treinadas por DOE

3.3.1 Planejamento pré-experimental

Nessa seção, a definição do problema, a seleção da variável de resposta e a escolha dos fatores e níveis serão discutidas. É importante salientar que para o treinamento de redes neurais, o método da tentativa e erro é o mais utilizado e apresenta desvantagens como o consumo elevado de tempo. A proposta nessa seção é a de apresentar uma metodologia bem estruturada para estimar os parâmetros da rede neural.

Para avaliar a eficiência da rede neural com relação à performance de sua classificação, a medida utilizada foi a área sob a curva ROC, indicador já bem detalhado na seção 2.5.1.

Diversos fatores podem ser considerados no treinamento das redes neurais artificiais. O Quadro 11 apresenta os fatores considerados em problemas de classificação utilizando o software Statistica. Os detalhes sobre os fatores são descritos na sequência. Para escolher os fatores e níveis, além das indicações do próprio software, foram utilizados como sugestões os procedimentos adotados nos trabalhos de Balestrassi *et al.* (2009) e Pontes (2011). Apesar dos esforços, pouca evidência foi encontrada para auxiliar na escolha dos níveis dos parâmetros, sendo que um dos objetivos da análise é servir justamente como um estudo exploratório para definir fatores e níveis significativos de problemas de classificação em RNA para os casos avaliados.

Quadro 11 - Parâmetros de treinamento da Rede Neural para Problemas de Classificação

FATOR	SÍMBOLO	NÍVEIS	NÚMERO DE NÍVEIS
Arquitetura da rede	---	MLP, RBF, SVM, ...	1
Número de camadas	NC	0,1,2,3	2
Função de erro	FE	Soma de Quadrados, Entropia Cruzada	2
Número de Neurônios	NN	<i>k x N ou k(N+1), onde N é o número de entradas e k = 1; 2</i>	2
Função de Ativação da Camada de Entrada	---	Identidade , Tangente Hiperbólica, Logística, Exponencial	1
Função de Ativação das Unidades da Camada Oculta	CO	Identidade, Tangente Hiperbólica, Logística , Exponencial	2
Função de Ativação das Unidades da Camada de Saída	CS	Identidade , Tangente Hiperbólica, Logística, Exponencial	2
Algoritmo de Treinamento – Fase 1	F1	<i>Backpropagation, Quickpropagation, Delta-bar-delta</i>	2
Algoritmo de Treinamento – Fase 2	F2	Gradiente Conjugado, Quasi-Newton (BFGS) , Levenberg-Marquardt	2
Épocas – Fase1	E1	50, 100	2
Épocas – Fase 2	E2	50, 100	2
Taxa de Aprendizagem	TA	0,01; 0,1	2
Método de Inicialização	MI	Unif(0,1), N(0,1)	2
Aplicação de Condições de Parada	CP	0; 0,1	2
Melhoria Mínima no Erro para Treinamento/Seleção	ET	0; -0,1	2
Melhoria Mínima no erro para épocas	EE	0; 10	2
Poda de Neurônio	PN	Sem poda, Poda de pesos inferiores a 0,05	2
Poda de Entrada	PE	Sem poda, poda de entradas com baixa influência após treinamento (taxa = 1)	2
Limiar de Classificação	---	Limiares especificados manualmente, Limiares com menor perda	1
Regularização de decaimento de peso – Fase 1	D1	Sem regularização; 0,01	2
Regularização de decaimento de peso – Fase 2	D2	Sem regularização; 0,01	2
Ajuste do <i>Backpropagation</i>	---	Sem ajustar parâmetros; Com ajuste de parâmetros	1
Ajuste do <i>Quickpropagation</i>	---	Sem ajustar parâmetros; Com ajuste de parâmetros	1
Método de Amostragem	---	Aleatório , validação cruzada, <i>bootstrap</i>	1

Arquitetura da rede: RNAs são algoritmos de modelagem não lineares. Exemplos de RNAs para problemas de classificação podem ser os Perceptron Multicamadas (em inglês, *Multilayer Perceptrons* ou *MLP*), Funções de Bases Radiais (em inglês, *Radial Basis Function* ou *RBF*), Máquinas de Vetores de Suporte (em inglês, *Support Vector Machines* ou *SVM*), dentre outros. MLPs são as redes mais populares e responsáveis, segundo Zhang (2000) e Zhang (2007), por cerca de 95% das aplicações de redes neurais encontradas na literatura, além de possuírem a melhor performance em uma grande variedade de aplicações, sendo assim escolhidas como a arquitetura padrão para as análises.

Número de camadas ocultas: Uma camada oculta é um grupo de neurônios que tem uma função específica e são processados como um todo. Normalmente, MLPs com uma única camada são capazes de aproximar qualquer função contínua.

- a) 1 camada oculta
- b) 2 camadas ocultas

Função de erro: Especifica a função de erro a ser utilizada no treinamento da rede neural.

- a) Entropia: Se a função de erro de entropia cruzada é escolhida, a rede realiza otimização de máxima verossimilhança, assumindo que os dados são retirados de distribuições exponenciais.
- b) Soma dos Quadrados: A abordagem alternativa é a utilização da soma dos quadrados.

Número de neurônios por camada (NN): Especifica o número de neurônios na camada oculta da rede. Quanto maior o número de neurônios, mais complexa e flexível a rede se torna. No trabalho, optou-se por usar uma quantidade que varia entre o número de entradas e o dobro delas ($k(N+1)$, em que N é o número de entradas e $k = 1; 2$).

- a) 1 (N+1) Neurônios
- b) 2 (N+1) Neurônios

Função de Ativação: A escolha da função de ativação, isto é, a função matemática específica, é crucial para construir um modelo de rede neural uma vez que está diretamente ligada com a performance do modelo. Deve ser escolhida tanto para as unidades da camada oculta quanto para as unidades da camada de saída. Via de regra, os neurônios de entrada não têm função de ativação, ou seja, eles usam uma função identidade em que os sinais de entrada não são transformados.

Função de Ativação para as Unidades da Camada Oculta (CO):

- a) Tangente hiperbólica. A função tangente hiperbólica é uma função simétrica em formato S (sigmoide) cujas saídas se encontram entre -1 e +1.
- b) Logística. Usa a função sigmoide logística, com a saída variando de 0 a 1.

Função de Ativação para as Unidades da Camada de Saída (CS):

- a) Identidade. Com essa função, o nível de ativação é passado diretamente para a saída.
- b) Logística. Usa a função sigmoide logística, com a saída variando de 0 a 1.

Algoritmo de Treinamento – Fase 1 (F1): Esse fator está relacionada com o algoritmo de treinamento para um MLP em um estágio inicial e assume dois níveis.

- a) Retropropagação ou *Backpropagation*. Um algoritmo simples com grande número de parâmetros de ajuste. Frequentemente, tem lenta convergência terminal mas boa convergência inicial.
- b) Propagação Rápida ou *Quickpropagation*. Algoritmo mais antigo com performance comparável com o *Backpropagation* na maioria dos casos, embora pareça consideravelmente melhor em algumas aplicações.

Algoritmo de Treinamento – Fase 2 (F2): Esse fator está relacionada com o algoritmo de treinamento para um MLP em um segundo estágio e assume dois níveis.

- a) Gradiente Conjugado. O algoritmo procede através de uma série de linhas de busca no espaço de erro, sendo um bom algoritmo genérico com rápida convergência.
- b) Quasi-Newton. Poderoso algoritmo de treinamento de segunda ordem com rápida convergência mas alta requisição de memória devido à armazenagem da matriz hessiana.

Épocas: Uma época é a apresentação de todo o conjunto de treinamento para a rede neural em uma dada fase. Ao aumentar esse número, o valor da acurácia do modelo também aumentará (porém com maior tempo). É relacionada com o tamanho da amostra.

Épocas – Fase 1 (E1)

- a) 50 épocas de treinamento
- b) 100 épocas de treinamento

Épocas – Fase 2 (E2)

- a) 50 épocas de treinamento
- b) 100 épocas de treinamento

Taxa de aprendizagem (TA): Um valor entre 0 e 1 que representa uma variável de ajuste para o treinamento dos algoritmos de *Backpropagation*, *Quickpropagation* e *Delta-bar-delta*. Baixas taxas de aprendizagem requerem maiores interações de treinamento. Uma maior taxa permite à conversão mais rápida da rede, porém as chances de uma solução não-ótima aumentam.

- a) Taxa de aprendizagem igual a 0,01.
- b) Taxa de aprendizagem igual a 0,1.

Método de inicialização (MI): Esse fator define como os pesos devem ser definidos no começo do treinamento e assume dois níveis:

- c) Uniforme. Os pesos são definidos através de um valor aleatório de uma distribuição uniforme, dentro de um intervalo em que os valores mínimos e máximos são dados. No caso os valores mínimos e máximos são 0 e 1, respectivamente.
- d) Normal. Os pesos são definidos através de um valor aleatório de uma distribuição normal, dentro de um intervalo cuja média e o desvio-padrão são dados. No caso, $N(0,1)$ foi adotada.

Aplicação de Condições de Parada (CP): Se o erro no treinamento ou teste de seleção fica abaixo dos valores alvo estimados, considera-se que a rede treinou suficientemente bem e o treinamento é finalizado.

- a) 0,0.
- b) 0,1.

Melhoria mínima no erro para treinamento/seleção: Esse fator representa a melhoria mínima no erro que deve ser feita. Se a taxa de melhoria cai abaixo desse nível o treinamento é encerrado. O valor 0 implica que o treinamento terminará se o erro se deteriorar. Pode-se especificar uma taxa negativa, o que é equivalente a especificar uma taxa máxima de deterioração que será tolerada. A melhoria é medida sobre um número de épocas, conhecida como “janela” (ver abaixo).

- a) 0,0.
- b) -0,1.

Melhoria mínima no erro para número de épocas: Especifica o número de épocas sobre as quais a melhoria é medida. Alguns algoritmos, incluindo *Backpropagation*, possuem ruído nos erros de treinamento e seleção e todos os algoritmos podem demonstrar ruído no erro de seleção. As “janelas” especificam um número de épocas nas quais as taxas de erro são monitoradas e melhoradas. O treinamento só é interrompido se o erro não melhorar em muitas épocas.

- a) 0.
- b) 10.

Poda de neurônios (PN). Um neurônio com peso de conexão de saída de baixa magnitude fornece baixa contribuição para as ativações da próxima camada e pode ser podado, levando a uma rede compacta e rápida com performance equivalente.

- a) Sem poda.
- b) Poda de pesos inferiores a 0,05.

Poda de entrada com baixa sensibilidade depois do treinamento (PE). Aplica análise de sensibilidade baseada em podas depois do treinamento. Uma análise de sensibilidade é efetuada após o treinamento da rede e variáveis de entrada com relações de sensibilidade de treinamento e seleção abaixo de um limiar são podadas.

- a) Sem poda.
- b) Poda de entradas com baixa influência após treinamento (relação=1).

Limiares de classificação. No caso de apenas um neurônio de saída, há dois níveis para serem escolhidos. Pode-se escolher manualmente os limiares ou através de uma forma automática em que um coeficiente de perda é escolhido. Um coeficiente de perda de 1 indica que as duas classes são igualmente importantes, situação escolhida para o estudo.

Regularização do Decaimento dos Pesos – Fase 1 (D1). Refere-se ao uso da regularização do decaimento dos pesos na fase 1 em que foram considerados os algoritmos *Backpropagation*, *Quickpropagation* e *Delta-bar-delta*. Serve para encorajar o desenvolvimento de pesos menores, o que tende a reduzir o problema de sobreajuste ou *overfitting*, assim aumentando o poder de generalização da rede.

- a) Sem regularização
- b) 0,01

Regularização do Decaimento dos Pesos – Fase 2 (D2). Refere-se ao uso da regularização do decaimento dos pesos na fase 2 em que foram considerados os algoritmos Gradiente Conjugado, Quasi-Newton e Levenberg-Marquardt. Serve para encorajar o desenvolvimento de pesos menores, o que tende a reduzir o problema de sobreajuste ou *overfitting*, assim aumentando o poder de generalização da rede.

