

LORENA SOPHIA CAMPOS DE OLIVEIRA

**REDES BAYESIANAS DINÂMICAS COM DEFINIÇÃO DE LIMIAR
APLICADAS AO ESTUDO DE CASO DETECÇÃO DE
EXTRASSÍSTOLE VENTRICULAR**

Proposta de Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Doutor em Engenharia Elétrica.

Orientador: Prof. Dr. Mario Sarcinelli Filho.

Co-orientador: Prof. Dr. Rodrigo Varejão Andreão.

VITÓRIA
2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

O48r Oliveira, Lorena Sophia Campos de, 1977-
Redes bayesianas dinâmicas com detecção de limiar aplicadas
ao estudo de caso detecção de extrassístole ventricular / Lorena
Sophia Campos de Oliveira. – 2010.
137 f : il.

Orientador: Mário Sarcinelli Filho.

Co-Orientador: Rodrigo Varejão Andreão.

Tese (Doutorado em Engenharia Elétrica) – Universidade
Federal do Espírito Santo, Centro Tecnológico.

1. Redes bayesianas. 2. Complexos ventriculares prematuros.
3. Técnicas de apoio para a decisão. 4. Inteligência artificial. 5.
Incerteza. 6. Teoria bayesiana de decisão estatística. I. Sarcinelli
Filho, Mário. II. Andreão, Rodrigo Varejão. III. Universidade
Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

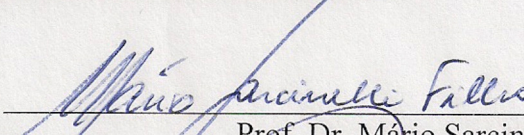
LORENA SOPHIA CAMPOS DE OLIVEIRA

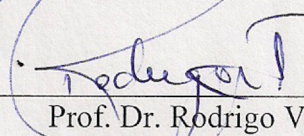
**REDES BAYESIANAS DINÂMICAS COM DEFINIÇÃO DE LIMIAR
APLICADAS AO ESTUDO DE CASO DETECÇÃO DE
EXTRASSÍSTOLE VENTRICULAR**

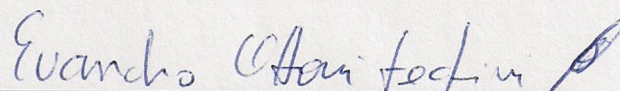
Tese submetida ao programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Doutor em Engenharia Elétrica.

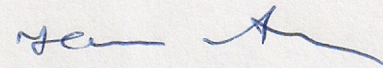
Aprovada em 10 de setembro de 2010.

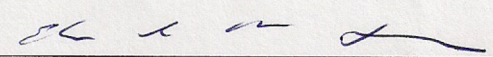
COMISSÃO EXAMINADORA

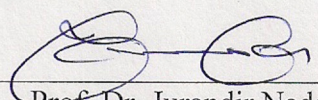

Prof. Dr. Mário Sarcinelli-Filho
Orientador - Universidade Federal do Espírito Santo


Prof. Dr. Rodrigo Varejão Andreão
Co-orientador - Instituto Federal de Educação de
Ciência e Tecnologia do Espírito Santo


Prof. Dr. Evandro Ottoni Teatini Salles
Universidade Federal do Espírito Santo


Prof. Dr. Hannu Tapio Ahonen
Universidade Federal do Espírito Santo


Prof. Dr. Edson de Paula Ferreira
Universidade Federal do Espírito Santo


Prof. Dr. Jurandir Nadal
Universidade Federal do Rio de Janeiro

VITÓRIA
2010

“Não há problema que não possa ser solucionado pela paciência.”

(Chico Xavier)

“Quem não compreende um olhar tampouco compreenderá uma longa explicação.”

(Mário Quintana)

Agradecimentos

Durante estes quatro anos senti que brinquei com minha imaginação, meu conhecimento, minha capacidade... e finalmente cheguei ao objetivo tão esperado: a Tese.

Nada na vida conquistamos sozinhos. Sempre precisamos de outras pessoas para alcançar os nossos objetivos. Enfim, este trabalho deve muito a algumas pessoas e instituições, por diferentes razões, e eu gostaria de agradecer especialmente:

Primeiramente a Deus, por me dar o dom da vida e estar sempre presente na minha existência, ao meu Anjo da Guarda que sempre me inspira e protege.

Ao meu orientador, Prof. Dr. Mario Sarcinelli Filho, pela orientação, pelo bom convívio nestes quatro anos de trabalho, principalmente, a percorrer novos caminhos, pela coragem de ousar trabalhar com novas idéias e conceitos, correndo os riscos inerentes a esta atitude.

Ao meu co-orientador Prof. Dr. Rodrigo Varejão Andreão, que para mim é considerado também orientador, que sempre demonstrou acreditar no meu potencial, pela oportunidade oferecida, pela orientação, por ouvir com interesse e ânimo todas as questões, dúvidas e problemas que surgiam durante o processo de reflexão.

Ao meu marido, Vitor, que sempre esteve ao meu lado nos momentos mais difíceis, aguentando o estresse, os fins de semana estudando, e principalmente por me apoiar durante um ano em que eu estive fora do país.

Aos meus pais Geraldo e Graça que me ensinaram a viver com dignidade, trabalho, amor, respeito, e acima de tudo, sempre estiveram ao meu lado e apoiaram-me.

Ao meu irmão Talles, pelo companheirismo, amizade, carinho e por sempre estar ao meu lado.

A todos os meus familiares, e principalmente as minhas tias Glorinha, Ieda e Sonia, que me receberam como mães neste período em que me arrisquei em uma nova cidade, num novo

trabalho e novos amigos. Aos meus sogros, Vera e Eraldo, que sempre estão torcendo muito por mim.

A grande amiga Luciana, com quem eu pude conversar sobre a tese durante estes anos, e que sempre me ajudou com conselhos muito importantes.

A CAPES, pelo apoio financeiro, que permitiu que eu me dedicasse durante estes quatro anos a esta Tese, bem como permitiu a bolsa de doutorado sanduíche a França, no TELECOM Sud-Paris.

A todos os professores da pós-graduação da UFES, que estiveram envolvidos direta ou indiretamente no meu aprendizado.

Aos colegas do LAI, que me receberam muito bem, e me aguentaram durante estes anos, ao Felipe, Cristiano, Márcia, Sandra, Mariana, Flávio, Daniel, Timótio, Wanderlei, André, Karin, enfim, a todos amigos do LAI I, II e Labtel.

Aos membros da banca de qualificação, que me deram ótimos conselhos e sugestões.

Merci à tous mes amis de Paris, principalement les enseignants Jérôme Boudy et Bernadette Dorizzi pour votre attention et orientation. Grand merci a tous de TELECOM Sud-Paris.

Resumo

Esta Tese propõe uma abordagem Bayesiana Dinâmica com definição de limiar para desenvolver um sistema de apoio à decisão médica. São empregadas tanto Redes Bayesianas estáticas quanto Redes Bayesianas Dinâmicas para classificação de um tipo específico de arritmia cardíaca, utilizado como estudo de caso. Estas metodologias são utilizadas por serem adequadas para o tratamento de incerteza, presente no raciocínio clínico e que por isto mesmo deve ser levada em conta em qualquer sistema de auxílio ao diagnóstico, pois são ferramentas de classificação probabilística. Várias topologias de Redes Bayesianas são implementadas e testadas, para que seja possível encontrar a estrutura mais adequada ao problema proposto. Especificamente, é considerada a detecção de Extrassístoles Ventriculares (ESV) que é a anormalidade do ritmo cardíaco em que os ventrículos se contraem mais cedo do que o esperado. Os resultados obtidos com a utilização de Rede Bayesiana Dinâmica com definição de limiar para detecção de Extrassístoles Ventriculares chegaram a 99,53%, 100%, 100% e 99,97% para os valores de Sensibilidade, Especificidade, Valor Preditivo Positivo e Valor Preditivo Negativo respectivamente, com intervalo de confiança de $\pm 0,4\%$ e $\pm 0\%$ para Sensibilidade e Valor Preditivo Positivo respectivamente.

Abstract

This work proposes a dynamic bayesian approach with threshold setting to develop a system to support a cardiologist in making a decision, in terms of classifying a heart beat. Dynamic Bayesian Networks (DBN) and static Bayesian Networks (BN) are adopted for performing such classification, since they are very suitable to deal with the uncertainties involved in the cardiologist's reasoning, thanks to their probabilistic and logic model. Different BN topologies are implemented and tested, aiming at finding the one more suitable to the problem under study. Specifically speaking, it is considered the detection of premature beats (PVC), which are one kind of arrhythmia related to the premature contraction of the ventricles. The results obtained with the use of Dynamic Bayesian Network with threshold setting for detection of premature beats reached 99.53%, 100%, 100% and 99.97% for the values of Sensitivity, Specificity, Positive Predictive Value and Negative Predictive Value respectively, and confidence interval of $\pm 0.4\%$ and $\pm 0\%$ for sensitivity and positive predictive value respectively.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1: Introdução	01
1.1 Objetivo Geral.....	03
1.2 O Eletrocardiograma	04
1.3 Componentes Individuais do Eletrocardiograma	06
1.4 Extrassístole Ventricular	11
1.5 Metodologia	16
1.6 Contribuições da Tese	18
1.7 Organização do Testo	18
Capítulo 2: Redes Bayesianas (RB)	20
2.1 Introdução	20
2.2 Probabilidade: Uma Visão Geral	24
2.3 Inferência Bayesiana	26
2.4 Aprendizado da Rede Bayesiana.....	32
2.5 Rede Bayesiana Dinâmica	36
2.5.1 Inferência em Rede Bayesiana Dinâmica	40
Capítulo 3: Redes Bayesianas Estáticas para a Detecção de Estra-Sístoles Ventriculares	46
3.1 Informações Preliminares à Implementação da Rede Bayesiana Estática	47
3.2 Primeiro Momento: Comparação entre RB e Outras Metodologias de Classificação .	56
3.2.1 Classificação Heurística de ESV (utilizando limiares).....	56
3.2.2 Redes Bayesianas com Estimação Empíricas das Tabelas de Probabilidade	59
3.2.3 Redes Bayesianas com aprendizado bayesiano	59
3.3 Segundo Momento: Fusão de Canais	61
3.3.1 Rede Bayesiana com Fusão de Canais e Nós que Armazenam Informações Ateriores	62

