

**UNIVERSIDADE DO EXTREMO SUL CATARINENSE - UNESC**

**CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**CAMILA DE FARIAS CRISPIM**

**UTILIZAÇÃO DE MÁQUINAS DE SUPORTE VETORIAL PARA A  
CLASSIFICAÇÃO DE DADOS DE SEPSE ASSOCIADA A DELIRIUM**

**CRICIÚMA**

**2014**

**CAMILA DE FARIAS CRISPIM**

**UTILIZAÇÃO DE MÁQUINAS DE SUPORTE VETORIAL PARA A  
CLASSIFICAÇÃO DE DADOS DE SEPSE ASSOCIADA A DELIRIUM**

Trabalho de Conclusão de Curso, apresentado para obtenção do Grau de Bacharel no Curso de Ciência da Computação da Universidade do Extremo Sul Catarinense, UNESC.

Orientador: Prof. MSc. Gustavo Bisognin  
Coorientadora: Prof. MSc. Merisandra Côrtes de Mattos Garcia.

**CRICIÚMA**

**2014**

**CAMILA DE FARIAS CRISPIM**

**UTILIZAÇÃO DE MÁQUINAS DE SUPORTE VETORIAL PARA A  
CLASSIFICAÇÃO DE DADOS DE SEPSE ASSOCIADA A DELIRIUM**

Trabalho de Conclusão de Curso aprovado  
pela Banca Examinadora para obtenção do  
Grau de bacharel, no Curso de Ciência da  
Computação da Universidade do Extremo Sul  
Catarinense, UNESC.

Criciúma, 24 de junho de 2014.

**BANCA EXAMINADORA**

  
Prof. Gustavo Bisognin - Mestre - (UNESC) - Orientador

  
Profa. Merisandra C. de Mattos Garcia - Mestre - (UNESC) - Coorientadora

  
Prof. Evânio Ramos Nicoleit - Mestre - (UNESC)

  
Prof. Kristian Madeira - Mestre - (UNESC)

## RESUMO

A sepse é uma doença que no decorrer dos anos vem se tornando umas das principais causas de óbitos em Unidades de Terapia Intensiva. Diversos fatores, como a dificuldade em seu diagnóstico, contribuem para que a taxa de mortalidade em função desta doença seja cada vez mais alta. A partir disso, foi desenvolvido um simulador computacional que apresenta ao usuário se o paciente foi classificado como um possível óbito ou não, utilizando como fonte, uma base de dados de sepse associada a delirium. A técnica de Inteligência Artificial conhecida como Máquina de Suporte Vetorial foi utilizada no desenvolvimento deste simulador, e é ela a responsável pela realização da classificação da base de dados utilizada. Para se obter um modelo adequado de classificador para este simulador, utilizando a própria Máquina de Suporte Vetorial e uma outra técnica denominada Redes Neurais Artificiais, foram feitos diversos testes a fim de se obter os melhores parâmetros para a realização da classificação. Neste processo, foram utilizados dois algoritmos de Máquina de Suporte Vetorial, denominados SMO e LIBSVM e duas funções de Redes Neurais Artificiais chamadas MLP e RBF. Para a classificação, também foi utilizado o método de treinamento e teste *Leave-One-Out* e para fins de comparação, aplicou-se da mesma maneira o método *k-fold*. Com estes testes, foram obtidas a matriz de confusão e a curva Característica de Operação do Receptor, popularmente conhecida como curva ROC. Estas medidas de análise de conhecimento foram aplicadas com o propósito de avaliar o conhecimento gerado a partir dos resultados dos testes realizados. Baseado em ambas as medidas, foram analisados os resultados obtidos na classificação. Nesta análise, pode-se verificar que o número de pacientes sépticos em óbito foi superior aos que não estavam em óbito, confirmando as pesquisas realizadas de trabalhos correlatos. Além disso, observou-se que em duas das três classificações possíveis da base de dados, a Máquina de Suporte Vetorial obteve resultados melhores que os de Redes Neurais Artificiais. O simulador computacional foi desenvolvido baseado na análise da melhor classificação obtida. Nele, são informados valores para os biomarcadores, que são as variáveis utilizadas na própria classificação. Desta maneira, ao término do projeto, obteve-se uma análise dos resultados gerados com os testes aplicados na base de dados e um simulador que reproduzia os resultados adquiridos com classificação da base utilizando o melhor resultado de Máquina de Suporte Vetorial analisado.

**Palavras-chave:** Máquina de Suporte Vetorial. Método *Leave-One-Out*. Classificação de Dados. Sepse. Redes Neurais Artificiais.

## ABSTRACT

Sepsis is a disease that over the years has become a leading cause of death in intensive care units. Several factors, such as difficulty in diagnosis, contributes to the mortality rate of this disease is increasingly high. From this, it was developed a computer simulator that presents to the user if the patient was classified as a possible death or not, using as a source, a database of sepsis associated with delirium. The technique of artificial intelligence known as Support Vector Machine, used to develop this simulator, and it is responsible for carrying out the classification of database used. In order to obtain an appropriate classifier for this simulation model, using the Support Vector Machine and another technique called Neural Networks; several tests were made in order to obtain the best parameters for conducting the classification. In this process, were used two algorithms for Support Vector Machine, called SMO and LIBSVM, and two functions for Artificial Neural Networks, called MLP and RBF. For classification, we also used the method of training Leave-One-Out test and for comparison purposes, we applied the same way, the k-fold method. With these tests, were obtained the confusion matrix and the Receiver Operating Characteristic curve, popularly known as ROC curve. These knowledge analysis measures were applied in order to assess the knowledge generated from the results of the tests. Based on both measures, the classification results obtained were analyzed. In this analysis, it can be verified that the number of death in septic patients was higher than those who were not in death, confirming the research carried out in related work. Additionally, it was observed that two of the three possible classifiers of the database, the Support Vector Machine achieved better results than those of Artificial Neural Networks. The computer simulation was developed based on analysis of the best classification obtained. In it were informed values for biomarkers, which are the variables used in the classification. Thus at the end of the project, was obtained an analysis of the results generated by the tests applied in the database and a simulator to reproduce the results obtained with the basic classification using the best result from Support Vector Machine analyzed.

**Keywords:** Support Vector Machine. Leave-One-Out method. Classification of Data. Sepsis. Artificial Neural Networks.

## **AGRADECIMENTOS**

Agradeço a todos que me auxiliaram neste período. Agradeço a minha família, a meus amigos e aos meus colegas de trabalho pela paciência que tiveram comigo. Ao meu orientador Gustavo e minha coorientadora Merisandra pelas contribuições e auxílio dados. Por fim, agradeço a Deus por esta oportunidade e por esta conquista.

## LISTA DE ILUSTRAÇÕES

|   |    |
|---|----|
| Figura 1 – Dados referentes a incidência de sepse .....                       | 15 |
| Figura 2 – Taxas de mortalidade por sepse no estado da Bahia .....            | 16 |
| Figura 3 – Hierarquia do aprendizado .....                                    | 19 |
| Figura 4 – Modelo utilizado pela MSV na classificação de dados .....          | 21 |
| Figura 5 – Hiperplano dividido na classificação linear .....                  | 22 |
| Figura 6 - Margem máxima e vetores de suporte.....                            | 23 |
| Figura 7 – Representação da classificação não-linear .....                    | 24 |
| Figura 8 – Distribuição do conjunto de dados Data Iris .....                  | 25 |
| Figura 9 – Utilização de função de kernel linear.....                         | 26 |
| Figura 10 – Utilização de função de kernel polinomial. ....                   | 26 |
| Figura 11 – Utilização da função de kernel de Base Radial Gaussiana .....     | 27 |
| Figura 12 – Arquitetura de uma rede neural simplificada .....                 | 28 |
| Figura 13 – Esquema de uma RNA MLP .....                                      | 30 |
| Figura 14 – Esquema de uma RNA RBF .....                                      | 30 |
| Figura 15 – Exemplo de vetores de suporte.....                                | 32 |
| Figura 18 – Treinamento da MSV utilizando o método <i>Leave-One-Out</i> ..... | 33 |
| Figura 19 – Exemplo de Base de Treinamento .....                              | 34 |
| Figura 20 – Exemplo de matriz de confusão .....                               | 34 |
| Figura 21 – Método <i>k-fold</i> .....  | 35 |
| Figura 16 – Modelo de Matriz de Confusão de duas classes .....                | 37 |
| Figura 17 – Exemplo de Curva ROC .....  | 38 |
| Figura 22 – Ferramenta WEKA .....   | 46 |
| Figura 23 – Visualização de atributos no WEKA .....                           | 48 |
| Figura 24 – Visualização de atributos no WEKA.....                            | 49 |
| Figura 25 – Esquema para classificação dos dados .....                        | 49 |
| Figura 26 – Esquema de testes na base de dados .....                          | 51 |
| Figura 27 – Funcionamento do simulador .....                                  | 53 |
| Figura 28 – Simulador desenvolvido .....                                      | 54 |
| Figura 29 – Treinamento e teste com SMO para classificação Sepse .....        | 55 |
| Figura 30 – Treinamento e teste com LIBSVM para classificação Sepse. ....     | 56 |
| Figura 31 – Treinamento e teste com MLP para classificação Sepse.....         | 57 |
| Figura 32 – Treinamento e teste com RBF para classificação Sepse .....        | 57 |

|  |    |
|--|----|
| Figura 33 – Treinamento e teste com SMO para classificação de SAD .....          | 58 |
| Figura 34 – Treinamento e teste com LIBSVM para classificação de SAD .....       | 58 |
| Figura 35 - Treinamento e teste com MLP e RBF para classificação de SAD .....    | 59 |
| Figura 36 - Treinamento e teste com SMO para classificação da Evolução .....     | 60 |
| Figura 37 - Treinamento e teste com LIBSVM para classificação da Evolução .....  | 60 |
| Figura 38 - Treinamento e teste com MLP e RBF para classificação da Evolução ... | 60 |
| Figura 39 – Classificação de Sepsis .....  | 61 |
| Figura 40 – Classificação de SAD .....   | 62 |
| Figura 41 – Classificação de Evolução .....                                      | 62 |

## LISTA DE TABELAS

|  |    |
|--|----|
| Tabela 1 – Exemplo de dados para classificação ..... | 32 |
|--|----|

## LISTA DE ABREVIATURAS E SIGLAS

|              |   |
|--------------|---|
| TNF          | Fator de Necrose Tumoral                        |
| BDNF         | Fator Neurotrófico Derivado do Cérebro          |
| ARRF         | Formato de Arquivo Atributo – Relação           |
| IL-1 $\beta$ | Interleucina 1 $\beta$                          |
| JSF          | <i>JavaServerFaces</i>                          |
| MSV          | Máquina de Suporte Vetorial                     |
| MLP          | <i>Multilayer Perceptron</i>                    |
| RBF          | <i>Radial Bases Function</i>                    |
| RNA          | Redes Neurais Artificiais                       |
| ROC          | Curva Característica de Operação de um Receptor |
| SAD          | Sepse Associada a Delirium                      |
| SMO          | <i>Sequential Minimal Optimization</i>          |
| UTI          | Unidade de Terapia Intensiva                    |

## SUMÁRIO

|  |           |
|--|-----------|
| <b>1 INTRODUÇÃO</b> .....  | <b>9</b>  |
| 1.1 OBJETIVO GERAL .....   | 10        |
| 1.2 OBJETIVOS ESPECÍFICOS .....  | 10        |
| 1.3 JUSTIFICATIVA .....  | 11        |
| 1.4 ESTRUTURA DO TRABALHO .....  | 12        |
| <b>2 SEPSE</b> .....   | <b>13</b> |
| 2.1 EPIDEMIOLOGIA .....  | 14        |
| <b>3 APRENDIZADO DE MÁQUINA</b> .....  | <b>18</b> |
| 3.1 CLASSIFICAÇÃO DE DADOS .....   | 19        |
| 3.2 MÁQUINA DE SUPORTE VETORIAL .....  | 20        |
| <b>3.2.1 Classificação Linear</b> .....  | <b>22</b> |
| <b>3.2.2 Classificação Não-Linear</b> .....  | <b>24</b> |
| <b>3.2.3 Funções de Kernel</b> .....   | <b>24</b> |
| 3.2.3.1 Utilização das Funções de Kernel na Base de Dados Data Iris .....                                  | 25        |
| 3.3 REDES NEURAS ARTIFICIAIS .....   | 28        |
| <b>3.3.1 Multilayer Perceptron</b> .....   | <b>29</b> |
| <b>3.3.2 Radial Bases Function</b> .....   | <b>30</b> |
| <b>4 TREINAMENTO E CLASSIFICAÇÃO COM MSV</b> .....   | <b>31</b> |
| 4.1 MÉTODO <i>LEAVE-ONE-OUT</i> .....  | 33        |
| <b>4.1.1 Base de Treinamento</b> .....   | <b>33</b> |
| <b>4.1.2 Base de Teste</b> .....   | <b>34</b> |
| 4.2 MÉTODO <i>K-FOLD</i> .....   | 35        |
| <b>5 MEDIDAS DE análise de conhecimento</b> .....  | <b>37</b> |
| 5.1 MATRIZ DE CONFUSÃO .....   | 37        |
| 5.2 CURVA ROC .....  | 38        |
| <b>6 TRABALHOS CORRELATOS</b> .....  | <b>40</b> |
| 6.1 MAPEAMENTO DE AMBIENTES EXTERNOS UTILIZANDO ROBOS MÓVEIS   | 40        |
| 6.2 MÁQUINAS DE VETORES DE SUPORTE NA CLASSIFICAÇÃO DE IMPRESSÕES DIGITAIS .....                           | 41        |
| 6.3 ANÁLISE DE WAVELETS COM MÁQUINA DE VETOR DE SUPORTE NO ELETRENCEFALOGRAMA DA DOENÇA DE ALZHEIMER ..... | 41        |

|  |           |
|--|-----------|
| 6.4 SUPPORT VECTOR MACHINES NA CLASSIFICAÇÃO DE IMAGENS HIPERESPECTRAIS.....                                       | 42        |
| 6.5 DETECÇÃO DE INTRUSÃO USANDO TÉCNICAS DE APRENDIZAGEM DE MÁQUINAS .....   | 43        |
| 6.6 TÉCNICAS DE PROCESSAMENTO DE IMAGENS PARA LOCALIZAÇÃO E RECONHECIMENTO DE FACES .....                          | 43        |
| <b>7 UTILIZAÇÃO DE MÁQUINAS DE SUPORTE VETORIAL PARA CLASSIFICAÇÃO DE DADOS DE SEPSE ASSOCIADA A DELIRIUM.....</b> | <b>45</b> |
| 7.1 METODOLOGIA.....   | 45        |
| <b>7.1.1 Descrição da Base de Dados .....</b>  | <b>45</b> |
| <b>7.1.2 Utilização da Ferramenta Weka.....</b>  | <b>46</b> |
| <b>7.1.3 Classificação dos Dados .....</b>   | <b>49</b> |
| 7.1.3.1 Pré-processamento da base de dados.....  | 50        |
| 7.1.3.2 Treinamento e teste da base de dados .....   | 50        |
| <b>7.1.4 Desenvolvimento do Simulador .....</b>  | <b>52</b> |
| 7.2 RESULTADOS OBTIDOS .....   | 55        |
| <b>7.2.1 Classificação de Sepsis.....</b>  | <b>55</b> |
| <b>7.2.2 Classificação de SAD.....</b>   | <b>57</b> |
| <b>7.2.3 Classificação de Evolução .....</b>   | <b>59</b> |
| <b>7.2.4 Análise dos resultados .....</b>  | <b>61</b> |
| <b>8 CONCLUSÃO .....</b>   | <b>64</b> |
| <b>REFERÊNCIAS.....</b>  | <b>66</b> |

## 1 INTRODUÇÃO

A sepse é uma síndrome complexa de origem infecciosa, causada pela resposta inflamatória sistêmica descontrolada do indivíduo. Caracteriza-se por ter manifestações múltiplas, podendo ocasionar a disfunção ou falência de um ou mais órgãos ou mesmo a morte do indivíduo (CARVALHO; TROTTA, 2003).

No decorrer dos anos a sepse vem sendo uma das principais causas de óbitos em Unidades de Tratamento Intensivo (UTI). Apesar do desenvolvimento crescente de diversos métodos diagnósticos, a taxa de mortalidade devido a essa doença permanece, há décadas, em torno de 30 a 40% (SILVA; VELASCO, 2007). É considerada a décima causa de morte na UTI e pode ocasionar danos cerebrais a curto e longo prazo (MARTIN et al, 2000, tradução nossa).

Em muitos casos, a sepse pode ser associada ao estado de delirium. Delirium é um estado confusional agudo, representando a manifestação da disfunção cerebral e relata-se que a prevalência em UTI desses casos varia de 28% a 73%, segundo Pitrowsky et al (2005 apud PANDHARIPANDE et al, 2010). O delirium pode ser um dos primeiros sintomas da sepse, e pode se manifestar em pacientes com sepse em até 70% dos casos, conforme Louis-Vincent (2010, tradução nossa).

O número de óbitos por sepse já é considerável, e quando associado a delirium, o número de óbitos é ainda maior. A taxa de mortalidade de pacientes sépticos com estado mental alterado é de 49% em comparação com uma taxa de 26% em pacientes sépticos com estado mental normal (ZAMPIERI, 2011, tradução nossa).

Baseado na necessidade de diagnosticar de forma ágil e precisa os casos de sepse associada a delirium, presentes nas UTIs propõe-se para tal a utilização da Máquina de Suporte Vetorial (MSV).

Proposta nos anos 90, a MSV, tinha como propósito inicial ser um método para classificação de padrões e de regressão não-linear (AMORIM, 2006). Segundo Wang (2005, tradução nossa), a MSV é um método de aprendizado supervisionado

que gera funções de mapeamento de entrada e saída de dados de treinamento rotulados. Sendo que essas funções de mapeamento podem ser funções de classificação de dados. Podem ser aplicadas na categorização de textos, na análise de imagens e em Bioinformática, de acordo com Lorena e Carvalho (2007).

Juntamente com a MSV, propõe-se a utilização do método de treinamento e teste *Leave-One-Out* e *k-fold*. Para comparar os resultados obtidos na classificação, pretende-se utilizar outra técnica de Inteligência Artificial, as Redes Neurais Artificiais.

Então, a partir dos conceitos vistos, busca-se identificar previamente a possibilidade de óbitos por sepse associada a delirium, utilizando um método de classificação e testes baseado na técnica de inteligência artificial conhecida como Máquina de Suporte Vetorial.

## 1.1 OBJETIVO GERAL

Disponibilizar um simulador computacional baseado em MSV para a predição da probabilidade de óbitos de pacientes sépticos associados a delirium.

## 1.2 OBJETIVOS ESPECÍFICOS

Para o desenvolvimento deste projeto de pesquisa foram selecionados os seguintes objetivos:

- a) conhecer o conceito de máquinas de suporte vetorial;
- b) conhecer os conceitos de sepsemia e delirium;
- c) aplicar o método de treinamento e teste *Leave-One-Out*;
- d) aplicar os algoritmos de MSV e RNA para a predição dos dados;
- e) analisar os dados resultantes por meio de medidas de análise de conhecimento;
- f) comparar os resultados obtidos com MSV e RNA.

### 1.3 JUSTIFICATIVA

A sepse é doença que vem desafiando a medicina, devido aos esforços feitos para a sua compreensão (HENKIN et al, 2009). De acordo com Zampieri et al (2011, tradução nossa), o reconhecimento e o diagnóstico precoce dos casos de sepse, poderiam prevenir as complicações geradas por ela, reduzindo, desta forma, a taxa de mortalidade de pacientes sépticos em UTI.

Apesar de sua importância e da demanda de recursos, seu reconhecimento muitas vezes ainda não ocorre em tempo hábil, deixando margem para a ocorrência de disfunção de múltiplos órgãos e sistemas (HENKIN et al, 2009).

Devido ao número expressivo de óbitos em UTI decorrentes de casos de sepse associada a delirium, identifica-se a necessidade de conhecer a probabilidade destes óbitos através da classificação de dados em MSV.

Segundo Mendoza (2009), os resultados obtidos através da MSV são comparáveis e muitas vezes superiores a outros algoritmos de aprendizado, como as RNAs.

Desta maneira, pretende-se aplicar o método de treinamento e teste *Leave-One-Out* e analisar os dados resultantes através da matriz de confusão e de outras medidas de análise de conhecimento, como a Curva Característica de Operação de um Receptor (ROC), que possibilita a visualização das taxas do classificador.

Com o método *Leave-One-Out* é possível treinar e testar uma base de dados a partir de um modelo gerado pela MSV e encontrar um padrão para estes dados. Com os resultados gerados, compõe-se a matriz de confusão. Além da aplicação do método *Leave-One-Out*, também será utilizado um outro método para avaliação do desempenho do classificador chamado *k-fold*. Este método, assim como o *Leave-One-Out* também treina e testa uma base de dados.

A partir dos mesmos dados, também serão utilizados dois algoritmos de RNA, MLP e RBF. Estes algoritmos serão utilizados para que seja possível comparar os resultados obtidos por eles com os resultados dos algoritmos de MSV, que foram

SMO e LIBSVM. Estes algoritmos foram utilizados no trabalho em função de os mesmos serem disponibilizados na ferramenta Weka, utilizada para mineração destes dados.

Com a aplicação destes métodos e da técnica da MSV, pretende-se realizar a predição destes óbitos por sepse associada a delirium, com base nos dados coletados nas Unidades de Tratamento Intensivo.

#### 1.4 ESTRUTURA DO TRABALHO

O trabalho é composto por sete capítulos. O primeiro capítulo tem como propósito mostrar qual é o tema do projeto e os objetivos selecionados para serem atingidos com o desenvolvimento dele. A definição do problema encontrado e justificativa para a escolha de tal projeto também são explorados.

A abordagem dos conceitos e da epidemiologia da sepse associada a delirium são apresentados no capítulo dois. Pretende-se com ele, conceituar a sepse e relatar a incidência da mesma, auxiliando na compreensão do projeto proposto.

O objetivo do terceiro capítulo é apresentar definições referentes ao aprendizado de máquina, enfatizando as Máquinas de Suporte Vetorial (MSV). Neste capítulo é abordada a utilização de MSV, os tipos de classificações possíveis e o uso das principais funções de kernel disponíveis na MSV na base de dados de exemplo. O funcionamento das Redes Neurais Artificiais (RNA) também pode ser visto no mesmo capítulo.

O capítulo quatro apresenta os métodos de treinamento e teste utilizados no trabalho, demonstrando a maneira com que eles foram aplicados nos dados. No capítulo cinco são apresentados alguns exemplos de utilização de MSV em diferentes áreas.

As etapas do trabalho desenvolvido são listadas no capítulo seis. No mesmo capítulo é relatado, por meio de etapas, cada parte desenvolvida. O trabalho realizado também é mostrado, juntamente com os resultados obtidos por meio dele.

No capítulo sete apresenta-se a conclusão da pesquisa feita e sugestões para trabalhos futuros.