- a) Sem regularização
- b) 0,01

Ajuste do Backpropagation (AB). Refere-se às opções adicionais para o treinamento do Backpropagation. Primeiramente, é possível especificar a taxa de aprendizagem e o momentum iniciais e finais utilizados. Uma taxa mais alta pode convergir mais rapidamente, mas pode exibir maior instabilidade. O momentum é usado para compensar uma convergência lenta. Em segundo lugar, é possível embaralhar a ordem dos casos para cada época de treinamento. E, por último, é possível adicionar ruído gaussiano, para reduzir a tendência de sobreajuste.

Ajuste do Quickpropagation (AQ). Refere-se às opções adicionais para o treinamento do Quickpropagation. Primeiramente, é possível especificar a taxa de aprendizagem e a aceleração utilizadas. A taxa de aprendizagem é especificada inicialmente, aplicada a primeira época. A aceleração especifica a máxima taxa de crescimento geométrico na mudança dos pesos que é permitida. E, por último, é possível adicionar ruído gaussiano, para reduzir a tendência de sobreajuste.

Método de amostragem (MA). Especifica o método de amostragem (aleatório, validação cruzada ou *bootstrap*). A título de simplificação, adotou-se a amostragem aleatória em que cada variável é colocada em um subconjunto (treinamento, seleção e teste), optou-se ainda por usar a proporção 8:1:1.

3.3.2 Arranjo Plackett-Burman

Considerando os fatores e níveis apresentados no Quadro 11, é interessante observar a complexidade de se experimentar todas as possibilidades de combinação, juntamente com a simulação das redes. Se considerássemos apenas dois níveis para cada fator, seriam necessários $2^{24} = 16.777.216$ experimentos, o que inviabilizaria seu desenvolvimento.

Neste trabalho, o arranjo de Plackett-Burman foi utilizado devido a sua flexibilidade em tratar um elevado número de parâmetros com poucos experimentos. Contudo, devido a esse baixo número de experimentos, a interação entre fatores é negligenciada. O objetivo é justamente verificar o desempenho do DOE mais simples na parametrização da rede e em seus resultados. Dos 24 fatores apresentados, 6 deles foram definidos a priori e excluídos do arranjo experimental (Arquitetura da Rede, Função de Ativação da Camada de Entrada, Limiar de Classificação, Ajuste do *Backpropagation*, Ajuste do *Quickpropagation* e o Método de Amostragem), como demonstrado na definição de cada parâmetro na seção 3.3.1. Com relação aos parâmetros restantes, existem 18 fatores que serão tratados em dois níveis. O

arranjo Plackett-Burman com 20 rodadas foi escolhido e optou-se por utilizar uma repetição para cálculo da média e da variação. Para cada rodada, as áreas sob a curva ROC foram registrada. A tabela com o arranjo experimental para o caso 1 (discutido na seção 3.1) bem como os resultados da AUC podem ser vistos na Tabela 11.

Tabela 11 - Arranjo de Plackett-Burman para o caso 1

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC1	AUC2
2	En	18	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,701	0,704
2	SQ	9	TH	Log	BP	CG	50	50	0,1	U	0,1	-0	10	S	S	S	N	0,782	0,816
1	SQ	18	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,694	0,702
1	En	18	TH	Id	QP	QN	50	50	0	U	0,1	-0	10	N	S	S	S	0,712	0,664
2	En	9	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,720	0,680
2	SQ	9	Log	Log	QP	CG	100	100	0	U	0	-0	10	N	S	N	S	0,743	0,734
2	SQ	18	Log	Id	QP	QN	50	100	0,1	U	0	-0	0	S	N	S	N	0,735	0,716
2	SQ	18	TH	Id	BP	QN	100	50	0,1	N	0	-0	0	N	S	N	S	0,505	0,572
1	SQ	18	TH	Log	BP	CG	100	100	0	N	0,1	-0	0	N	N	S	N	0,788	0,796
2	En	18	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,725	0,711
1	SQ	9	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,720	0,754
2	En	18	Log	Log	QP	QN	100	50	0	N	0,1	-0	10	S	N	N	N	0,662	0,701
1	SQ	9	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,677	0,654
1	En	18	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,734	0,721
1	En	9	TH	Id	QP	CG	100	100	0,1	N	0	-0	10	S	N	S	S	0,580	0,601
1	En	9	Log	Log	BP	QN	50	100	0,1	N	0,1	-0	0	S	S	N	S	0,719	0,730
2	En	9	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,705	0,728
2	SQ	9	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,683	0,711
1	SQ	18	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,756	0,727
1	En	9	Log	Id	BP	CG	50	50	0	U	0	-0	0	N	N	N	N	0,677	0,680

De posse desses resultados, calculou-se a média das áreas para cada uma das vinte observações, bem como o valor absoluto da diferença, ou amplitude da diferença, entre AUC1 e AUC2. Os resultados desses cálculos são apresentados na Tabela 12. Esses dois valores foram calculados porque além de uma posterior maximização nos valores da média de AUC, espera-se uma maior robustez da rede, que pode ser traduzida como uma minimização da amplitude da diferença encontrada.

Tabela 12 - Média e amplitude da diferença para cada rodada experimental

Observação	1	2	3	4	5	6	7	8	9	10
Média	0,702	0,799	0,698	0,688	0,700	0,738	0,725	0,577	0,792	0,718
Diferença	0,003	0,034	0,008	0,047	0,040	0,009	0,019	0,010	0,008	0,015
Observação	11	12	13	14	15	16	17	18	19	20
Média	0,737	0,681	0,665	0,727	0,590	0,724	0,716	0,697	0,742	0,678
Diferença	0,034	0,039	0,023	0,014	0,021	0,011	0,023	0,028	0,029	0,002

A Figura 13 apresenta os gráficos dos efeitos principais tanto para a média da área sob a curva quanto para a diferença (em valor absoluto) das áreas sob a curva ROC. Para o gráfico da esquerda, quanto maior a inclinação das retas, mais o fator influencia na alteração do valor da área. Assim, o fator que mais influência é a função de ativação da camada de saída (CS). Fatores que não influenciam possuem a reta de maneira horizontal como número de camadas na camada oculta (NC) e número de neurônios (NN). Para o gráfico da direita, quanto maior a inclinação da reta, maior a influência na amplitude da diferença entre os valores. Os fatores com grande influência foram o algoritmo de treinamento da fase 2 (F2) e a época de treinamento da fase 2 (E2). Fatores como taxa de aprendizado (TA) demonstraram pouca influência.

Para a seleção dos parâmetros e níveis da rede neural, fez-se uso da influência dos efeitos principais nas duas saídas observadas, média da AUC e valor absoluto da diferença, observados na Figura 13. Como o objetivo foi o da escolha da maior média possível para o valor da AUC e o menor valor possível para a diferença entre os valores calculados, procedeu-se da seguinte maneira: quando as informações para a escolha do nível do parâmetro convergiam para o mesmo resultado, ou seja, quando as informações obtidas por ambos os gráficos apontassem para o mesmo nível de determinado parâmetro, esse último seria o escolhido para a rede final. É o caso do parâmetro FE (função de erro), ao observar os gráficos, constata-se que o nível SQ (Soma de Quadrados) representa tanto uma maior AUC quanto uma menor diferença, sendo assim o nível indicado para esse parâmetro. Por outro lado, quando as informações para a escolha do parâmetro eram conflitantes entre si, ou seja, quando o gráfico das médias da AUC indicassem um determinado nível para um parâmetro e o gráfico das diferenças indicassem outro nível para o mesmo parâmetro, optou-se por observar o grau de inclinação da reta (e, portanto, em qual dos resultados - média ou diferença - o parâmetro exerce maior influência). É o caso do parâmetro NC (número de camadas), ao se observar seus efeitos para a média, o nível “2 camadas” resultaria em uma maior AUC, mas com uma inclinação suave, quase imperceptível. No entanto, ao se observar seus efeitos

para a diferença, o nível “1 camada” resultaria em uma menor diferença, sendo que a inclinação da reta é mais acentuada, fazendo deste o nível escolhido. A configuração da rede obtida, retratando todos os níveis dos parâmetros, é apresentada na Tabela 13.

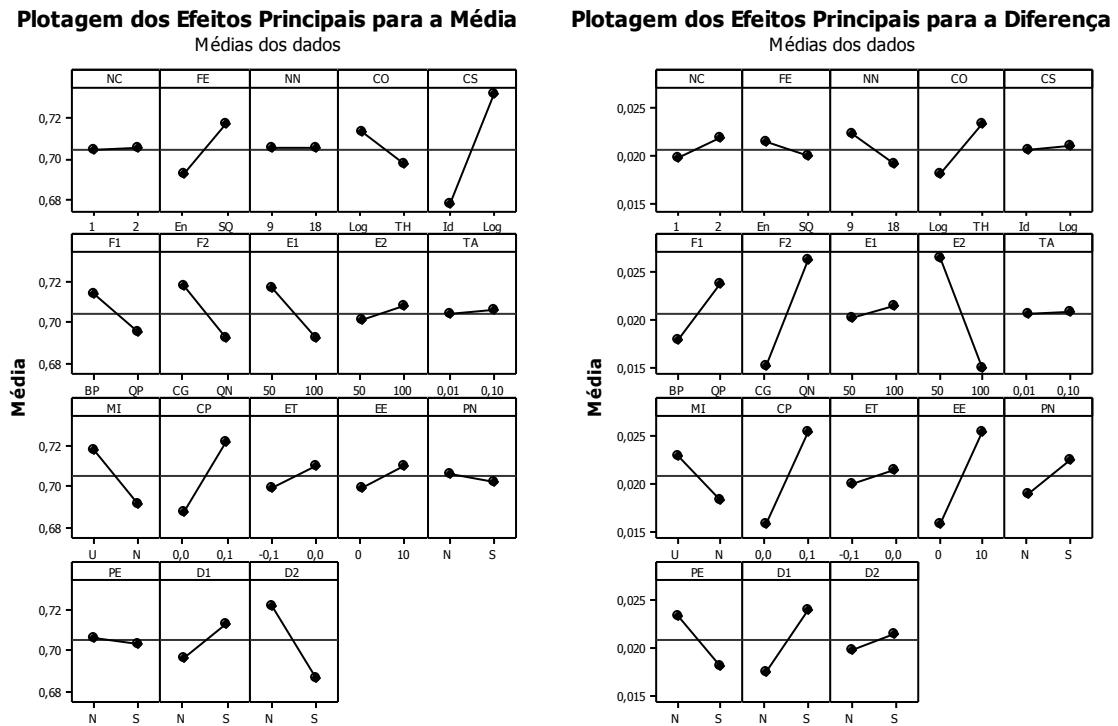


Figura 13 - Efeitos principais para média e amplitude da diferença

Tabela 13 - Rede escolhida para o caso 1

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2
1	SQ	18	Log	Log	BP	CG	50	100	0,1	U	0,1	0	0	N	S	S	N

O valor encontrado para a AUC, utilizando-se os parâmetros da rede obtida, foi de 0,7768. Ao se repetir os mesmos parâmetros para configurar novas redes, os valores encontrados foram muito similares indicando boa robustez.

Os mesmos procedimentos foram realizados para os demais conjuntos de dados e encontrou-se as configurações das redes apresentadas na Tabela 14. Além disso, de acordo com a configuração estipulada, pôde-se obter os respectivos valores das áreas sob as curvas ROC obtidas pelas redes mostrados na Tabela 15.