3.3.2 Rede Bayesiana com Apenas um canal	63
3.3.3 Redes Bayesianas com Fusão de Canais.....	64
3.3.4 Conclusão Referentes à Rede Bayesiana e à Base de Dados QT <i>database</i>	65
3.3.5 Conclusão e Perspectivas	66
3.4 Terceiro Momento: RB Utilizando Diferentes Algoritmos de Segmentação do Sinal	67
3.5 Quarto Momento: RB Utilizando a Onda P e Fusão de Resultados	70
3.5.1 Impacto da onda P na RB.....	70
3.5.2 Rede Bayesiana com Fusão de Resultados (FR)	72
3.6 Considerações Sobre o Uso de RB Estática.....	75
Capítulo 4: Rede Bayesiana Dinâmica com Definição de Limiar	77
4.1 <i>Data Warehouse</i> (DW)	78
4.2 <i>Data Mart</i> (DM).....	80
4.3 <i>Data Mining</i>	81
4.4 <i>Data Mart</i> para detecção de ESVs	83
4.4.1 Método de Seleção de Registros para Classificação de ESVs	83
4.4.2 Rede Bayesiana Estática para Classificação de ESVs.....	87
4.4.3 Rede Bayesiana Dinâmica (RBD) para detecção de ESV	88
Capítulo 5: Conclusão	101
5.1 Perspectivas para a Continuidade do Trabalho	104
Referências Bibliográficas.....	105
Anexo 1	116
Anexo 2.....	120
Anexo 3.....	121

Lista de Figuras

Figura 1.1	Ilustração de um batimento cardíaco observado em um eletrocardiograma, com a indicação de suas formas elementares.....	05
Figura 1.2	Relação entre a simulação cardíaca e o instante de aparição das ondas no eletrocardiograma: 1) Nodo sinusal; 2) Átrio; 3) Nodo AV; 4) Feixe de His; 5) Fibras de Purkinje; 6) Músculo ventricular (Andreão, 2004).....	07
Figura 1.3	As três derivações bipolares. (a) Local de colocação dos eletrodos. (b) representação vetorial das derivações e formas de ondas observadas durante a despolarização e repolarização ventricular (Andreão, 2004).....	09
Figura 1.4	Representação vetorial das derivações unipolares.....	10
Figura 1.5	Posição dos eletrodos das derivações precordiais (V1 a V6).....	11
Figura 1.6	Exemplo de propagação elétrica com foco ectópico em um dos ventrículos.....	13
Figura 1.7	Batimentos contidos em um sinal ECG, identificados segundo sua classe (N – normal; ESV – Extrassístole ventricular).....	14
Figura 1.8	Alguns Exemplos de ESVs.....	15
Figura 1.9	Exemplo de uma situação especial de ESV, Bigeminismo.....	16
Figura 2.1	Exemplo gráfico de Rede Bayesiana com nós discretos e tabelas de probabilidades associadas a cada nó.....	30
Figura 2.2	Exemplo de uma Rede Bayesiana demonstrando seus valores quantitativos e qualitativos.....	31
Figura 2.3	Exemplo de uma Bayesiana que calcula a probabilidade de ocorrência do nó ‘ESV’, dado que a probabilidade dos nós ‘BatimentoPrematuro’ e ‘BatimentoVentricular’ é de 100% V.....	32
Figura 2.4	Exemplo de uma Rede Bayesiana Clássica.....	39
Figura 2.5	Representação gráfica de uma Rede Bayesiana Dinâmica em dois instantes de tempo. (2TRB).....	40
Figura 2.6	Rede Bayesiana Dinâmica representada em 4 instantes de tempo.....	40
Figura 2.7	Exemplo de uma ‘janela móvel’ com duas fatias de tempo em uma RBD. A seta pontilhada indica o decorrer do tempo e a direção em que a janela se move. O segundo retângulo pontilhado da figura representa a janela após um certo tempo.....	43
Figura 3.1	Diagrama que representa os passos seguidos para a criação da Rede Bayesiana.....	48
Figura 3.2	Modelo geral da Rede Bayesiana.....	51
Figura 3.3	(a) Gaussianas normalizadas dos batimentos Normal e Ventricular, representando o valor da verossimilhança da onda QRS. (b) Gaussianas normalizadas dos batimentos Normal e Prematuro, representando a distância entre dois picos R sucessivos.....	53

Figura 3.4 Algoritmo em pseudocódigo que representa o cálculo da normalização na fase de Teste.....	54
Figura 3.5 Diagrama que representa o trabalho realizado nesta Tese. A Camada 0 segmenta o sinal identificando as ondas, enquanto a Camada 1 classifica o batimento de acordo com a abordagem Bayesiana.....	57
Figura 3.6 (a) Gaussianas do batimento ‘Outros’ e Ventricular (valor da verossimilhança da onda QRS); e (b) Gaussianas do batimento ‘Outros’ e Prematuro, todas com seus respectivos limiares..	58
Figura 3.7 Rede Bayesiana com fusão de canais, isto é, utilizando o canal 1 e o canal 2, bem como valores correspondentes ao batimento cardíaco anterior..	62
Figura 3.8 Rede Bayesiana sem fusão de canais, (a) utilizando o canal 1 e (b) utilizando o canal 2..	63
Figura 3.9 : Rede Bayesiana com fusão de canais.....	65
Figura 3.10 Topologia referente ao SAS.....	69
Figura 3.11 Rede Bayesiana com a onda P.	71
Figura 4.1 Arquitetura do Sistema de Auxílio ao Diagnóstico proposto, utilizando RBD.....	78
Figura 4.2 Esquema de construção de <i>Data Marts</i> a partir do <i>Data Warehouse</i>	80
Figura 4.3 Arquitetura do <i>Data Mart</i> para a classificação de ESVs.....	83
Figura 4.4 Modelo de Rede Bayesiana..	89
Figura 4.5 Rede Bayesiana Dinâmica com representação em três instantes de tempo.....	90
Figura 4.6 Rede Bayesiana Dinâmica com representação em 3 instantes de tempo.....	91
Figura 4.7 Arquitetura Final do Sistema Proposto na Tese.....	99

Lista de Tabelas

Tabela 3.1	Representação de uma Tabela de Contingência para classificação de batimentos ‘Outros’ e ‘ESV’.....	54
Tabela 3.2	Resultado da Classificação Heurística..	59
Tabela 3.3	Valores dos nós discretos, estimação empírica..	59
Tabela 3.4	Resultados obtidos com a Rede Bayesiana com estimação empírica.....	60
Tabela 3.5	Resultados obtidos para a Rede Bayesiana com a fase de aprendizado bayesiano.....	61
Tabela 3.6	Resultados para a classificação modelada pela Figura 3.7.....	62
Tabela 3.7	Resultado da classificação modelada pela Figura 3.8 (a).....	64
Tabela 3.8	Resultado da classificação modelada pela Figura 3.8 (b)..	64
Tabela 3.9	Resultado da classificação modelada pela Figura 3.9, com fusão de canais.....	65
Tabela 3.10	Resultados referentes a Base de Dados de MIT-BIH <i>Arrhythmia Database</i> e algoritmo de segmentação de Aguiar (2008) - SAS.....	69
Tabela 3.11	Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, com a onda P	72
Tabela 3.12	Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, para o conjunto de treinamento TT.....	73
	Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, para TP.....	73
Tabela 3.14	Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, ‘FR-OndaP’	75
Tabela 4.1	Tabela de Contingência para representação de três níveis de classificação..	86
Tabela 4.2	Tabela de Contingência com os resultados da classificação da RB estática..	86
Tabela 4.3	Resultado dos valores de Se, Es, VPP e VPN, baseados na Tabela de Contingência..	87
Tabela 4.4	Tabela de Contingência e resultados obtidos com a Rede Bayesiana estática após o <i>Data Mart</i> ..	88
Tabela 4.5	Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o <i>Data Mart</i> , como apresentada na Figura 4.5.....	90
Tabela 4.6	Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica apresentada na Figura 4.6, após o <i>Data Mart</i>	91
Tabela 4.7	Valores de probabilidade para 13 batimentos classificados incorretamente tanto pela RBD quanto pela RB (eles foram classificados como ‘Outros’, embora sejam ‘ESV’)..	93

Tabela 4.8 Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o <i>Data Mart</i> , com limiar de 65%.....	94
Tabela 4.9 Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o <i>Data Mart</i> , com limiar de 75%.....	95
Tabela 4.10 Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o <i>Data Mart</i> , com limiar de 90%.....	96
Tabela 4.11 : Tabela de Contingência e resultados obtidos com para a Rede Bayesiana Dinamica após o <i>Data Mart</i> , com limiar de 95% e 80%.....	97

Capítulo 1

Introdução

O objetivo da Inteligência Artificial (IA) é prover um modelo computacional de comportamento inteligente, e, sobretudo, raciocínio de senso comum (Pearl, 1991). O estudo de IA é muito antigo, não se sabendo exatamente quando se iniciou. Sabe-se com certeza que por volta de 1950 Alan Turing começou a pesquisar sobre IA. A máquina teórica de Turing indicava que sistemas poderosos poderiam ser construídos, e tornou possível o processamento de símbolos, ligando a abstração de sistemas cognitivos e a realidade concreta dos números. Turing é considerado como precursor da IA, mas foi em 1956 que a Inteligência Artificial começou a ser reconhecida como ciência (Barreto, 1999).