## **2 SEPSE**

Sepse é uma resposta do organismo para uma infecção. Desenvolve-se quando uma infecção toma grandes proporções e acaba ficando desregulada (COHEN, 2002, tradução nossa). Em 1991, na Conferência de Consenso sobre Sepse, ela foi definida como uma síndrome de resposta inflamatória sistêmica com uma fonte de infecção (DANIELS; NUTBEAM, 2009, tradução nossa). Conforme descrevem Carvalho e Trotta (2003), a sepse possui diferentes maneiras de se manifestar, e pode ocasionar a disfunção ou falência de órgãos.

No século XIX já haviam sido identificados casos de sepse, porém não existia nada comprovado até então, apenas pesquisas e resultados científicos que eram vistos com ceticismo por parte da comunidade médica. Somente na metade da

década de 90, foram encontradas na literatura as características vistas em pacientes sépticos (SILVA; VELASCO, 2007).

Existem diferentes denominações para a sepse, conforme suas características. Quando a sepse se manifesta juntamente com disfunção orgânica ou hipotensão, é chamada de sepse grave. Choque séptico é o nome dado aos casos em que há a sepse grave e o paciente desenvolve também, hipotensão arterial, sendo preciso fazer uso de drogas vasopressoras para a regulação da pressão arterial (BAUDOUIN, 2009, tradução nossa).

Dentre os sintomas identificados da sepse, pode-se citar a febre, confusão mental e a diminuição da produção da urina (COHEN, 2002, tradução nossa).

Conforme destaca Brito e Zuse (2010), é necessário um atendimento preciso, eficaz e imediato para pacientes sépticos. O controle precoce pode reduzir sua taxa de mortalidade de 10 a 15%, comparados a pacientes que não recebem um tratamento rápido e correto (SILVA; VELASCO, 2007). A falta de tratamento adequado acarreta, segundo Cohen (2002, tradução nossa) diversas disfunções, como anormalidades de coagulação e insuficiência respiratória ou renal, como também, de acordo com Carvalho e Trotta (2003), pode ocasionar falência orgânica, choque séptico ou a própria morte do paciente.

A sepse possui uma elevada taxa de mortalidade fazendo com que seja um grande desafio em todo o mundo (SALES JÚNIOR et al, 2006). Atualmente, ela vem sendo uma fonte de preocupação e vem demandando muita atenção, já que é a principal causa de mortes nas UTIs (PENINCK; MACHADO, 2012). Além da alta mortalidade associada a ela, dos valores despendidos para manutenção destes pacientes nos hospitais, ainda há as sequelas deixadas por ela, provocando a diminuição da qualidade de vida dessas pessoas (REINKE; RITTER; DAL PIZZOL, 2004).

## 2.1 EPIDEMIOLOGIA

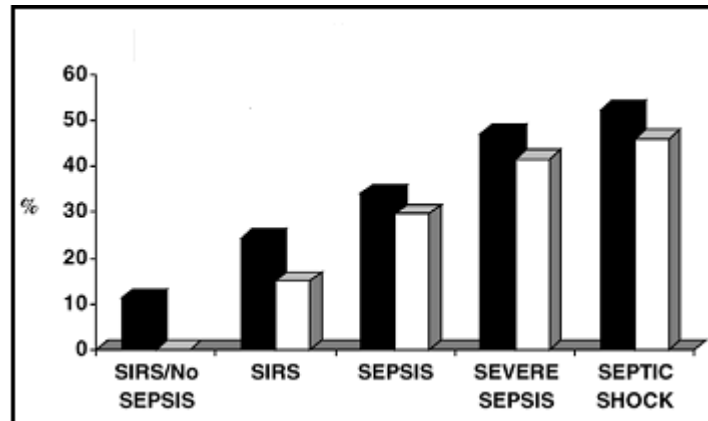
Apesar de avanços no diagnóstico da doença, do uso de métodos que facilitam o início do tratamento, ainda não se tem total compreensão desta doença (SALES JÚNIOR et al, 2006).

Conforme os estudos realizados por Angus et al (2001, tradução nossa), nos Estados Unidos a incidência de casos com sepse está na faixa de 750 mil por ano, sendo que aproximadamente metade deles receberam tratamentos em UTIs. Ainda com base nestes estudos, foi calculada uma taxa de mortalidade de aproximadamente 30% em pacientes normais, correspondendo a mais de 215 mil mortes no país em decorrência de sepse. A partir de dados coletados de hospitais dos Estados Unidos, durante os anos de 1979 a 2000, chegou-se a mais de 10 milhões de casos de sepse nesse período, com aumento de 8,7% ao ano (MARTIN et al, 2000, tradução nossa).

Na Europa, os estudos feitos indicaram uma incidência de sepse em 35% dos pacientes internados nas UTIs. A taxa de mortalidade foi de 27%, elevando para 50% nos casos de choque séptico (VINCENT et al, 2006).

Durante um ano foram coletados dados de internações de pacientes de dois estados brasileiros, São Paulo e Santa Catarina. A incidência de sepse foi de quase 60%, com os números de óbitos variando entre 24,2% e 52,2% conforme o caso (SILVA et al, 2004, tradução nossa). De acordo com estes estudos, pode-se ilustrar os dados obtidos demonstrando as taxas de mortalidade resultantes (figura 1).

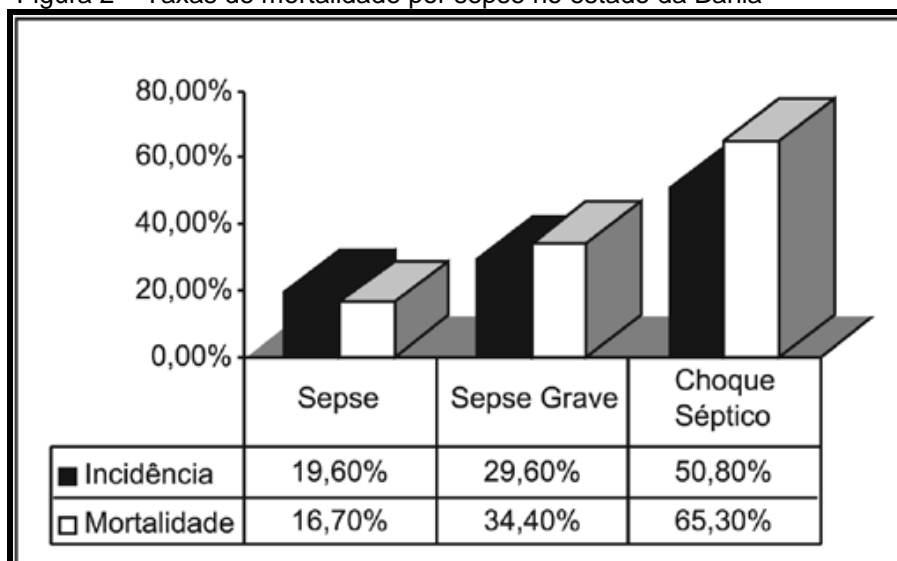
Figura 1 – Dados referentes a incidência de sepse



Fonte: Silva et al (2004).

Outro estudo envolvendo setenta e cinco UTIs comprovou o elevado número de óbitos de pacientes sépticos no Brasil. As taxas de mortalidade obtidas variam de aproximadamente 16% a 65%, como apresenta a figura 2 (SALES JÚNIOR et al, 2006). De acordo com Juncal et al (2011), baseado nos estudos realizados, no estado da Bahia foram vistos que quase metade das internações na UTI eram decorrentes de casos de sepse.

Figura 2 – Taxas de mortalidade por sepse no estado da Bahia



Fonte: Juncal et al (2011).

Tendo em vista todas as pesquisas realizadas e o número expressivo de casos e óbitos consequentes de sepse, é fundamental que se procure maneiras de

se agilizar e tornar mais preciso o processo de identificação da doença, para evitar que esses números cresçam ainda mais.

## 2.2 SEPSE ASSOCIADA A *DELIRIUM*

De acordo com Association (2000, tradução nossa), delirium é uma perturbação de consciência associada a uma alteração na cognição, desenvolvendo-se em um curto período de tempo.

Conforme Kress (2010, tradução nossa) é comum encontrar pacientes sépticos com disfunções cerebrais. Mesmo sendo frequente, muitas vezes não se atribui ao delirium a devida importância (SONNEVILLE et al, 2013, tradução nossa). Segundo afirma Ebersoldt, Sharshar e Annane (2007, tradução nossa), ainda não há um tratamento para sepse associada a delirium.

Em estudos feitos por Martin et al (2010, tradução nossa), ficou evidente a associação da sepse ao *delirium*. Em 30,8% dos pacientes com sepse que foram submetidos ao estudo, foi diagnosticado o delirium, o que indicou para ele, que o delirium pode ser um precedente da sepse.

A taxa de mortalidade em hospitais e UTI é elevada quando os pacientes estão em delirium (ZAMPIERI et al, 2011, tradução nossa). Delirium geralmente é uma manifestação de sepse, e de acordo com Papadopoulos et al (2000, tradução nossa), esses pacientes têm maior mortalidade do que os que não apresentam esse sintoma.

### 3 APRENDIZADO DE MÁQUINA

Aprendizado de Máquina é uma área da Inteligência Computacional em que são estudadas formas de extração de conhecimento a partir de dados (LIBRALAO et al, 2005).

Os métodos de aprendizado de máquina que são utilizados, buscam obter formas de reconhecimento de padrões para um conjunto de dados, baseando-se na indução (LORENA; CARVALHO, 2007). Essas técnicas tornam-se importantes, pois podem solucionar tarefas que as programações normalmente utilizadas, não conseguem solucionar (SANTOS, 2002).

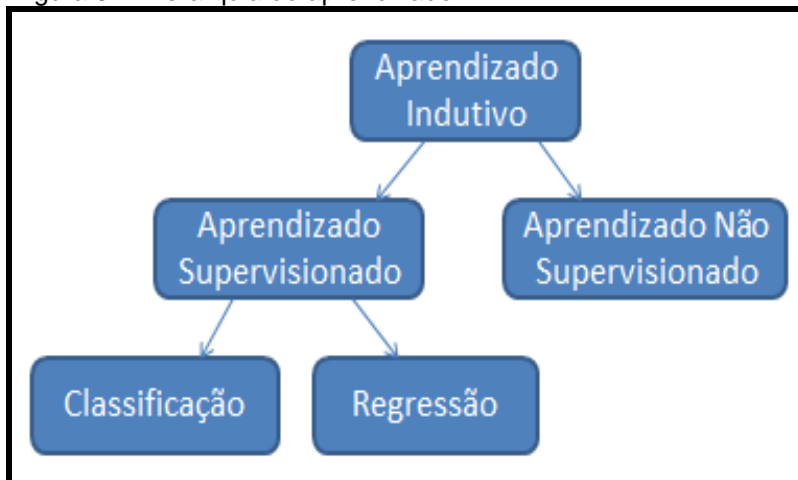
A aprendizagem de máquina pode ocorrer de duas maneiras: supervisionada e não supervisionada (figura 3). Na aprendizagem de máquina supervisionada, conforme descreve Lorena e Carvalho (2007), existem dados de entrada para auxílio no treinamento e tem como finalidade, indicar uma saída correta para outras entradas não reconhecidas previamente. O objetivo deste aprendizado é a indução de um mapeamento de valores, que origine um modelo em que seja possível prever novos valores (BATISTA, 2003).

De acordo com Batista (2003), o aprendizado supervisionado possui alguns paradigmas:

- a) simbólico: este paradigma procura aprender a partir da construção de representações simbólicas de um conceito de uma análise de exemplos. Tendo como exemplo a árvore de decisão;
- b) estatístico: as técnicas estatísticas procuram focar em tarefas em que os atributos possuam valores contínuos. Busca-se utilizar modelos estatísticos para conseguir uma aproximação do conceito induzido;
- c) *instance-based*: neste paradigma, utiliza-se um exemplo mais próximo para realizar a classificação;
- d) conexionista: relacionado com a forma como o cérebro humano se conecta. Busca-se uma aproximação de funções por meio de uma regressão não-linear.

Já o aprendizado não supervisionado, não possui dados de entrada já rotulados. Ele aprende a representação desses dados, com base em medidas de qualidade.

Figura 3 – Hierarquia do aprendizado



Fonte: Batista (2003).

Uma forma de aprendizado supervisionado é justamente a classificação de dados, que será utilizada no trabalho desenvolvido.

### 3.1 CLASSIFICAÇÃO DE DADOS

A mineração de dados, *Data Mining*, procura resolver problemas com base em dados já existentes, através da descoberta de padrões nos dados (WITTEN; FRANK; HALL, 2011). É uma área de pesquisa que inclui técnicas de banco de dados, estatística, inteligência artificial, aprendizado de máquina, entre outras (CARDOSO; MACHADO, 2008). Ela possui duas tarefas principais: Tarefas de Previsão e Tarefas Descritivas. A primeira procura chegar ao valor de um atributo a partir de outro já existente. Já a outra, deriva padrões, estabelece relacionamentos e tendências (TAN; STEINBACH; KUMAR, 2009). Com base nessas análises, é possível aplicar as técnicas de *data mining* em diversas áreas, como telefonia, educação, finanças e também, na área médica (GOLDSCHMIDT; PASSOS, 2005).

Na mineração de dados busca-se gerar novas informações a partir dos dados que já existem. Sendo que estas novas informações quando interpretadas, devem poder ser aplicadas na prática. Essas aplicações podem variar desde a gestão de relacionamento com cliente, mineração de texto até a prevenção de doenças (STEINWART; CHRISTMANN, 2008, tradução nossa).

Para criar um modelo que faça a predição de novos dados, com base nos já existentes, é necessário utilizar um algoritmo que realize a classificação dos dados. A escolha deste algoritmo é relacionada com a eficiência que o mesmo obtém quando aplicado nos dados (BATISTA, 2003).

Existem maneiras para realizar a classificação de dados. Pode-se utilizar Classificadores Bayesianos, usados quando o relacionamento entre os atributos e a variável classe não é determinístico. Existem as RNAs, criada em tentativas de simular a estrutura do cérebro humano e há também a MSV (TAN; STEINBACH; KUMAR, 2009).

Tanto as RNAs quanto as MSV serão utilizadas no trabalho desenvolvido.

### 3.2 MÁQUINA DE SUPORTE VETORIAL

A Máquina de Suporte Vetorial é um método de aprendizado supervisionado que gera funções para treinamento e classificação de dados (WANG, 2005, tradução nossa). Foi desenvolvida por Vladimir Vapnik e colaboradores, e desde a sua criação vem se destacando nas situações em que é aplicada (CRISTIANINI; SHAW-TAYLOR, 2000, tradução nossa).

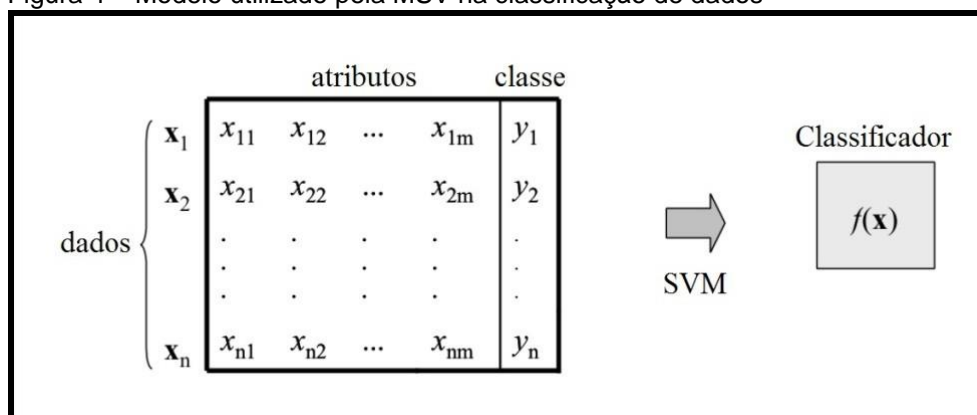
Conforme explica Campbell e Yiming (2011, tradução nossa), a MSV é uma máquina de aprendizado abstrata, que utiliza um conjunto de dados para treinamento e a partir deste treinamento, tenta generalizar e prever, de forma correta, o comportamento do restante dos dados.

A MSV baseia-se nas Teorias de Aprendizagem Estatística e Otimização Matemática. A teoria de Aprendizagem Estatística no uso da MSV procura controlar a generalização feita por ela, ou seja, a classificação dos dados de maneira precisa. Já a Teoria da Otimização Matemática procura encontrar soluções a partir de

funções dadas previamente (CRISTIANINI; SHAWE-TAYLOR, 2000, tradução nossa).

De acordo com Ventura (2007), as MSVs utilizam a aprendizagem estatística e a indutiva, onde é possível encontrar uma forma genérica para um determinado grupo de dados. Utilizando esse conceito, caso tenha-se, por exemplo, um conjunto de dados no formato  $(x_i, y_i)$ , no qual  $x_i$  indica o dado (atributo) e  $y_i$  a sua classificação (classe), a MSV deve produzir um classificador que consiga reconhecer a qual formato pertence os dados submetidos, constituindo desta maneira o treinamento desses dados (figura 4).

Figura 4 – Modelo utilizado pela MSV na classificação de dados



Fonte: Ventura (2007).

Na classificação, pode ocorrer situações em que mesmo uma boa taxa de acerto não indica que a função irá generalizar bem o restante dos exemplos. Nos casos em que o classificador não tem capacidade de generalizar, acontece o chamado *overfitting* (sobre ajuste), e também existe o inverso, denominado *underfitting* (sub-ajuste), que é quando a função não tem uma complexidade suficiente para o problema (OLIVEIRA JUNIOR, 2010).

Para adquirir do classificador obtido uma taxa de predição correta, divide-se o grupo de dados em um conjunto de teste e um de treinamento. No conjunto de treinamento, os dados servirão apenas para o aprendizado da máquina. Já no conjunto de teste, os dados serão usados para verificação do aprendizado,

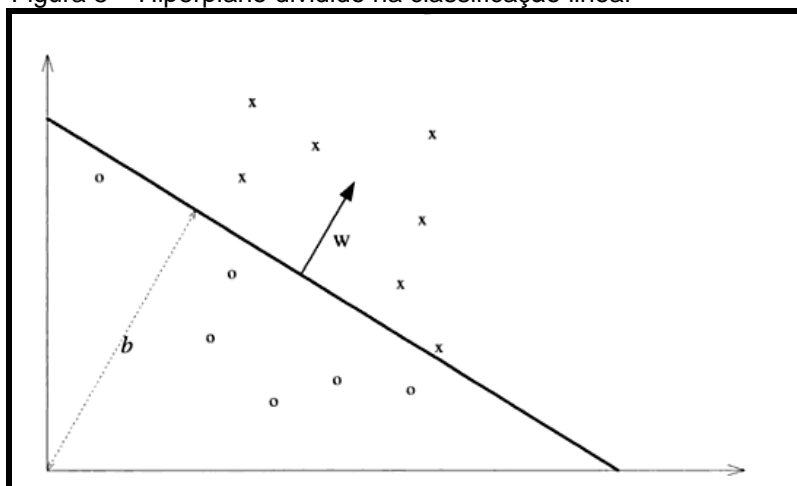
indicando se a predição de novos dados será realizada de forma correta (LORENA; CARVALHO, 2007).

Conforme destaca Ventura (2007), a MSV ultimamente está recebendo grande atenção, devido aos resultados que são obtidos com o seu uso, visto que em alguns casos podem ser superiores aos obtidos com o uso de outros algoritmos de aprendizagem, como as RNAs. Os algoritmos de RNA são baseados na minimização do risco empírico, que é a minimização do erro de treinamento, enquanto a máquina de suporte vetorial baseia-se também na minimização do erro estrutural, que refere-se a minimização do erro de generalização que pode ocorrer na classificação dos dados. Isso acaba tornando a MSV mais atrativa e mais completa (SANTOS, 2002).

### 3.2.1 Classificação Linear

Em uma máquina de suporte vetorial utilizada em problemas linearmente separáveis, tem-se o conjunto de dados de treinamento que é dividido em duas partes em um hiperplano (figura 5). Neste modelo, além da separação do conjunto, é possível identificar os vetores de suporte utilizados (CRISTIANINI; SHAWE - TAYLOR, 2000, tradução nossa).

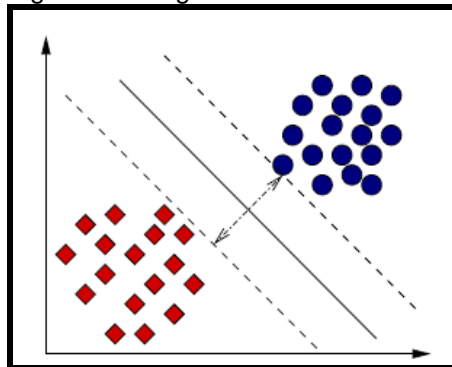
Figura 5 – Hiperplano dividido na classificação linear



Fonte: Cristianini e Shawe-Taylor (2000, tradução nossa).

Nesta classificação, busca-se a produção de um classificador genérico, que funcione bem para todas as situações. Dentre todos os classificadores possíveis, há aquele que aumenta a distância entre os classificados e os pontos mais próximos dos dados de cada subconjunto dividido. Este classificador é chamado de Hiperplano Ótimo de Separação (figura 6). Eles usam os vetores de suporte para encontrar essa maior distância procurada (MENDOZA, 2009).

Figura 6 - Margem máxima e vetores de suporte.



Fonte: Santos (2002).

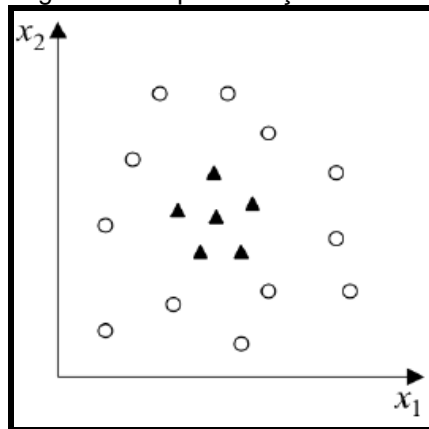
Os vetores de suporte são os pontos das classes mais próximos ao separador. Eles são utilizados pelos algoritmos de treinamento para a geração da classificação dos dados (ALBUQUERQUE, 2011).

Para conjuntos de dados linearmente separáveis, as MSVs lineares obtêm eficiência em seus resultados. Mas existem situações em que não é possível separar estes dados de treinamento em um hiperplano. Este tipo de classificação denomina-se classificação não-linear.

### 3.2.2 Classificação Não-Linear

Na classificação não-linear, mostrada na figura 7, segue-se o teorema de Cover, que afirma que um conjunto de dados não-lineares quando transformados em um espaço de maior dimensão chamado espaço de características, é possível encontrar características nestes dados que os tornam linearmente separáveis (LORENA; CARVALHO, 2007).

Figura 7 – Representação da classificação não-linear



Fonte: Lorena e Carvalho (2007).

Sendo assim, observa-se que o fator que diferencia estas duas classificações é a possibilidade de encontrar o hiperplano ótimo de separação, utilizado pela MSV para classificação de dados, sem a necessidade de outros métodos, como o uso de um espaço de características.

Para realizar a classificação de dados não-lineares, utiliza-se as funções de kernel.