Tabela 14 - Configurações das redes para cada caso investigado

	NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2
Caso 1	1	SQ	18	Log	Log	BP	CG	50	100	0,1	U	0,1	0	0	N	S	S	N
Caso 2	2	SQ	20	Log	Id	BP	QN	50	100	0,01	U	0	0	0	S	S	S	N
Caso 3	1	SQ	14	Log	Log	BP	CG	50	100	0,01	U	0	0	0	S	S	S	N
Caso 4	1	SQ	26	Log	Log	BP	QN	50	100	0,01	U	0,1	0	10	N	N	S	N
Caso 5	1	SQ	7	Log	Log	BP	QN	50	50	0,01	U	0	-0,1	10	S	N	N	N
Caso 6	1	SQ	16	Log	Log	BP	CG	100	100	0,01	U	0	-0,1	0	S	S	N	N
Caso 7	1	En	20	TH	Log	BP	CG	50	50	0,01	U	0	-0,1	0	S	S	S	N
Caso 8	2	SQ	13	Log	Log	BP	CG	100	100	0,01	U	0	-0,1	0	S	S	S	N

Tabela 15 - AUC das redes para cada caso investigado

Caso	AUC
Caso 1	0,777
Caso 2	0,908
Caso 3	0,760
Caso 4	0,908
Caso 5	0,967
Caso 6	0,889
Caso 7	0,713
Caso 8	0,739

3.4 Comparação dos resultados

Após a obtenção dos valores das áreas sob as curvas ROC, é importante verificar qual a significância da diferença encontrada entre os dois classificadores utilizados em cada caso. Em termos absolutos, os resultados das redes superaram aqueles encontrados pela regressão logística em 6 dos 8 casos analisados como demonstrado na Tabela 16. Mas até que ponto essa diferença é substancial em termos estatísticos?

Tabela 16 - AUC de ambos os classificadores

Caso	AUC (RL)	AUC (RNA)
Caso 1	0,747	0,777
Caso 2	0,845	0,908
Caso 3	0,720	0,760
Caso 4	0,911	0,908
Caso 5	0,965	0,967
Caso 6	0,888	0,889
Caso 7	0,710	0,713
Caso 8	0,739	0,739

Para responder à essa questão, analisa-se se os resultados obtidos pelas redes neurais artificiais foram estatisticamente superiores àqueles obtidos pela regressão logística binária, através de um teste t para dados emparelhados, em que as médias dos oito conjunto de dados para a AUC de ambas as metodologias foram confrontadas.

A estatística t é dada, portanto, por:

$$t = \frac{\bar{d} - \mu_0}{s_d / (\sqrt{n})} \quad (9)$$

em que:

μ_0 = média hipotética das diferenças da população (=0)

\bar{d} = média das diferenças das amostras emparelhadas

s_d = desvio padrão das diferenças das amostras emparelhadas

n = tamanho da amostra

Testando a hipótese nula de que as médias das populações são iguais, com a hipótese alternativa de que a média das AUC para as RNAs é superior à média encontrada pela regressão logística binária, obtém-se um valor de $t = 1,97$ e $p\text{-value} = 0,045$, conforme demonstrado pela tabela 17.

Tabela 17 - Teste t para dados emparelhados

	N	Média	Desvio Padrão
RNA	8	0,8326	0,0956
RL	8	0,8156	0,0988
Diferença	8	0,01700	0,02444
<i>T-value = 1,97</i>		<i>P-Value = 0,045</i>	

Dessa forma, com nível de significância de 0,05, rejeita-se a hipótese nula de que as médias são iguais ($p\text{-value} < 0,05$), ou seja, a média das áreas sob a curva ROC das redes foi estatisticamente superior às áreas calculadas pelo modelo de regressão logística binária.

4. CONCLUSÕES

O presente trabalho visou comparar duas das técnicas mais populares na abordagem de problemas de classificação. Através da revisão realizada, analisou-se a hipótese de que a metodologia das redes neurais artificiais treinadas por DOE obteria uma maior capacidade de poder discriminatório, analisado através das áreas sob a curva ROC, sendo assim um melhor classificador quando comparado à regressão logística e preferivelmente utilizada nesses casos.

Os resultados alcançados com os oito casos clássicos analisados, todos discutidos no livro de Hosmer, Lemeshow e Sturdivant (2013), vão ao encontro do exposto no parágrafo anterior, uma vez que, em termos absolutos, as áreas sob as curvas ROC dos casos modelados por RNAs superaram àqueles modelados por RL na maioria dos estudos. Similarmente, ao se analisar as diferenças em termos estatísticos, através da análise do *p-value* do teste t para dados emparelhados, as redes neurais artificiais obtiveram uma maior área sob a curva, isto é, um melhor poder discriminatório quando comparadas à regressão logística binária. Dessa forma, para os exemplos considerados, as redes neurais provaram ser um melhor classificador.

Todavia, ainda há a possibilidade do modelador buscar por características intrínsecas de cada metodologia ao optar por cada uma das técnicas. A regressão logística, por exemplo, tem por objetivo destacar relações causais entre a variável dependente e as variáveis independentes e com seu uso pode-se entender facilmente o efeito das variáveis predictoras na variável de saída, uma vez que a equação do modelo é fornecida. O uso das redes neurais é particularmente útil quando o objetivo principal é a própria classificação, principalmente quando contém parâmetros com interações, além de influências não-lineares presentes no conjunto de dados. Além disso, RNAs requerem menos treinamento estatístico formal e podem ser desenvolvidas com o uso de diferentes algoritmos.

É importante destacar que, apesar da superioridade encontrada através do uso das redes neurais artificiais, ambas as metodologias cumpriram com sucesso os objetivos a que se propuseram, de maneira a garantir área sob a curva todas maiores que 70%, indicando um poder classificatório razoável e chegando a mais de 90% em alguns casos, indicando um excelente poder de classificação.

4.1 Considerações finais

É interessante ressaltar ao fim deste trabalho que todos os objetivos específicos propostos inicialmente foram cumpridos:

- Criou-se um modelo de regressão logística para cada um dos casos analisados e

através deles pôde-se inferir, dentre outros, a influência de cada parâmetro na variável de resposta.

- De maneira estruturada, através do uso do planejamento de experimentos para obtenção dos melhores parâmetros, criou-se um modelo de rede neural artificial para cada um dos casos analisados.
- Obtiveram-se as áreas sob a curva ROC para cada um dos casos, utilizando ambos os métodos.
- Utilizou-se testes de hipóteses, disponíveis na literatura, para verificar se há diferença no desempenho encontrado em ambas as metodologias.

4.2 Trabalhos futuros

Para a elaboração deste trabalho, fez-se uso de uma série de delimitações necessárias para sua finalização em tempo hábil. Dessa forma, algumas sugestões para trabalhos futuros são descritas a seguir:

- Extrapolar a metodologia para novos conjuntos de dados e verificar se os desempenhos obtidos continuam similares.
- Inserir outras metodologias de classificação de dados além das utilizadas nesse trabalho como a análise discriminante ou as máquinas de vetores de suporte.
- Com relação à regressão logística, verificar a presença de interações entre as variáveis preditoras.
- Com relação ao planejamento ou delineamento de experimentos, explorar, alternativa ou adicionalmente ao Plackett-Burman (notadamente um delineamento de seleção), outros tipos de delineamento de seleção e aperfeiçoamento.
- Com relação à parametrização da rede, utilizar um procedimento que viabilize sua posterior otimização. Uma sugestão seria, após descobrir os efeitos principais, utilizar o fatorial completo (livre de confundimento) para gerar um modelo matemático que se possa otimizar.
- Com relação às redes neurais, verificar o desempenho de outras arquiteturas de rede além da MLP, além de testar interações e outros níveis para os parâmetros da rede.
- Com relação ao método de comparação, uma alternativa à análise das áreas sob as curvas ROC seria a análise de atributos concordantes como a estatística de Kappa.

APÊNDICE A - Banco de dados de cada caso abordado

Caso 2

STA	AGE	GEN	RAC	SER	CAN	CRN	INF	CPR	SYS	HRA	PRE	TYP	FRA	PO2	PH	PCO	BIC	CRE	LOC
0	27	1	1	0	0	0	1	0	142	88	0	1	0	0	0	0	0	0	0
0	59	0	1	0	0	0	0	0	112	80	1	1	0	0	0	0	0	0	0
0	77	0	1	1	0	0	0	0	100	70	0	0	0	0	0	0	0	0	0
0	54	0	1	0	0	0	1	0	142	103	0	1	1	0	0	0	0	0	0
0	87	1	1	1	0	0	1	0	110	154	1	1	0	0	0	0	0	0	0
0	69	0	1	0	0	0	1	0	110	132	0	1	0	1	0	0	1	0	0
0	63	0	1	1	0	0	0	0	104	66	0	0	0	0	0	0	0	0	0
0	30	1	1	0	0	0	0	0	144	110	0	1	0	0	0	0	0	0	0
0	35	0	2	0	0	0	0	0	108	60	0	1	0	0	0	0	0	0	0
0	70	1	1	1	1	0	0	0	138	103	0	0	0	0	0	0	0	0	0
0	55	1	1	1	0	0	1	0	188	86	1	0	0	0	0	0	0	0	0
0	48	0	2	1	1	0	0	0	162	100	0	0	0	0	0	0	0	0	0
0	66	1	1	1	0	0	0	0	160	80	1	0	0	0	0	0	0	0	0
0	61	1	1	0	0	1	0	0	174	99	0	1	0	0	1	0	1	1	0
0	66	0	1	0	0	0	0	0	206	90	0	1	0	0	0	0	0	1	0
0	52	0	1	1	0	0	1	0	150	71	1	0	0	0	0	0	0	0	0
...
1	75	0	1	0	0	1	1	0	130	119	0	1	0	0	1	0	1	1	0
1	65	1	1	0	0	0	1	1	104	150	0	1	0	0	0	1	0	0	2
1	49	0	1	0	0	0	1	1	140	108	0	1	0	0	0	0	1	0	0
1	75	1	1	0	0	1	1	1	150	66	0	1	0	0	0	0	0	1	2
1	72	1	1	0	0	0	0	0	90	160	0	1	0	0	0	0	0	0	0
1	69	0	1	0	0	1	0	0	80	81	0	1	0	0	0	0	0	0	2
1	64	0	1	0	1	0	1	0	80	118	0	1	0	1	0	0	0	1	0
1	60	0	1	0	0	0	1	0	56	114	1	1	0	0	1	0	1	0	0
1	60	0	3	1	0	1	1	0	130	55	0	1	0	0	0	0	0	0	1
1	50	1	2	0	0	0	0	0	256	64	0	1	0	0	0	0	0	0	1

Caso 3

SITE	PHY	PRIF	AGE	WEI	HEI	BMI	PRE	MOM	ARM	SMO	RAT	SCO	FRAC
1	14	0	62	70,3	158	28,161	0	0	0	0	2	1	0
4	284	0	65	87,1	160	34,023	0	0	0	0	2	2	0
6	305	1	88	50,8	157	20,609	0	1	1	0	1	11	0
6	309	0	82	62,1	160	24,258	0	0	0	0	1	5	0
1	37	0	61	68	152	29,432	0	0	0	0	2	1	0
5	299	1	67	68	161	26,234	0	0	0	1	2	4	0
5	302	0	84	50,8	150	22,578	0	0	0	0	1	6	0
1	36	1	82	40,8	153	17,429	0	0	0	0	2	7	0
1	8	1	86	62,6	156	25,723	0	0	0	0	2	7	0
4	282	0	58	63,5	166	23,044	0	0	0	0	1	0	0
6	315	0	67	67,6	153	28,878	0	1	0	1	1	4	0
1	34	0	56	117,9	167	42,275	0	0	1	1	2	3	0
6	315	0	59	67,1	162	25,568	0	0	0	1	1	1	0
1	33	0	72	57,6	165	21,157	0	0	1	0	1	5	0
1	23	0	64	61,2	160	23,906	1	0	0	0	2	1	0
3	179	0	68	78	161	30,091	0	0	1	0	1	4	0
...
6	316	0	58	72,6	160	28,359	0	0	0	0	2	0	1
6	313	0	56	72,6	162	27,663	0	1	0	0	2	1	1
1	22	1	59	54	143	26,407	0	0	0	0	1	2	1
3	166	0	88	81,6	158	32,687	0	0	0	0	1	6	1
2	70	0	87	64,4	158	25,797	0	1	1	0	2	9	1
5	287	1	79	63,5	157	25,762	0	1	1	0	3	8	1
5	296	0	64	48,1	149	21,666	0	0	0	0	3	2	1
5	287	1	61	70,8	161	27,314	1	0	1	0	3	4	1
3	181	1	81	77,6	153	33,15	0	0	1	0	2	8	1
6	317	0	63	74,8	165	27,475	0	0	0	0	1	1	1