Em termos etimológicos, a palavra *inteligência* vem do latim *inter* (entre) e *legere* (escolher), ou seja, quer dizer ‘*escolher entre*’ uma coisa e outra. Então, é correto dizer que a palavra *inteligência* significa tudo o que se pode escolher. A palavra *artificial* vem do latim *artificiale*, e significa algo produzido pelo homem, isto é, não natural. Então, Inteligência Artificial é a simulação de algum tipo de inteligência, produzida pelo homem, para dotar as máquinas de habilidade que reproduza a inteligência humana. Estudar e desenvolver técnicas que tornam os computadores capazes de tomar decisões, de forma parecida com o ser humano, é o objetivo da Inteligência Artificial. Ela disponibiliza um conjunto amplo de modelos lógicos matemáticos, que se forem bem definidos e aplicados podem expressar decisões bem sucedidas.

A Inteligência Artificial abrange diversas áreas, cujas aplicações potenciais são inúmeras. Métodos probabilísticos são muito discutidos neste contexto. Segundo Pearl (1991), baseado em sistemas de Inteligência Artificial, existem três escolas para lidar com a incerteza: a logística, a neo-calculista e a neo-probabilista. A logística utiliza técnicas numéricas para lidar com incerteza, como lógica não monotônica. A segunda utiliza representações numéricas de incerteza, porém considera o cálculo de probabilidades inapropriado para esta tarefa, sugerindo assim um novo cálculo, como junção de crenças de Dempster-Shafer (Dempster, 1968), lógica Fuzzy (Yuan e Klir, 1995; Zadeh, 1975) e fatores de certeza. A terceira escola,

os neo-probabilistas, adotam a teoria da probabilidade, aplicando esta teoria juntamente com as facilidades computacionais necessárias para realizar tarefas relacionadas a Inteligência Artificial. As Redes Bayesianas (Pearl, 1987), ou RB, fazem parte desta última escala.

‘A principal vantagem do raciocínio probabilístico sobre o raciocínio lógico¹ é o fato de que agentes podem tomar decisões racionais mesmo quando não existe informação suficiente para se provar que uma ação funcionará’ (Charniak, 1991). Lidar com falta de informação significa lidar com incertezas. Simular o comportamento do especialista, então, significa tratar a incerteza, pois de alguma forma ele o faz.

Na visão informática da Inteligência Artificial, esta é definida como um campo de estudo que procura explicar e emular o comportamento inteligente, o aprendizado, e a adaptação em processos computacionais. Porém, nas diversas áreas que abordam este comportamento inteligente a incerteza está presente, e a teoria de probabilidades é um formalismo muito útil para a representação da incerteza.

A procura por uma metodologia para o tratamento da incerteza é muito antiga, e a Teoria da Probabilidade objetiva alcançar um resultado mais aproximado do que se entende por exato. Para determinados problemas, como o de diagnóstico médico, por exemplo, é inevitável ter que lidar com incerteza. Nos primeiros anos do estudo de probabilidade aplicada a IA, a única técnica disponível, ainda com suas limitações, era o método probabilista clássico (às vezes denominado *Naïve Bayes*² (Mitchell, 1997)). Com ele foram construídos os primeiros sistemas de diagnóstico médico, e que tiveram sucesso relativo em problemas que, no contexto atual, podem parecer pequenos em tamanho.

Além de representar a incerteza, através do uso de variáveis aleatórias, uma das principais vantagens da teoria da probabilidade é sua capacidade para exprimir relações qualitativas úteis entre crenças e processar estas relações a fim de produzir, intuitivamente, conclusões plausíveis. Ou por razões de economia de armazenamento ou por razões de generalidade, usualmente as pessoas esquecem as experiências atuais e retêm suas impressões mentais na forma de pesos, médias ou relações qualitativas abstratas, impressões estas que as ajudam a determinar ações futuras (Castillo *et al*, 1997).

¹ Refere-se ao raciocínio lógico clássico.

² O classificador *Naïve Bayes* é denominado ingênuo (*naïve*) por assumir que os atributos são condicionalmente independentes, ou seja, a informação de um evento não é relacionada com nenhum outro (Chakrabarti, 2002).

Atualmente, o uso de inferência probabilística para o desenvolvimento de sistemas inteligentes tornou-se um foco de interesse para a comunidade acadêmica. Neste trabalho, esta inferência probabilística é calculada utilizando a RB, permitindo tirar conclusões à medida que novas informações ou evidências são conhecidas. Por exemplo, na área médica pode-se concluir um diagnóstico baseando-se em sintomas, que são evidências. É possível, também, fazer inferência para estimar estados de variáveis que não foram observadas.

A construção de modelos de conhecimento bayesianos envolve três aspectos: a aprendizagem da estrutura da rede, a aprendizagem das relações de causalidade e as probabilidades associadas. Este trabalho engloba estes três aspectos, os quais serão amplamente discutidos e estudados no seu decorrer.

A representação visual das RB é feita através de grafos, um tipo de representação que permite uma melhor comunicação com o especialista do domínio. Redes Bayesianas, em geral, são intuitivas e são representações diretas do mundo. Desta forma, admitem uma maior rapidez na definição, desenvolvimento ou modificação do modelo, o que ajuda a compreender melhor o problema.

1.1 Objetivo Geral

Esta Tese se refere à utilização de técnicas de Inteligência Artificial, em particular Redes Bayesianas, desenvolvendo um sistema capaz de diagnosticar episódios de extrassístoles ventriculares (ESV) a partir dos registros de ECG, ponderando as incertezas inerentes ao problema de diagnóstico médico. O interesse particular pelo estudo do ECG deve-se ao fato de que o mesmo representa um exame simples e eficaz para o diagnóstico das arritmias cardíacas (Kadish *et al.*, 2001), além de ser imprescindível para avaliar a atividade do músculo cardíaco e detectar possíveis alterações e distúrbios. Nesse contexto, o objetivo desse trabalho consiste em propor um sistema de classificação probabilística (baseado nas abordagens de Redes Bayesianas e Redes Bayesianas Dinâmicas), de forma que os algoritmos de aprendizado sejam usados para refinar a análise do ECG ambulatorial, explorando uma abordagem que combine, da melhor maneira possível, as potencialidades de uma estratégia estatística no problema de auxílio ao diagnóstico de ESV.

1.2 O Eletrocardiograma

O eletrocardiograma, comumente conhecido como ECG, é, sem dúvida, o exame mais empregado em cardiologia, destacando-se por ser rápido, barato e não invasivo (Gawande, 2002). Além disso, a disponibilidade do ECG permite ao médico diagnosticar uma ampla variedade de doenças do coração.

O ECG é o registro dos fenômenos elétricos que se originam durante a atividade cardíaca (é gerado pelo músculo do coração), e a caracterização de cada cardiopatia se manifesta em modificações específicas da forma de onda do sinal (ressalvando-se que algumas anomalias cardíacas não alteram o ECG). Algumas destas modificações são facilmente identificáveis e associadas a doenças. Entretanto, existem aquelas que são de difícil percepção, mesmo para um médico treinado. Estudos demonstraram que de 2 a 8 % dos pacientes que são examinados em pronto-socorros são equivocadamente diagnosticados (sendo encaminhados às suas casas) e um quarto dessas pessoas morre ou sofre uma parada cardíaca completa (Gawande, 2002). Assim sendo, é importante o desenvolvimento de sistemas que possam auxiliar o médico na tomada de decisão sobre o diagnóstico correto a respeito de um sinal de ECG, minimizando o erro humano.

A análise automática do ECG é um tema de pesquisa que surgiu há quatro décadas, e que é muito disseminado no domínio biomédico. O interesse particular pelo estudo do ECG deve-se ao fato de que o mesmo representa um exame eficaz para o diagnóstico de arritmias cardíacas, de problemas de condução do impulso elétrico responsável pela contração do músculo cardíaco e de doenças do coração (Kadish *et al.*, 2001).

Considerando a abordagem efetuada pelo especialista médico, pode-se identificar três diferentes níveis de análise do sinal ECG (Andreão *et al.*, 2004), a saber:

- *detecção do complexo QRS e segmentação do batimento*: necessária para a determinação da frequência cardíaca, além de servir de referência para o alinhamento dos batimentos. A detecção das formas elementares do ECG fornece as características fundamentais (amplitude e intervalos) que são empregadas nas análises subsequentes de classificação;
- *classificação do batimento*: após isolar cada batimento detectado e segmentado, efetua-se a sua classificação;

- *análise a longo termo*: após classificar uma sequência longa de batimentos, procede-se uma análise do comportamento global da atividade elétrica do coração do indivíduo.