### 3.2.3 Funções de Kernel

As funções de Kernel usam uma projeção de dados para que seja possível a realização da classificação de dados não-linearmente separáveis (SANTOS, 2002). Conforme Cristianini e Shawe-Taylor (2000, tradução nossa) a escolha da função de kernel na MSV é fundamental, pois deve-se escolher a que melhor atende as necessidades identificadas.

Dentre as mais usadas, tem-se: Polinômios, Funções de Base Radial Gaussiana e Rede Neural Sigmóide de duas camadas (SANTOS, 2002).

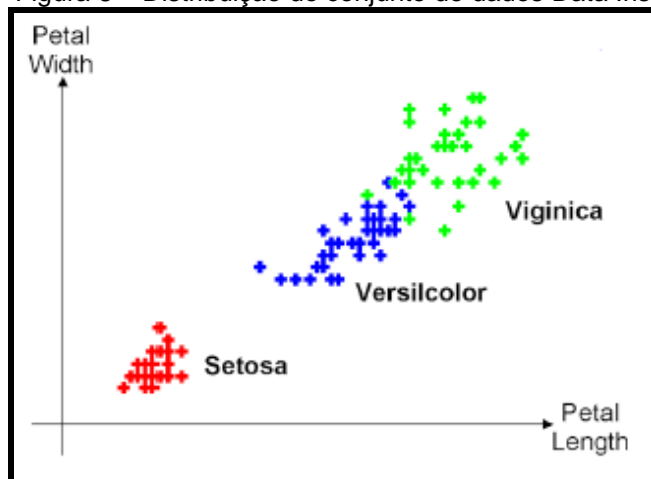
### 3.2.3.1 Utilização das Funções de Kernel na Base de Dados Data Iris

Para demonstrar o uso das funções de kernel para separação de classes na MSV, utilizou-se como exemplo as demonstrações retiradas de outros trabalhos, em que foi utilizada a base de dados Iris

A base de dados Iris é muito utilizada para classificação de dados. Este conjunto de dados contém três classes (espécies de gênero Iris), com cinquenta casos por cada classe (LEE; VERRI, 2002, tradução nossa). A partir dela, é possível avaliar desempenhos de algoritmos desenvolvidos para classificação. Ela é composta por quatro características (comprimento e largura da pétala e comprimento e largura da sépala) (figura 8) (GUNN, 1998, tradução nossa).

Conforme destaca Lima (2004), a primeira vez em que se utilizou as MSVs juntamente com este conjunto de dados, foi com o objetivo de avaliar a generalização diferentes configurações.

Figura 8 – Distribuição do conjunto de dados Data Iris

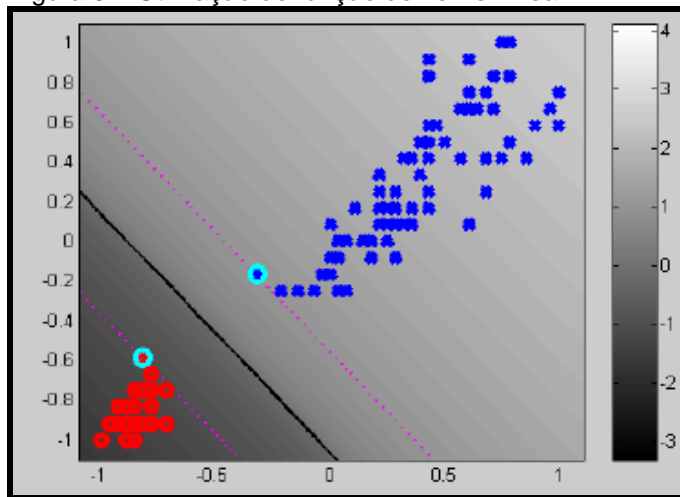


Fonte: Lima (2004).

De acordo com o mostrado por Lima (2004), foi possível a visualização da aplicação destas funções na base de dados. A seguir, serão mostrados os exemplos das funções, com base nos resultados obtidos por ele.

Para a separação das classes Setosa e Versicolor, pode-se utilizar uma classificação linear, como mostra a figura 9.

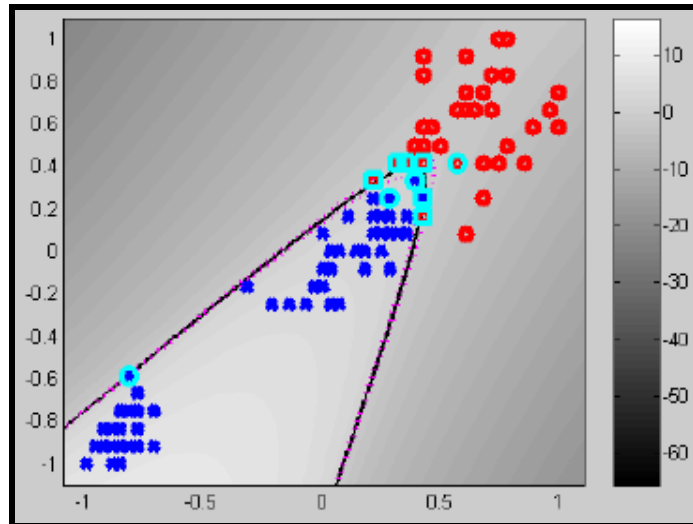
Figura 9 – Utilização de função de kernel linear.



Fonte: Lima (2004).

Na separação da espécie Virginica das demais, não é possível o uso de uma função de kernel linear, já que pode haver nas outras espécies valores iguais para os comprimentos das pétalas. Na figura 10, foi usada a função de kernel Polinomial.

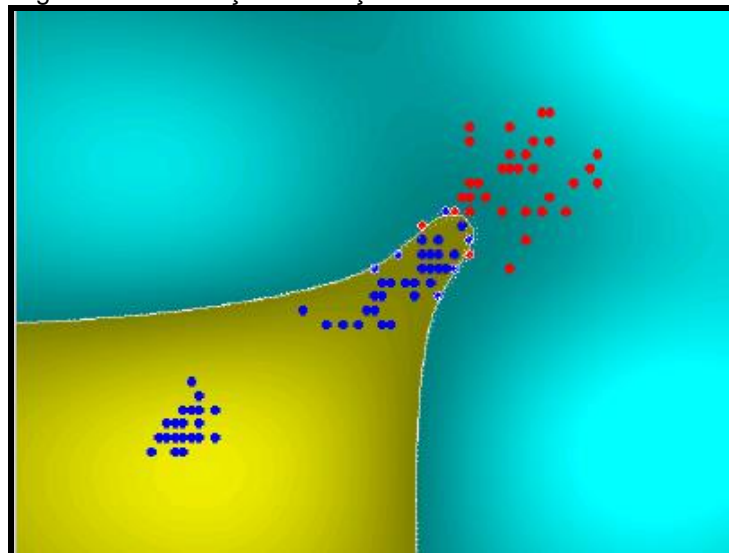
Figura 10 – Utilização de função de kernel polinomial.



Fonte: Lima (2004).

Com base nos resultados encontrados por Gunn (1998, tradução nossa), tem-se mais exemplos de uso de outras funções de kernel. Utilizando uma função de kernel de Base Radial Gaussiana, observa-se um resultado semelhante ao encontrado utilizando uma função polinomial (figura 11).

Figura 11 – Utilização da função de kernel de Base Radial Gaussiana



Fonte: Gunn (1998, tradução nossa).

Conforme os exemplos apresentados consegue-se verificar a aplicação das funções de kernel na classificação de dados. Pode-se notar também, que

quando utilizadas em dados que não são linearmente separáveis, são elas que desempenham esta função de separação.

### 3.3 REDES NEURAIIS ARTIFICIAIS

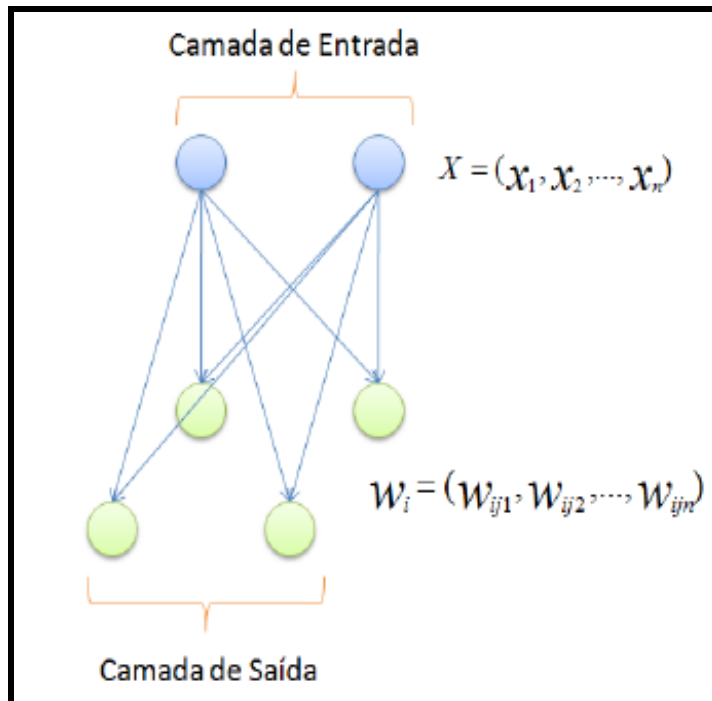
O estudo das Redes Neurais Artificiais partiu da tentativa de imitar o cérebro humano, que consiste principalmente nas células nervosas denominadas neurônios, que são ligados por meio de axiônios. Estes, transmitem impulsos nervosos de neurônio em neurônio (TAN; STEINBACH; KUMAR, 2009).

Haykin (2002) define uma rede neural como um processador que é composto por unidades de processamento, que têm como habilidade o armazenamento do conhecimento gerado.

De acordo com Moll (2010), as redes neurais são muito utilizadas em tarefas de aproximação de funções e classificação e reconhecimento de padrões. Elas fazem parte do paradigma conexionista, mostrado anteriormente.

A RNA consiste basicamente em neurônios que recebem diretamente entradas da rede, chamando assim de camada de entrada. Cada camada recebe como entrada a saída da outra camada até chegar a camada de saída. As camadas que não são nem de saída nem de entrada, são denominadas camadas ocultas (figura 12) (KOVÁCS, 2006).

Figura 12 – Arquitetura de uma rede neural simplificada



Fonte: Moll (2010).

Além dos neurônios e das camadas, as redes são compostas também por pesos, conforme destaca Moll (2010). Esses pesos representam as conexões entre os neurônios de cada camada. Os pesos inicialmente são definidos aleatoriamente. No decorrer do treinamento, com a detecção dos erros, a rede pode ir voltando camada por camada e corrigindo os pesos (CERQUEIRA et al, 2001).

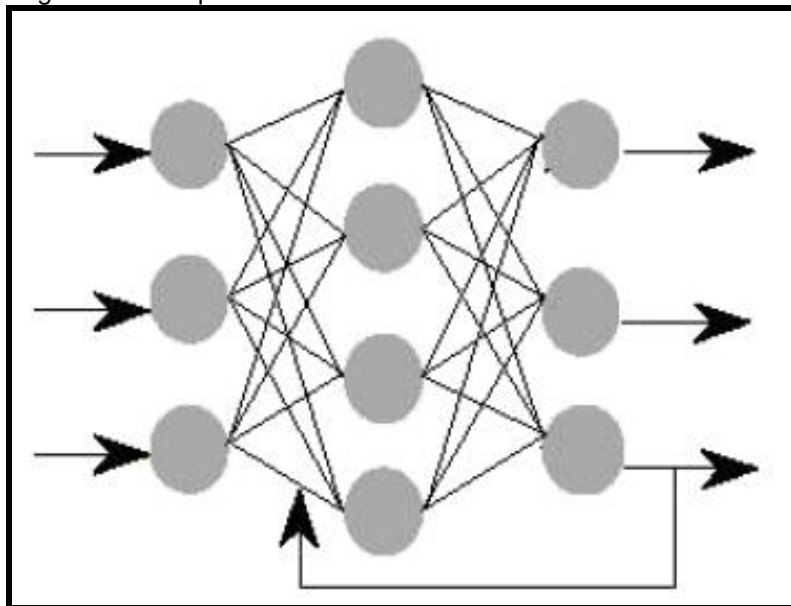
Existem dois tipos de redes neurais: supervisionado e não supervisionado. Quando é não supervisionado, a rede final deve fazer as representações internas somente com base nos padrões apresentados. Já nas supervisionadas, já se tem a informação de quais classes devem ser escolhidas como vetores de entrada, como exemplo, tem-se a *Multilayer Perceptron* (MLP) e a *Radial Bases Function* (RBF) (MOLL, 2010). Ambas as redes foram utilizadas na classificação realizada no trabalho desenvolvido.

### 3.3.1 *Multilayer Perceptron*

As MLP têm sido aplicadas na resolução de problemas com alto grau de dificuldade, devido a utilização do algoritmo de *backpropagation* (HAYKIN, 2002).

Uma MLP é formada por uma camada de neurônios de entrada, uma ou mais camadas intermediárias e a camada de saída. Quando se utiliza uma técnica chamada *backpropagation*, se envia um sinal no sentido inverso, para que seja possível fazer a rede aprender com o erro (figura 13) (RAMALHO, 2008).

Figura 13 – Esquema de uma RNA MLP



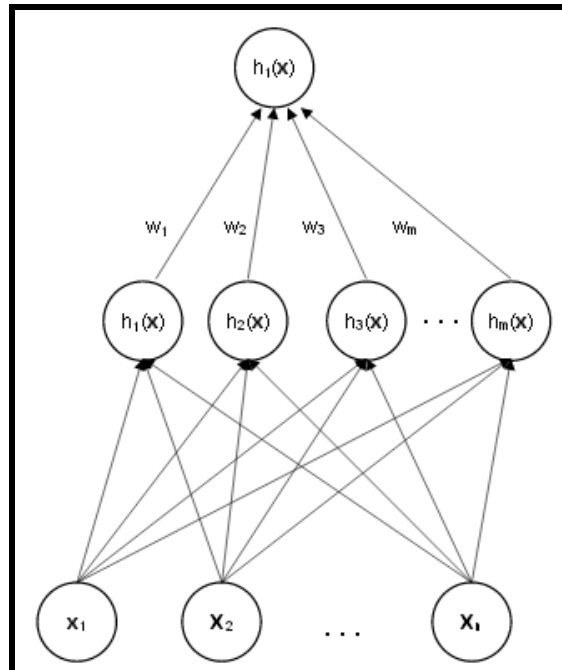
Fonte: Ramalho (2008).

Conforme afirma Bonifácio (2010), o uso de camadas ocultas juntamente com a capacidade de aprender com o próprio treinamento, torna a MLP muito mais poderosa computacionalmente que os demais tipos de RNA.

### 3.3.2 Radial Bases Function

As *Radial Bases Function* (RBF) são constituídas por apenas uma camada de entrada, uma camada intermediária e uma camada de saída (figura 14).

Figura 14 – Esquema de uma RNA RBF



Fonte: Cerqueira et al (2001).

Cada camada da rede RBF possui uma função bem definida. A camada de entrada serve para agrupar os dados, tornando os dados não linearmente separáveis em linearmente separáveis. Já a camada de saída, busca classificar a camada anterior (BONIFÁCIO, 2010).

Tanto MSV quando RNA serão utilizadas no projeto de pesquisa para que seja possível estabelecer uma comparação entre estas duas técnicas de aprendizagem de máquina.

#### 4 TREINAMENTO E CLASSIFICAÇÃO COM MSV

Para simular o processo de treinamento e teste de uma máquina de suporte vetorial, foram atribuídos valores a três variáveis,  $v_1, v_2$  e  $v_3$ . Sendo que estas variáveis poderão integrar duas classes, as classes A e B. Esta estrutura pode ser visualizada na tabela 1.

Com base nestes valores e através da MSV, será possível realizar a predição de outros dados pertencentes a esta estrutura, mas que ainda não possuíam classificação.

Tabela 1 – Exemplo de dados para classificação

|    | var1 | var2 | var3 | Classificação |
|----|------|------|------|---------------|
| V1 | 1    | 1.5  | 2    | A             |
| V2 | 1.1  | 1.8  | 1.9  | A             |
| V3 | 3    | 2.5  | 4    | B             |

Fonte: Do autor.

Na tabela apresentada, são mostrados os valores das variáveis em cada vetor de suporte encontrado, v1, v2 e v3, e a classificação que ela tem.

Os vetores de suporte variam conforme a característica de cada variável utilizada. A figura 15 apresenta os vetores que foram utilizados no exemplo mostrado.

Figura 15 – Exemplo de vetores de suporte

|     |     |     |
|-----|-----|-----|
| 1   | 1.5 | 2   |
| 1.1 | 1.8 | 1.9 |
| 3   | 2.5 | 4   |

Fonte: Do autor.

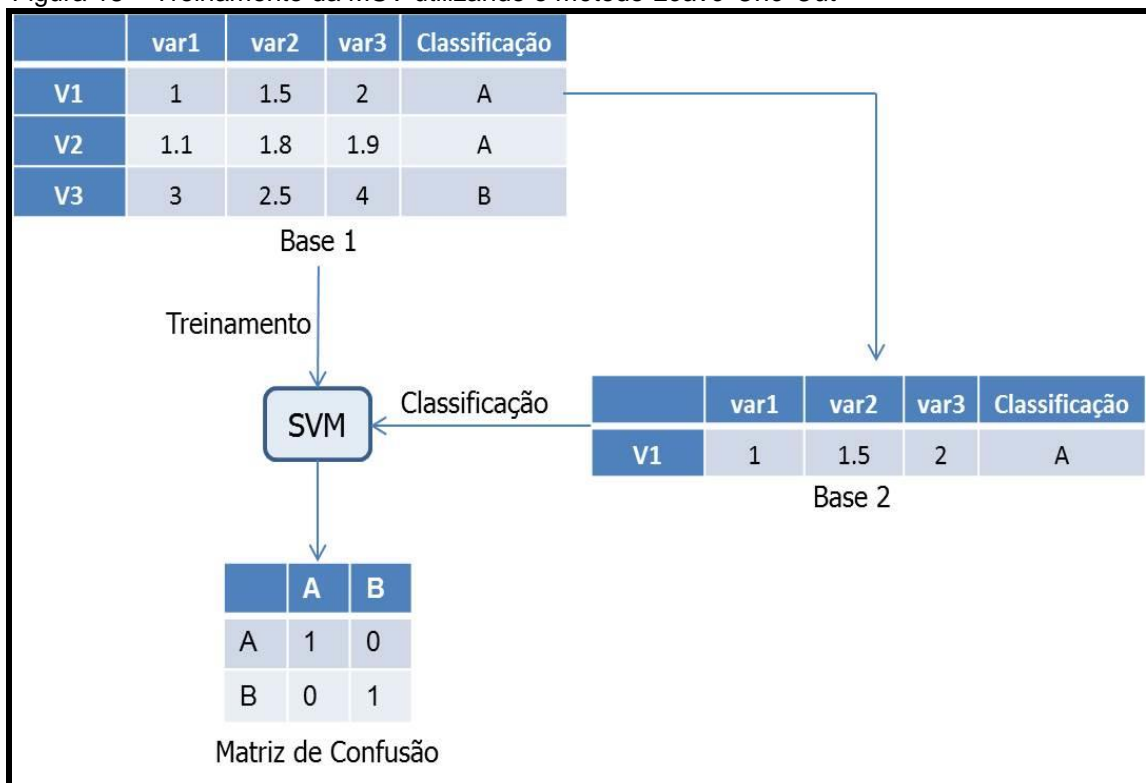
Os valores atribuídos aos vetores de suporte serão utilizados no treinamento do modelo da MSV. Para realizar o treinamento e a classificação destes dados, serão utilizados os métodos *Leave-One-Out* e *k-fold*. Ao término do processo de classificação, será obtida uma matriz de confusão e um valor para curva ROC que serão utilizados para analisar os resultados.

#### 4.1 MÉTODO *LEAVE-ONE-OUT*

Este método foi proposto em 1999 por Weston, e baseia-se em margens adaptativas, enquanto as MSV utilizam margens fixas para cada padrão de treinamento. O procedimento feito neste método consiste na remoção de um conjunto de treinamento, treinar a máquina e realizar os testes (WESTON, 1999, tradução nossa).

Utilizando o exemplo mostrado, a figura 18 descreve o processo de treinamento da MSV utilizando o método *Leave-One-Out*.

Figura 18 – Treinamento da MSV utilizando o método *Leave-One-Out*



Fonte: Do autor.

Como é possível observar na figura 18, a base de dados 1 é utilizada para o treinamento da MSV. Já a base 2 é utilizada para o teste.

##### 4.1.1 Base de Treinamento

A base de treinamento é composta por todos os vetores menos o utilizado para o teste. Ela é utilizada para treinar a MSV, e produzir, com base no teste, os pontos que serão inseridos na matriz de confusão.

Utilizando como exemplo de predição a figura 18 e supondo que o vetor testado será o v1, a base de treinamento irá conter os vetores v2 e v3, conforme ilustra a figura 19.

Figura 19 – Exemplo de Base de Treinamento

|    | var1 | var2 | var3 | Classificação |
|----|------|------|------|---------------|
| V2 | 1.1  | 1.8  | 1.9  | A             |
| V3 | 3    | 2.5  | 4    | B             |

Fonte: Do autor.

Conforme o método *Leave-One-Out* orienta, ao retirar o vetor v1 da base de treinamento, deve-se inseri-lo na base de teste. Este processo é realizado em todos os vetores da base de dados.

#### 4.1.2 Base de Teste

A base de teste irá conter o vetor utilizado para o teste da MSV. Este vetor servirá para testar a MSV treinada com os vetores restantes, os inseridos na base de treinamento, e resultará em pontos para a classe encontrada que serão armazenados na matriz de confusão. Após a classificação, o vetor usado para o teste é inserido novamente na base de treinamento e outro vetor é utilizado para o teste, para reiniciar o processo de treinamento da MSV.

Com o término do processo, será obtida uma matriz de confusão com os números de acertos encontrados durante a realização do método (figura 20).

Figura 20 – Exemplo de matriz de confusão

|    | C1 | C2 | C3 | C4 | C5 | ← Classes                     |
|----|----|----|----|----|----|-------------------------------|
| C1 | 5  | 0  | 0  | 0  | 0  |                               |
| C2 | 0  | 5  | 0  | 0  | 0  |                               |
| C3 | 0  | 0  | 4  | 1  | 0  |                               |
| C4 | 0  | 1  | 2  | 3  | 0  |                               |
| C5 | 0  | 0  | 1  | 3  | 1  | ← Número de acertos da classe |

↑ Classes

↑ Erro do Classificador

Fonte: Do autor.

A figura 20 apresenta um exemplo de uma matriz e confusão. A partir do que é gerado por ela, pode-se verificar as informações da classificação, como o número de acertos, número de erros e as classes em que tiveram esses erros.