Caso 4

PLA	AGE	RAC	GEN	NEU	EMO	DAN	ELO	LOS	BEH	CUS	VIO
0	15,953	0	0	3	1	0	0	14	0	0	0
1	14,571	1	1	0	0	1	0	44	7	0	1
0	15,819	1	0	0	0	1	0	11	4	0	0
0	15,598	0	1	0	0	3	1	4	6	1	1
0	16,35	0	1	2	1	3	0	5	7	0	1
0	13,552	0	1	0	0	3	1	8	6	0	1
1	14,448	0	0	0	1	1	0	6	4	1	1
0	14,338	1	0	0	0	3	0	3	5	0	1
1	13,509	1	0	0	0	1	1	45	6	0	1
1	12,559	1	1	0	0	3	1	8	7	1	1
0	14,004	0	1	0	0	3	1	5	5	0	0
0	16,802	1	1	0	1	3	0	13	8	0	1
0	15,546	0	0	0	0	0	0	4	2	0	0
0	15,68	1	1	1	1	3	0	5	7	0	1
1	15,023	1	0	0	0	1	1	4	6	1	1
0	13,739	0	0	0	0	3	0	4	8	0	1
...
0	12,159	1	1	0	0	2	1	14	4	0	1
1	15,028	0	1	0	0	1	0	6	6	1	1
0	14,259	0	0	1	0	1	1	10	6	0	1
0	11,776	0	0	0	1	3	0	9	4	0	1
1	11,819	0	0	0	0	2	1	6	8	1	1
1	11,526	1	1	0	0	3	1	230	7	1	1
1	15,003	1	1	0	1	2	0	62	6	1	1
0	16,586	0	0	0	0	1	0	16	5	0	1
1	17,539	0	0	0	0	3	1	20	6	1	1
0	13,465	0	0	0	1	3	0	7	7	0	1

Caso 5

DEA	AGE	GEN	RAC	TBS	INH	FLA
0	26,6	1	1	25,3	0	1
0	2	0	0	5	0	0
0	22	0	0	2	0	0
0	37,3	1	1	2	0	0
0	52,1	1	1	6	0	1
0	50,2	1	1	7	0	0
0	2,5	0	0	7	0	0
0	53,8	0	1	0,9	0	1
0	31,9	1	1	2	0	0
0	41,1	1	1	22	0	1
0	3,8	1	0	0,5	0	0
0	5,9	1	0	30,5	0	0
0	1,9	1	0	12	0	0
0	56,2	1	1	7	0	1
0	67,3	1	1	17	0	1
0	15,5	1	1	2	0	1
...
1	54,9	1	0	15	1	1
1	52,6	1	0	76,5	1	1
1	84,5	1	1	29,5	0	1
1	42,8	0	1	98	0	1
1	83,7	0	1	50,5	0	0
1	34,2	1	1	91	1	1
1	59	1	1	37,5	1	1
1	85,5	1	1	4,6	1	1
1	46,8	1	0	47	1	1
1	40,8	1	1	1,2	1	1

Caso 6

MYO	AGE	GEN	SPH	AL	ACD	LT	VCD	SPO	REA	COM	STU	TVH	DIO	MOM	DAD
1	6	1	-0,052	21,89	3,69	3,498	14,7	45	8	0	0	10	34	1	1
0	6	1	0,608	22,38	3,702	3,392	15,29	4	0	1	1	7	12	1	1
0	6	1	1,179	22,49	3,462	3,514	15,52	14	0	2	0	10	14	0	0
1	6	1	0,525	22,2	3,862	3,612	14,73	18	11	0	0	4	37	0	1
0	5	0	0,697	23,29	3,676	3,454	16,16	14	0	0	0	4	4	1	0
0	6	0	1,744	22,14	3,224	3,556	15,36	10	6	2	1	19	44	0	1
0	6	1	0,683	22,33	3,186	3,654	15,49	12	7	2	1	8	36	0	1
0	6	1	1,272	22,39	3,732	3,584	15,08	12	0	0	0	8	8	0	0
0	7	0	1,396	22,62	3,464	3,408	15,74	4	0	3	1	3	12	0	0
0	6	1	0,972	22,74	3,504	3,696	15,54	30	5	1	0	10	27	0	0
0	6	0	0,595	22,17	3,358	3,528	15,29	8	8	10	12	10	90	1	1
0	6	1	0,509	24,5	3,864	3,466	17,17	5	5	0	0	5	20	1	1
0	6	0	2,185	22,19	3,49	3,57	15,13	8	2	3	0	3	15	0	0
0	7	0	0,987	22,88	3,702	3,266	15,92	14	3	0	1	4	16	0	0
0	6	0	0,169	22,98	3,504	3,556	15,92	10	0	2	0	20	24	1	0
0	7	0	0,466	24,03	3,704	3,438	16,89	8	2	1	1	3	14	0	0
...
0	6	1	4,372	21,55	3,704	3,598	14,25	5	3	10	2	10	45	0	1
0	6	1	2,121	22,63	3,716	3,682	15,24	20	1	1	0	14	19	0	0
0	6	0	0,439	23,3	3,73	3,322	16,25	10	2	15	15	20	101	1	1
1	8	0	-0,149	22,88	3,876	3,366	15,64	23	5	0	2	4	25	0	1
0	9	0	-0,032	22,96	3,758	3,336	15,86	9	9	5	0	16	53	1	0
1	6	0	0,678	22,4	3,663	3,803	14,93	2	0	7	3	14	37	1	0
0	6	1	0,665	22,5	3,57	3,378	15,56	6	0	1	0	8	10	1	1
0	6	0	1,834	22,94	3,624	3,424	15,89	8	0	0	0	4	4	1	1
0	6	1	0,665	21,92	3,688	3,598	14,64	12	2	1	0	15	23	0	0
0	6	0	0,802	22,26	3,53	3,484	15,25	25	0	2	0	10	14	1	1

Caso 7

GEN	AGE	MAR	SAM	PSU	STR	TCH	HDL	SYS	DBP	WT	HT	BMI	VIG	MOD	WLK	VIG	MOD	SED	OBE
0	34	1	80101	1	9	135	50	114	88	87,4	165	32,2	1	1	1	1	1	480	0
0	16		13953	1	10	192	60	112	62	72,3	181	22	1	1	0	1	0	240	0
1	60	2	20090	2	1	202	45	154	70	117	166	42,4	1	1	1	1	1	240	1
0	26	1	22538	1	14	160	45	102	50	97,6	173	32,6	1	1	0	0	1	720	0
1	49	6	74212	2	11	259	45	118	82	86,7	168	30,6	1	1	1	1	1	240	0
0	80	1	11998	1	3	182	75	142	62	79,1	174	26	1	0	0	1	0	60	0
0	80	2	21807	1	5	148	49	126	62	89,6	180	27,6	1	0	1	1	1	540	0
0	17		11445	1	7	178	66	122	76	74,7	170	26	1	0	1	0	0	480	0
1	42	1	39087	2	1	170	35			108	164	39,9	1	1	1	1	1	30	1
0	66	1	9286	1	1	217	40	146	68	82,9	171	28,3	1	1	1	1	1		0
1	45	1	86636	2	4	225	82	106	62	75,7	167	27,2	1	0	1	0	0	720	0
0	28	5	23137	1	2	299	51	108	62	79,4	163	29,7	1	1	0	0	0	30	0
1	19		18878	2	8	185	48			112	166	40,7	1	1	0	1	0		1
0	44	3	90640	2	5	197	37	118	90	102	188	29,1	1	1	1	1	1	600	0
0	66	1	56053	2	12	193	26	124	64	68	170	23,7	1	0	1	1	0	180	0
1	49	1	18367	1	8	202	58	140	80	76,5	164	28,4	1	1	1	1	1	60	0
...
1	16		51175	1	14	194	57	102	64	67,9	165	24,9	0	0	1	0	1	480	0
1	74	2	18562	2	6			144	72	97,6	166	35,4	1	1	1	1	1	600	1
1	66	1	60892	1	11			140	82	50,7	154	21,3	1	1	1	1	1	600	0
0	52	3	12838	1	15	210	29	106	80	144	179	44,8	1	0	0	1	1	120	1
0	24	5	19614	2	10	155	57	126	80	75,1	172	25,5	0	0	0	0	0	240	0
0	32	1	17152	1	15	198	38			94,3	164	35,3	1	0	1	1	0	240	1
1	48	3	21110	2	6	246	76	112	74	87,1	157	35,4	1	0	1	1	1	600	1
0	27	6	23193	1	10	121	51	112	54	57	164	21,1	0	0	0	1	0	360	0
0	75	1	11984	2	7	176	59	126	48	75,1	163	28,4	1	1	1	1	1	480	0
1	63	1	6257	2	7	181	38	138	70	71,3	157	28,8	1	1	1	1	0	120	0

Caso 8

POL	MHV	INP	YEA	GRO	URB	COM	ANY	NUM	GEN	RAC	ETH	AGE
0	0	0	2002	1	0	1	1	1	0	0	0	4,67
0	1	0	2003	1	0	1	1	1	0	0	0	5,67
0	0	0	2004	1	0	0	0	0	0	0	0	6
0	1	0	2005	1	0	1	1	1	0	0	0	7,08
0	0	0	2006	1	0	1	1	1	0	0	0	8
0	1	0	2007	2	0	1	1	1	0	0	0	9,92
0	2	0	2008	2	0	1	1	1	0	0	0	10,67
0	2	0	2002	2	0	1	1	1	1	1	0	7,58
0	2	0	2003	2	0	1	1	1	1	1	0	8,08
0	3	0	2004	2	0	0	1	1	1	1	0	9,83
0	3	0	2005	2	0	1	1	1	1	1	0	10,33
0	3	0	2006	1	0	1	1	1	1	1	0	12,67
0	2	0	2007	1	0	0	0	0	1	1	0	12,5
0	2	0	2008	1	0	0	1	1	1	1	0	13,33
1	3	0	2002	2	0	0	0	0	1	0	0	11,75
1	3	2	2003	2	0	0	1	1	1	0	0	12,58
...
0	3	0	2006	2	0	0	0	0	1	0	0	15,17
0	3	0	2007	2	0	0	1	1	1	0	0	16,75
0	3	0	2008	2	0	0	1	1	1	0	0	17
0	2	0	2002	2	0	0	1	1	0	0	0	2,67
0	1	0	2003	2	0	1	1	1	0	0	0	3,58
0	1	0	2004	2	0	1	1	1	0	0	0	4,92
0	1	0	2005	2	0	1	1	1	0	0	0	5
0	1	0	2006	2	0	1	1	1	0	0	0	6,92
0	2	0	2007	2	0	1	1	1	0	0	0	7,5
0	2	0	2008	2	0	1	1	1	0	0	0	8,5

APÊNDICE B - Resultados da Regressão Logística

Caso 2

Link Function: *Logit*

Response Information

Variable	Value	Count	
STA_1	1	31	(Event)
	0	129	
	Total	160	

Factor Information

Factor	Levels	Values
TYP_1	2	0; 1
LOC_1	2	0; 1

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-2,83397	1,55536	-1,82	0,068			
AGE_1	0,0285708	0,0127411	2,24	0,025	1,03	1,00	1,06
SYS_1	-0,0174713	0,0078511	-2,23	0,026	0,98	0,97	1,00
TYP_1							
1	1,98674	0,804647	2,47	0,014	7,29	1,51	35,30
LOC_1							
1	3,38223	1,01090	3,35	0,001	29,44	4,06	213,48