O objetivo do eletrocardiograma é registrar a variação dos potenciais elétricos gerados pela atividade elétrica do coração, medida por eletrodos, a qual representa uma sequência de batimentos cardíacos. O sinal de ECG padrão é formado por segmentos consecutivos no tempo, cada um deles similar àquele representado na Figura 1.1, que corresponde a uma batida do coração (um batimento cardíaco). Pode-se dizer que o batimento é constituído por formas elementares (ou ondas elementares), sendo que tal conjunto de ondas elementares se repete ao longo do tempo, e o estudo das amplitudes e da morfologia destes segmentos elementares constitui a base da análise do sinal ECG.

Hoje em dia é possível contar com computadores cada vez menores, de grande capacidade de processamento e baixo consumo de energia, permitindo assim sua portabilidade. Neste sentido, o registro ambulatorial³ de longa duração se desenvolveu. A melhor qualidade da captura do sinal ECG e maior capacidade de interpretação das arritmias através do computador abriam novas perspectivas a este exame.

O ECG ambulatorial divide-se, geralmente, em duas categorias: gravação contínua e gravação intermitente (Crawford, 1999). Na primeira categoria, a gravação dura, tipicamente,

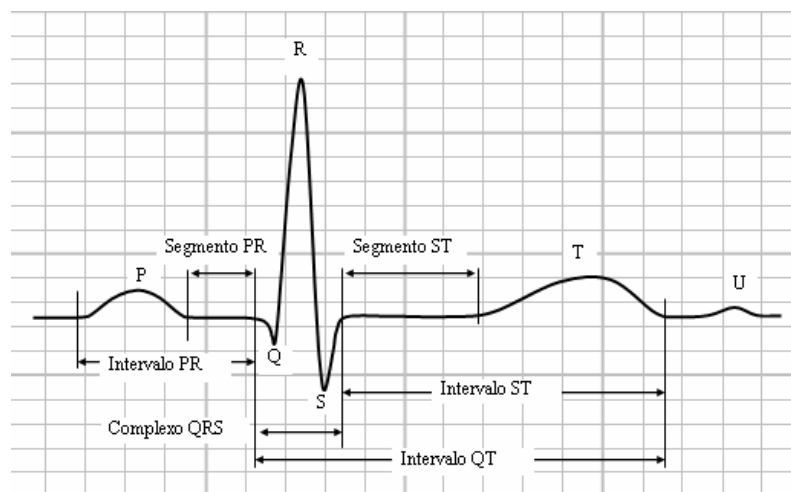


Figura 1.1. Ilustração de um batimento cardíaco observado em um eletrocardiograma, com a indicação de suas formas elementares.

³ Entende-se por registro ambulatorial todo sistema portátil de aquisição de ECG de longa duração baseado no método de Holter, sem a necessidade do paciente se deslocar a uma clínica ou hospital para fazê-lo.

de 24 a 48 horas. Para a segunda categoria, a gravação é ativada no momento em que um acontecimento é assinalado, quer pelo paciente, quer por um programa. Por esta razão, é possível atingir períodos de utilização do aparelho de até várias semanas. Atualmente, uma nova categoria de emprego do ECG se apresenta: o registro contínuo do ECG no contexto da telemedicina. Neste caso, os dados do ECG ambulatorial são disponibilizados em tempo-real ao especialista médico para que, a qualquer momento, um diagnóstico possa ser efetuado. Tendo em vista o telemonitoramento da atividade elétrica do coração, o sistema aqui desenvolvido será empregado no problema de auxílio ao diagnóstico médico, podendo ser utilizado para a identificação precoce de condições de emergência do paciente.

A análise automática do ECG ambulatorial pode ajudar consideravelmente o trabalho do médico, reduzindo o tempo gasto durante a releitura dos registros de 24 ou 48 horas de duração. A maior parte dos aparelhos de reprodução do ECG ambulatorial possui programas de análise do sinal. Contudo, os resultados obtidos são frequentemente considerados incorretos quando avaliados por um especialista experiente (Gawande, 2002). Além disso, os sistemas atuais funcionam como verdadeiras caixas pretas, impedindo que o especialista médico possa configurá-los, no âmbito da classificação das arritmias. Outro ponto fraco diz respeito à característica rígida dos mesmos, os quais são incapazes de se adaptarem ao sinal de ECG do paciente. Por estas razões, os sistemas atuais precisam ser aperfeiçoados para serem considerados um instrumento de apoio à decisão médica, que é sempre soberana.

1.3 Componentes Individuais do Eletrocardiograma

Um ciclo cardíaco consiste na combinação de dois períodos, o primeiro consiste na despolarização (contração ou estimulação) e o segundo na repolarização (recuperação ou repouso) cardíacas, fenômenos que se realizam sucessivamente e continuamente, em ritmo e frequência (batimentos por minuto) regulares (em condições normais).

A Figura 1.2 mostra, resumidamente, a condução do impulso cardíaco pelo coração humano. O impulso elétrico que provoca a contração cardíaca (ou despolarização cardíaca) origina-se no nodo sinusal (nodo SA, número 1 na Figura 1.2). Este nodo se localiza no átrio direito do coração, e é o responsável por marcar o passo do ritmo cardíaco. A partir do nodo sinusal, o estímulo elétrico se propaga em todas as direções, como uma onda que se espalha na superfície de uma lagoa quando se atira uma pedra (Gonçalves, 1995).

Após a ativação do impulso elétrico pelo nodo sinusal, este impulso é propagado através dos músculos atriais, e produz a contração de ambos os átrios. Essa onda de despolarização é registrada no ECG pela onda P. Esta onda representa a despolarização atrial.

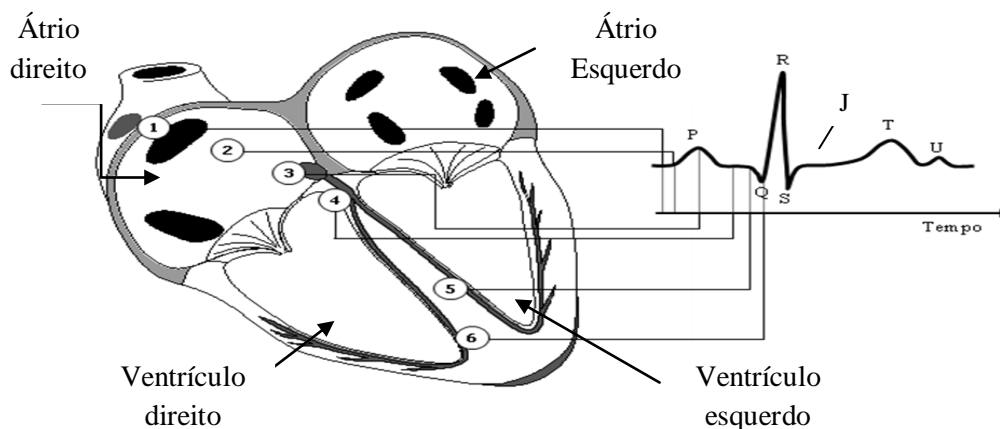


Figura 1.2: Relação entre a simulação cardíaca e o instante de aparição das ondas no eletrocardiograma: 1) Nodo sinusal; 2) Átrio; 3) Nodo AV; 4) Feixe de His; 5) Fibras de Purkinje; 6) Músculo ventricular (Andreão, 2004).

Ela é uma saliência de morfologia arredondada, simétrica, de pequena amplitude, e que precede imediatamente todos os complexos QRS. Tem ápice arredondado, com duração normal que varia de 0,09 a 0,11 s (Zipes, Camm, Borggreffe, 2006). A identificação da onda P é importante para a diferenciação entre ritmo sinusal e ritmos ectópicos, pois se ela está presente no batimento cardíaco significa que o impulso elétrico teve início no nodo sinusal. Se ela não é registrada ou não está bem posicionada no ECG, significa que o impulso elétrico não foi produzido pelo nodo sinusal, mas por um foco ectópico do coração.

Assim que o impulso elétrico alcança o nodo AV (nodo atrioventricular, representado na Figura 1.2 pelo número 3) ocorre um retardo proposital para que todo o átrio consiga se despolarizar. Este ‘atraso’ permite que o sangue passe para os ventrículos através das válvulas atrioventriculares. Assim que esta pausa termina, o nodo AV é estimulado e o impulso desce para o *feixe de His* (ramo direito e ramo esquerdo) e percorre simultaneamente ambos os ramos e termina na finas *fibras de Purkinje*.

Este impulso, que parte do nodo AV e se propaga pelo *feixe de His* e pelas *fibras de Purkinje*, é registrado no eletrocardiograma como o complexo QRS, que corresponde à despolarização ventricular. O pico ascendente deste complexo é a onda R. O pico descendente que precede a onda R é chamado de onda Q, e o pico descendente que se sucede à onda R é

chamado de onda S. A duração, medida entre o início e o término do complexo QRS, varia de 60 a 100 ms, em condições normais.

O intervalo PR ou PQ vai do início de P ao início de R ou de Q. Corresponde ao intervalo de tempo que o estímulo leva para alcançar os ventrículos, após a despolarização atrial. É uma linha horizontal, cuja duração normal varia de 0,12 a 0,18 ou 0,20 s.

No coração humano, a repolarização ventricular ocorre na mesma seqüência que a despolarização. A onda T do eletrocardiograma corresponde à repolarização ventricular, em sua quase totalidade. Tem a forma ligeiramente arredondada e assimétrica, com a fase ascendente mais lenta (ramo inicial mais longo) e a fase descendente mais rápida. Além disso, sua duração média é de 0,20 s. Qualquer condição que interfira com a repolarização normal pode provocar inversão nos registros dessa onda.