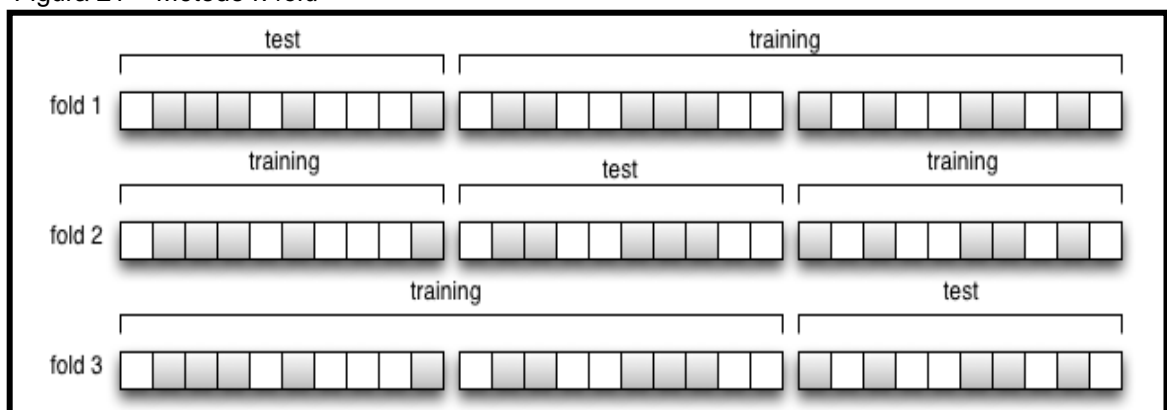
Para compor essa matriz, utiliza-se o primeiro vetor de suporte e a classe em que foi classificado. A classe irá receber um ponto. Isso é repetido  $n$  vezes, sendo que  $n$  corresponde ao número da amostra.

Com esta demonstração, é possível ver a aplicação do método *Leave-One-Out* no treinamento e no teste de dados de uma MSV.

## 4.2 MÉTODO *K-FOLD*

O método *k-fold* particiona um determinado conjunto em  $k$  partes. Durante o treinamento, uma das partes é utilizada para teste, enquanto as outras são utilizadas no treinamento (TAN; STEINBACH; KUMAR, 2009) (figura 21).

Figura 21 – Método *k-fold*



Fonte: Martin (2014).

De acordo com o que foi exemplificado na figura 21, existem três partes, cada uma contém dez registros. O número escolhido para o  $k$  foi três, determinando assim, que durante três execuções, será retirado uma parte para a realização do teste e as outras duas para o treino. No término das três execuções, o método *k-fold* terá sido aplicado na base de dados.

Pode-se então, verificar todo o processo que precisará ser executado para encontrar um padrão em um conjunto de dados.

Foi apresentado também, como será feito o tratamento dos dados, desde a definição dos vetores de suporte, obtidos através das classificações e funções de kernel, até o treinamento, os testes, a classificação e análise da classificação encontrada.

## 5 MEDIDAS DE ANÁLISE DE CONHECIMENTO

As medidas de análise de conhecimento gerado são utilizadas para avaliação dos resultados obtidos. No trabalho desenvolvido, foram utilizadas como medidas, a matriz de confusão e a curva ROC.

### 5.1 MATRIZ DE CONFUSÃO

A matriz de confusão é composta pelo número de classes em que resulta a classificação realizada. A diagonal principal representa o número de acertos das classes identificadas, já o restante dos números fora da diagonal, o número de erros (figura 16) (PESSOA et al, 2010).

Figura 16 – Modelo de Matriz de Confusão de duas classes

|   |  | Predição |    |                |
|---|--|----------|----|----------------|
|   |  | 0        | 1  |                |
| 0 |  | VN       | FP | Especificidade |
| 1 |  | FN       | VP | Sensibilidade  |
|   |  |          |    | Acurácia       |

Fonte: Pessoa et al (2010).

A figura 16 exemplifica uma matriz de confusão que possui duas classes. De acordo com Pessoa et al (2010), pode-se identificar as classes resultantes da classificação, 0 e 1, além de outros conceitos utilizados para medir a efetividade da classificação feita. VN refere-se aos verdadeiros negativos e VP são os verdadeiros positivos, que juntos compõe a diagonal principal e expressam o número de acertos. FN e FP são falsos negativos e falsos positivos respectivamente, que indicam os erros. As medidas de análise de conhecimento gerado são calculadas a partir destes números. Conforme apresenta a figura utilizada por Pessoa et al, são elas:

- a) especificidade: expressa o número de acertos relacionados à identificação dos verdadeiros negativos. Obtida a partir do cálculo:

$$NV/(NV + FP);$$

- b) sensibilidade: mostra o número de acertos relacionados aos verdadeiros positivos. Obtida pelo cálculo:  $VP / (VP + FN)$ ;
- c) acurácia: mostra a taxa de acertos global. Obtida a partir do cálculo:  $(VP + VN) / (VP + VN + FP + FN)$ .

As medidas citadas são relevantes na análise dos resultados. Com a classificação realizada através no método *Leave-One-Out* e a geração da Matriz de Confusão, poderá ser analisado se a classificação foi ou não bem sucedida.

## 5.2 CURVA ROC

A curva ROC é uma forma gráfica para exibição das taxas de positivos verdadeiros e positivos falsos do classificador. Ela é de grande utilidade na comparação entre modelos de classificação (TAN; STEINBACH; KUMAR, 2009).

De acordo com Witten e Frank (2005, tradução nossa), uma forma de se obter a curva ROC, seria uma comparação entre o que foi predito e o que realmente é verdadeiro.

Como pode ser verificado na figura 17, os valores relacionados a sensibilidade da classificação ficam no eixo x, enquanto os valores relacionados a especificidade se posicionam no eixo y, formando a curva a partir dos valores obtidos com a classificação.

17 –

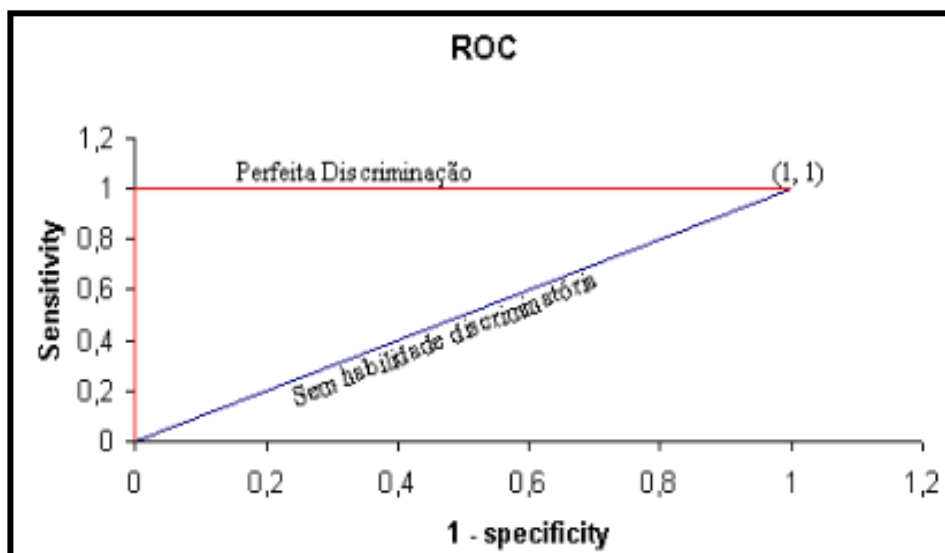


Figura  
Exemplo de  
Curva ROC

Fonte: Pessoa et al (2010).

A linha vermelha demonstra o que seria discriminação ideal, que é quando o classificador consegue realizar a separação das classes perfeitamente. E os valores sob a curva, indicam o quanto a classificação aproximou-se dessa discriminação ideal. Desta forma, quanto mais próxima do valor um, melhor a classificação.

Assim como demonstra a figura 17, a curva ROC proporciona uma visualização do desempenho do classificador, permitindo que se sejam feitas comparações entre modelos.

## 6 TRABALHOS CORRELATOS

A MSV oferece possibilidades de classificação aplicáveis em diversas áreas. Serão descritos alguns trabalhos em que foi utilizada a MSV para o atendimento das necessidades encontradas.

### 6.1 MAPEAMENTO DE AMBIENTES EXTERNOS UTILIZANDO ROBOS MÓVEIS

Esta dissertação foi desenvolvida por Alberto Yukinobu Hata e apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC – USP, em 2010. O objetivo principal do autor foi desenvolver uma solução para o mapeamento de ambientes, como parques e regiões urbanas. Para isso, ele elaborou um sistema de classificação de navegabilidade de terrenos irregulares. Foram feitos algoritmos para realizar o mapeamento destes terrenos. Sendo que um deles gerava um mapa de navegabilidade, onde foram utilizados os métodos de MSV e RNA para classificação dos terrenos em navegável, não-navegável e parcialmente navegável.

Conforme o Hata (2010), para o uso da MSV, foi utilizado como auxílio a biblioteca SVM Light na criação dos classificadores. Para a classificação, foram escolhidos quatro cenários distintos. Foi gerado um conjunto de padrões de treinamento e outro de validação. Para a criação do conjunto de dados, extraiu-se amostras das células de navegabilidade e as mesmas foram rotuladas manualmente. Foram utilizadas quatro funções de kernel diferentes (linear, polinomial, base radial e *sigmoid*).

O autor relata que os resultados obtidos por este estudo foram a classificação dos cenários e a comparação entre os métodos utilizados. As RNAs obtiveram um desempenho superior em função de que a MSV precisa de ajustes nos parâmetros utilizados para que funcione corretamente, o que não foi feito, segundo relatou o autor (HATA, 2010).

## 6.2 MÁQUINAS DE VETORES DE SUPORTE NA CLASSIFICAÇÃO DE IMPRESSÕES DIGITAIS

Este trabalho constituiu uma dissertação de mestrado apresentada à Universidade Federal do Ceará em 2002, por Allan Reffson Granja Lima. Neste trabalho foi proposta a classificação de impressões digitais para facilitar o gerenciamento de grandes bancos de dados e deixar o processo de identificação mais rápido.

Foi utilizada a base de dados NIST-4 que contém 4.000 imagens de impressões digitais. Este conjunto de dados foi dividido em dois, para o treinamento e teste. Utilizou-se um kernel gaussiano, e classificou-se os dados em sem rejeição de imagens e com rejeição de imagens. Para auxiliar a classificação, utilizou-se a implementação SVM Light.

Segundo o autor, os resultados obtidos foram superiores aos esperados. Foi verificado que a MSV oferece uma infraestrutura rápida para a geração dos modelos e para classificação das impressões digitais (LIMA, 2002).

## 6.3 ANÁLISE DE WAVELETS COM MÁQUINA DE VETOR DE SUPORTE NO ELETRENCEFALOGRAMA DA DOENÇA DE ALZHEIMER

Esta tese foi desenvolvida em 2013 por Paulo Afonso Medeiros Kanda, na Faculdade de Medicina da Universidade de São Paulo. Neste estudo foi proposto o uso da função Wavelet Morlet juntamente com MSV para a definição de um padrão que diferencie os pacientes com doença de Alzheimer.

Os dados foram aplicados no software WEKA, desenvolvido pela Universidade de Waikato, na Nova Zelândia. Para o treinamento e teste da classificação foi utilizado o método *Leave-One-Out*, juntamente com uma validação cruzada. Nela, os dados foram divididos em grupos de dez, e realizado outro treinamento.

Como resultado, o autor do estudo considerou a MSV uma ferramenta importante para classificação de dados. Segundo ele, no método *Leave-One-Out*, a acurácia chegou a 81%. O autor ainda destaca que a MSV demonstrou ter alta especificidade e sensibilidade no seu diagnóstico.

Nesta tese, foram utilizados cento e cinquenta e oito indivíduos, originando cento e setenta e dois mil atributos a serem analisados pela MSV. Obteve-se 90,74% de acurácia na classificação final dos dados (KANDA, 2013).

#### 6.4 SUPPORT VECTOR MACHINES NA CLASSIFICAÇÃO DE IMAGENS HIPERESPECTRAIS

Esta dissertação foi desenvolvida em 2009 por Rafaela Andreola, na Universidade Federal do Rio Grande do Sul. Foi proposta a análise do desempenho da MSV quando aplicada a imagens hiperespectrais de sensoriamento remoto, como as que são obtidas por sistemas de sensores a bordo de satélites para estudo da superfície da Terra.

De acordo com Andreola (2009), os dados utilizados foram os coletados pelo sistema sensor AVIRA sobre uma área agrícola localizada no noroeste do Estado da Indiana (EUA). Neste conjunto de dados, separou-se seis classes: milho cultivo mínimo, milho plantio direto, pastagens e árvores, soja cultivo convencional, soja cultivo mínimo e soja plantio direto, totalizando sete mil e sessenta e cinco amostras. Para o treinamento, os dados foram divididos em grupos de 50, 99, 200 e 300 amostras de cada classe. Foram feitos experimentos com funções de kernel polinomial e RBF e avaliados os seus desempenhos. Com o kernel polinomial obteve-se uma acurácia de 86,2%, enquanto com o kernel RBF a acurácia foi de 85,9%.

A autora do estudo relata que os experimentos realizados confirmaram a eficácia da metodologia utilizada. Os resultados mostraram a eficiência da MSV, sendo superiores aos do outro método que também foi utilizado, o classificador paramétrico (ANDREOLA, 2009).

## 6.5 DETECÇÃO DE INTRUSÃO USANDO TÉCNICAS DE APRENDIZAGEM DE MÁQUINAS

Este trabalho foi uma dissertação desenvolvida em 2010 por Vinícius Moll, submetida à Universidade Federal de Santa Catarina. Foi proposto o uso de MSV para detectar e classificar intrusões e ataques em redes ou em máquinas.

Na implementação do protótipo, foi utilizado software WEKA. Para a classificação utilizou-se as funções de kernel de base radial e polinomial. Foram coletadas 126.722 entradas de tráfego normal em um servidor da internet e 31.758 entradas de ataques e utilizadas no treinamento e no teste da MSV.

Nos testes feitos por ele, aplicou-se na MSV 41.835 entradas de tráfego normal e 10.480 entradas de ataques, diferentes das usadas no treinamento. Foi obtida uma taxa de 83,90% de acertos.

Conforme o autor, o uso desta técnica foi fundamental para a redução da taxa de falsos positivos e para o aumento das taxas de detecção de ataques, comparado aos outros detectores já existentes (MOLL, 2010).

## 6.6 TÉCNICAS DE PROCESSAMENTO DE IMAGENS PARA LOCALIZAÇÃO E RECONHECIMENTO DE FACES

Esta dissertação foi proposta em 2006 por Osvaldo César Pinheiro de Almeida no Instituto de Ciências Matemáticas e de Computação, da Universidade de São Paulo. O objetivo deste trabalho foi a localização e rastreamento de faces e o reconhecimento da face, onde foi utilizado a MSV.

Para a realização do reconhecimento de faces, utilizaram-se grupos de teste e treinamento. Selecionou-se a MSV multiclases, devido ao fato de que cada indivíduo é uma classe e as imagens das faces deste indivíduo são os exemplos de cada classe. A função de kernel utilizada foi a RBF. Para encontrar os parâmetros

usou-se a ferramenta *gridplay*, sendo que os parâmetros definidos foram o custo e a gama. A base de dados utilizada foi obtida a partir da compilação da base FERET.

Com este estudo, o melhor resultado obtido foi a taxa de 91.94% de acertos. Segundo o autor, os resultados foram satisfatórios, e ele ainda destaca que se o número de amostras fosse superior, a taxa de acertos seria maior (ALMEIDA, 2006).

## **7 UTILIZAÇÃO DE MÁQUINAS DE SUPORTE VETORIAL PARA CLASSIFICAÇÃO DE DADOS DE SEPSE ASSOCIADA A DELIRIUM**

Com base nos estudos realizados, foram feitos diversos treinamentos e testes, com diferentes funções de kernel e algoritmos para a obtenção de uma melhor classificação. Foi desenvolvido também, um simulador que utiliza um algoritmo de MSV para fazer a classificação, treinamento e teste dos dados, conforme será apresentado.

### **7.1 METODOLOGIA**

Com o objetivo de encontrar o algoritmo mais adequado para a classificação, a fim de aplica-lo no simulador, foi utilizada uma base de dados para auxiliar o processo de treinamento, teste e classificação dos dados.

Para o desenvolvimento do trabalho, foram feitas as seguintes etapas:

- a) descrição da base de dados;
- b) utilização da ferramenta WEKA;
- c) classificação dos dados;
- d) desenvolvimento do simulador.

#### **7.1.1 Descrição da Base de Dados**

No trabalho realizado foi utilizada uma base de dados coletada em uma Unidade de Terapia Intensiva na cidade de Criciúma, Santa Catarina.

A base de dados utilizada contém setenta e oito registros e é composta por dezesseis atributos: BNDF (dia um e dia dois), IL1B (dia um e dia dois), TNF (dia um e dia dois), STNFR1 (dia um e dia dois), STNFR2 (dia um e dia dois), adiponectina (dia um e dia dois), IL6, IL10, média de IL6 e IL10 neopterina. A base também possui três classes: sepse, sepse associada a delirium e evolução. Os atributos são biomarcadores que foram colhidos, alguns durante mais de um dia, e são descritos abaixo:

- a) BDNF: Fator Neurotrófico Derivado do Cérebro (BNDF). É uma proteína que tem influência no desenvolvimento, fisiologia e patologia do sistema nervoso, e também na memória e no aprendizado (MARTINHO, 2012);
- b) IL - 1 $\beta$ : Interleucina 1 $\beta$  (IL - 1 $\beta$ ). Proteína ligada a resposta inflamatória e acredita-se que ela seja ligada a traumas, convulsões e doenças como Parkinson e Alzheimer (ALCALDE, 2009);
- c) TNF: Fator de Necrose Tumoral (TNF). Também é uma proteína relacionada a processos inflamatórios (VITALE, 2007);
- d) adiponectina: a adiponectina é um hormônio expelido pelo tecido adiposo e influencia expressivamente em distúrbios metabólicos (COCATE, 2011);
- e) neopterin: é um biomarcador de ativação imunitária, relacionado a imunidade celular (PINTO, 2006).

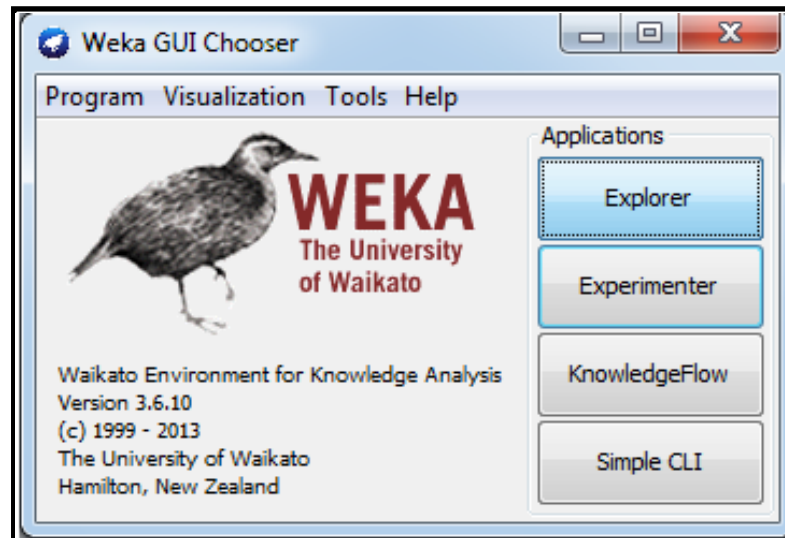
As classificações possíveis para os registros são: Sepsis (sim ou não), Sepsis Associada a Delírio (sim ou não), Evolução (morte ou não morte).

### 7.1.2 Utilização da Ferramenta Weka

Para classificação dos dados, foi utilizada a ferramenta de mineração de dados chamada WEKA (figura 22). Ela foi escolhida devido a sua praticidade, confiabilidade e por ser amplamente utilizada no meio acadêmico em projetos de *data mining*. Além disso, ela oferece todas as funções necessárias para o desenvolvimento do que foi proposto no trabalho.

O WEKA é feito pela Universidade de Waikato, situada na Nova Zelândia e é um programa de código aberto e gratuito (HALL, 2009, tradução nossa).

Figura 22 – Ferramenta WEKA

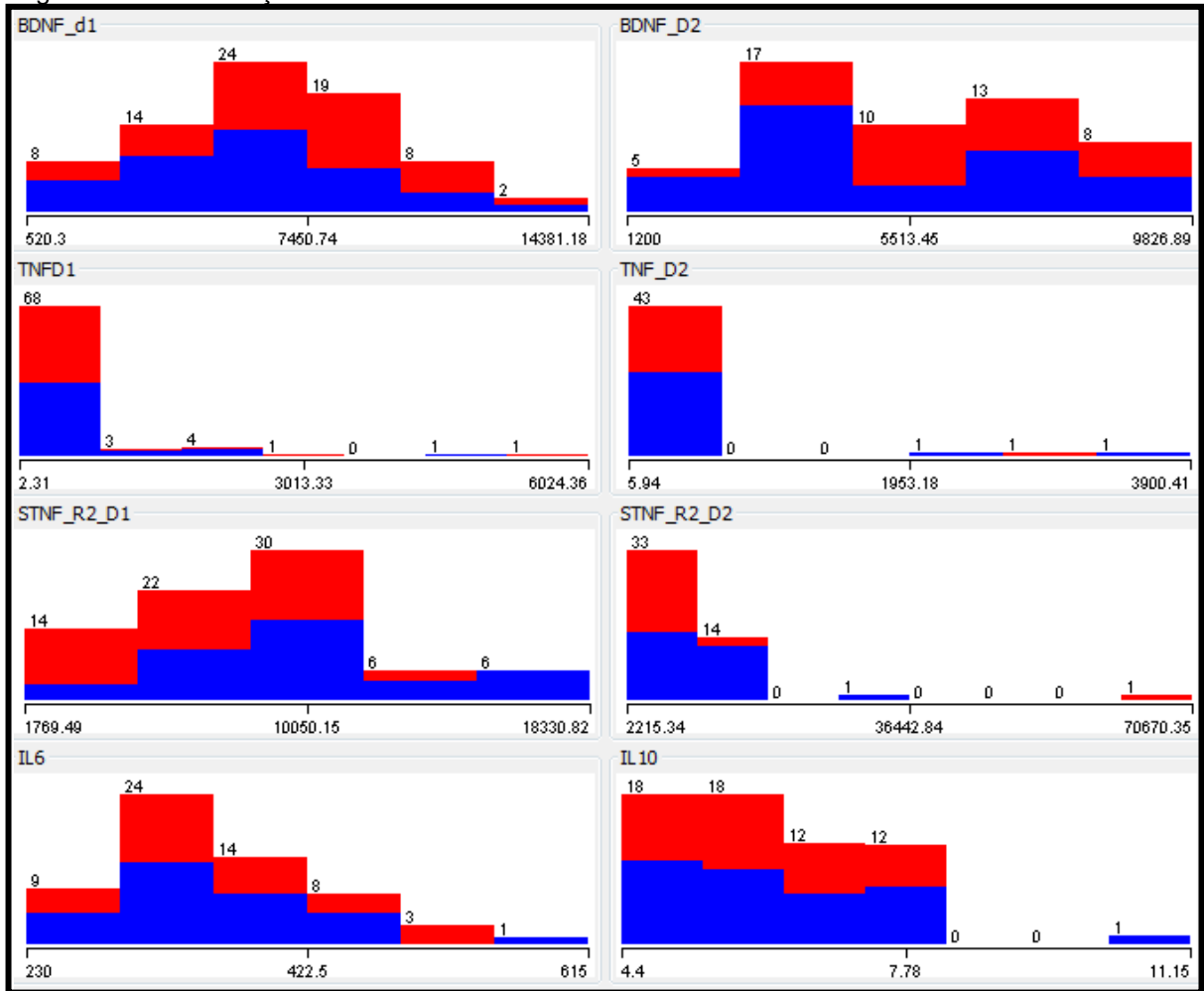


Fonte: Do autor.