Log-Likelihood = -57,380

Test that all slopes are zero: G = 42,557, DF = 4, P-value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	163,814	153	0,260
Deviance	114,760	153	0,991
Hosmer-Lemeshow	14,976	8	0,060

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
1												
Obs	0	0	3	0	0	2	4	2	9	11		31
Exp	0,3	0,5	0,9	1,2	1,6	2,2	3,1	4,1	5,4	11,6		
0												
Obs	16	16	13	16	16	14	12	14	7	5		129
Exp	15,7	15,5	15,1	14,8	14,4	13,8	12,9	11,9	10,6	4,4		
Total	16	16	16	16	16	16	16	16	16	16		160

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	3392	84,8	Somers' D 0,70
Discordant	590	14,8	Goodman-Kruskal Gamma 0,70
Ties	17	0,4	Kendall's Tau-a 0,22
Total	3999	100,0	

Caso 3

Link Function: *Logit*

Response Information

Variable	Value	Count	
FRACTURE_1	1	98	(Event)
	0	302	
	Total	400	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	3,05533	3,59363	0,85	0,395			
PRIORFRAC_1							
1	0,761184	0,278058	2,74	0,006	2,14	1,24	3,69
AGE_1	0,0227245	0,0145660	1,56	0,119	1,02	0,99	1,05
HEIGHT_1	-0,0409262	0,0205293	-1,99	0,046	0,96	0,92	1,00
MOMFRAC_1							
1	0,611246	0,329387	1,86	0,063	1,84	0,97	3,51
ARMASSIST_1							
1	0,379508	0,259535	1,46	0,144	1,46	0,88	2,43
RATERISK_1							
2	0,440837	0,307996	1,43	0,152	1,55	0,85	2,84
3	0,543139	0,331985	1,64	0,102	1,72	0,90	3,30

Log-Likelihood = -204,236

Test that all slopes are zero: G = 36,949, DF = 7, P-value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	377,412	373	0,426
Deviance	399,107	373	0,169
Hosmer-Lemeshow	7,984	8	0,435

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	2	4	4	4	10	12	12	13	19	18	98
Exp	3,7	5,0	5,9	6,8	7,5	8,6	10,5	13,1	15,9	20,9	
0											
Obs	38	36	36	37	30	28	28	27	21	21	302
Exp	36,3	35,0	34,1	34,2	32,5	31,4	29,5	26,9	24,1	18,1	
Total	40	40	40	41	40	40	40	40	40	39	400

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	21058	71,2	Somers' D 0,43
Discordant	8402	28,4	Goodman-Kruskal Gamma 0,43
Ties	136	0,5	Kendall's Tau-a 0,16
Total	29596	100,0	

Caso 4

Link Function: *Logit*

Response Information

Variable	Value	Count	
PLACE4_1	1	194	(Event)
	0	212	
	Total	406	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Constant	-6,61196	1,57709	-4,19	0,000			
AGE_1	0,176378	0,0959710	1,84	0,066	1,19	0,99	1,44
LOS_1	0,0889503	0,0164670	5,40	0,000	1,09	1,06	1,13
BEHAV_1	0,261197	0,0828162	3,15	0,002	1,30	1,10	1,53
CUSTD_1							
1	3,55963	0,338636	10,51	0,000	35,15	18,10	68,26

Log-Likelihood = -142,826

Test that all slopes are zero: G = 276,386, DF = 4, P-value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	353,938	401	0,956
Deviance	285,651	401	1,000
Hosmer-Lemeshow	16,470	8	0,036

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
1												
Obs	0	5	3	3	17	19	34	35	37	41	194	
Exp	1,9	3,2	4,5	6,5	9,2	21,1	32,9	35,2	38,7	40,9		
0												
Obs	40	36	37	38	24	21	7	5	4	0	212	
Exp	38,1	37,8	35,5	34,5	31,8	18,9	8,1	4,8	2,3	0,1		
Total	40	41	40	41	41	40	41	40	41	41	406	

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	37881	92,1	Somers' D 0,84
Discordant	3184	7,7	Goodman-Kruskal Gamma 0,84
Ties	63	0,2	Kendall's Tau-a 0,42
Total	41128	100,0	

Caso 5

Link Function: *Logit*

Response Information

Variable	Value	Count
DEATH_1	1	121 (Event)
	0	679
Total		800

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Constant	-7,94380	0,694399	-11,44	0,000			
AGE_1	0,0825153	0,0091688	9,00	0,000	1,09	1,07	1,11
TBSA_1	0,0952398	0,0104352	9,13	0,000	1,10	1,08	1,12
INH_INJ_1							
1	1,26485	0,381368	3,32	0,001	3,54	1,68	7,48

Log-Likelihood = -136,866

Test that all slopes are zero: G = 406,062, DF = 3, P-value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	383,058	771	1,000
Deviance	273,732	771	1,000
Hosmer-Lemeshow	6,834	8	0,555

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
1												
Obs	0	0	0	0	0	0	4	15	34	68	121	
Exp	0,0	0,1	0,2	0,3	0,8	1,8	4,0	10,2	32,8	70,7		
0												
Obs	80	80	81	80	80	80	76	65	46	11	679	
Exp	80,0	79,9	80,8	79,7	79,2	78,2	76,0	69,8	47,2	8,3		
Total	80	80	81	80	80	80	80	80	80	79	800	

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	79289	96,5	Somers' D 0,93
Discordant	2808	3,4	Goodman-Kruskal Gamma 0,93
Ties	62	0,1	Kendall's Tau-a 0,24
Total	82159	100,0	

Caso 6

Link Function: *Logit*

Response Information

Variable	Value	Count	
MYOPIC_1	1	61	(Event)
	0	433	
	Total	494	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-7,90855	2,87472	-2,75	0,006			
GENDER_1							
1	0,941116	0,350036	2,69	0,007	2,56	1,29	5,09
SPHEQ_1	-3,45550	0,479562	-7,21	0,000	0,03	0,01	0,08
ACD_1	1,88583	0,772578	2,44	0,015	6,59	1,45	29,97
SPORTHR_1	-0,0442522	0,0224627	-1,97	0,049	0,96	0,92	1,00
MOMMY_1							
1	0,437365	0,338447	1,29	0,196	1,55	0,80	3,01
DADMY_1							
1	1,07210	0,351142	3,05	0,002	2,92	1,47	5,81

Log-Likelihood = -124,431

Test that all slopes are zero: G = 120,457, DF = 6, P-value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	325,609	487	1,000
Deviance	248,862	487	1,000
Hosmer-Lemeshow	5,138	8	0,743

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
1												
Obs	0	0	0	0	1	5	4	10	14	27	61	
Exp	0,0	0,2	0,6	1,1	1,8	2,8	4,2	7,7	14,4	28,2		
0												
Obs	49	49	50	49	49	44	45	40	35	23	433	
Exp	49,0	48,8	49,4	47,9	48,2	46,2	44,8	42,3	34,6	21,8		
Total	49	49	50	49	50	49	49	50	49	50	494	

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	23315	88,3	Somers' D 0,77
Discordant	3030	11,5	Goodman-Kruskal Gamma 0,77
Ties	68	0,3	Kendall's Tau-a 0,17
Total	26413	100,0	

Caso 7

Link Function: *Logit*

Response Information

Variable	Value	Count
OBESE_1	1	617 (Event)
	0	3315
Total		3932

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Constant	-3,06542	0,426320	-7,19	0,000			
GENDER_1							
1	1,04504	0,101464	10,30	0,000	2,84	2,33	3,47
HDL_1	-0,0387389	0,0035510	-10,91	0,000	0,96	0,96	0,97
SYSBP_1	0,0057080	0,0026170	2,18	0,029	1,01	1,00	1,01
DBP_1	0,0112336	0,0041662	2,70	0,007	1,01	1,00	1,02
WLKBIK_1							
1	0,401603	0,114765	3,50	0,000	1,49	1,19	1,87
MODRECEXR_1							
1	0,313087	0,100818	3,11	0,002	1,37	1,12	1,67
VIGRECEXR_1							
1	0,382955	0,147358	2,60	0,009	1,47	1,10	1,96
SEDMIN_1	0,0012137	0,0002424	5,01	0,000	1,00	1,00	1,00

Log-Likelihood = -1561,635

Test that all slopes are zero: G = 293,819, DF = 8, P-value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	3797,88	3915	0,908
Deviance	3112,18	3915	1,000
Hosmer-Lemeshow	15,77	8	0,046

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total
	1	2	3	4	5	6	7	8	9	10	
1											
Obs	12	24	20	32	47	74	71	94	107	136	617
Exp	12,6	23,7	31,8	39,8	48,4	57,5	68,3	82,9	104,1	147,9	
0											
Obs	381	369	373	361	347	319	322	299	286	258	3315
Exp	380,4	369,3	361,2	353,2	345,6	335,5	324,7	310,1	288,9	246,1	
Total	393	393	393	393	394	393	393	393	393	394	3932

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	1455178	71,1	Somers' D 0,43
Discordant	583002	28,5	Goodman-Kruskal Gamma 0,43
Ties	7175	0,4	Kendall's Tau-a 0,11
Total	2045355	100,0	

Caso 8

Link Function: *Logit*

Response Information

Variable	Value	Count
POLYPHARMACY_1	1	637 (Event)
	0	2163
Total		2800

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	-239,731	68,6992	-3,49	0,000			
MHV4_1							
1	0,493019	0,223615	2,20	0,027	1,64	1,06	2,54
2	1,35077	0,210549	6,42	0,000	3,86	2,56	5,83
3	2,02659	0,207307	9,78	0,000	7,59	5,05	11,39
INPTMHV3_1							
1	1,08762	0,222296	4,89	0,000	2,97	1,92	4,59
2	0,768474	0,346864	2,22	0,027	2,16	1,09	4,26
GROUP_1							
2	0,346382	0,110056	3,15	0,002	1,41	1,14	1,75
3	0,332115	0,149473	2,22	0,026	1,39	1,04	1,87
COMORBID_1							
1	-0,388135	0,151799	-2,56	0,011	0,68	0,50	0,91
GENDER_1							
1	0,423989	0,126696	3,35	0,001	1,53	1,19	1,96
AGE_1	0,0252412	0,0244632	1,03	0,302	1,03	0,98	1,08
YEAR_1	0,117954	0,0343553	3,43	0,001	1,13	1,05	1,20
RACE_1							
1	-0,541710	0,144262	-3,76	0,000	0,58	0,44	0,77
2	-0,170404	0,647552	-0,26	0,792	0,84	0,24	3,00

Log-Likelihood = -1315,225

Test that all slopes are zero: G = 372,483, DF = 13, P-value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	2606,63	2548	0,205
Deviance	2424,18	2548	0,960
Hosmer-Lemeshow	3,56	8	0,894

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
1	Obs	15	20	24	43	41	63	75	89	111	156	637
	Exp	13,8	21,5	28,7	37,4	47,3	58,8	73,0	89,2	112,6	154,9	
0	Obs	265	260	256	237	239	217	205	191	170	123	2163
	Exp	266,2	258,5	251,3	242,6	232,7	221,2	207,0	190,8	168,4	124,1	
Total		280	280	280	280	280	280	280	280	281	279	2800

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	1018321	73,9	Somers' D 0,48
Discordant	355591	25,8	Goodman-Kruskal Gamma 0,48
Ties	3919	0,3	Kendall's Tau-a 0,17
Total	1377831	100,0	