O segmento ST corresponde à fase inicial da repolarização ventricular. É o segmento de linha que se segue e une o complexo QRS à onda T. Ele começa ao término da onda S e continua até o início da onda T, e tem a forma ligeiramente curva, com a concavidade superior.

A pequena saliência que, às vezes, se segue à onda T é chamada de onda U, e tem, habitualmente, pouca significação clínica.

O ponto que marca a junção entre o final da deflexão QRS e o início do segmento ST, linha de repouso, é chamado de ponto J.

Assim, durante um ciclo cardíaco, a atividade elétrica do coração caracteriza-se por cinco ondas (denominadas também de deflexões) distintas. Estas ondas são registradas pelo ECG e rotuladas pelas letras P, Q, R, S e T (estas letras foram selecionadas arbitrariamente, e portanto não estão relacionadas aos nomes dos fenômenos que representam).

O estímulo da despolarização e repolarização das contrações das fibras musculares do coração gera um campo elétrico que se espalha pelo corpo humano em diversas direções. Este campo elétrico é forte o bastante para excitar eletrodos de superfície fixados na pele, que captam a atividade elétrica ali presente, registrando o fluxo dessa atividade em diferentes planos do corpo do indivíduo. Os diferentes registros das ondas cardíacas em pontos distintos do corpo humano constituem as derivações do ECG. Dessa forma, os registros do ECG obtidos dependem da derivação em que as ondas serão captadas (Gonçalves, 1995).

A atividade elétrica do coração é sempre a mesma, mas na superfície do corpo surgem distintas diferenças de potencial, como consequência dos fenômenos elétricos gerados durante

a excitação cardíaca, de forma que as ondas registradas (os registros de ECG) diferem em aparência nas diversas ‘derivações’ (Figura 1.3 – (b)). Estas diferenças podem ser medidas e registradas, obtendo-se desta forma as ‘derivações’, que podem ser definidas de acordo com a posição dos eletrodos (Figura 1.3 – (a)). Assim, para se obterem os registros de um ECG é preciso colocar eletrodos em vários pontos distintos do corpo do indivíduo, e esses eletrodos irão captar a diferença de potencial elétrico em diversos pontos do corpo humano.

Os primeiros ECGs foram obtidos historicamente por eletrodos colocados no braço direito, braço esquerdo e perna esquerda, formando o chamado Triângulo de Einthoven⁴. A disposição dos eletrodos e as três derivações bipolares representam o Triângulo de Einthoven, que une as extremidades (braços e perna) formando um triângulo equilátero, de forma que o coração, referenciado como um ponto, ocupa o centro do triângulo (tronco). Cada lado do triângulo é delimitado por dois eletrodos, um positivo e outro negativo, para que seja possível medir a diferença de potencial entre os dois pontos. As derivações são registradas por cada par de eletrodos.

A derivação é dita bipolar quando mede a diferença de potencial entre dois pontos da

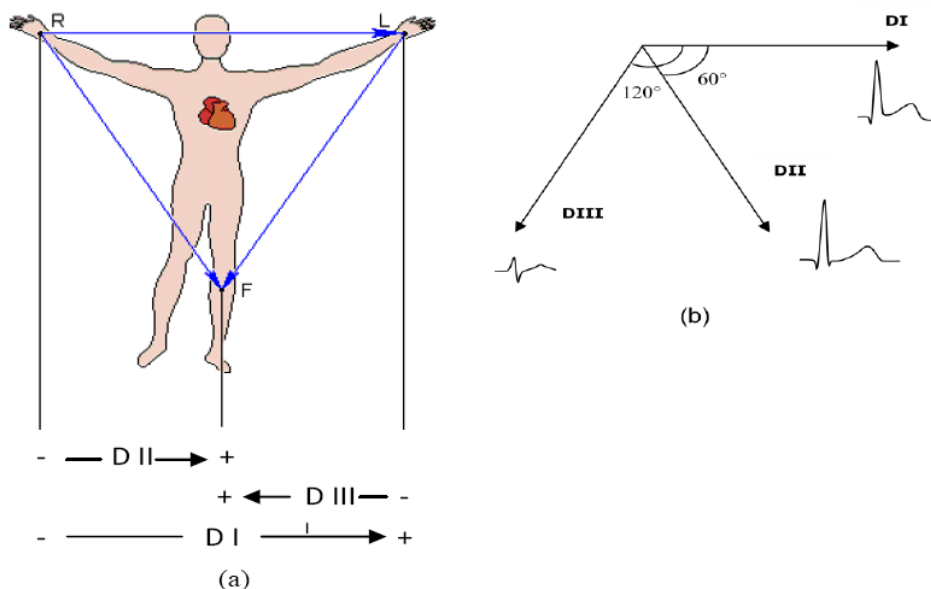


Figura 1.3: As três derivações bipolares. (a) Local de colocação dos eletrodos. (b) representação vetorial das derivações e formas de ondas observadas durante a despolarização e repolarização ventricular (Andreão, 2004).

⁴ Hipótese feita por Einthoven (Andreão, Nov 2004).

superfície corporal. A derivação padrão I, ou simplesmente D1, corresponde a $D1 = VL - VR$, ou seja, é a diferença de potencial entre o eletrodo do punho direito (pólo negativo) e o

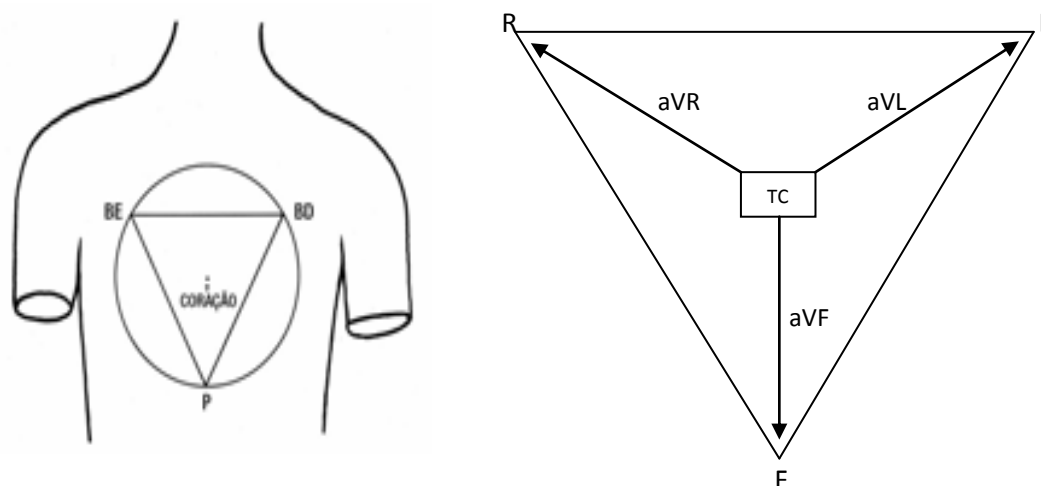


Figura 1.4: Representação vetorial das derivações unipolares.

eletrodo do punho esquerdo (pólo positivo). Já a derivação padrão II, ou simplesmente D2, corresponde a $D2 = VF - VR$, ou seja, é a diferença de potencial entre o eletrodo no punho direito (pólo negativo) e o eletrodo do tornozelo esquerdo (pólo positivo). Por fim, a derivação padrão III, ou simplesmente D3, corresponde a $D3 = VF - VR$, ou seja, é a diferença de potencial entre o eletrodo no punho esquerdo (pólo negativo) e o eletrodo do tornozelo esquerdo (pólo positivo).

Cada uma destas derivações registra a diferença de potencial elétrico em planos diferentes. Portanto, o ECG obtido em cada derivação tem um aspecto diferente (Figura 1.3).

As derivações unipolares dos membros são aquelas em que o traçado obtido se deve às variações de potencial recolhidas usando um eletrodo situado em um dos membros (braço direito, braço esquerdo e tornozelo esquerdo) e um ponto (próximo de zero) localizado no aparelho (derivações unipolares ou de Wilson⁵, Figura 1.4). Foi Frank Wilson quem descobriu que, para se registrarem esses tipos de derivações era necessário amplificar a voltagem no eletrocardiógrafo, para que se pudessem obter registros com traçado semelhante aos das derivações DI, DII e DIII. Por conta disso, as siglas que indicam as derivações modificadas são formadas pelas iniciais de expressões em inglês: A (aumentada), V (de voltagem), R (de

⁵ Estas derivações foram introduzidas por Wilson após a hipótese de Einthoven (Andreão, 2004).

right, braço direito), L (de *left*, braço esquerdo) e F (de *foot*, pé). As derivações aumentadas são:

aVR – um eletrodo no braço direito e o outro no potencial zero do aparelho;

aVL- um eletrodo no braço esquerdo e o outro no potencial zero do aparelho;

aVF- um eletrodo no tornozelo esquerdo e o outro no potencial zero do aparelho.

Quando o eletrodo é colocado em algum ponto da região torácica, as derivações unipolares são denominadas derivações precordiais, as quais têm o intuito de medir os potenciais próximo do coração. Habitualmente em número de seis, e designadas pela letra V, variando de V1 a V6, conforme a posição do eletrodo na superfície do tórax, assim Wilson introduziu as derivações do plano horizontal (ver figura 1.5).

As doze derivações eletrocardiográficas são, portanto, três derivações clássicas ou padrão (D1, D2 e D3), três derivações unipolares dos membros (aVR, aVL e aVF) e seis derivações unipolares precordiais (V1,V2,V3,V4,V5 e V6).