Dentre as funcionalidades oferecidas por ele estão a regressão e a classificação de dados. O WEKA disponibiliza diversas funções para mineração de dados, entre elas, funções de MSV e RNA.

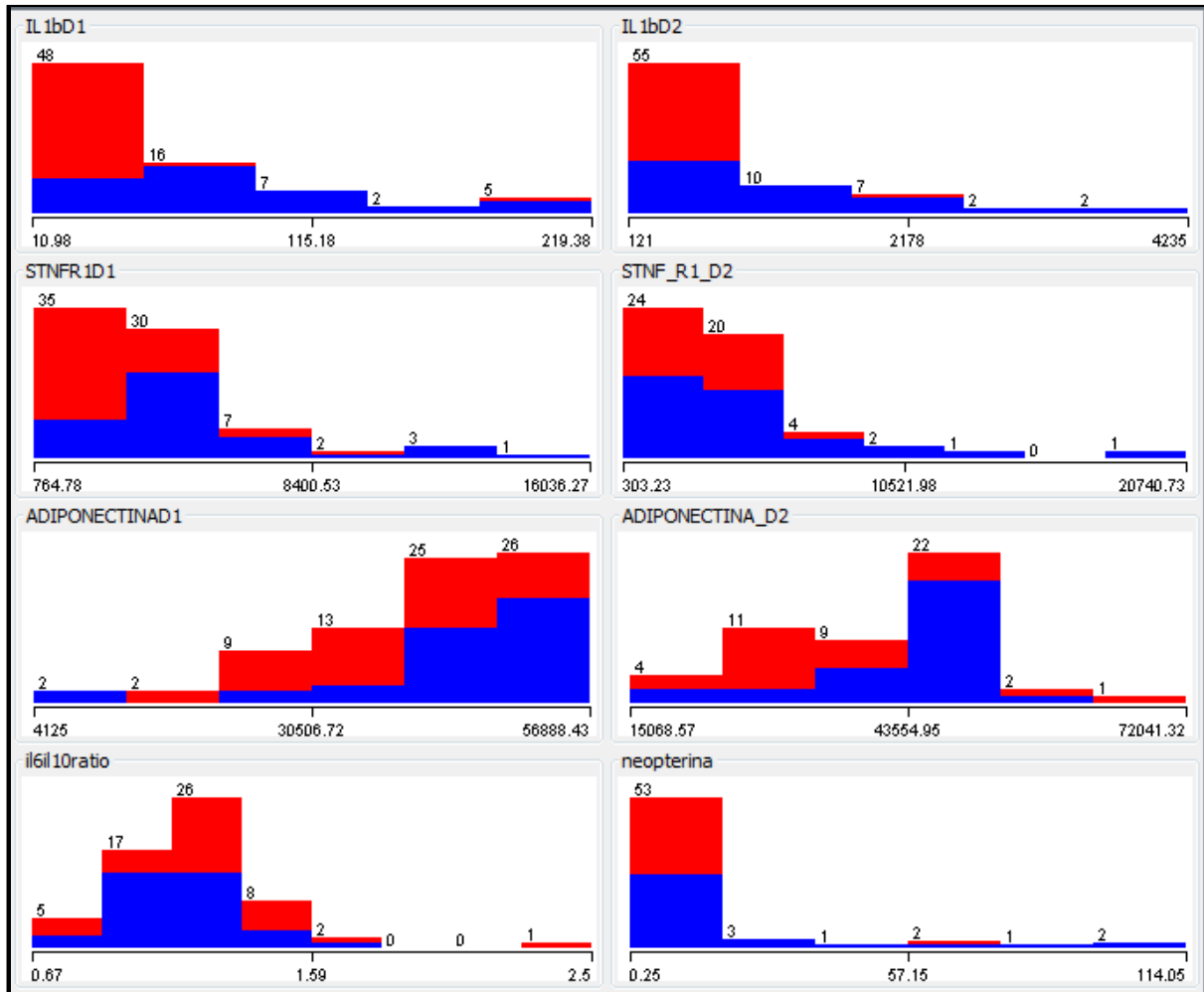
Para carregar os dados no WEKA, é necessário convertê-los para o formato ARRF (Formato de Arquivos de Atributo - Relação). Ao inserir o arquivo ARRF, o WEKA fornece uma visualização gráfica dos atributos e de seus valores, conforme mostram as figuras 23 e 24.

Figura 23 – Visualização de atributos no WEKA



Fonte: Do autor.

Figura 24 – Visualização de atributos no WEKA

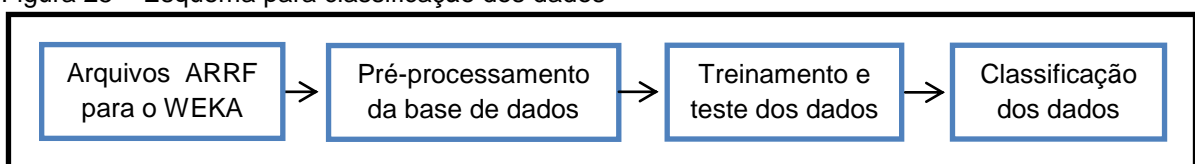


Fonte: Do autor.

### 7.1.3 Classificação dos Dados

Para realizar a classificação dos dados no WEKA, seguiu-se o esquema representado na figura 25.

Figura 25 – Esquema para classificação dos dados



Fonte: Do autor.

A primeira etapa foi a de preparação dos arquivos ARRF. Para isso, foram criados três arquivos que continham os dezesseis atributos da base, mais um atributo que indicava a classe, podendo ser Sepsis, Sepsis Associada a Delirium (SAD) ou Evolução.

Em seguida foi realizado o pré-processamento dos dados, etapa em que se verificou o algoritmo que seria utilizado, para que fosse possível fazer os ajustes necessários para a próxima etapa, que seria o treinamento e teste da base de dados, utilizando o algoritmo de MSV e de RNA.

Após a classificação, os resultados obtidos foram tabulados para que a análise das classificações e das taxas pudesse ser feita.

#### 7.1.3.1 Pré-processamento da base de dados

O pré-processamento de dados tem como objetivo melhorar a mineração de dados. É constituído por técnicas que influenciam os dados ou os próprios atributos (TAN; STEINBACH; KUMAR, 2009).

Nesta etapa, foram inseridos os arquivos ARRF no WEKA e observadas as particularidades de cada algoritmo.

Alguns registros da base de dados possuíam valores faltantes. O algoritmo LIBSVM, de MSV, exige a utilização de filtros para evitar esses valores faltantes. Para isso, foram utilizados os filtros *ReplaceMissingValues*, para substituir esses valores, e o filtro *NominalToBinary* fornecidos pela própria ferramenta. Os demais algoritmos utilizados já possuem esses filtros implementados internamente.

Em todos os algoritmos foi aplicada a opção de normalização de dados. Em alguns casos, foram retirados determinados atributos para verificar a influência destes sobre o resultado final. Estes testes serão descritos a seguir.

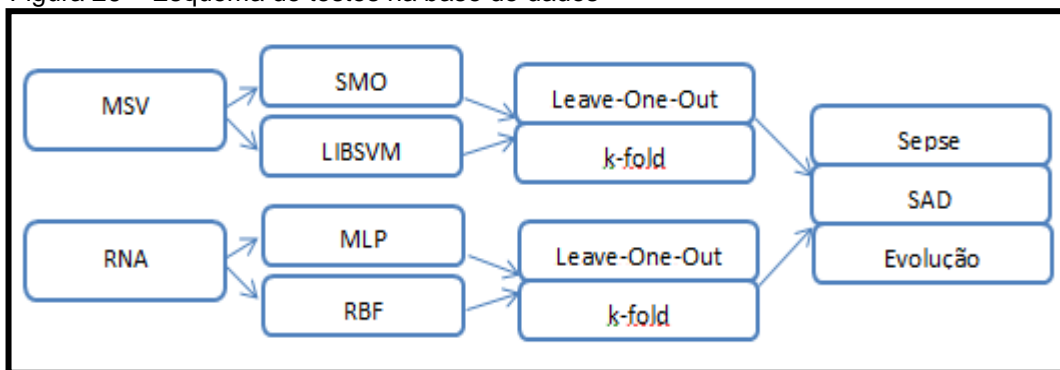
#### 7.1.3.2 Treinamento e teste da base de dados

Na classificação com MSV, foram utilizados dois algoritmos: LIBSVM e *Sequential Minimal Optimization* (SMO). LIBSVM é uma função que contém o algoritmo do próprio software LIBSVM implementado no WEKA, o que é chamado de

WLSVM. Este software implementa funções de MSV, e já inclui o algoritmo SMO (CHANG; LIN, 2014, tradução nossa). SMO é um outro algoritmo de MSV disponibilizado no WEKA, conhecido pela rapidez com que realiza a classificação. Já na classificação com RNA, foram utilizadas as redes *Multilayer Perceptron* (MLP) e *Radial Bases Function* (RBF).

Os testes foram feitos utilizando os algoritmos de MSV e as RNA, as três classificações possíveis conforme a base de dados (Sepse, SAD e Evolução) e os métodos de treinamento e teste *Leave-One-Out* e *k-fold*, como ilustra a figura 26.

Figura 26 – Esquema de testes na base de dados



Fonte: Do autor.

A base de dados continha três classificações: sepse, SAD e evolução. O treinamento foi feito separadamente para cada classificação, gerando resultados e análises diferentes para cada uma delas.

Nas classificações com MSV, foram utilizadas duas funções de kernel distintas para cada algoritmo. Para o algoritmo SMO, foram aplicadas as funções Polykernel e RBF, enquanto no algoritmo LIBSVM, utilizou-se as funções Polynomial e RBF. Para cada função de kernel, foram testados e selecionados parâmetros que melhoravam o resultado final da classificação. Os seguintes parâmetros foram modificados:

- a) parâmetro C: controla a relação entre os erros do treinamento e a complexidade do modelo criado (JOACHIMS, 2002, tradução nossa);

b) parâmetro Gamma: define o quanto um elemento irá influenciar a classificação (OLIVEIRA JUNIOR, 2010).

Para as funções de RNA utilizadas, MLP e RBF, foi necessário testar os seguintes parâmetros:

- a) constante *momentum*: é um valor que previne que a rede pare em soluções não ideais. Utilizado também para regular a velocidade de treinamento da rede neural (RAMALHO, 2008).
- b) taxa de aprendizado: conforme descreve Ramalho (2008), este valor está ligado a constante *momentum*, e também está relacionado ao tempo de convergência da rede;
- c) número de camadas: indica o número de camadas que será utilizado no treinamento.

Desta maneira, em cada teste, foram trocados os parâmetros e ajustados de forma com que chegassem a uma taxa de acertos maior. Após a realização da classificação, os resultados foram organizados conforme suas classes, para que fossem novamente testados, mas com alguns biomarcadores retirados. A retirada de biomarcadores foi com o objetivo de verificar o quão influentes eles eram no resultado.

#### **7.1.4 Desenvolvimento do Simulador**

Com o propósito de oferecer uma forma de prever dados de SAD, baseada na base de dados, foi desenvolvido um simulador que a partir das informações inseridas pelo usuário, informasse ao mesmo usuário a probabilidade do paciente vir a óbito ou não. Para isso, foi analisado qual classificador obteve melhores resultados com base nos testes feitos da maneira descrita anteriormente.

A ferramenta utilizada para o desenvolvimento foi o Netbeans IDE 7.3. O Netbeans pode ser adquirido através do site de seu fabricante.

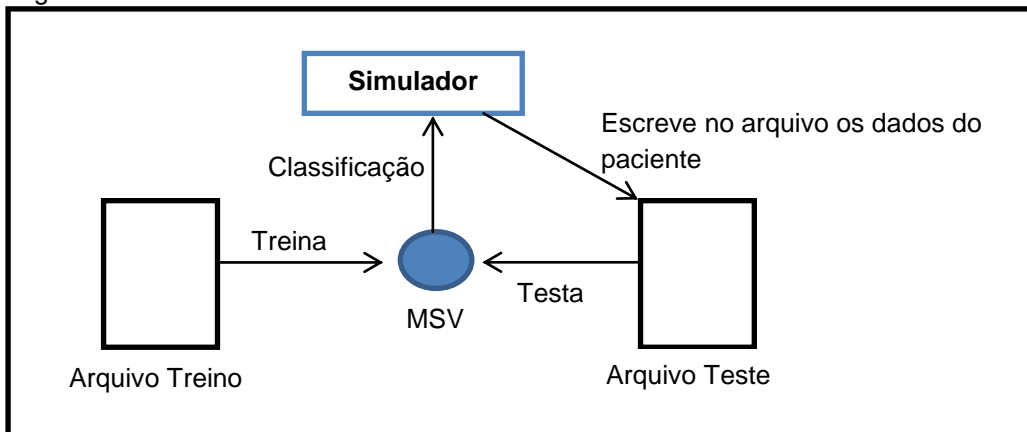
O projeto do Netbeans foi criado como uma aplicação Web, devido as vantagens oferecidas por ela e por ser o que melhor se enquadra no trabalho que foi

proposto. O projeto criado possui a página WEB referente a interface de interação com o usuário e as classes referentes a modelagem e a própria classificação dos registros.

Ao abrir a página Web no navegador, o usuário deve informar o nome e a idade do paciente, e os dados necessários para classificação, valores do tipo numérico relacionados aos biomarcadores. Ao clicar no botão “Classificar”, serão chamados internamente os métodos de classificação, e será retornado ao usuário, se o paciente que detém as informações inseridas na página, possui probabilidade de morte.

Para classificação, foi utilizada uma biblioteca do próprio WEKA. Com ela, é possível acessar as funções oferecidas por ele no próprio ambiente de desenvolvimento. A figura 27 ilustra o funcionamento do simulador.

Figura 27 – Funcionamento do simulador



Fonte: Do autor.

No treinamento do modelo do classificador, foi usado um arquivo ARRF, o mesmo utilizado nos testes. Este arquivo é composto pelos dezesseis atributos e pela classe evolução. Ao submeter os dados do paciente, o simulador lê as instâncias (registros) do arquivo e treina e testa a MSV com estas instâncias. Com isso, será gerado um modelo referente ao treinamento com os dados extraídos da base de dados.

Os valores inseridos pelo usuário são colocados em outro arquivo ARRF de teste e no próprio arquivo de treino para ser usado posteriormente. A classe evolução, para esta instância, já é inserida com o valor inicial de “Morte”, para que se pudesse montar corretamente a linha dos registros, contendo todos os valores para biomarcadores. Após ter o modelo do treinamento, este arquivo é lido, e as instâncias contidas nele, são testadas. Em seguida, é verificado se o que foi predito estava correto e essa classificação é retornada ao usuário.

O simulador desenvolvido pode ser visualizado na figura 28.

Figura 28 – Simulador desenvolvido

|                    |            |                 |                 |
|--------------------|------------|-----------------|-----------------|
| Paciente           |            |                 |                 |
| João da Rosa       |            |                 |                 |
| BNDF_d1            | BNDF_d2    | IL1b_d1         | IL1b_d2         |
| 5736.832           | 8782.398   | 205.1923        | 2342.0          |
| TNF_D1             | TNF_D2     | STNF_R1_D1      | STNF_R1_D2      |
| 36.357             | 32.424     | 3716.185        | 2507.254        |
| STNF_R2_D1         | STNF_R2_D2 | ADIPONECTINA_D1 | ADIPONECTINA_D2 |
| 12303.78           | 6390.651   | 48182.03        | 57038.58        |
| IL6                | IL10       | IL6IL10ratio    | Neoptarina      |
| 447.5              | 7.8        | 1.15            | 13.31           |
| <b>Classificar</b> |            |                 |                 |
| Não Morte          |            |                 |                 |

Fonte: Do autor.

A parte gráfica do simulador foi desenvolvida com auxílio foi *jQueryMobile*, que é utilizado na criação de aplicações Web móveis. Para o controle da página Web, foi utilizado o *JavaServerFaces* (JSF), que corresponde a outro *framework* utilizado em aplicações Web.

## 7.2 RESULTADOS OBTIDOS

A base de dados utilizadas possuía três classes: Sepsis, SAD e Evolução. Desta maneira, após os testes feitos na ferramenta WEKA com cada classe, utilizando as funções de kernel e as redes já mencionadas, obteve-se as taxas de cada classificação. Cada simulação realizada será descrita conforme a classe utilizada.

### 7.2.1 Classificação de Sepsis

Na base de dados utilizada, dos setenta e oito pacientes, metade deles estavam diagnosticados como sépticos. A partir disso, foram aplicados os algoritmos de MSV e RNA para classificação dos dados, tendo como classe desejada “Sepsis”.

No total foram vinte testes, sendo doze com algoritmos de MSV e oito com RNA. Os métodos de treinamento e teste utilizados, conforme já mostrado, foram o *leave-one-out* e *k-fold*. Serão exibidos os resultados mais significativos, contudo para se chegar nos valores utilizados para esses parâmetros, foram realizados outros testes.

Com algoritmo SMO de MSV realizaram-se seis testes. Os parâmetros utilizados para a função de kernel Polykernel foram  $C = 1$  e  $C = 2$ , enquanto que para a função de kernel RBF utilizou-se os parâmetros  $C = 5$  e  $\text{Gamma} = 0.5$  (figura 29).

Figura 29 – Treinamento e teste com SMO para classificação Sepsis

| Algoritmo | Função de Kernel   | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|-----------|--------------------|-------------------------------|---------|---------|--------|---------|
|           |                    |                               | Número  | %       | Número | %       |
| SMO       | Polykernel (C = 1) | Leave-One-Out                 | 60      | 76,9321 | 18     | 23,0769 |
| SMO       | Polykernel (C = 1) | k-fold                        | 61      | 78,2051 | 17     | 21,7949 |
| SMO       | Polykernel (C = 2) | Leave-One-Out                 | 63      | 80,7692 | 15     | 19,2308 |
| SMO       | Polykernel (C = 2) | k-fold                        | 62      | 79,4872 | 16     | 20,5128 |
| SMO       | RBF                | Leave-One-Out                 | 63      | 80,7692 | 15     | 19,2308 |
| SMO       | RBF                | k-fold                        | 62      | 79,4872 | 16     | 20,5128 |

Fonte: Do autor.

Conforme mostra a figura 29, maior taxa de acertos foi 80,7692% obtida com as funções de kernel Polykernel e RBF. Observa-se que a diferença entre os métodos *k-fold* e *leave-one-out* não foi significativa nestes testes.

Com LIBSVM também foram feitos seis testes. As funções de kernel Polynomial e RBF foram utilizadas com os parâmetros  $C = 5$  para ambas, e  $\gamma = 0.5$  para RBF. Utilizou-se também os métodos *leave-one-out* e *k-fold*, sendo que o *k-fold* foi testado duas vezes. Na primeira, separou-se em dez subconjuntos e na segunda, em trinta. Como demonstra a figura 30, para este algoritmo o *leave-one-out* obteve resultados inferiores ao resultados do *k-fold*. Percebe-se também, que a melhor taxa de acerto foi com a função RBF, com 63 registros classificados corretamente.

Figura 30 – Treinamento e teste com LIBSVM para classificação Sepse.

| Algoritmo | Função de Kernel | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|-----------|------------------|-------------------------------|---------|---------|--------|---------|
|           |                  |                               | Número  | %       | Número | %       |
| LIBSVM    | Polynomial       | Leave-One-Out                 | 59      | 75,6410 | 19     | 24,359  |
| LIBSVM    | Polynomial       | k-fold (N = 10)               | 60      | 76,9321 | 18     | 23,0769 |
| LIBSVM    | Polynomial       | k-fold (N = 30)               | 60      | 76,9321 | 18     | 23,0769 |
| LIBSVM    | RBF              | Leave-One-Out                 | 62      | 79,4872 | 16     | 20,5128 |
| LIBSVM    | RBF              | k-fold (N = 10)               | 61      | 79,4872 | 17     | 21,7949 |
| LIBSVM    | RBF              | k-fold (N = 30)               | 63      | 80,7692 | 15     | 19,2308 |

Fonte: Do autor.

Os resultados obtidos com a função MLP, de RNA, são mostrados na figura 31. Na figura 32, as taxas resultantes do treinamento e teste utilizando a função RBF são apresentadas. Os parâmetros utilizados foram o momentum = 0.4,

taxa de aprendizado = 0.8 e o número de treinamentos = 1000. O parâmetro que mais foi modificado foi o número de camadas da rede.

Figura 31 – Treinamento e teste com MLP para classificação Sepse

| Rede | Número de Camadas | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|------|-------------------|-------------------------------|---------|---------|--------|---------|
|      |                   |                               | Número  | %       | Número | %       |
| MLP  | 8                 | <i>Leave-One-Out</i>          | 62      | 79,4872 | 16     | 20,5128 |
| MLP  | 8                 | <i>k-fold</i> (N = 10)        | 60      | 76,9321 | 18     | 23,0769 |
| MLP  | 2                 | <i>Leave-One-Out</i>          | 57      | 73,0769 | 21     | 26,9231 |
| MLP  | 4                 | <i>Leave-One-Out</i>          | 57      | 73,0769 | 21     | 26,9231 |

Fonte: Do autor.

Figura 32 – Treinamento e teste com RBF para classificação Sepse

| Rede | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|------|-------------------------------|---------|---------|--------|---------|
|      |                               | Número  | %       | Número | %       |
| RBF  | <i>Leave-One-Out</i>          | 55      | 70,5128 | 23     | 29,4872 |
| RBF  | <i>k-fold</i> (N = 10)        | 53      | 67,9487 | 25     | 32,0513 |
| RBF  | <i>k-fold</i> (N = 30)        | 53      | 67,9487 | 25     | 32,0513 |
| RBF  | <i>k-fold</i> (N = 50)        | 49      | 62,8205 | 29     | 37,1795 |

Fonte: Do autor.

De acordo com o que foi apresentado, o desempenho com a rede MLP foi melhor do que com a rede RBF, já a primeira obteve uma taxa de acertos de aproximadamente 79.49%, enquanto a segunda só chegou a 70.51%.

Ao analisar os testes realizados com a base de dados, com o objetivo de classificar corretamente os pacientes sépticos e os não sépticos, observa-se que o número de acertos utilizando MSV foi superior ao número de acertos utilizando RNA. A diferença entre os resultados foi mínima, porém em uma base de dados com mais registros, essa diferença pode tornar-se significativa. A taxa de acertos com o algoritmos de MSV chegou a aproximadamente 80%, enquanto que com os algoritmos de RNA, chegou a 79%.