APÊNDICE C - Planejamento experimental das RNAs

Caso 2

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC 1	AUC 2
2	En	40	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,8730	0,8912
2	SQ	20	TH	Log	BP	CG	50	50	0,1	U	0,1	-0,1	10	S	S	S	N	0,9045	0,8917
1	SQ	40	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,8074	0,7980
1	En	40	TH	Id	QP	QN	50	50	0	U	0,1	-0,1	10	N	S	S	S	0,5322	0,5152
2	En	20	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,9173	0,9021
2	SQ	20	Log	Log	QP	CG	100	100	0	U	0	-0,1	10	N	S	N	S	0,9059	0,9311
2	SQ	40	Log	Id	QP	QN	50	100	0,1	U	0	-0,1	0	S	N	S	N	0,9225	0,9339
2	SQ	40	TH	Id	BP	QN	100	50	0,1	N	0	-0,1	0	N	S	N	S	0,5188	0,5264
1	SQ	40	TH	Log	BP	CG	100	100	0	N	0,1	-0,1	0	N	N	S	N	0,8246	0,7720
2	En	40	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,8623	0,8214
1	SQ	20	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,8931	0,8469
2	En	40	Log	Log	QP	QN	100	50	0	N	0,1	-0,1	10	S	N	N	N	0,8225	0,8050
1	SQ	20	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,9313	0,9322
1	En	40	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,9209	0,9055
1	En	20	TH	Id	QP	CG	100	100	0,1	N	0	-0,1	10	S	N	S	S	0,5323	0,5808
1	En	20	Log	Log	BP	QN	50	100	0,1	N	0,1	-0,1	0	S	S	N	S	0,8325	0,8725
2	En	20	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,8069	0,7882
2	SQ	20	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,8815	0,8792
1	SQ	40	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,7943	0,7678
1	En	20	Log	Id	BP	CG	50	50	0	U	0	-0,1	0	N	N	N	N	0,8520	0,8084

Caso 3

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC 1	AUC 2
2	En	28	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,7797	0,7253
2	SQ	14	TH	Log	BP	CG	50	50	0,1	U	0,1	-0,1	10	S	S	S	N	0,7530	0,7346
1	SQ	28	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,7085	0,7127
1	En	28	TH	Id	QP	QN	50	50	0	U	0,1	-0,1	10	N	S	S	S	0,5660	0,6201
2	En	14	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,7911	0,8057
2	SQ	14	Log	Log	QP	CG	100	100	0	U	0	-0,1	10	N	S	N	S	0,6943	0,6213
2	SQ	28	Log	Id	QP	QN	50	100	0,1	U	0	-0,1	0	S	N	S	N	0,7246	0,6543
2	SQ	28	TH	Id	BP	QN	100	50	0,1	N	0	-0,1	0	N	S	N	S	0,5126	0,5492
1	SQ	28	TH	Log	BP	CG	100	100	0	N	0,1	-0,1	0	N	N	S	N	0,7672	0,7733
2	En	28	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,5543	0,5713
1	SQ	14	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,6862	0,7348
2	En	28	Log	Log	QP	QN	100	50	0	N	0,1	-0,1	10	S	N	N	N	0,6461	0,7128
1	SQ	14	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,7393	0,7197
1	En	28	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,6956	0,6895
1	En	14	TH	Id	QP	CG	100	100	0,1	N	0	-0,1	10	S	N	S	S	0,5416	0,5741
1	En	14	Log	Log	BP	QN	50	100	0,1	N	0,1	-0,1	0	S	S	N	S	0,7257	0,6736
2	En	14	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,6774	0,6872
2	SQ	14	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,6290	0,6967
1	SQ	28	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,6368	0,6102
1	En	14	Log	Id	BP	CG	50	50	0	U	0	-0,1	0	N	N	N	N	0,6916	0,7035

Caso 4

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC 1	AUC 2
2	En	26	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,8551	0,9279
2	SQ	13	TH	Log	BP	CG	50	50	0,1	U	0,1	-0,1	10	S	S	S	N	0,9259	0,9414
1	SQ	26	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,8671	0,8575
1	En	26	TH	Id	QP	QN	50	50	0	U	0,1	-0,1	10	N	S	S	S	0,8676	0,9025
2	En	13	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,9214	0,9389
2	SQ	13	Log	Log	QP	CG	100	100	0	U	0	-0,1	10	N	S	N	S	0,9238	0,9145
2	SQ	26	Log	Id	QP	QN	50	100	0,1	U	0	-0,1	0	S	N	S	N	0,9156	0,9179
2	SQ	26	TH	Id	BP	QN	100	50	0,1	N	0	-0,1	0	N	S	N	S	0,5417	0,6302
1	SQ	26	TH	Log	BP	CG	100	100	0	N	0,1	-0,1	0	N	N	S	N	0,9217	0,9320
2	En	26	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,5691	0,5445
1	SQ	13	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,9239	0,9315
2	En	26	Log	Log	QP	QN	100	50	0	N	0,1	-0,1	10	S	N	N	N	0,8841	0,9204
1	SQ	13	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,9203	0,8865
1	En	26	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,9124	0,9244
1	En	13	TH	Id	QP	CG	100	100	0,1	N	0	-0,1	10	S	N	S	S	0,6330	0,5306
1	En	13	Log	Log	BP	QN	50	100	0,1	N	0,1	-0,1	0	S	S	N	S	0,8752	0,9067
2	En	13	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,6773	0,8250
2	SQ	13	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,9088	0,9103
1	SQ	26	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,8161	0,8648
1	En	13	Log	Id	BP	CG	50	50	0	U	0	-0,1	0	N	N	N	N	0,8911	0,8888

Caso 5

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC1	AUC2
2	En	14	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,9508	0,9710
2	SQ	7	TH	Log	BP	CG	50	50	0,1	U	0,1	-0,1	10	S	S	S	N	0,9669	0,9689
1	SQ	14	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,8999	0,9043
1	En	14	TH	Id	QP	QN	50	50	0	U	0,1	-0,1	10	N	S	S	S	0,9065	0,8806
2	En	7	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,9690	0,9686
2	SQ	7	Log	Log	QP	CG	100	100	0	U	0	-0,1	10	N	S	N	S	0,9694	0,9692
2	SQ	14	Log	Id	QP	QN	50	100	0,1	U	0	-0,1	0	S	N	S	N	0,9667	0,9636
2	SQ	14	TH	Id	BP	QN	100	50	0,1	N	0	-0,1	0	N	S	N	S	0,9383	0,9586
1	SQ	14	TH	Log	BP	CG	100	100	0	N	0,1	-0,1	0	N	N	S	N	0,9660	0,9484
2	En	14	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,7517	0,6170
1	SQ	7	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,9600	0,9605
2	En	14	Log	Log	QP	QN	100	50	0	N	0,1	-0,1	10	S	N	N	N	0,9469	0,9633
1	SQ	7	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,9652	0,9635
1	En	14	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,9693	0,9684
1	En	7	TH	Id	QP	CG	100	100	0,1	N	0	-0,1	10	S	N	S	S	0,4807	0,5000
1	En	7	Log	Log	BP	QN	50	100	0,1	N	0,1	-0,1	0	S	S	N	S	0,9509	0,9696
2	En	7	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,6331	0,8223
2	SQ	7	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,9557	0,9561
1	SQ	14	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,9629	0,9638
1	En	7	Log	Id	BP	CG	50	50	0	U	0	-0,1	0	N	N	N	N	0,9601	0,9623

Caso 6

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC1	AUC2
2	En	32	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,6424	0,7128
2	SQ	16	TH	Log	BP	CG	50	50	0,1	U	0,1	-0,1	10	S	S	S	N	0,8839	0,8840
1	SQ	32	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,7301	0,7226
1	En	32	TH	Id	QP	QN	50	50	0	U	0,1	-0,1	10	N	S	S	S	0,4578	0,5408
2	En	16	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,8039	0,8306
2	SQ	16	Log	Log	QP	CG	100	100	0	U	0	-0,1	10	N	S	N	S	0,8632	0,8784
2	SQ	32	Log	Id	QP	QN	50	100	0,1	U	0	-0,1	0	S	N	S	N	0,6027	0,6278
2	SQ	32	TH	Id	BP	QN	100	50	0,1	N	0	-0,1	0	N	S	N	S	0,6596	0,7190
1	SQ	32	TH	Log	BP	CG	100	100	0	N	0,1	-0,1	0	N	N	S	N	0,8681	0,8708
2	En	32	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,5055	0,6233
1	SQ	16	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,5609	0,5906
2	En	32	Log	Log	QP	QN	100	50	0	N	0,1	-0,1	10	S	N	N	N	0,5904	0,5163
1	SQ	16	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,8681	0,8461
1	En	32	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,8899	0,8919
1	En	16	TH	Id	QP	CG	100	100	0,1	N	0	-0,1	10	S	N	S	S	0,4472	0,5049
1	En	16	Log	Log	BP	QN	50	100	0,1	N	0,1	-0,1	0	S	S	N	S	0,8863	0,8896
2	En	16	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,5702	0,5384
2	SQ	16	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,6840	0,5743
1	SQ	32	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,8006	0,7168
1	En	16	Log	Id	BP	CG	50	50	0	U	0	-0,1	0	N	N	N	N	0,8639	0,8697

Caso 7

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC1	AUC2
2	En	40	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,5882	0,6415
2	SQ	20	TH	Log	BP	CG	50	50	0,1	U	0,1	-0,1	10	S	S	S	N	0,6941	0,6841
1	SQ	40	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,6031	0,6124
1	En	40	TH	Id	QP	QN	50	50	0	U	0,1	-0,1	10	N	S	S	S	0,5478	0,5620
2	En	20	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,6939	0,7041
2	SQ	20	Log	Log	QP	CG	100	100	0	U	0	-0,1	10	N	S	N	S	0,4850	0,5181
2	SQ	40	Log	Id	QP	QN	50	100	0,1	U	0	-0,1	0	S	N	S	N	0,5084	0,5994
2	SQ	40	TH	Id	BP	QN	100	50	0,1	N	0	-0,1	0	N	S	N	S	0,5836	0,5112
1	SQ	40	TH	Log	BP	CG	100	100	0	N	0,1	-0,1	0	N	N	S	N	0,7023	0,7005
2	En	40	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,5317	0,5115
1	SQ	20	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,5033	0,5620
2	En	40	Log	Log	QP	QN	100	50	0	N	0,1	-0,1	10	S	N	N	N	0,5369	0,4732
1	SQ	20	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,4997	0,5448
1	En	40	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,6013	0,6537
1	En	20	TH	Id	QP	CG	100	100	0,1	N	0	-0,1	10	S	N	S	S	0,5000	0,4772
1	En	20	Log	Log	BP	QN	50	100	0,1	N	0,1	-0,1	0	S	S	N	S	0,6365	0,6478
2	En	20	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,5978	0,5000
2	SQ	20	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,6035	0,6089
1	SQ	40	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,4633	0,4703
1	En	20	Log	Id	BP	CG	50	50	0	U	0	-0,1	0	N	N	N	N	0,6926	0,6696