O presente trabalho utiliza estas derivações como parâmetros de entrada das Redes Bayesianas propostas, sendo que elas são encontradas em registros selecionados, os quais serão discutidos e explicados posteriormente.

1.4 Extrassístole Ventricular

O objetivo específico desta Tese é verificar a viabilidade da utilização de RB para detecção de arritmias cardíacas. Para se estudar esta tecnologia e sua viabilidade, foi selecionada uma arritmia em particular, a Extrassístole Ventricular (ESV), utilizada neste

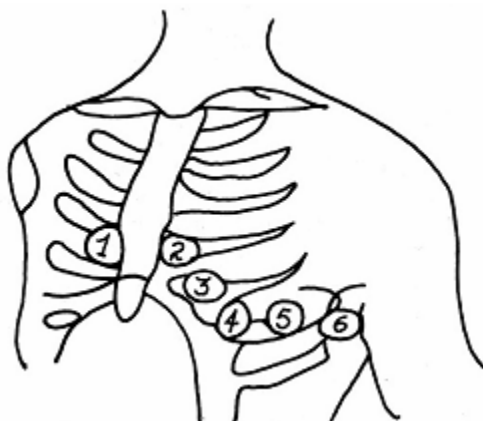


Figura 1.5: Posição dos eletrodos das derivações precordiais (V1 a V6).

trabalho como um estudo de caso específico.

Esta arritmia cardíaca é descrita de forma sucinta neste capítulo, objetivando-se adquirir uma visão geral sobre este problema e suas características.

O termo arritmia é usado na área médica para designar todos os distúrbios na frequência, no ritmo e na propagação dos impulsos elétricos cardíacos. O ECG oferece um meio eficiente para diagnosticar esses diferentes distúrbios rítmicos do coração. Dentre as arritmias cardíacas a ESV é a mais comum, principalmente quando associada ao infarto agudo do miocárdio. Porém, ela também pode ocorrer associada a hipopotassemia, intoxicação digitalica e estresse.

Este trabalho é voltado para a detecção de ESVs causadas por contrações prematuras do coração, que interrompem brevemente o compasso normal das batidas. A ESV, enquanto uma contração ventricular prematura, é um batimento cardíaco detectado a partir do ECG, sendo provocado pela ativação elétrica antecipada dos ventrículos, em relação ao batimento cardíaco normal. Em particular, a análise deste tipo de arritmia cardíaca recebeu muita atenção, como é mostrado em Strintzis (1992) e Maglaveras (1998).

Existem três tipos de extrassístoles:

- as atriais (originadas nos átrios, que são as câmaras cardíacas superiores),
- as juncionais (originadas na região intermediária do sistema elétrico, nodo AV),
- as ventriculares (ESVs, originadas nas células dos ventrículos).

As duas primeiras são chamadas de extrassístoles (ES) supraventriculares (originadas acima dos ventrículos). As ES são batimentos precoces que se originam fora do marca passo sinusal. As manifestações clínicas podem ser assintomáticas, mas também podem produzir palpitações, com sensações de ‘falhas’ e ‘soco no peito’.

A importância da arritmia ESV é que ela poderá ser a precursora de arritmias cardíacas ventriculares mais graves (extrassistolia ventricular maligna), que dão origem a taquicardia e a fibrilação ventriculares, as quais são causas de parada cardíaca e morte.

A ocorrência desta arritmia cardíaca se distingue pela produção de uma contração ventricular prematura, isto é, um foco ectópico localizado em um dos ventrículos provoca uma descarga elétrica antecipada. Este impulso elétrico não é proveniente do nodo sinusal, produzindo, desta forma, uma contração ventricular.

Esta arritmia provoca a despolarização de um ventrículo antes do outro, por ter origem num foco situado em um dos ventrículos. Enfim, as ESV são batimentos cuja sequência de

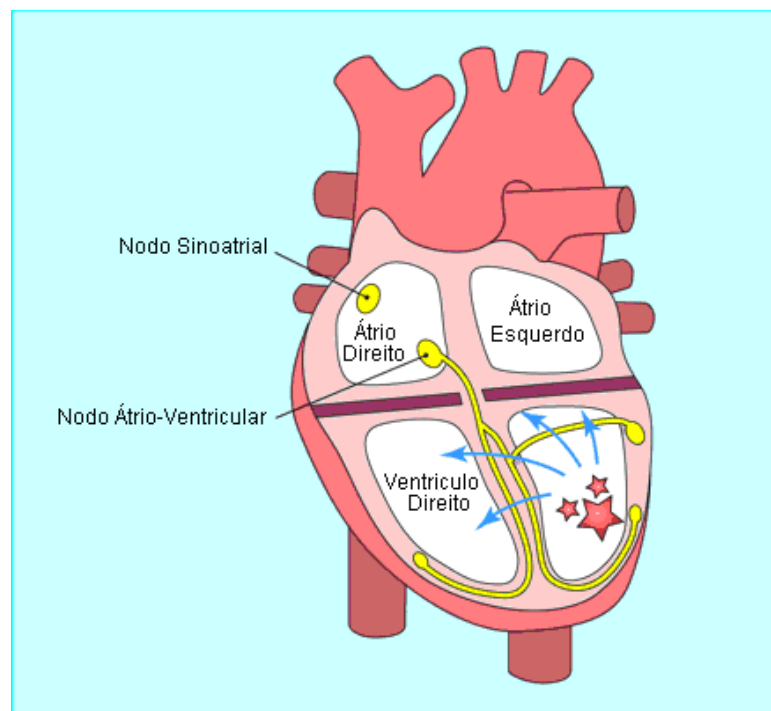


Figura 1.6: Exemplo de propagação elétrica com foco ectópico em um dos ventrículos.

Fonte: <http://bmj.bmjournals.com>

ativação cardíaca é alterada, e os impulsos não seguem mais o caminho pelo *feixe de His*, como seria no sistema de condução normal. Por este motivo a propagação elétrica será mais lenta do que o normal, causando assim um ‘alargamento’ no complexo QRS (o ECG registra uma alteração nas ondas Q, R e S, muito maiores do que em um QRS normal). Após a ocorrência deste episódio arritmico ocorre uma pausa compensatória, durante a qual o coração se encontra ‘eletricamente silencioso’. A Figura 1.6 mostra um exemplo de um foco ectópico de geração de impulso elétrico situado no ventrículo esquerdo, causando, desta forma, uma ESV.

Sendo assim, o batimento classificado como ESV é caracterizado pela presença de um complexo QRS prematuro, seguido, na maior parte dos casos, de uma pausa compensatória. Outra característica é a ausência de uma onda P antes do complexo QRS prematuro. Isto ocorre por não ser o nodo sinusal o causador deste impulso elétrico, já que ele é o gerador da onda P no registro do ECG. Por fim, uma terceira característica dessa arritmia é a deformação da onda QRS, que fica mais ‘larga’. Pode-se observar estas características no exemplo de batimentos cardíacos apresentado na Figura 1.7.

Sendo assim, as características eletrocardiográficas apresentadas pela ESV são as seguintes (Gonçalves, 1995):

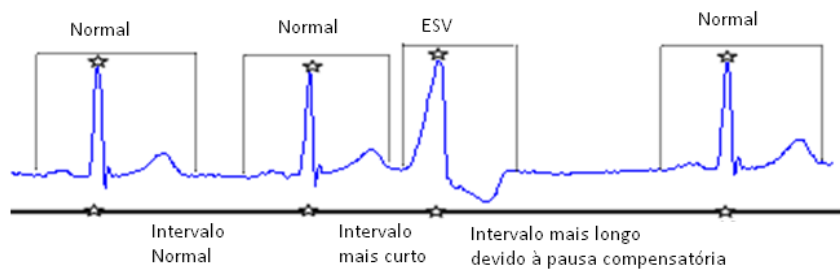


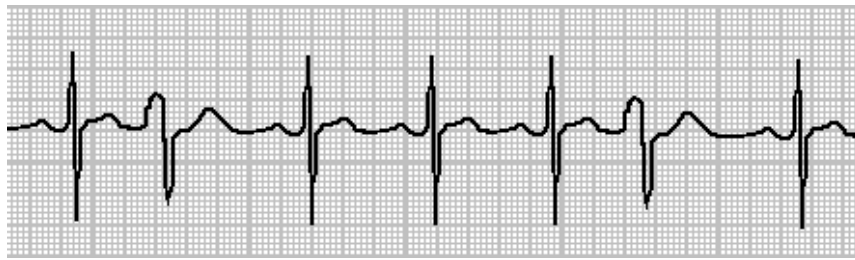
Figura 1.7: Batimentos contidos em um sinal ECG, identificados segundo sua classe (N – normal; ESV – Extrassístole ventricular).

- ritmo irregular no momento do batimento ectópico, e uma pausa compensatória que se segue;
- onda P ausente, pois o impulso inicia-se no ventrículo;
- complexo QRS ‘deformado/alargado’, pois sua forma depende da origem do impulso elétrico;
- onda T se inicia diretamente no complexo QRS.

Abordagens automáticas baseadas nestas características são muito atrativas, pois comportam a detecção do complexo QRS, reconhecimento de formas e classificação do batimento (Andreão, 2004).