## 7.2.2 Classificação de SAD

Para esta classe, cinquenta e sete pacientes estavam com sepse associada a delirium, e vinte e um não. A partir disso, foram testados os algoritmos de MSV e RNA, a fim de encontrar uma melhor classificação e conseguir comparar os resultados de ambos. Os parâmetros utilizados nesta classificação foram semelhantes aos utilizados para classificação de sepse: para as funções de kernel RBF, utilizou-se  $C = 5$  e  $\gamma = 0.5$ , e nas funções Polykernel,  $C = 1$ , e Polynomial  $C = 5$ .

As figuras 33 e 34 mostram o uso dos algoritmos SMO e LIBSVM respectivamente.

Figura 33 – Treinamento e teste com SMO para classificação de SAD

| Algoritmo | Função de Kernel | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|-----------|------------------|-------------------------------|---------|---------|--------|---------|
|           |                  |                               | Número  | %       | Número | %       |
| SMO       | Polykernel       | Leave-One-Out                 | 68      | 87,1795 | 10     | 12,8205 |
| SMO       | Polykernel       | <i>k-fold</i>                 | 68      | 87,1795 | 10     | 12,8205 |
| SMO       | RBF              | Leave-One-Out                 | 68      | 87,1795 | 10     | 12,8205 |
| SMO       | RBF              | <i>k-fold</i>                 | 69      | 88,4615 | 9      | 11,5385 |

Fonte: Do autor.

Figura 34 – Treinamento e teste com LIBSVM para classificação de SAD

| Algoritmo | Função de Kernel | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|-----------|------------------|-------------------------------|---------|---------|--------|---------|
|           |                  |                               | Número  | %       | Número | %       |
| LIBSVM    | Polynomial       | Leave-One-Out                 | 64      | 82,0513 | 14     | 17,9487 |
| LIBSVM    | Polynomial       | <i>k-fold</i>                 | 64      | 82,0513 | 14     | 17,9487 |
| LIBSVM    | RBF              | Leave-One-Out                 | 69      | 88,4615 | 9      | 11,5385 |
| LIBSVM    | RBF              | <i>k-fold</i>                 | 69      | 88,4615 | 9      | 11,5385 |

Fonte: Do autor.

Como pode ser visto, os resultados obtidos nesta classificação foram bem superiores aos resultados da classificação anterior (sepse). Pode-se verificar também, que os algoritmos que obtiveram um melhor desempenho foram o SMO, com *k-fold* e LIBSVM com a função de kernel RBF, ambos com sessenta e nove acertos. Neste teste, os métodos de treinamento e teste escolhidos não demonstraram uma interferência significativa na taxa de acertos.

A figura 35 apresenta os resultados da classificação realizada utilizando as redes MLP e RBF. Para esta classificação, na rede MLP, utilizou-se o parâmetro momentum = 0.4 e a taxa de aprendizado = 0.3.

Figura 35 - Treinamento e teste com MLP e RBF para classificação de SAD

| Rede | Número de Camadas | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|------|-------------------|-------------------------------|---------|---------|--------|---------|
|      |                   |                               | Número  | %       | Número | %       |
| MLP  | 8                 | <i>Leave-One-Out</i>          | 64      | 82,0513 | 14     | 17,9487 |
| MLP  | 8                 | <i>k-fold</i>                 | 65      | 83,3333 | 13     | 16,6667 |
| RBF  |                   | <i>Leave-One-Out</i>          | 61      | 78,2051 | 17     | 21,7949 |
| RBF  |                   | <i>k-fold</i>                 | 65      | 83,3333 | 13     | 16,6667 |

Fonte: Do autor.

Conforme mostra a figura 35, as duas redes obtiveram resultados semelhantes, sendo que o máximo de acertos foi sessenta e cinco.

Comparando os resultados de RNA e MSV, nota-se que para a classificação de SAD, a MSV gerou uma taxa de acertos superior a RNA. Com MSV o número de acertos foi sessenta e nove. Já com RNA, o número de acertos máximo foi sessenta e cinco.

### 7.2.3 Classificação de Evolução

Esta classe indica se o paciente foi a óbito ou não. Na base de dados, cinquenta e oito pacientes morreram e apenas vinte não.

Na classificação em que aplicou-se o algoritmo SMO com a função de kernel Polykernel, o parâmetro C foi configurado com o valor 2. Já na função de kernel RBF, os parâmetros C e gamma foram configurados com os valores 2 e 0.1 respectivamente. Como pode ser visualizado na figura 36, a função de kernel RBF gerou resultados melhores que a função de kernel Polykernel. O número de acertos máximo foi sessenta, com o método de treinamento e teste *k-fold*.

Figura 36 - Treinamento e teste com SMO para classificação da Evolução

| Algoritmo | Função de Kernel | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|-----------|------------------|-------------------------------|---------|---------|--------|---------|
|           |                  |                               | Número  | %       | Número | %       |
| SMO       | Polykernel       | <i>Leave-One-Out</i>          | 56      | 71,7949 | 22     | 28,2051 |
| SMO       | Polykernel       | <i>k-fold</i>                 | 55      | 70,5028 | 23     | 29,4872 |
| SMO       | RBF              | <i>Leave-One-Out</i>          | 59      | 75,6410 | 19     | 24,359  |
| SMO       | RBF              | <i>k-fold</i>                 | 60      | 76,9231 | 18     | 23,0769 |

Fonte: Do autor.

Utilizando o algoritmo LIBSVM, com a função Polynomial configurada com os valores  $C = 2$  e  $\gamma = 0.1$  e a função RBF configurada com os valores  $C = 5$  e  $\gamma = 0.1$  obteve-se os resultados demonstrados na figura 37. Observa-se que os acertos foram semelhantes para as duas funções de kernel e para os dois métodos de treinamento. O número de acertos máximo com este algoritmo foi sessenta e nove.

Figura 37 - Treinamento e teste com LIBSVM para classificação da Evolução

| Algoritmo | Função de Kernel  | Método de Treinamento e Teste | Acertos |         | Erros  |         |
|-----------|-------------------|-------------------------------|---------|---------|--------|---------|
|           |                   |                               | Número  | %       | Número | %       |
| LIBSVM    | <i>Polynomial</i> | <i>Leave-One-Out</i>          | 58      | 74,359  | 20     | 25,641  |
| LIBSVM    | <i>Polynomial</i> | <i>k-fold</i>                 | 58      | 74,359  | 20     | 25,641  |
| LIBSVM    | RBF               | <i>Leave-One-Out</i>          | 58      | 74,359  | 20     | 25,641  |
| LIBSVM    | RBF               | <i>k-fold</i>                 | 57      | 73,0769 | 21     | 26,9231 |

Fonte: Do autor.

A figura 38 apresenta os resultados dos testes com as redes MLP e RBF. Nela, pode ser verificado que o desempenho da rede MLP foi superior ao da rede RBF. Para esta classificação, o número máximo de acertos foi sessenta e três, obtido com a rede MLP utilizando duas camadas e o método *Leave-One-Out*.

Figura 38 - Treinamento e teste com MLP e RBF para classificação da Evolução

| Rede | Número de Camadas | Método de Treinamento e | Acertos | Erros |
|------|-------------------|-------------------------|---------|-------|
|------|-------------------|-------------------------|---------|-------|

|     |   | Teste                | Número | %       | Número | %       |
|-----|---|----------------------|--------|---------|--------|---------|
| MLP | 8 | <i>Leave-One-Out</i> | 60     | 76,9231 | 18     | 23,0769 |
| MLP | 2 | <i>Leave-One-Out</i> | 63     | 80,7692 | 15     | 19,2308 |
| MLP | 2 | <i>k-fold</i>        | 58     | 72,3590 | 20     | 25,541  |
| RBF |   | <i>Leave-One-Out</i> | 59     | 75,6410 | 19     | 24,359  |
| RBF |   | <i>k-fold</i>        | 56     | 71,7949 | 22     | 28,2051 |

Fonte: Do autor.

Na classificação de Evolução, pode-se observar que o melhor desempenho foi com RNA, na rede MLP. O número de acertos máximo com MSV foi sessenta, utilizando o algoritmo SMO e a função de kernel RBF. Com RNA, o máximo foi sessenta e três, utilizando a rede MLP.

#### 7.2.4 Análise dos resultados

A partir dos resultados obtidos com os testes de cada classe, verificou-se quais foram as melhores taxas para cada função com o método *Leave-One-Out*, e novamente os resultados foram tabulados. A curva ROC também foi adicionada, para que fosse possível analisá-la juntamente com os outros dados. As figuras 39, 40 e 41, mostram essas informações.

Figura 39 – Classificação de Sepsis

| Classe | Técnica | Algoritmo | Função de Kernel | Acertos |       | Erros |       | Curva ROC |
|--------|---------|-----------|------------------|---------|-------|-------|-------|-----------|
|        |         |           |                  | Nº      | %     | Nº    | %     |           |
| Sepsis | MSV     | SMO       | Polykernel       | 63      | 80,77 | 15    | 19,23 | 0.808     |
| Sepsis | MSV     | SMO       | RBF              | 63      | 80,77 | 15    | 19,23 | 0.808     |
| Sepsis | MSV     | LIBSVM    | Polynomial       | 59      | 75,64 | 19    | 24,36 | 0.756     |
| Sepsis | MSV     | LIBSVM    | RBF              | 62      | 79,49 | 16    | 20,51 | 0.795     |
| Sepsis | RNA     | MLP       |                  | 62      | 79,49 | 16    | 20,51 | 0.795     |
| Sepsis | RNA     | RBF       |                  | 55      | 70,51 | 23    | 29,49 | 0.649     |

Fonte: Do autor.

De acordo com o mostrado na figura 39, ao observar a curva ROC e as taxas de acertos, pode-se considerar esta classificação como uma boa classificação, já que ambas as medidas de análises tiveram bons resultados.

Figura 40 – Classificação de SAD

| Classe | Técnica | Algoritmo | Função de Kernel | Acertos |       | Erros |       | Curva ROC |
|--------|---------|-----------|------------------|---------|-------|-------|-------|-----------|
|        |         |           |                  | Nº      | %     | Nº    | %     |           |
| SAD    | MSV     | SMO       | Polykernel       | 68      | 87,18 | 10    | 12,82 | 0.822     |
| SAD    | MSV     | SMO       | RBF              | 68      | 87,18 | 10    | 12,82 | 0.822     |
| SAD    | MSV     | LIBSVM    | Polynomial       | 69      | 88,46 | 9     | 11,54 | 0.846     |
| SAD    | MSV     | LIBSVM    | RBF              | 64      | 82,05 | 14    | 17,95 | 0.787     |
| SAD    | RNA     | MLP       |                  | 64      | 82,05 | 14    | 17,95 | 0.787     |
| SAD    | RNA     | RBF       |                  | 65      | 83,33 | 13    | 16,67 | 0.85      |

Fonte: Do autor.

A figura 40, apresenta a melhor classificação obtida dentre as três possíveis. Pode-se notar que a classe SAD, obteve os melhores resultados do classificador.

Figura 41 – Classificação de Evolução

| Classe   | Técnica | Algoritmo | Função de Kernel | Acertos |       | Erros |       | Curva ROC |
|----------|---------|-----------|------------------|---------|-------|-------|-------|-----------|
|          |         |           |                  | Nº      | %     | Nº    | %     |           |
| Evolução | MSV     | SMO       | Polykernel       | 56      | 71,79 | 22    | 28,21 | 0.581     |
| Evolução | MSV     | SMO       | RBF              | 59      | 75,64 | 19    | 24,36 | 0.558     |
| Evolução | MSV     | LIBSVM    | Polynomial       | 58      | 74,36 | 20    | 25,64 | 0.582     |
| Evolução | MSV     | LIBSVM    | RBF              | 58      | 74,36 | 20    | 25,64 | 0.582     |
| Evolução | RNA     | MLP       |                  | 63      | 80,77 | 15    | 19,23 | 0.737     |
| Evolução | RNA     | RBF       |                  | 59      | 75,64 | 19    | 24,36 | 0.718     |

Fonte: Do autor.

A partir do que pode ser verificado nas figuras 39, 40 e 41, percebe-se que a melhor taxa de acertos foi na classificação de SAD, e a pior foi na de Evolução. Observando a curva ROC, valor obtido com base no número de acertos e de erros de cada classificação, pode-se notar que ela que na maioria das classificações ela ficou próxima de 1, indicando um bom resultado. Na classificação de Evolução, os valores da curva ficaram com um valor mais baixo, já que a classificação também não resultou em um número de acertos muito alto.

Em novos testes, verificou-se que o biomarcador TNF, referente a segunda coleta, não influenciava nos resultados das classificações. Já o biomarcador BNDF, ao ser retirado da classificação, reduziu o número de acertos.

Ao testar sem o biomarcador STNF observou-se que a taxa de acertos da classificação aumentou.

No simulador desenvolvido foi utilizada a classe Evolução, para que fosse possível fazer uma predição a partir dos dados do paciente que foram informados. Para isso, foi utilizado o algoritmo SMO e a função de kernel RBF para realizar esta classificação. O método *Leave-One-Out* foi utilizado para o treinamento e teste. Na classificação feita com este algoritmo e com esta função, obteve-se aproximadamente 75% de acertos.

Os resultados gerados pelo projeto proposto englobam os resultados gerados a partir das classificações feitas com as classes disponibilizadas na base de dados, o simulador desenvolvido para a predição de novos registros e a própria análise realizada acerca destas classificações.

Com o trabalho desenvolvido, espera-se que outras análises possam ser realizadas a partir do conhecimento gerado com ele. Com as análises realizadas no projeto e as que ainda podem ser feitas, juntamente com os aprimoramentos necessários, deseja-se que o diagnóstico de sepse associada a delirium possa ser auxiliado.

## 8 CONCLUSÃO

A mineração de dados se faz extremamente importante para auxiliar as pessoas em determinadas tarefas. A extração de conhecimento e a análise que se pode fazer a partir de uma classificação de dados tem muita utilidade em diversas áreas.

Dentre as técnicas de mineração de dados disponíveis, foi utilizada a MSV. Com ela, buscou-se desenvolver um simulador que fizesse uma predição a partir da classificação de uma base de dados de SAD. A Sepsé vem sendo, nos últimos anos, uma das principais causas de morte em UTI. O desenvolvimento do simulador foi justamente para auxiliar no diagnóstico desta doença, o que poderia agilizar a detecção da gravidade da situação do paciente.

Para o desenvolvimento deste simulador, foi preciso conhecer o funcionamento da MSV, para aplicá-la nesta base de dados de SAD. Conhecendo a maneira como era feita a classificação, utilizou-se a ferramenta WEKA para realizar os testes com os registros da base de dados.

Com a base de dados, foi feita uma análise prévia dos registros. Esta análise confirmou as pesquisas realizadas sobre sepsé, quando diziam que a mortalidade da doença era alta.

Foram realizados diversos testes a fim de encontrar os melhores parâmetros para classificação e nestes testes, foram utilizados os métodos de treinamento e teste *k-fold* e principalmente, o *Leave-One-Out*.

Para selecionar a melhor classificação, a partir dos dados tabulados, foi feita uma análise dos dados por meio da matriz de confusão e da curva ROC. Ambas indicaram qual classificação obteve os melhores resultados, indicando assim, qual que deveria ser utilizada no simulador. Com a matriz de confusão, pode-se fazer a análise dos resultados e pode-se comparar os resultados de MSV com os de RNA. Ao comparar estas duas técnicas, observou-se que para duas classes da base, a MSV obteve resultados superiores aos da RNA.

Com as análises devidas, foi possível desenvolver o simulador baseado na melhor classificação com MSV e o *Leave-One-Out* da evolução dos pacientes. Uma das dificuldades do projeto foi na integração das funções do WEKA com o ambiente de desenvolvimento escolhido, já que mesmo com a documentação disponibilizada pelos desenvolvedores, ainda era necessário encontrar uma forma de realizar o *Leave-One-Out* e gerar o mesmo modelo que era gerado no próprio WEKA. Entretanto, essa e as outras dificuldades foram contornadas e os objetivos selecionados para este projeto, foram alcançados.

Como sugestão para trabalhos futuros, considerando o projeto de pesquisa realizado, podem ser listados os seguintes tópicos:

- a) analisar os biomarcadores de forma mais efetiva, para procurar identificar o motivo de alguns influenciarem diretamente nos resultados e outros não;
- b) aplicar os algoritmos de MSV e as redes novamente quando a base estiver com um maior número de registros, com o objetivo de verificar se as conclusões obtidas neste projeto, continuam mesmo com uma base maior;
- c) acrescentar novas funções no simulador, como gravação de registros em um banco de dados e possibilidade de informar outros dados do paciente;
- d) desenvolver o simulador para dispositivos móveis.

## REFERÊNCIAS

- ALBUQUERQUE, Rafael Walter de. **Monitoramento da cobertura do solo no entorno de hidrelétricas utilizando o classificador SVM (Support Vector Machine)**. 2011. 95 f. Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2011. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3138/tde-06062012-164051>>. Acesso em: 19 out. 2013.
- ALCALDE, Taís Frederes Krämer et al . Ausência de associação entre polimorfismo do gene da interleucina-1 beta e o prognóstico de pacientes com traumatismo crânio-encefálico grave. **Rev. Bras. Ter. Intensiva**, São Paulo, v. 21, n. 4, dez. 2009. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-507X2009000400002&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-507X2009000400002&lng=en&nrm=iso)>. Acesso em: 25 maio 2014.
- ALMEIDA, Osvaldo Cesar Pinheiro de. **Técnicas de processamento de imagens para localização e reconhecimento de faces**. 2006. Dissertação (Mestrado) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-22012007-160023/>>. Acesso em: 15 nov. 2013.
- AMORIM, Dinani Gomes. **Redes ART com categorias internas de geometria irregular**. 2006. 352 f. Tese (Doutorado) - Universidade de Santiago de Compostela, Santiago de Compostela, 2006. Disponível em: <[http://www.gsi.dec.usc.es/system/files/tesis\\_dinani.pdf](http://www.gsi.dec.usc.es/system/files/tesis_dinani.pdf) >. Acesso em: 27 abr. 2013.
- ANDREOLA, Rafaela. **Support Vector Machines na Classificação de Imagens Hiperespectrais**. 2009. 117 f. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/17894>>. Acesso em: 13 nov. 2013.
- ANGUS, Derek C. et al. Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. **Critical Care Medicine**, Mount Prospect, v. 29, n. 7, p.1303-1310, 2001. Disponível em: <[http://www.pulsion.com/fileadmin/redaktion/Literature/Costeffectiveness/Angus\\_CCM\\_2001\\_1303.pdf](http://www.pulsion.com/fileadmin/redaktion/Literature/Costeffectiveness/Angus_CCM_2001_1303.pdf)>. Acesso em: 25 ago. 2013.
- ASSOCIATION, American Psychiatric. **Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR**. American Psychiatric Pub, 2000.

BATISTA, Gustavo Enrique de Almeida Prado Alves. **Pré-processamento de dados em aprendizado de máquina supervisionado**. 2003. Tese (Doutorado) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-06102003-160219/>>. Acesso em: 15. Nov. 2013.

BAUDOUIIN, Simon. **Sepsis**. Londres: Springer, 2009. 111 p.

BONIFÁCIO, Fernando Nunes. **Comparação entre as Redes Neurais Artificiais MLP, RBF e LVQ na Classificação de Dados**. 2010. 94 f. TCC (Graduação) - Curso de Ciência de Computação, Unioeste – Universidade Estadual do Oeste do Paraná, Cascavel, 2010. Disponível em: <[www.inf.unioeste.br/~tcc/2010/TCC-FernandoNunes.pdf](http://www.inf.unioeste.br/~tcc/2010/TCC-FernandoNunes.pdf)>. Acesso em: 03 jun. 2014.

BRITO, Cândida Márcia de; ZUSE, Carmen Lucia. Demandas de enfermagem ao paciente em sepse, sepse severa e choque séptico na unidade de emergência. **Nursing (São Paulo)**, São Paulo, v. 14, n. 149 , p.529-535, out. 2010.

CAMPBELL, Colin; YING, Yiming. **Learning with Support Vector Machines**. Morgan & Claypool Publishers, 2011.

CARDOSO, Olinda Nogueira Paes; MACHADO, Rosa Teresa Moreira. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**, Rio de Janeiro, v. 42, n. 3, p.495-528, jun. 2008. Disponível em: <<http://www.scielo.br/pdf/rap/v42n3/a04v42n3.pdf>>. Acesso em: 06 out. 2013.

CARVALHO, Paulo R. A.; TROTTA, Eliana de A.. Avanços no diagnóstico e tratamento da sepse. **Jornal de Pediatria**, Porto Alegre, v. 79, 2003. Disponível em: <[http://www.scielo.br/scielo.php?pid=S0021-75572003000800009&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0021-75572003000800009&script=sci_arttext)>. Acesso em: 17 ago. 2013.

CERQUEIRA, Eduardo O. de et al . Redes neurais e suas aplicações em calibração multivariada. **Quím. Nova**, São Paulo , v. 24, n. 6, Dec. 2001 . Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-40422001000600025&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422001000600025&lng=en&nrm=iso)>. Acesso em: 02 jun. 2014.

CHANG, Chih-chung; LIN, Chih-jen. **LIBSVM -- A Library for Support Vector Machines**. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>>. Acesso em: 04 jun. 2014.

COCATE, Paula Guedes; DOMINGUES, Sabrina Fontes; NATALI, Antônio José. Concentrações sanguíneas de adiponectina e exercício físico: associações com a sensibilidade insulínica. **Rev. Bras. Ciênc. Esporte**, Porto Alegre, v. 33, n. 3, set. 2011. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0101-32892011000300018&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-32892011000300018&lng=en&nrm=iso)>. Acesso em: 25 maio 2014.

COHEN, Jonathan. The immunopathogenesis of sepsis. **Nature**, Washington Dc, v. 420, p.19-26, 19 dez. 2012. Disponível em: <[http://www.math.pitt.edu/~cbsg/Materials/Cohen\\_nature01326.pdf](http://www.math.pitt.edu/~cbsg/Materials/Cohen_nature01326.pdf)>. Acesso em: 02 set. 2013.

CRISTIANINI, Nello; SHAWE-TAYLOR, John. **An introduction to Support Vector Machines and other Kernel-based learning methods**. Cambridge University Press, 2000.

DANIELS, Ron; NUTBEAM, Tim. **ABC of Sepsis**. Oxford: Blackwell Publishing Ltd, 2010.

EBERSOLDT, Marion; SHARSHAR, Tarek; ANNANE, Djillali. Sepsis-associated delirium. **Intensive Care Med**, Ss, p. 941-950. 05 abr. 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17410344>>. Acesso em: 14 set. 2013.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel Lopes. **Data Mining: um guia prático: conceitos, técnicas, ferramentas, orientações e aplicações**. Rio de Janeiro: Elsevier, 2005

GUNN, Steve R.. **Support Vector Machines for Classification and Regression**. Southampton, 1998. Disponível em: <<http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf>>. Acesso em: 03 nov. 2013.