Caso 8

NC	FE	NN	CO	CS	F1	F2	E1	E2	TA	MI	CP	ET	EE	PN	PE	D1	D2	AUC1	AUC2
2	En	26	TH	Id	BP	CG	50	100	0	N	0	0	10	S	S	N	N	0,6948	0,7221
2	SQ	13	TH	Log	BP	CG	50	50	0,1	U	0,1	-0,1	10	S	S	S	N	0,7420	0,7405
1	SQ	26	Log	Log	QP	CG	50	50	0	N	0	0	0	S	S	S	S	0,7211	0,7219
1	En	26	TH	Id	QP	QN	50	50	0	U	0,1	-0,1	10	N	S	S	S	0,5353	0,5000
2	En	13	TH	Log	BP	QN	100	50	0	U	0	0	0	S	N	S	S	0,7450	0,7412
2	SQ	13	Log	Log	QP	CG	100	100	0	U	0	-0,1	10	N	S	N	S	0,7387	0,7426
2	SQ	26	Log	Id	QP	QN	50	100	0,1	U	0	-0,1	0	S	N	S	N	0,4834	0,7371
2	SQ	26	TH	Id	BP	QN	100	50	0,1	N	0	-0,1	0	N	S	N	S	0,4948	0,6950
1	SQ	26	TH	Log	BP	CG	100	100	0	N	0,1	-0,1	0	N	N	S	N	0,7444	0,7472
2	En	26	TH	Log	QP	CG	50	100	0,1	U	0,1	0	0	N	N	N	S	0,4904	0,6210
1	SQ	13	TH	Log	QP	QN	50	50	0,1	N	0	0	10	N	N	N	N	0,4172	0,5664
2	En	26	Log	Log	QP	QN	100	50	0	N	0,1	-0,1	10	S	N	N	N	0,7232	0,7486
1	SQ	13	TH	Id	QP	QN	100	100	0	U	0,1	0	0	S	S	N	N	0,7290	0,5924
1	En	26	Log	Log	BP	QN	100	100	0,1	U	0	0	10	N	S	S	N	0,6739	0,7253
1	En	13	TH	Id	QP	CG	100	100	0,1	N	0	-0,1	10	S	N	S	S	0,6835	0,4887
1	En	13	Log	Log	BP	QN	50	100	0,1	N	0,1	-0,1	0	S	S	N	S	0,7295	0,7388
2	En	13	Log	Id	QP	CG	100	50	0,1	N	0,1	0	0	N	S	S	N	0,5000	0,6197
2	SQ	13	Log	Id	BP	QN	50	100	0	N	0,1	0	10	N	N	S	S	0,7133	0,6945
1	SQ	26	Log	Id	BP	CG	100	50	0,1	U	0,1	0	10	S	N	N	S	0,6781	0,5007
1	En	13	Log	Id	BP	CG	50	50	0	U	0	-0,1	0	N	N	N	N	0,7392	0,7435

REFERÊNCIAS BIBLIOGRÁFICAS

ADYA, M.; COLLOPY, F. How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation. **Journal of Forecasting**, v. 17, p. 481–495, 1998.

AGGARWAL, N.; KUMAR, A.; KHATTER, H.; AGGARWAL, V. Analysis the effect of data mining techniques on database. **Advances in Engineering Software**, v. 47, n. 1, p. 164–169, 2012.

AHMAD, J.; SHAFI, I.; ANSARI, S.; ISMAIL SHAH, S. Neural network-based approach for the non-invasive diagnosis and classification of hepatotropic viral disease. **IET Communications**, v. 6, n. 18, p. 3265–3273, 2012.

AL-SHAYEA, Q. K. Artificial Neural Networks in Medical Diagnosis. , v. 8, n. 2, p. 150–154, 2011.

ASKIN, O. E.; GOKALP, F. Comparing the Predictive and Classification Performances of Logistic Regression and Neural Networks: A Case Study on Timss 2011. **Procedia - Social and Behavioral Sciences**, v. 106, p. 667–676, 2013.

BAGLEY, S. C.; WHITE, H.; GOLOMB, B. A. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. **Journal of clinical epidemiology**, v. 54, n. 10, p. 979–85, 2001.

BALESTRASSI, P. P.; POPOVA, E.; PAIVA, A. P.; MARANGON LIMA, J. W. Design of experiments on neural network's training for nonlinear time series forecasting. **Neurocomputing**, v. 72, n. 4-6, p. 1160–1178, 2009.

BEN-DAVID, A. About the relationship between ROC curves and Cohen's kappa. **Engineering Applications of Artificial Intelligence**, v. 21, n. 6, p. 874–882, 2008.

BERTRAND, J. W. M.; FRANSOO, J. C. Operations management research methodologies using quantitative modeling. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 241–264, 2002.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. Springer, 2006.

BORQUE, A.; SANZ, G.; ALLEPUZ, C.; PLAZA, L.; GIL, P.; RIOJA, L. A. The use of neural networks and logistic regression analysis for predicting pathological stage in men undergoing radical prostatectomy: a population based study. **The Journal of urology**, v. 166, n. 5, p. 1672–8, 2001.

BRAGA, A. DE P.; CARVALHO, A. P. DE L. F.; LUDEMIR, T. B. **Redes Neurais Artificiais - Teoria e Aplicações**. 2nd ed. LTC Livros Técnicos Editora S.A., 2007.

CAOCCI, G.; BACCOLI, R.; VACCA, A.; MASTRONUZZI, A.; BERTAINA, A.; PIRAS, E.; LITTERA, R.; LOCATELLI, F.; CARCASSI, C.; LA NASA, G. Comparison between an artificial neural network and logistic regression in predicting acute graft-vs-host disease after unrelated donor hematopoietic stem cell transplantation in thalassemia patients. **Experimental hematology**, v. 38, n. 5, p. 426–33, 2010.

CHEN, H.; ZHANG, J.; XU, Y.; CHEN, B.; ZHANG, K. Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans. **Expert Systems with Applications**, v. 39, n. 13, p. 11503–11509, 2012.

CHIANG, W. K.; ZHANG, D.; ZHOU, L. Predicting and explaining patronage behavior toward web and traditional stores using neural networks: a comparative analysis with logistic regression. **Decision Support Systems**, v. 41, n. 2, p. 514–531, 2006.

CHOI, J.; OH, H.-J.; LEE, H.-J.; LEE, C.; LEE, S. Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. **Engineering Geology**, v. 124, p. 12–23, 2012.

CHUN, F. K.-H.; GRAEFEN, M.; BRIGANTI, A.; GALLINA, A.; HOPP, J.; KATTAN, M.; HULAND, H.; KARAKIEWICZ, P. Initial biopsy outcome prediction--head-to-head comparison of a logistic regression-based nomogram versus artificial neural network. **European urology**, v. 51, n. 5, p. 1236–40; discussion 1241–3, 2007.

CHUNG, C. A. **SIMULATION MODELING HANDBOOK A Practical Approach**. CRC Press, 2004.

COLEMAN, D. E.; MONTGOMERY, D. C. A Systematic Approach to Planning for a Designed Industrial Experiment. **Technometrics**, v. 35, n. 1, p. 1–12, 1993.

COURVOISIER, D. S.; COMBESURE, C.; AGORITSAS, T.; GAYET-AGERON, A.; PERNEGER, T. V. Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. **Journal of clinical epidemiology**, v. 64, n. 9, p. 993–1000, 2011.

COUTO, A. F. DO. **Análise de Correlação no Reconhecimento de Interações em Arranjos Plackett-Burman**. Dissertação de Mestrado - UNIFEI, 2012.

DAS, I.; SAHOO, S.; VAN WESTEN, C.; STEIN, A.; HACK, R. Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). **Geomorphology**, v. 114, n. 4, p. 627–637, 2010.

DEHURI, S.; CHO, S.-B. Evolutionarily optimized features in functional link neural network for classification. **Expert Systems with Applications**, v. 37, n. 6, p. 4379–4391, 2010.

DIRUSSO, S. M.; CHAHINE, A. A.; SULLIVAN, T.; RUSICCI, D.; NEALON, P.; CUFF, S.; SAVINO, J.; SLIM, M. Development of a model for prediction of survival in pediatric trauma patients: Comparison of artificial neural networks and logistic regression. **Journal of Pediatric Surgery**, v. 37, n. 7, p. 1098–1104, 2002.

DIXON, B. A case study using support vector machines, neural networks and logistic regression in a GIS to identify wells contaminated with nitrate-N. **Hydrogeology Journal**, v. 17, n. 6, p. 1507–1520, 2009.

DREISEITL, S.; OHNO-MACHADO, L. Logistic regression and artificial neural network classification models: a methodology review. **Journal of Biomedical Informatics**, v. 35, n. 5-6, p. 352–359, 2002.

ERGÜN, U. U.; SERHATLIOĞLU, S.; HARDALAÇ, F.; GÜLER, I. Classification of carotid artery stenosis of patients with diabetes by neural network and logistic regression. **Computers in biology and medicine**, v. 34, n. 5, p. 389–405, 2004.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **American Association for Artificial Intelligence**, p. 37–54, 1996.

FIGUEIRA, C. V. **Modelos de Regressão Logística**. Dissertação de Mestrado - UFRGS, 2006.

FREEMAN, R. V.; EAGLE, K. A.; BATES, E. R.; WERNS, S. W.; KLINE-ROGERS, E.; KARAVITE, D.; MOSCUCCI, M. Comparison of artificial neural networks with logistic regression in prediction of in-hospital death after percutaneous transluminal coronary angioplasty. **American heart journal**, v. 140, n. 3, p. 511–20, 2000.

GREEN, M.; BJÖRK, J.; FORBERG, J.; EKELUND, U.; EDENBRANDT, L.; OHLSSON, M. Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. **Artificial intelligence in medicine**, v. 38, n. 3, p. 305–18, 2006.

GUTIÉRREZ, P. A.; LÓPEZ-GRANADOS, F.; PEÑA-BARRAGÁN, J. M.; JURADO-EXPÓSITO, M.; HERVÁS-MARTÍNEZ, C. Logistic regression product-unit neural networks for mapping *Ridolfia segetum* infestations in sunflower crop using multitemporal remote sensed data. **Computers and Electronics in Agriculture**, v. 64, n. 2, p. 293–306, 2008.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Multivariate Data Analysis**. 2006.

HAJMEER, M.; BASHEER, I. Comparison of logistic regression and neural network-based classifiers for bacterial growth. **Food Microbiology**, v. 20, n. 1, p. 43–55, 2003.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 2012.

HAND, D. J.; TILL, R. J. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. **Machine Learning**, v. 45, p. 171–186, 2001.

HAYASHI, Y.; SETIONO, R. Combining neural network predictions for medical diagnosis. **Computers in biology and medicine**, v. 32, n. 4, p. 237–46, 2002.

- HAYKIN, S. **Neural Networks and Learning Machines**. 3rd ed. Pearson, 2009.
- HOSMER, D. W.; HOSMER, T.; LE CESSIE, S.; LEMESHOW, S. A comparison of goodness-of-fit tests for the logistic regression model. **Statistics in Medicine**, v. 16, n. May 1995, p. 965–980, 1997.
- HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied Logistic Regression**. 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013.
- AL HOUSSEINI, A.; NEWMAN, T.; COX, A.; DEVOE, L. D. Prediction of risk for cesarean delivery in term nulliparas: a comparison of neural network and multiple logistic regression models. **American journal of obstetrics and gynecology**, v. 201, n. 1, p. 113.e1–6, 2009.
- ITCHHAPORIA, D.; SNOW, P. B.; ALMASSY, R. J.; OETGEN, W. J. Artificial neural networks: current status in cardiovascular medicine. **Journal of the American College of Cardiology**, v. 28, n. 2, p. 515–21, 1996.
- JIANG, J.; TRUNDLE, P.; REN, J. Medical image analysis with artificial neural networks. **Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society**, v. 34, n. 8, p. 617–31, 2010.
- JUNIOR, F. H. F. DE C. **Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais**. Dissertação de Mestrado - USP, 2003.
- KAWAKAMI, S.; NUMAO, N.; OKUBO, Y.; KOGA, G.; YAMAMOTO, S.; KAZUTAKA, S.; YASUHISA, F.; YONESE, J.; MASUDA H.; KIHARA, K.; FUKUI, I. Development, validation, and head-to-head comparison of logistic regression-based nomograms and artificial neural network models predicting prostate cancer on initial extended biopsy. **European urology**, v. 54, n. 3, p. 601–11, 2008.
- KIM, S. Y. Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. **The Service Industries Journal**, v. 31, n. 3, p. 441–468, 2011.
- KURT, I.; TURE, M.; KURUM, A. T. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. **Expert Systems with Applications**, v. 34, n. 1, p. 366–374, 2008.

KURUVILLA, J.; GUNAVATHI, K. Lung cancer classification using neural networks for CT images. **Computer methods and programs in biomedicine**, v. 113, n. 1, p. 202–9, 2014.

LANGER, D. L.; VAN DER KWAST, T. H.; EVANS, A. J.; TRACHTENBER, J.; WILSON, B. C.; HAIDER, M.A. Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. **Journal of magnetic resonance imaging : JMRI**, v. 30, n. 2, p. 327–34, 2009.