Extrassístoles ventriculares, quando encontradas na ausência de doença cardiovascular, podem não ter nenhum significado (extrassístolia ventricular benigna), mas quando encontradas após infarto agudo do miocárdio ou na insuficiência cardíaca, por exemplo, acarretam implicações. Quando associada a medicamentos, por exemplo a intoxicação digitalica, pode levar a um ritmo letal. A identificação desta arritmia também tem grande importância para a correta classificação de episódios isquêmicos. Observam-se alguns exemplos de ESVs na Figura 1.8 que apresenta as diferentes formas que esta arritmia pode apresentar. Dependendo da forma (Figura 1.8, partes a e b) que apresentam numa derivação do eletrocardiograma, são chamadas de monomórficas (quando apresentam a mesma morfologia) ou polimórficas (quando apresentam diferentes morfologias) (Porto, 2001).

As ESVs indicam irritabilidade do ventrículo, e podem iniciar uma salva (sequência rápida) sob a forma de taquicardia ou fibrilação ventricular. A taquicardia ventricular (TV) acontece quando ocorre uma ‘salva’ de ESVs. Como o impulso elétrico não percorre o músculo cardíaco da forma normal, o coração não se contrai normalmente. Seus batimentos se tornam mais rápidos. À medida que o coração bate mais rápido, bombeia menos sangue a



(a) Extrassístoles ventriculares monomórficas



(b) Extrassístoles ventriculares polimórficas



(c) Extrassístole ventricular precoce iniciando fibrilação ventricular.

Figura 1.8: Alguns exemplos de ESVs.

cada contração, pois não há tempo suficiente para o seu enchimento entre dois batimentos. Se este ritmo rápido continuar, o corpo não receberá sangue e oxigenação suficientes. É importante ressaltar que este tipo de arritmia em especial (TV), não é objetivo deste estudo: o foco definido é apenas a identificação de ESV isolada.

Outras situações especiais de ESV também podem ser observadas, como Bigeminismo (Figura 1.9), Trigemínismo e Quadrigeminismo, casos estes que também são tratados neste trabalho. O Bigeminismo ocorre quando uma ESV se acopla a um batimento normal e se repete a cada batimento normal (alternância entre um batimento normal e uma extrassístole). O Trigemínismo ocorre quando uma ESV se acopla a dois batimentos normais, criando um padrão que se repete por muitas vezes (dois batimentos normais e uma extrassístole), e assim sucessivamente.

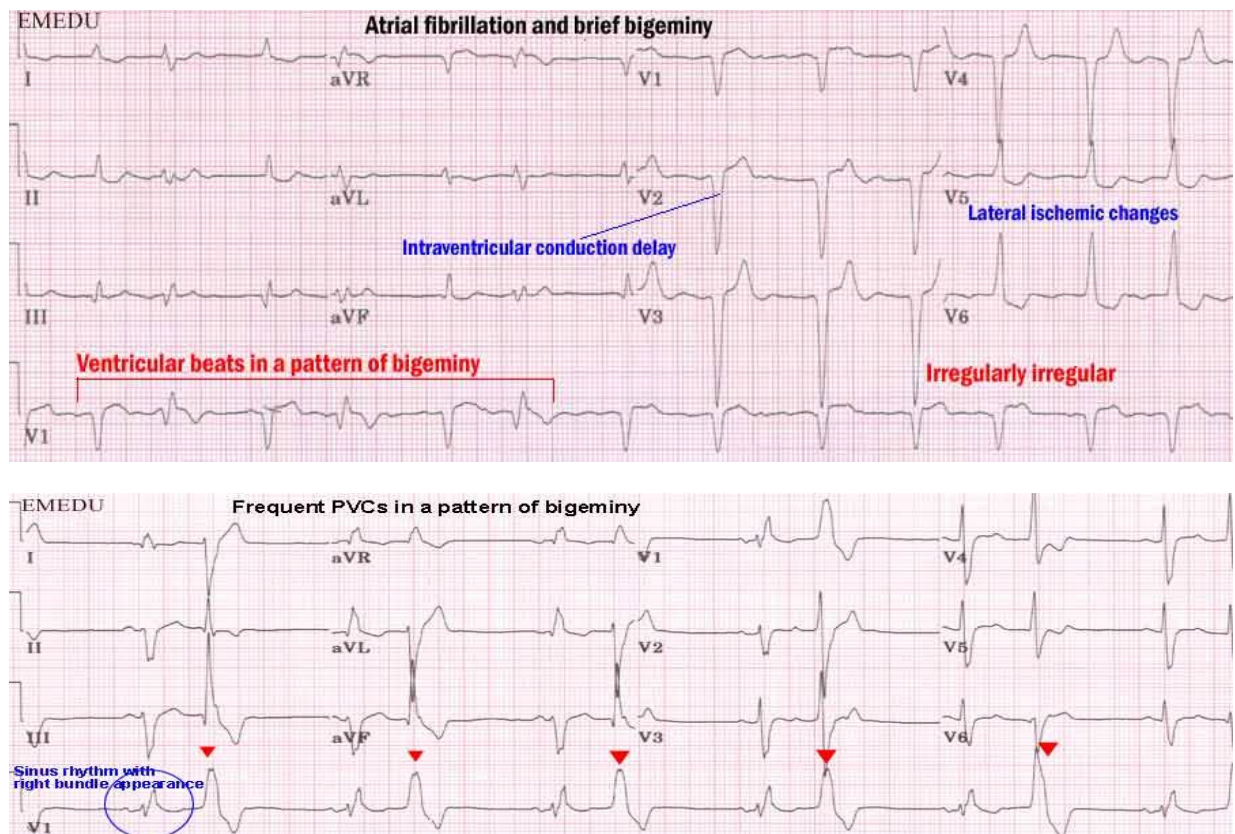


Figura 1.9: Exemplo de uma situação especial de ESV, Bigeminismo.

As extrasístoles ventriculares podem ser isoladas ou ocorrer em grupos. Duas extrasístoles ventriculares consecutivas são denominadas acoplamento. Três ou mais extrasístoles ventriculares consecutivas, a uma velocidade de 100 batimentos por minuto ou mais, são denominadas taquicardia ventricular. Observa-se que para a caracterização de TV é necessário fazer uma relação com outra variável, que é a frequência cardíaca (100 bpm⁶).

Como será discutido no Capítulo 4, estas alterações do batimento cardíaco, apesar de não serem o foco da classificação aqui descrita, estão presentes em alguns registros utilizados tanto na fase de treinamento quanto na fase de teste da rede, como a taquicardia ventricular, embora sejam casos que não estão presentes de forma representativa.

Em muitos trabalhos pode-se observar a pertinência do estudo de ECG bem como de ESV, como pode ser observado em Sayadi *et al* (2010) e Lim (2009).

1.5 Metodologia

⁶ bpm = batimentos por minuto.

Pretende-se mostrar, com este trabalho, a eficácia da utilização das abordagens baseadas em Redes Bayesianas. A rede apresentada refere-se a um problema de diagnóstico de doenças cardíacas. A principal questão respondida é se a RB apresenta um percentual de acerto aceitável para ser considerada uma ferramenta importante para auxílio ao diagnóstico de arritmias cardíacas. Para responder esta pergunta foi utilizado um estudo de caso específico, que é a identificação de episódios arrítmicos denominados ESVs.

São realizados vários estudos da disposição dos nós da rede, isto é, as relações de causalidade entre cada variável em questão. Após a escolha da relação de causalidade, são implementadas várias topologias da RB, como será apresentado nos capítulos seguintes. Utiliza-se, inicialmente, a implementação da RB estática, isto é, sem levar em consideração o tempo. Para isto, é empregado o algoritmo de *Maximum Likelihood* (ML), disponível na biblioteca do MATLAB denominada BNT (2007) (ver o Anexo 1).

A aplicação da RB estática é testada utilizando duas bases de dados, a QT *database* (1997) e a MIT-BIH *database* (1997).

Esta rede estática também é implementada utilizando dois algoritmos de segmentação, os quais fornecem os valores das variáveis de entrada da rede (as evidências). Estes algoritmos geram duas variáveis fundamentais para a construção da rede: a primeira refere-se ao tempo entre picos R (tempo RR) sucessivos (para que se possa identificar se um batimento é prematuro), e a segunda indica o grau de semelhança do complexo QRS (para que se verifique se este é ‘alargado’). Estes dois algoritmos de segmentação serão discutidos mais detalhadamente nos próximos capítulos, bem como a motivação da escolha do algoritmo de segmentação. A decisão de utilizar dois algoritmos de segmentação visa demonstrar que a Rede Bayesiana é um classificador adequado que apresenta um desempenho que ‘independe’⁷ do algoritmo utilizado para segmentar o sinal de ECG.

Finalmente, após a escolha da base de dados (MIT-BIH *database*), e do algoritmo de segmentação (baseado no trabalho de Andreão (2004)), realiza-se a construção da Rede Bayesiana Dinâmica (RBD). Neste caso, também são testadas várias topologias, e após a escolha da topologia apropriada estudou-se a diferença do valor resultante de probabilidade, e este resultado é comparado com o mesmo resultado da rede estática.

⁷ A palavra ‘independe’ apresentada nesta frase refere-se à equivalência dos resultados utilizando diferentes algoritmos de segmentação, sendo que estes algoritmos geram resultados satisfatórios. Isto não significa que se um algoritmo qualquer que segmenta o sinal gerar resultados com baixa qualidade a rede apresentará resultados satisfatórios. Isto é, a palavra ‘independe’ é relacionada diretamente a ‘algoritmos de segmentação satisfatórios’.