HALL et al. The WEKA Data Mining Software: An Update. **Acm Sigkdd Explorations Newsletter**, Nova York, v. 11, n. 1, p.10-18, jun. 2009. Disponível em: <<http://www.sigkdd.org/sites/default/files/issues/11-1-2009-07/p2V11n1.pdf>>. Acesso em: 31 maio 2014.

HATA, Alberto Yukinobu. **Mapeamento de ambientes externos utilizando robôs móveis**. 2010. 103 f. Dissertação (Mestrado) - Curso de Dd, Instituto de Ciências Matemáticas e Computação, São Carlos, 2010.

HAYKIN, Simon. **Redes Neurais: Princípios e Prática**. Bookman, 2002.

HENKIN, Caroline Schwartz et al. Sepsis: uma visão atual. **Scientia Medica**, Porto Alegre, v. 19, n. 3, p.135-145, jul. 2009. Disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/scientiamedica/article/viewFile/4716/4285>>. Acesso em: 26 ago. 2013.

JUNCAL, Verena Ribeiro et al. Impacto clínico do diagnóstico de sepse à admissão em UTI de um hospital privado em Salvador, Bahia. **J Bras Pneumol.**, São Paulo, v. 37, n. 1, p.85-92, fev. 2011. Disponível em: <[http://www.scielo.br/scielo.php?pid=S1806-37132011000100013&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S1806-37132011000100013&script=sci_arttext)>. Acesso em: 09 set. 2013.

KANDA, Paulo Afonso Medeiros. **Análise de wavelets com máquina de vetor de suporte no eletrencefalograma da doença de Alzheimer**. 2012. Tese - Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/5/5138/tde-12042013-105244/>>. Acesso em: 10 nov. 2013.

KÓVACS, Zsolt. **Redes Neurais Artificiais: fundamentos e aplicações**. 4. ed. São Paulo: Editora Livraria da Física, 2006.

KRESS, John P. The complex interplay between delirium, sepsis and sedation. **Critical Care**, v. 14, n. 164, 14 jun. 2010. Disponível em: <<http://ccforum.com/content/14/3/164>>. Acesso em: 14 set. 2013.

LIBRALAO, Giampaolo Luiz et al. Determinação de vícios refrativos oculares utilizando Support Vector Machines. **Sba Controle & Automação**, Campinas, v. 16, n. 2, jun. 2005. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-17592005000200004&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-17592005000200004&lng=en&nrm=iso)>. Acesso em: 15 nov. 2013.

LIMA, Allan Reffson Granja. **Máquinas de Vetores de Suporte na Classificação de Impressões Digitais**. 2002. 81 f. Dissertação (Mestrado). Universidade Federal do Ceará, Fortaleza, 2002. Disponível em: <[http://mdcc.ufc.br/teses/doc\\_download/43-044-allan-reffson-granja-lima](http://mdcc.ufc.br/teses/doc_download/43-044-allan-reffson-granja-lima)>. Acesso em: 10 nov. 2013.

LIMA, Clodoaldo Aparecido de Moraes. **Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte**. 2004. 342 f. Tese (Doutorado) - Universidade Estadual de Campinas, Campinas, 2004. Disponível em: <[ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/moraes\\_dout/tese\\_dout.pdf](ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/moraes_dout/tese_dout.pdf)>. Acesso em: 03 nov. 2013.

LORENA, Ana Carolina; CARVALHO, André C. P. L. F. de. Uma Introdução às Support Vector Machines. **Rita**, Porto Alegre, v. 14, n. 2 , p.43-67, 2007. Disponível em: <[seer.ufrgs.br/rita/article/download/rita\\_v14\\_n2\\_p43-67/3543](http://seer.ufrgs.br/rita/article/download/rita_v14_n2_p43-67/3543)>. Acesso em: 28 abr. 2013.

LOUIS-VINCENT, Jean. **Intensive Care Medicine**. New York: Springer Verlag Ny, 2010.

MARQUES, Filipe José Monteiro Morgado. **Pré-Processamento de Bases Clínicas para Modelação: Classificação de Sobrevivência de Doentes com Sépsis**. 2011. 115 f. Dissertação (Mestrado) - Curso de Engenharia Mecânica, Universidade Técnica de Lisboa, Lisboa, 2011. Disponível em: <<https://fenix.tecnico.ulisboa.pt/downloadFile/395142731914/dissertacao.pdf>>. Acesso em: 29 jun. 2014.

MARTIN, Billie-jean et al. Delirium as a predictor of sepsis in post-coronary artery bypass grafting patients: a retrospective cohort study. **Critical Care**, v. 14, p.1-6, 27 set. 2010. Disponível em: <<http://ccforum.com/content/14/5/R171>>. Acesso em: 10 set. 2013.

MARTIN, Greg S. et al. The Epidemiology of Sepsis in the United States from 1979 through 2000. **New England Journal Of Medicine**, p. 1546-1554. 17 abr. 2003. Disponível em: <<http://www.nejm.org/doi/full/10.1056/NEJMoa022139>>. Acesso em: 14 set. 2013.

MARTIN, Paul. **Dispel Tutorial**. Disponível em: <[http://homepages.inf.ed.ac.uk/pmartin/tutorial/case\\_studies.html](http://homepages.inf.ed.ac.uk/pmartin/tutorial/case_studies.html)>. Acesso em: 03 jun. 2014.

MARTINHO JR, Eduardo et al . Polimorfismo do gene do BDNF, cognição e gravidade dos sintomas em uma amostra de base populacional brasileira de indivíduos apresentando o primeiro episódio psicótico. **Rev. Bras. Psiquiatr.**, São Paulo, v. 34, supl. 2, out. 2012. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1516-44462012000600007&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-44462012000600007&lng=en&nrm=iso)>. Acesso em: 25 maio 2014.

MENDOZA, Leonardo Alfredo Forero. **Redes Neurais e Máquinas de Vetores de Suporte no reconhecimento de locutor usando coeficientes MFC e características do sinal glotal**. 2009. 128 f. Dissertação (Mestrado) - Universidade Federal Fluminense, Niterói, 2009. Disponível em: <[http://www.btdtd.ndc.uff.br/tde\\_arquivos/38/TDE-2009-11-17T161830Z-](http://www.btdtd.ndc.uff.br/tde_arquivos/38/TDE-2009-11-17T161830Z-)

2282/Publico/Dissert%20Leonardo%20Alfredo%20Forero%20Mendoza.pdf>. Acesso em: 19 out. 2013.

MOLL, Vinícius. **Detecção de intrusão usando técnicas de aprendizagem de máquinas**. 2010. 150 f. Dissertação (Mestrado). Universidade Federal de Santa Catarina, Florianópolis, 2010. Disponível em: <<https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/94110/280714.pdf?sequence=1>>. Acesso em: 15 nov. 2013.

OLIVEIRA JUNIOR, Gilson Medeiros de. **Máquina de Vetores Suporte: estudo e análise de parâmetros para otimização de resultado**. 2010. 46 f. TCC (Graduação) - Curso de Ciência de Computação, Universidade Federal de Pernambuco, Pernambuco, 2010. Disponível em: <<http://www.cin.ufpe.br/~tg/2010-2/gmoj.pdf>>. Acesso em: 31 maio 2014.

PAPADOPOULOS, Marios C. et al. Pathophysiology of septic encephalopathy: A review. **Critical Care Medicine**, v. 28, n. 8, p.3019-3024, 2000. Disponível em: <[http://www.uthsc.edu/pulmonary/documents/Pathophysiology\\_of\\_Septic\\_Encephalopathy\\_A\\_Review.pdf](http://www.uthsc.edu/pulmonary/documents/Pathophysiology_of_Septic_Encephalopathy_A_Review.pdf)>. Acesso em: 14 set. 2013.

PENINCK, Paula Pedroso; MACHADO, Regimar Carla. APLICAÇÃO DO ALGORITMO DA SEPSE POR ENFERMEIROS NA UNIDADE DE TERAPIA INTENSIVA. **Revista da Rede de Enfermagem do Nordeste**, Fortaleza, v. 13, n. 1, 2012. Disponível em: <<http://www.revistarene.ufc.br/revista/index.php/revista/article/view/30>>. Acesso em: 09 set. 2013.

PESSOA et al. Mineração de Dados Meteorológicos Associada a Eventos Severos no Pantanal Sul Matogrossense. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 33., 2010, Águas de Lindóia, SP. **Anais...** São Carlos, SP: SBMAC, 2010. p. 652-658. Disponível em: <[http://www.sbmac.org.br/eventos/cnmac/xxxiii\\_cnmac/pdf/306.pdf](http://www.sbmac.org.br/eventos/cnmac/xxxiii_cnmac/pdf/306.pdf)> Acesso em: 10 nov. 2013.

PINTO, A Mota et al. Determinação da neopterina e de defesas antioxidantes na asma de evolução arrastada. **Revista Portuguesa de Pneumologia**, Lisboa, v. 12, n. 6, nov. 2006. Disponível em <[http://www.scielo.gpeari.mctes.pt/scielo.php?script=sci\\_arttext&pid=S0873-21592006000600003&lng=pt&nrm=iso](http://www.scielo.gpeari.mctes.pt/scielo.php?script=sci_arttext&pid=S0873-21592006000600003&lng=pt&nrm=iso)>. Acesso em: 31 maio 2014.

PITROWSKY, Melissa Tassano et al. Importância da monitorização do delirium na unidade de terapia intensiva. **Revista Brasileira de Terapia Intens**

iva, São Paulo, v. 22, n. 3, set. 2010. Disponível em:<[http://www.scielo.br/scielo.php?pid=S0103-507X2010000300010&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0103-507X2010000300010&script=sci_arttext)>. Acesso em: 15 set. 2013.

RAMALHO, José Pinto. **Oxicorte**: estudo da transferência de calor e modelamento por redes neurais artificiais de variáveis do processo. 2008. Tese (Doutorado em Engenharia Metalúrgica e de Materiais) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2008. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3133/tde-30092008-150619/>>. Acesso em: 01 jun. 2014.

REINKE, Adalisa; RITTER, Cristiane; PIZZOL, Felipe Dal. Estresse oxidativo e desenvolvimento de sepse. **Revista de Iniciação Científica**, Criciúma, v. 2, n. 1, p. 117-130, 04 jun. 2004.

SALES JÚNIOR, João Andrade L. et al. Sepsis Brasil: Estudo Epidemiológico da Sepsis em Unidades de Terapia Intensiva Brasileiras. **Revista Brasileira Terapia Intensiva**, São Paulo, v. 18, n. 1, 2006. Disponível em: <[http://www.scielo.br/scielo.php?pid=S0103-507X2006000100003&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0103-507X2006000100003&script=sci_arttext)>. Acesso em: 09 set. 2013.

SANTOS, Eulanda Miranda Dos. **Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência**. 2002. 111 f. Dissertação (Mestrado) - Universidade Federal da Paraíba, Campina Grande, 2002. Disponível em: <[http://docs.computacao.ufcg.edu.br/posgraduacao/dissertacoes/2002/Dissertacao\\_EulandaMirandadosSantos.pdf](http://docs.computacao.ufcg.edu.br/posgraduacao/dissertacoes/2002/Dissertacao_EulandaMirandadosSantos.pdf)>. Acesso em: 21 out. 2013.

SILVA, Eliézer et al. Brazilian Sepsis Epidemiological Study (BASES study). **Critical Care Medicine**, v. 8, n. 4, p.251-260, 2004. Disponível em: <[http://www.scielo.br/scielo.php?pid=S0103-507X2006000100003&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0103-507X2006000100003&script=sci_arttext)>. Acesso em: 25 ago. 2013.

SILVA, Fabiano Pinheiro da; VELASCO, Irineu Tadeu. **Sepsis**. Barueri: Manole, 2007.

SONNEVILLE, Romain et al. Understanding brain dysfunction in sepsis. **Annals Of Intensive Care**, p. 1-11. 29 maio 2013. Disponível em: <<http://www.annalsofintensivecare.com/content/3/1/15>>. Acesso em: 21 set. 2013.

STEINWART, Ingo; CHRISTMANN, Andreas. **Support Vector Machines**. Springer, 2008

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining - Mineração de Dados**. Ciência Moderna, 2009.

VENTURA, André Rodrigues. **Análise da Heterogeneidade Intra-tumoral em Imagens PET-FDG**. 2010. 55 f. Dissertação (Mestrado) - Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Coimbra, 2010. Disponível em: <[https://estudogeral.sib.uc.pt/bitstream/10316/13813/1/Tese%20MIEB\\_Andr%C3%A9%20Ventura.pdf](https://estudogeral.sib.uc.pt/bitstream/10316/13813/1/Tese%20MIEB_Andr%C3%A9%20Ventura.pdf)>. Acesso em: 15 out. 2013.

VINCENT, Jean-Louis et al. Sepsis in European intensive care units: Results of the SOAP study\*. **Critical Care Medicine**, v. 34, n. 2, p.344-353, 2006. Disponível em: <[http://www.ccmpitt.com/ebm/sepsis/Sepsis%20in%20European%20intensive%20care%20units\\_results%20of%20the%20SOAP%20study..pdf](http://www.ccmpitt.com/ebm/sepsis/Sepsis%20in%20European%20intensive%20care%20units_results%20of%20the%20SOAP%20study..pdf)>. Acesso em: 14 set. 2013.

VITALE, Rodrigo Faller; RIBEIRO, Fernando de Andrade Quintanilha. O papel do Fator de Necrose Tumoral Alfa (TNF-alfa) no processo de erosão óssea presente no colesteatoma adquirido da orelha média. **Rev. Bras. Otorrinolaringol.**, São Paulo, v. 73, n. 1, fev. 2007. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-72992007000100020&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-72992007000100020&lng=en&nrm=iso)>. Acesso em: 25 maio 2014.

WANG, Lipo. **Support Vector Machines: Theory and Applications**. Springer, 2005.

WESTON, Jason A. E.. **Extensions to the Support Vector Method**. 1999. 127 f. Tese (Doutorado) - University Of London, Eghan, Surrey, England, 1999. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.2796>>. Acesso em: 03 nov. 2013.

WITTEN, Ian H.; FRANK, Eibe. **Data Mining: Practical Machine Learning Tools and Techniques**. 2. ed. San Francisco: Elsevier, 2005. 558 p.

WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A.. **Data Mining: Practical Machine Learning Tools and Techniques**. Elsevier, 2011.

WHAN-LEE, Seong; VERRI, Alessandro. **Pattern Recognition with Support Vector Machines**. Springer, 2002.

ZAMPIERI, Fernando Godinho et al. Sepsis-associated encephalopathy: not just delirium. **Clinics**, São Paulo, v. 66, n. 10, p.1825-1831, 2011. Disponível em:

<[http://www.scielo.br/scielo.php?pid=S1807-59322011001000024&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S1807-59322011001000024&script=sci_arttext)>. Acesso em: 14 set. 2013.

**APÊNDICE(S)**

## APÊNDICE A – Artigo Científico

### Utilização de Máquinas de Suporte Vetorial para a Classificação de Dados de Sepses Associada a Delirium

Camila F. Crispim<sup>1</sup>, Gustavo Bisognin<sup>1</sup>, Merisandra C. de Mattos Garcia<sup>1</sup>

<sup>1</sup>Universidade do Extremo Sul Catarinense (UNESC) – Criciúma, SC - Brasil

camilafarias@outlook.com, {gustavo,mem}@unesc.net

**Abstract.** Sepsis is a disease that is becoming a major cause of death in intensive care units. The difficulty in the diagnosis of sepsis is one of the factors that contribute to its high mortality. Thus, a computer simulation was developed to present the patient was classified as a possible death or not, according to a database of sepsis associated with delirium. The Artificial Intelligence technique called Support Vector Machine to perform classification was used. For the tests in the database was also used Artificial Neural Networks. The results obtained from these tests were analyzed for the simulator could be used.

**Resumo.** A sepse é uma doença que está se tornando uma das principais causas de óbito em Unidades de Terapia Intensiva. A dificuldade no diagnóstico da sepse é um dos fatores que contribuem para sua alta mortalidade. Desta maneira, foi desenvolvido um simulador computacional que apresente se o paciente foi classificado como um possível óbito ou não, de acordo com uma base de dados de sepse associada a delirium. Foi utilizada a técnica de Inteligência Artificial chamada Máquina de Suporte Vetorial para realizar a classificação. Para a realização dos testes na base de dados utilizou-se também Redes Neurais Artificiais. Os resultados obtidos com estes testes foram analisados para que pudessem ser utilizados no simulador.

#### 1. Introdução

Nos últimos anos, a sepse continua sendo uma das principais causas de mortes em Unidades de Terapia Intensiva (UTI), apesar do desenvolvimento de métodos para o seu diagnóstico e dos esforços feitos para modificar esta realidade [SILVA e VELASCO 2007]. A sepse consiste em uma infecção e pode se manifestar de vários modos [CARVALHO e TROTTA 2003]. Em alguns casos, a sepse pode estar associada ao estado de delirium, que representa uma disfunção no cérebro [PITROWSKY 2005 apud PANDHARIPANDE et al, 2010]. Nesta situação, de sepse associada a delirium, a taxa de mortalidade cresce ainda mais [ZAMPIERI 2011, tradução nossa]. Em função destas altas taxas e na dificuldade de seu diagnóstico foi proposta a classificação de dados de sepse associada a delirium utilizando a técnica de inteligência artificial conhecida como Máquina de Suporte Vetorial (MSV).

A MSV é uma técnica de aprendizado supervisionado, que busca classificar dados novos a partir de outros que já existem [WANG 2005, tradução nossa]. Esta técnica está

presente em diversos estudos, como em categorização de textos, em análise de imagens e na bioinformática [LORENA e CARVALHO 2007].

A MSV possibilita a predição de dados novos de sepse associada a delirium utilizando a base de dados como fonte para a classificação. Além da realização da classificação dos dados de sepse associada a delirium com MSV, foi prevista também, a utilização de outra técnica de Inteligência Artificial, chamada Redes Neurais Artificiais (RNA), para que fosse possível realizar uma comparação entre as técnicas a fim de identificar qual técnica melhor classificava a base de dados utilizada.

Desta maneira, utilizando a classificação realizada, foi desenvolvido um simulador computacional, que faria a predição desses novos dados, utilizando internamente a MSV. Para se obter o melhor modelo de classificação, foram realizados testes na base de dados, onde se utilizava as técnicas de MSV e RNA, proporcionando também uma análise dos resultados adquiridos com estes testes.

Com os testes efetuados na base de dados, foi possível realizar a comparação entre MSV e RNA, possibilitando a identificação da melhor técnica para esta base de dados. A partir desta comparação, o simulador computacional desenvolvido conseguiu realizar a predição dos dados de sepse associada a delirium, utilizando a base de dados já existente de sepse associada a delirium como modelo na classificação.

## **2. Sepse**

A sepse pode ser considerada como uma resposta do organismo para uma infecção [COHEN 2002, tradução nossa]. Ela possui três classificações: sepse, sepse grave e choque séptico. Elas referem-se a gravidade do quadro do paciente, de acordo com os sintomas apresentados por ele [BAUDOIN 2009, tradução nossa].

Esta doença exige um atendimento eficaz e preciso [BRITO e SUZE 2010]. Quando um controle precoce é feito, a taxa de mortalidade pode ser reduzida de 10 a 15% [SILVA e VELASCO 2007]. Quando não tratados corretamente, os pacientes acometidos por sepse podem sofrer com problemas relacionados a disfunções nos órgãos, bem como vir a óbito [CARVALHO e TROTTA 2003].

Mesmo com tantos avanços, a taxa de mortalidade por sepse ainda permanece elevada. Estudos mostram que esta taxa pode variar de 16,70%, nos casos de apenas sepse até 65,30%, nos casos de choque séptico [JUNCAL et al 2011].

De acordo com os estudos feitos por Zampieri [2011, tradução nossa], esta taxa pode tornar-se ainda maior quando a sepse está associada a delirium. Conforme ele descreve, esta taxa pode aumentar 49%. O delirium é uma perturbação de consciência associada a uma alteração na cognição do paciente [ASSOCIATION 2000, tradução nossa]. É comum encontrar pacientes sépticos em estado de delirium, segundo destaca Kress [2010, tradução nossa], contudo não se atribui a devida importância a isso [SONEVILLE et al 2013, tradução nossa]. Desta maneira, ainda não há tratamento adequado para sepse associada a delirium [EBERSOLDT SHARSHAR e ANNANE, 2007, tradução nossa].

### 3. Classificação de Dados

A mineração de dados busca resolver problemas com bases de dados já existentes, por meio da descoberta de padrões nos dados já existentes [WITTEN FRANK e HALL 2011]. É uma área que abrange técnicas de estatística, inteligência artificial, banco de dados, em tre outras [CARDOSO e MACHADO 2008]. Com ela, torna-se possível a geração de novas informações a partir de dados que já existem. Sendo que, estas novas informações, quando interpretadas podem ser aplicadas na prática [STEINWART e CHRISTMANN 2008, tradução nossa].

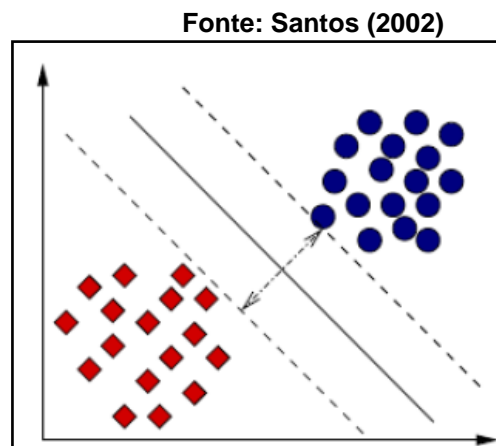
Para a criação de um modelo que faça a predição de novos dados, é necessário utilizar um algoritmo que faça a classificação destes dados. Para escolher o algoritmo da classificação, deve-se avaliar o que se mostra mais eficiente para os dados utilizados [BATISTA 2003].

No trabalho desenvolvido, utilizou-se dois métodos de classificação de dados, a MSV e a RNA.

#### 3.1. Máquina de Suporte Vetorial

A MSV é uma máquina de aprendizado abstrata, que utiliza um conjunto de dados como treinamento e a partir deste treinamento, tenta generalizar e prever de forma correta, o comportamento do restante dos dados [CAMPBELL e YIMING 2000, tradução nossa].

Na classificação, procura-se um classificador que consiga a maior separação entre as classes da base de dados. Esse classificador é conhecido como Hiperplano Ótimo de Separação, como mostra a figura 1 [MENDOZA 2009].



**Figura 1 – Hiperplano Ótimo de Separação na Máquina de Suporte Vetorial**

Existem classificações, chamadas Classificações Lineares, em que é possível fazer a distinção das classes de forma simples, como a mostrada na figura 1. Quando a classificação não é linear, que é o que acontece com a maior parte das classificações, utiliza-se as funções de kernel para auxiliar a classificação [LORENA e CARVALHO 2007]. Essas funções tem como objetivo projetar os dados em um espaço maior, permitindo que a classificação seja realizada [SANTOS 2002].