LARASATI, A.; DEYONG, C.; SLEVITCH, L. The Application of Neural Network and Logistics Regression Models on Predicting Customer Satisfaction in a Student-Operated Restaurant. **Procedia - Social and Behavioral Sciences**, v. 65, p. 94–99, 2012.

LEE, S. H.; YU, D.; BACHMAN, A. H.; LIM, J.; ARDEKANI, B. A. Application of fused lasso logistic regression to the study of corpus callosum thickness in early Alzheimer's disease. **Journal of neuroscience methods**, v. 221, p. 78–84, 2014.

LEUNG, P.; TRAN, L. T. Predicting shrimp disease occurrence: artificial neural networks vs. logistic regression. **Aquaculture**, v. 187, n. 1-2, p. 35–49, 2000.

LIAO, S.-H.; CHU, P.-H.; HSIAO, P.-Y. Data mining techniques and applications – A decade review from 2000 to 2011. **Expert Systems with Applications**, v. 39, n. 12, p. 11303–11311, 2012.

LIEW, P.-L.; LEE, Y.-C.; LIN, Y.-C.; LEE, T.-S.; WANG, W.; CHIEN, C.-W. Comparison of artificial neural networks with logistic regression in prediction of gallbladder disease among obese patients. **Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver**, v. 39, n. 4, p. 356–62, 2007.

LIN, C.-C.; OU, Y.-K.; CHEN, S.-H.; LIU, Y.-C.; LIN, J. Comparison of artificial neural network and logistic regression models for predicting mortality in elderly patients with hip fracture. **Injury**, v. 41, n. 8, p. 869–73, 2010.

LISBOA, P. J. G. A review of evidence of health benefit from artificial neural networks in medical intervention. **Neural Networks**, v. 15, p. 11–39, 2002.

LISBOA, P. J.; TAKTAK, A. F. G. The use of artificial neural networks in decision support in cancer: a systematic review. **Neural networks : the official journal of the International Neural Network Society**, v. 19, n. 4, p. 408–15, 2006.

MADADI, M.; ZHANG, S.; YEARY, K. H. K.; HENDERSON, L. M. Analyzing factors associated with women's attitudes and behaviors toward screening mammography using design-based logistic regression. **Breast cancer research and treatment**, v. 144, n. 1, p. 193–204, 2014.

MAGALLANES, J. F.; OLIVIERI, A. C. The effect of factor interactions in Plackett–Burman experimental designs. **Chemometrics and Intelligent Laboratory Systems**, v. 102, n. 1, p. 8–14, 2010.

MARTINS, R. A.; MELLO, C. H. P.; TURRIONI, J. B. **Guia para elaboração de Monografia e TCC em Engenharia de Produção**. Primeira ed. Atlas, 2013.

MCLAREN, C. E.; CHEN, W.-P.; NIE, K.; SU, M.-Y. Prediction of malignant breast lesions from MRI features: a comparison of artificial neural network and logistic regression techniques. **Academic radiology**, v. 16, n. 7, p. 842–51, 2009.

MELLINGER, M. Multivariate data analysis: Its methods. **Chemometrics and Intelligent Laboratory Systems**, v. 2, n. 1-3, p. 29–36, 1987.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem prática**. 2005.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 7th ed. John Wiley & Sons Inc., 2009.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 2011.

MORTEZA, A.; NAKHJAVANI, M.; ASGARANI, F.; CARVALHO, F.; KARIMI, R.; ESTEGHAMATI, A. Inconsistency in albuminuria predictors in type 2 diabetes: a comparison between neural network and conditional logistic regression. **Translational research : the journal of laboratory and clinical medicine**, v. 161, n. 5, p. 397–405, 2013.

NGUYEN, T.; MALLEY, R.; INKELIS, S.; KUPPERMANN, N. Comparison of prediction models for adverse outcome in pediatric meningococcal disease using artificial neural network and logistic regression analyses. **Journal of clinical epidemiology**, v. 55, n. 7, p. 687–95, 2002.

NGUYEN, V. Q.; ABE, S.; SUN, G.; MATSUOKA, A.; NISHIMURA, H.; ISHIHARA, M.; MATSUI, T. Rapid screening for influenza using a multivariable logistic regression model to save labor at a clinic in Iwaki, Fukushima, Japan. **American journal of infection control**, v. 42, n. 5, p. 551–3, 2014.

OTTENBACHER, K. J.; LINN, R. T.; SMITH, P. M.; ILLIG, A. B.; MACUSO, M.; GRANGER, C. V. Comparison of logistic regression and neural network analysis applied to predicting living setting after hip fracture. **Annals of epidemiology**, v. 14, n. 8, p. 551–9, 2004.

OTTENBACHER, K. J.; OTTENBACHER, H. R.; TOOTH, L.; OSTIR, G. V. A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions. **Journal of clinical epidemiology**, v. 57, n. 11, p. 1147–52, 2004.

OTTENBACHER, K. J.; SMITH, P. M.; ILLIG, S. B.; LINN, R. T.; FIEDLER, R. C.; GRANGER, C. V. Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. **Journal of clinical epidemiology**, v. 54, n. 11, p. 1159–65, 2001.

ÖZEL, T.; KARPAT, Y. Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks. **International Journal of Machine Tools and Manufacture**, v. 45, n. 4-5, p. 467–479, 2005.

PALIWAL, M.; KUMAR, U. A. Neural networks and statistical techniques: A review of applications. **Expert Systems with Applications**, v. 36, n. 1, p. 2–17, 2009.

PARDOE, I. **Applied Regression Modeling A Business Approach**. 2006.

PAVLEKOVIC, M.; BENSIC, M.; ZEKIC-SUSAC, M. Modeling children's mathematical gift by neural networks and logistic regression. **Expert Systems with Applications**, v. 37, n. 10, p. 7167–7173, 2010.

PEPE, M. S. Evaluating technologies for classification and prediction in medicine. **Statistics in medicine**, v. 24, n. 24, p. 3687–96, 2005.

PLACKETT, R. L.; BURMAN, J. P. The design of optimum multifactorial experiments. **Biometrika**, v. 33, n. 4, p. 305–325, 1946.

PONTES, F. J. **Projeto Otimizado de Redes Neurais Artificiais para Predição da Rugosidade em Processos de Usinagem com a Utilização da Metodologia de Projeto de Experimentos**. Tese de Doutorado - UNESP, 2011.

POURSHAHRIAR, H. Correct vs. accurate prediction: A comparison between prediction power of artificial neural networks and logistic regression in psychological researches. **Procedia - Social and Behavioral Sciences**, v. 32, n. 2011, p. 97–103, 2012.

PRADHAN, B.; LEE, S. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. **Environmental Modelling & Software**, v. 25, n. 6, p. 747–759, 2010.

PREMACHANDRA, I. M.; BHABRA, G. S.; SUEYOSHI, T. DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. **European Journal of Operational Research**, v. 193, n. 2, p. 412–424, 2009.

RAZI, M.; ATHAPPILLY, K. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. **Expert Systems with Applications**, v. 29, n. 1, p. 65–74, 2005.

EL SANHARAWI, M.; NAUDET, F. [Understanding logistic regression]. **Journal français d'ophtalmologie**, v. 36, n. 8, p. 710–5, 2013.

SCHUMACHER, M.; ROBNER, R.; VACH, W. Neural networks and logistic regression : Part I. **Computational Statistics & Data Analysis**, v. 21, p. 661–682, 1996.

SHEPHERD, A. J. **Second-Order Methods for Neural Network**. Springer, 1997.

SHI, H.-Y.; HWANG, S.-L.; LEE, K.-T.; LIN, C.-L. In-hospital mortality after traumatic brain injury surgery: a nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. **Journal of neurosurgery**, v. 118, n. 4, p. 746–52, 2013.

SICK, B. on-Line and Indirect Tool Wear Monitoring in Turning With Artificial Neural Networks: a Review of More Than a Decade of Research. **Mechanical Systems and Signal Processing**, v. 16, n. 4, p. 487–546, 2002.

SONG, J. H.; VENKATESH, S. S.; CONANT, E. A.; ARGER, P. H.; SEHGAL, C. M. Comparative analysis of logistic regression and artificial neural network for computer-aided diagnosis of breast masses. **Academic radiology**, v. 12, n. 4, p. 487–95, 2005.

SØREIDE, K. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. **Journal of clinical pathology**, v. 62, n. 1, p. 1–5, 2009.

STEYERBERG, E. W.; EIJKEMANS, M. J.; HARRELL, F. E.; HABBEMA, J. D. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. **Statistics in medicine**, v. 19, n. 8, p. 1059–79, 2000.

STEYERBERG, E. W.; HARRELL, F. E.; BORSBOOM, G. J.; EIJKEMANS, M. J. C.; VERGOUWE, Y.; HABBEMA, J. D. F. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. **Journal of clinical epidemiology**, v. 54, n. 8, p. 774–81, 2001.

SWIDERSKI, B.; KUREK, J.; OSOWSKI, S. Multistage classification by using logistic regression and neural networks for assessment of financial condition of company. **Decision Support Systems**, v. 52, n. 2, p. 539–547, 2012.

TANG, Z.-H.; LIU, J.; ZENG, F.; LI, Z.; YU, X.; ZHOU, L. Comparison of prediction model for cardiovascular autonomic dysfunction using artificial neural network and logistic regression analysis. **Plos one**, v. 8, n. 8, p. e70571, 2013.

TIMMERMAN, D.; TESTA, A. C.; BOURNE, T.; FERRAZI, E.; AMEYE, L.; KONSTANTINOVIC, M. L.; CALSTER, B. V.; COLLINS, W.; VERGOTE, I. HUFFEL, S.;

VALENTIN, L. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the International Ovarian Tumor Analysis Group. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 23, n. 34, p. 8794–801, 2005.

TRTICA-MAJNARIC, L.; ZEKIC-SUSAC, M.; SARLIJA, N.; VITALE, B. Prediction of influenza vaccination outcome by neural networks and logistic regression. **Journal of biomedical informatics**, v. 43, n. 5, p. 774–81, 2010.

TU, J. V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. **Journal of clinical epidemiology**, v. 49, n. 11, p. 1225–31, 1996.

VACH, W.; ROBNER, R.; SCHUMACHER, M. Neural networks and logistic regression: Part II. **Computational Statistics & Data Analysis**, v. 21, p. 683-705, 1996.

VALENZUELA, T.; ROE, D.; CRETIN, S.; SPAITE, D.; LARSEN, M. Estimating Effectiveness of Cardiac Arrest Interventions A Logistic Regression Survival Mode. **Circulation**, v. 96, n. 10, p. 3308–3313, 1997.

VALLEJOS, J. A.; MCKINNON, S. D. Logistic regression and neural network classification of seismic records. **International Journal of Rock Mechanics and Mining Sciences**, v. 62, p. 86–95, 2013.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining - Practical Machine Learning Tools and Techniques**. 3rd ed. 2011.

WOUTERS, M.; BENSCHOP, A.; VAN LAAR, M.; KORF, D. J. Cannabis use and proximity to coffee shops in the Netherlands. **European Journal of Criminology**, v. 9, n. 4, p. 337–353, 2012.

YESILNACAR, E.; TOPAL, T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). **Engineering Geology**, v. 79, n. 3-4, p. 251–266, 2005.

YILMAZ, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). **Computers & Geosciences**, v. 35, n. 6, p. 1125–1138, 2009a.

YILMAZ, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. **Environmental Earth Sciences**, v. 61, n. 4, p. 821–836, 2009b.

ZHANG, G. P. Neural networks for classification: a survey. **IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)**, v. 30, n. 4, p. 451–462, 2000.

ZHANG, G. P. Avoiding Pitfalls in Neural Network Research. **IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)**, v. 37, n. 1, p. 3–16, 2007.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks : The state of the art. **International Journal of Forecasting**, v. 14, p. 35–62, 1998.