No trabalho desenvolvido, foram utilizados os algoritmos SMO e LIBSVM para a classificação dos dados. O algoritmo SMO é conhecido pela rapidez com que realiza a classificação, e o algoritmo LIBSVM, é uma função de um software de mesmo nome, onde já possui o algoritmo SMO implementado internamente.

### 3.2. Redes Neurais Artificiais

A RNA consiste em neurônios que recebem diretamente entradas da rede, chamando assim de camada de entrada. Cada camada recebe como entrada a saída da outra camada até chegar a camada de saída, como mostra a figura 2 [KÓVACS 2006].

Conforme destaca Moll (2010), as redes também são compostas por pesos. Estes pesos representam as conexões entre os neurônios de cada camada.

Para o desenvolvimento do trabalho, utilizou-se dois tipos de rede: *Multilayer Perceptron* (MLP) e *Radial Bases Function* (RBF). A primeira é constituída por uma camada de neurônios de entrada e uma ou mais camadas intermediárias e a camada de saída [RAMALHO 2008]. Já a segunda, é composta por apenas uma camada de entrada, uma intermediária e uma de saída [BONIFÁCIO 2010].

Estas redes foram utilizadas com o intuito de que os seus resultados pudessem ser comparados posteriormente com os resultados obtidos por meio de MSV.

## 4. Métodos de Treinamento e Teste

Para realizar o treinamento e teste do classificador, utiliza-se métodos de treinamento e teste. Os métodos *Leave-One-Out* e *k-fold* foram os escolhidos para serem utilizados neste projeto, devido a sua popularidade e sua efetividade nos resultados. O método *Leave-One-Out* foi o mais utilizado. Neste método, tem-se duas bases, a de treinamento e a de teste, a de treinamento é composta por todos os registros da base, menos um, que é o que está inserido da base de teste. Este processo é realizado para cada registro, realizando, desta maneira, o método *Leave-One-Out* [WESTOM 1999, tradução nossa].

Para comparar os resultados obtidos com o método *Leave-One-Out*, utilizou-se também o método *k-fold*. Neste método, a base é subdividida em  $k$  partes. Durante o processo de treinamento, uma das partes é utilizada para o teste, enquanto as outras são utilizadas para treinamento [TAN STEINBACH e KUMAR, 2009].

## 5. Medidas de Análise do Conhecimento Gerado

Com o propósito de avaliar os resultados obtidos através da classificação de dados, pode-se utilizar medidas de análise de conhecimento gerado. No trabalho, foram utilizadas a matriz de confusão e a Curva Característica de Operação de um Receptor (ROC), popularmente conhecida como curva ROC.

### 5.1. Matriz de Confusão

É composta pelo número de classes em que resulta a classificação realizada. A diagonal principal representa o número de acertos das classes identificadas, já o restante dos números da matriz, referem-se ao número de erros da classificação [PESSOA et al 2010].

Com a matriz de confusão é possível visualizar algumas medidas de análise como sensibilidade, especificidade e acurácia [PESSOA et al 2010]. A partir destas medida, é possível avaliar a qualidade da classificação realizada.

## 5.2. Curva ROC

A curva ROC consiste em uma forma gráfica para exibição das taxas de um classificador. Ela é de grande utilidade na comparação de modelos de classificador [TAN, STEINBACH e KUMAR 2009].

Uma maneira de obtenção da curva ROC, seria a realização de um a comparação entre o que foi predito e o que realmente é verdadeiro (WITTEN; FRANK, 2005).

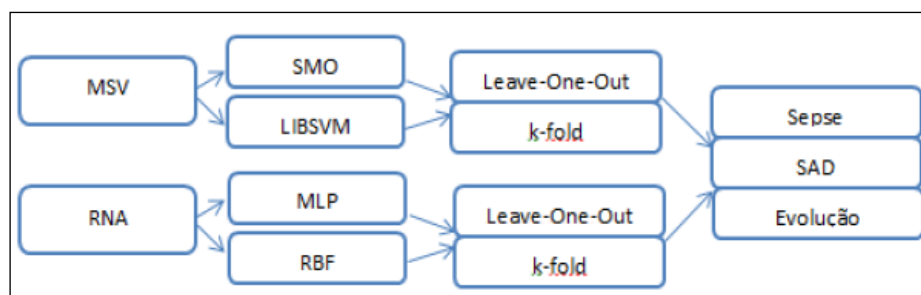
## 6. Trabalho Desenvolvido

No trabalho realizado, utilizou-se uma base de dados coletada em uma Unidade de Terapia Intensiva, localizada na cidade de Criciúma, Santa Catarina. A base de dados contém setenta e oito registros e é composta por dezesseis atributos e três classes.

Os atributos são biomarcadores coletados dos pacientes e as classes são classificações possíveis na base dados, que são: Sepse, Sepse Associada a Delirium (SAD) e Evolução, que indica se o paciente veio a óbito ou não.

Para a realização do trabalho, utilizou-se o software Weka. Ele é desenvolvido pela universidade de Waikato, situada na Nova Zelândia, e é um programa de código aberto e gratuito [HALL 2009, tradução nossa]. Ele foi escolhido devido a sua praticidade e por ser amplamente utilizado no meio acadêmico.

Para a realização dos testes na base de dados, utilizou-se o esquema apresentado na figura 2. Nas classificações de MSV, foram utilizadas duas funções de kernel para cada algoritmo. Para o algoritmo SMO, aplicou-se as funções Polykernel e RBF, enquanto para o algoritmo LIBSVM, utilizou-se as funções Polynomial e RBF. Nas classificações com RNA, utilizaram-se as redes MLP e RBF. Em ambas as classificações aplicou-se os métodos de treinamento e teste *Leave-One-Out* e *k-fold*.



**Figura 2. Esquema de testes na base de dados**

A partir do esquema demonstrado, foram realizados os testes na base, alterando os parâmetros de cada função. Estes parâmetros influenciavam na relação entre os erros de treinamento e a complexidade do modelo gerado, definiam o quanto um único registro

influiu sobre o resultado da classificação, o quanto uma rede iria aprender e o número de camadas de cada rede.

Desta maneira, em cada teste, foram modificados os parâmetros e armazenados os resultados de cada classificação.

Para o desenvolvimento do simulador, utilizou-se a função de kernel RBF e o algoritmo SMO, de MSV. Este simulador foi desenvolvido no ambiente de desenvolvimento chamado Netbeans IDE 7.3. O projeto foi criado como uma aplicação Web, devido as vantagens oferecidas pela mesma e por melhor se enquadrar no que foi proposto. O projeto criado, possui uma página Web referente a interface de interação com o usuário e as classes referentes a modelagem e a própria classificação dos registros.

Ao abrir o navegador, o usuário deve informar os dados necessários para classificação, valores do tipo numérico, que estão relacionados aos biomarcadores. Quando o usuário solicita a classificação, são chamados internamente os métodos que farão a classificação do quadro do paciente, de acordo com os valores informados (figura 4).

**Figura 4. Simulador**

| Paciente           |            |                 |                 |
|--------------------|------------|-----------------|-----------------|
| João da Rosa       |            |                 |                 |
| BNDF_d1            | BNDF_d2    | IL1b_d1         | IL1b_d2         |
| 5736.832           | 8782.398   | 205.1923        | 2342.0          |
| TNF_D1             | TNF_D2     | STNF_R1_D1      | STNF_R1_D2      |
| 36.357             | 32.424     | 3716.185        | 2507.254        |
| STNF_R2_D1         | STNF_R2_D2 | ADIPONECTINA_D1 | ADIPONECTINA_D2 |
| 12303.78           | 6390.651   | 48182.03        | 57038.58        |
| IL6                | IL10       | IL6IL10ratio    | Neopterin       |
| 447.5              | 7.8        | 1.15            | 13.31           |
| <b>Classificar</b> |            |                 |                 |
| Não Morte          |            |                 |                 |

Para esta classificação, utilizou-se uma biblioteca do Weka, que permite acessar as funções do próprio software. A classificação é feita utilizando um arquivo que contém a base de treinamento, e um arquivo que possui a base de teste. O classificador é treinado com o arquivo de treinamento e é testado com o arquivo de teste, sendo que este sempre irá ter os valores inseridos pelo usuário no simulador.

## 7. Resultados Obtidos

A partir dos resultados obtidos com os testes de cada classe, verificou-se quais foram as melhores taxas para cada função com o método *Leave-One-Out*, para que os dados pudessem ser tabulados para análise.

2 Tabela 1. Classificação de Sepses

| Classe | Técnica | Algoritmo | Função de Kernel | Acertos |       | Erros |       | Curva ROC |
|--------|---------|-----------|------------------|---------|-------|-------|-------|-----------|
|        |         |           |                  | Nº      | %     | Nº    | %     |           |
| Sepses | MSV     | SMO       | Polykernel       | 63      | 80,77 | 15    | 19,23 | 0.808     |
| Sepses | MSV     | SMO       | RBF              | 63      | 80,77 | 15    | 19,23 | 0.808     |
| Sepses | MSV     | LIBSVM    | Polynomial       | 59      | 75,64 | 19    | 24,36 | 0.756     |
| Sepses | MSV     | LIBSVM    | RBF              | 62      | 79,49 | 16    | 20,51 | 0.795     |
| Sepses | RNA     | MLP       |                  | 62      | 79,49 | 16    | 20,51 | 0.795     |
| Sepses | RNA     | RBF       |                  | 55      | 70,51 | 23    | 29,49 | 0.649     |

Com base nos dados apresentados na tabela 1, pode-se verificar que a classificação de sepsis pode ser considerada boa. A curva ROC está próxima do valor 1, o que indica que realmente foi uma boa classificação. Nota-se também, que os melhores resultados foram obtidos com MSV, chegando a uma taxa de acertos de 80%.

Tabela 2. Classificação de SAD

| Classe | Técnica | Algoritmo | Função de Kernel | Acertos |       | Erros |       | Curva ROC |
|--------|---------|-----------|------------------|---------|-------|-------|-------|-----------|
|        |         |           |                  | Nº      | %     | Nº    | %     |           |
| SAD    | MSV     | SMO       | Polykernel       | 68      | 87,18 | 10    | 12,82 | 0.822     |
| SAD    | MSV     | SMO       | RBF              | 68      | 87,18 | 10    | 12,82 | 0.822     |
| SAD    | MSV     | LIBSVM    | Polynomial       | 69      | 88,46 | 9     | 11,54 | 0.846     |
| SAD    | MSV     | LIBSVM    | RBF              | 64      | 82,05 | 14    | 17,95 | 0.787     |
| SAD    | RNA     | MLP       |                  | 64      | 82,05 | 14    | 17,95 | 0.787     |
| SAD    | RNA     | RBF       |                  | 65      | 83,33 | 13    | 16,67 | 0.85      |

Conforme mostra a tabela 2, na classificação de SAD, observa-se que novamente a MSV obteve um desempenho melhor, totalizando sessenta e nove acertos. Analisando a curva ROC, percebe-se que esta classificação também pode ser considerada boa, superando até os resultados gerados pela classificação de Sepsis.

Tabela 3. Classificação de Evolução

| Classe   | Técnica | Algoritmo | Função de Kernel | Acertos |       | Erros |       | Curva ROC |
|----------|---------|-----------|------------------|---------|-------|-------|-------|-----------|
|          |         |           |                  | Nº      | %     | Nº    | %     |           |
| Evolução | MSV     | SMO       | Polykernel       | 56      | 71,79 | 22    | 28,21 | 0.581     |
| Evolução | MSV     | SMO       | RBF              | 59      | 75,64 | 19    | 24,36 | 0.558     |
| Evolução | MSV     | LIBSVM    | Polynomial       | 58      | 74,36 | 20    | 25,64 | 0.582     |
| Evolução | MSV     | LIBSVM    | RBF              | 58      | 74,36 | 20    | 25,64 | 0.582     |
| Evolução | RNA     | MLP       |                  | 63      | 80,77 | 15    | 19,23 | 0.737     |
| Evolução | RNA     | RBF       |                  | 59      | 75,64 | 19    | 24,36 | 0.718     |

A tabela 3, apresenta os resultados da classificação de Evolução, que foi a classe utilizada para o desenvolvimento do simulador. Observa-se que esta classificação foi a que gerou resultados menos satisfatórios. Analisando a curva ROC, entende-se que esta classificação pode ser considerada a menos efetiva que as outras mostradas anteriormente.

Pode-se notar também, que nesta classe, as classificações com RNA obtiveram resultados melhores que as feitas com MSV.

Ainda com a realização dos testes, verificou-se a possível influência de alguns biomarcadores no resultado da classificação. Percebeu-se que existiam alguns biomarcadores

que não exerciam influência sobre a classificação, já que a sua retirada não alterou as taxas da classificação. Enquanto outros, modificavam a taxa de acerto para um valor superior, demonstrando assim, uma influência nos resultados obtidos.

## 8. Conclusão

A mineração de dados se faz extremamente importante para auxiliar as pessoas em determinadas tarefas. A extração de conhecimento e a análise que se pode fazer a partir de uma classificação de dados tem muita utilidade em diversas áreas. Dentre as técnicas de mineração de dados disponíveis, foi utilizada a MSV. Com ela, buscou-se desenvolver um simulador que fizesse uma predição a partir da classificação de uma base de dados de SAD.

A Sepsis vem sendo, nos últimos anos, uma das principais causas de morte em UTI. O desenvolvimento do simulador foi justamente para auxiliar no diagnóstico desta doença, o que poderia agilizar a detecção da gravidade da situação do paciente.

Para o desenvolvimento deste simulador, foi preciso conhecer o funcionamento da MSV, para aplicá-la nesta base de dados de SAD. Conhecendo a maneira como era feita a classificação, utilizou-se a ferramenta WEKA para realizar os testes com os registros da base de dados.

Para selecionar a melhor classificação, a partir dos dados tabulados, foi feita uma análise dos dados por meio da matriz de confusão e da curva ROC. Ambas indicaram qual classificação obteve os melhores resultados, indicando assim, qual que deveria ser utilizada no simulador. Com a matriz de confusão, pode-se fazer a análise dos resultados e pode-se comparar os resultados de MSV com os de RNA. Ao comparar estas duas técnicas, observou-se que para duas classes da base, a MSV obteve resultados superiores aos da RNA. Com as análises devidas, foi possível desenvolver o simulador baseado na melhor classificação com MSV da evolução dos pacientes.

A pesquisa demonstrou-se importante em função da base de dados utilizada ser de extrema importância. Além disso, as diversas análises que esta pesquisa possibilitou também motivaram ainda mais o desenvolvimento deste trabalho.

## 9. Referências

- ASSOCIATION, American Psychiatric (2000). **Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR**. American Psychiatric Pub, 2000.
- BATISTA, Gustavo Enrique de Almeida Prado Alves. **Pré-processamento de dados em aprendizado de máquina supervisionado**. 2003. Tese (Doutorado) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-06102003-160219/>>. Acesso em: 15. Nov. 2013.
- BAUDOUIN, Simon (2009). **Sepsis**. Londres: Springer. 111 p.
- BONIFÁCIO, Fernando Nunes. **Comparação entre as Redes Neurais Artificiais MLP, RBF e LVQ na Classificação de Dados**. 2010. 94 f. TCC (Graduação) - Curso de Ciência de

- Computação, Unioeste – Universidade Estadual do Oeste do Paraná, Cascavel, 2010. Disponível em: <[www.inf.unioeste.br/~tcc/2010/TCC-FernandoNunes.pdf](http://www.inf.unioeste.br/~tcc/2010/TCC-FernandoNunes.pdf)>. Acesso em: 03 jun. 2014.
- BRITO, Cândida Márcia de e ZUSE, Carmen Lucia (2010). Demandas de enfermagem ao paciente em sepse, sepse severa e choque séptico na unidade de emergência. **Nursing (São Paulo)**, São Paulo, v. 14, n. 149 , p.529-535, out. 2010.
- CAMPBELL, Colin e YING, Yiming. **Learning with Support Vector Machines**. Morgan & Claypool Publishers, 2011.
- CARDOSO, Olinda Nogueira Paes e MACHADO, Rosa Teresa Moreira. (2008) Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**, Rio de Janeiro, v. 42, n. 3, p.495-528, jun. 2008. Disponível em: <<http://www.scielo.br/pdf/rap/v42n3/a04v42n3.pdf>>. Acesso em: 06 out. 2013.
- CARVALHO, Paulo R. A. e TROTTA, Eliana de A. (2003) Avanços no diagnóstico e tratamento da sepse. **Jornal de Pediatria**, Porto Alegre, v. 79. Disponível em: <[http://www.scielo.br/scielo.php?pid=S0021-75572003000800009&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0021-75572003000800009&script=sci_arttext)>. Acesso em: 17 ago. 2013.
- COHEN, Jonathan. (2002) The immunopathogenesis of sepsis. **Nature**, Washington Dc, v. 420, p.19-26, 19 dez. Disponível em: <[http://www.math.pitt.edu/~cbsg/Materials/Cohen\\_nature01326.pdf](http://www.math.pitt.edu/~cbsg/Materials/Cohen_nature01326.pdf)>. Acesso em: 02 set. 2013.
- EBERSOLDT, Marion; SHARSHAR, Tarek e ANNANE, Djillali (2007). Sepsis-associated delirium. **Intensive Care Med**, Ss, p. 941-950. 05 abr. 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17410344>>. Acesso em: 14 set. 2013.
- HALL et al. The WEKA Data Mining Software: An Update. **Acm Sigkdd Explorations Newsletter**, Nova York, v. 11, n. 1, p.10-18, jun. 2009. Disponível em: <<http://www.sigkdd.org/sites/default/files/issues/11-1-2009-07/p2V11n1.pdf>>. Acesso em: 31 maio 2014.
- JUNCAL, Verena Ribeiro et al. (2011) Impacto clínico do diagnóstico de sepse à admissão em UTI de um hospital privado em Salvador, Bahia. **J Bras Pneumol.**, São Paulo, v. 37, n. 1, p.85-92, fev. 2011. Disponível em: <[http://www.scielo.br/scielo.php?pid=S1806-37132011000100013&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S1806-37132011000100013&script=sci_arttext)>. Acesso em: 09 set. 2013.
- KÓVACS, Zsolt (2006). **Redes Neurais Artificiais: fundamentos e aplicações**. 4. ed. São Paulo: Editora Livraria da Física, 2006.
- KRESS, John P. (2010) The complex interplay between delirium, sepsis and sedation. **Critical Care**, v. 14, n. 164, 14 jun. 2010. Disponível em: <<http://ccforum.com/content/14/3/164>>. Acesso em: 14 set. 2013.
- LORENA, Ana Carolina e CARVALHO, André C. P. L. F. de. (2007) Uma Introdução às Support Vector Machines. **Rita**, Porto Alegre, v. 14, n. 2 , p.43-67. Disponível em: <[seer.ufrgs.br/rita/article/download/rita\\_v14\\_n2\\_p43-67/3543](http://seer.ufrgs.br/rita/article/download/rita_v14_n2_p43-67/3543) >. Acesso em: 28 abr. 2013.
- MENDOZA, Leonardo Alfredo Forero. **Redes Neurais e Máquinas de Vetores de Suporte no reconhecimento de locutor usando coeficientes MFC e características do sinal glotal**.

2009. 128 f. Dissertação (Mestrado) - Universidade Federal Fluminense, Niterói, 2009. Disponível em: <[http://www.btdt.ndc.uff.br/tde\\_arquivos/38/TDE-2009-11-17T161830Z-2282/Publico/Dissert%20Leonardo%20Alfredo%20Forero%20Mendoza.pdf](http://www.btdt.ndc.uff.br/tde_arquivos/38/TDE-2009-11-17T161830Z-2282/Publico/Dissert%20Leonardo%20Alfredo%20Forero%20Mendoza.pdf)>. Acesso em: 19 out. 2013.
- MOLL, Vinícius. **Detecção de intrusão usando técnicas de aprendizagem de máquinas**. 2010. 150 f. Dissertação (Mestrado). Universidade Federal de Santa Catarina, Florianópolis, 2010. Disponível em: <<https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/94110/280714.pdf?sequence=1>>. Acesso em: 15 nov. 2013.
- PESSOA et al. Mineração de Dados Meteorológicos Associada a Eventos Severos no Pantanal Sul Matogrossense. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 33., 2010, Águas de Lindóia, SP. **Anais...** São Carlos, SP: SBMAC, 2010. p. 652-658. Disponível em:
- PITROWSKY, Melissa Tassano et al. (2010) Importância da monitorização do delirium na unidade de terapia intensiva. **Revista Brasileira de Terapia Intensiva**, São Paulo, v. 22, n. 3, set. Disponível em:<[http://www.scielo.br/scielo.php?pid=S0103-507X2010000300010&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0103-507X2010000300010&script=sci_arttext)>. Acesso em: 15 set. 2013.
- RAMALHO, José Pinto. **Oxicorte**: estudo da transferência de calor e modelamento por redes neurais artificiais de variáveis do processo. 2008. Tese (Doutorado em Engenharia Metalúrgica e de Materiais) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2008. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3133/tde-30092008-150619/>>. Acesso em: 01 jun. 2014.
- SANTOS, Eulanda Miranda Dos. **Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência**. 2002. 111 f. Dissertação (Mestrado) - Universidade Federal da Paraíba, Campina Grande, 2002. Disponível em: <[http://docs.computacao.ufcg.edu.br/posgraduacao/dissertacoes/2002/Dissertacao\\_EulandaMirandadosSantos.pdf](http://docs.computacao.ufcg.edu.br/posgraduacao/dissertacoes/2002/Dissertacao_EulandaMirandadosSantos.pdf)>. Acesso em: 21 out. 2013.
- SILVA, Fabiano Pinheiro da e VELASCO, Irineu Tadeu (2007) **Sepse**. Barueri: Manole.
- SONNEVILLE, Romain et al. (2013) Understanding brain dysfunction in sepsis. **Annals Of Intensive Care**, p. 1-11. 29 maio 2013. Disponível em: <<http://www.annalsofintensivecare.com/content/3/1/15>>. Acesso em: 21 set. 2013.
- STEINWART, Ingo e CHRISTMANN, Andreas. (2008) **Support Vector Machines**. Springer, 2008
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining - Mineração de Dados**. Ciência Moderna, 2009.
- WANG, Lipo. (2005) **Support Vector Machines: Theory and Applications**. Springer.
- WESTON, Jason A. E.. **Extensions to the Support Vector Method**. 1999. 127 f. Tese (Doutorado) - University Of London, Eghan, Surrey, England, 1999. Disponível em:

<<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.2796>>. Acesso em: 03 nov. 2013.

WITTEN, Ian H. FRANK, Eibe e HALL, Mark A. (2011) **Data Mining: Practical Machine Learning Tools and Techniques**. Elsevier, 2011.

ZAMPIERI, Fernando Godinho et al. (2011) Sepsis-associated encephalopathy: not just delirium. **Clinics**, São Paulo, v. 66, n. 10, p.1825-1831. Disponível em: <[http://www.scielo.br/scielo.php?pid=S1807-59322011001000024&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S1807-59322011001000024&script=sci_arttext)>. Acesso em: 14 set. 2013.