1.6 Contribuições da Tese

Esta tese traz três contribuições principais ao estudo de classificação de arritmias cardíacas, em particular ao estudo da arritmia ESV:

- apresentar a Rede Bayesiana e a Rede Bayesiana Dinâmica, e demonstrar a viabilidade desta abordagem para a classificação da arritmia cardíaca em questão, podendo, futuramente, ser aplicada para identificação de outras arritmias a partir dos registros de ECG;
- apresentar um modelo flexível que permite vários tipos de expansões, não só acrescentando níveis de padrões (ou complexidade) ao modelo, como também permitindo o uso de diferentes sub-modelos e métodos para modelagem e síntese;
- apresentar uma representação gráfica bastante interativa e intuitiva para o especialista médico, estrutura esta que também explicita relações de dependências e constitui uma ferramenta poderosa na aquisição de conhecimentos e no processo de verificação.

1.7 Organização do Texto

O texto da tese é dividido em cinco capítulos. O capítulo 2 trata de métodos estatísticos, particularmente de RB estática e Rede Bayesiana Dinâmica (RBD). Esse capítulo apresenta uma introdução ao tópico de Redes Bayesianas e tem o objetivo de tornar a tese mais auto-contida, e também possibilitar que a descrição do método, nos capítulos 3 e 4, fique mais sintética.

No capítulo 3, por sua vez, são apresentadas as RB estáticas que foram implementadas. Estas redes são expostas de forma detalhada, e suas estruturas e variáveis são descritas. São tecidas, também, algumas considerações sobre a captura, análise e segmentação dos dados. Por fim, é indicada e justificada a topologia, a base de dados e o método de segmentação que será empregado no desenvolvimento da RBD.

Já no capítulo 4 a Rede Bayesiana Dinâmica que foi desenvolvida é discutida e apresentada, demonstrando sua viabilidade através dos resultados obtidos. Gera-se uma interação direta com o especialista, visando uma melhor interpretação da incerteza tratada na RBD. A incerteza não é somente considerada através dos valores das variáveis aleatórias propagadas na rede, mas também pelo resultado do valor de probabilidade que a rede produz.

Finalmente, no capítulo 5 são apresentadas a conclusão e algumas sugestões de continuidade do trabalho.

Capítulo 2

Redes Bayesianas (RB)

2.1 Introdução

As pessoas sempre precisam lidar com a incerteza das informações apresentadas, tanto em sua vida pessoal quanto em sua vida profissional. Assim, quando se pensa em modelar a abordagem do especialista médico através de um Sistema Computacional de Auxílio ao Diagnóstico é necessário levar em consideração a incerteza. Existem algumas teorias voltadas para o tratamento da incerteza, as distribuições de probabilidade e os conjuntos difusos, devendo-se escolher a teoria que se enquadra melhor no caso em questão. De acordo com alguns autores (Dubois e Prade, 1994; Hadjali, Dubois et al., 2004), probabilidades e conjuntos difusos são duas formas conceituais e comportamentais distintas, direcionadas para a representação e tratamento da incerteza. Ainda baseado no trabalho destes autores, a incerteza não está diretamente relacionada com a seleção ponderada dos processos de descrição de um sistema, uma vez que o modo como estes processos se manifestam e o método pelo qual podem ser captados corretamente baseiam-se nas observações do analista. Conseqüentemente, não existe nenhuma forma universal simples para enfrentar com êxito a incerteza. Neste sentido, as probabilidades e os conjuntos difusos podem ser considerados teorias complementares, ao invés de antagônicas. Esta complementariedade entre as duas metodologias é bastante discutida em Nunes (2005). Porém, nesta Tese, a utilização somente da abordagem baseada em probabilidade mostrou-se adequada para o tratamento de incerteza, não se revelando necessário construir um sistema híbrido baseado nestas duas teorias. Diante disso, é proposta neste trabalho uma abordagem estatística baseada em inferência probabilística, com o intuito de se modelar a incerteza dos dados e da decisão a ser tomada. A abordagem calcula a probabilidade de um evento ocorrer, dada toda a evidência disponível. A base para efetuar este cálculo é o teorema de Bayes, e a tecnologia implementada é baseada na utilização das Redes Bayesianas.

Muitos trabalhos direcionados a auxílio ao diagnóstico médico, além das abordagens probabilísticas, utilizam também regras heurísticas e Redes Neurais para a construção de tais sistemas. As abordagens heurísticas simulam, de certa maneira, o raciocínio do especialista

médico. As regras são estabelecidas de maneira hierárquica, e traduzem, então, a experiência do especialista. Como consequência, elas são altamente dependentes da maneira pela qual o mesmo trata o problema. As relações de causa e consequência entre as regras podem ser representadas graficamente, permitindo ao especialista acompanhar a lógica de decisão e tomar uma posição crítica sobre os resultados. Porém, este tipo de metodologia não leva em consideração a idéia de incerteza, que é um ponto chave para uma boa classificação. Um exemplo deste tipo de metodologia é apresentado em Andreão, Dorizzi e Boudy (2006), que, após realizar uma análise preliminar via HMM (Modelos Ocultos de Markov), utilizou heurística para classificação das doenças apresentadas no ECG. Outra abordagem heurística é apresentada no capítulo 3, a qual foi construída utilizando um limiar para a classificação tanto de batimentos cardíacos normais quanto de extrassístoles ventriculares (ESV). Neste exemplo, verifica-se que o resultado da abordagem Bayesiana é melhor que o resultado da abordagem heurística. Ademais, as abordagens heurísticas dependem de ajustes nos critérios de decisão pré-estabelecidos sempre que novos casos reais são encontrados, isto é, a intervenção do especialista neste caso é fundamental, diferente da abordagem via RB, onde a intervenção do especialista é mínima, pois o treinamento do sistema é normalmente efetuado automaticamente.

A Rede Neural (RN) é um classificador muito empregado na análise de ECG, como mostram Farrugia, Yee e Niclolls (1991), que apresentam um trabalho sobre uma rede neural artificial testada para a classificação de ritmos cardíacos usando o sinal de ECG. Também em Kuppuraj (1993) é desenvolvido um sistema que emprega rede neural para detectar ritmos anormais obtidos do ECG. Em Maglaveras e Strintzis (1998) é apresentada uma Rede Neural supervisionada para detecção automática de episódios isquêmicos, considerando elevação ou depressão do segmento ST do sinal de ECG. O algoritmo usado para treinar a RN foi um algoritmo *backpropagation* adaptativo.

Porém, estes sistemas funcionam como verdadeiras caixas pretas, impedindo que o especialista médico possa configurá-los conforme a arritmia analisada. Por outro lado, as Redes Bayesianas permitem uma representação gráfica interativa (diferentemente das Redes Neurais, cujo conteúdo é inacessível), e possuem flexibilidade na modelagem de problemas mais complexos. Outro ponto que mostra a diferença entre os dois tipos de rede é que cada *nó* (neste trabalho, sempre que nos referimos a *nós* da Rede Bayesiana estamos lidando com variáveis que serão representadas por *nós* na topologia da rede) em uma Rede Bayesiana

possui em si próprio uma interpretação do domínio do sistema, o que não acontece na Rede Neural, onde os *nós* interiores (não de entrada e/ou saída) cumprem apenas um papel estrutural na montagem da rede, sem um significado próprio.

De maneira inversa aos casos de regras heurísticas, as abordagens estatísticas são resultado de uma etapa de aprendizagem sobre dados disponíveis. Desta forma, a capacidade de classificação dessas abordagens é muito dependente das informações aprendidas *a priori*. No entanto, o sistema pode evoluir a partir da concepção de novas bases de treinamento. O interesse na abordagem estatística se deve à sua capacidade de se adaptar automaticamente às variações de morfologia do sinal (Andreão, 2004).

Já existem nesta área alguns trabalhos, como aquele de Ono *et al.* (2004), que propõe um método de classificação baseado nas regras de Bayes, utilizando o magnetocardiograma (MCG). O teorema de Bayes e as características da onda são o suporte do trabalho, no qual foi desenvolvido um procedimento de parametrização que consiste em uma aproximação bidimensional *wavelet* e uma análise conjunta dos mapas de campo magnético, sendo a probabilidade de classificação do grupo considerado doente e do grupo considerado normal estimada pelo teorema de Bayes. Já a abordagem proposta por Gao *et al.* (2005), que esboça um sistema para a detecção de arritmias cardíacas no sinal ECG, é baseada em uma inferência Bayesiana e Rede Neural. Nesta mesma linha de pesquisa, Wiggins *et al.* (2005) desenvolvem um sistema para classificação da idade dos pacientes, com base em informações extraídas do sinal de ECG, cuja abordagem se baseia em Redes Bayesianas evolutivas, sendo a evolução feita via um algoritmo genético. Outro trabalho, apresentado por Przytula, Dash e Tompson (2003) mostra a evolução das Redes Bayesianas para diagnóstico. Ali se descrevem todas as fases envolvidas na implementação do sistema, e cada uma destas fases é discutida e ilustrada com um exemplo, podendo-se visualizar os resultados da evolução da RB, os quais são mostrados em formato gráfico, em matrizes 2-D e 3-D. Em Heart Disease Program (2010), o Programa de Doenças do Coração é um sistema de computador que auxilia o médico na tarefa de diagnóstico no domínio das doenças cardiovasculares, no qual, para o auxílio ao diagnóstico, é aplicado o raciocínio probabilístico que utiliza Rede Bayesiana e as relações de causalidade temporal.

Finalmente, em Przytula e Thompson (2000) são descritos todos os passos para o processo de projeto da rede, especialmente detalhes de aquisição de conhecimento e da integração de informações vindas de diferentes fontes. São desenvolvidas, primeiramente,

formas simples da Rede Bayesiana, aumentando sua complexidade o quanto seja preciso, equilibrando cuidadosamente a precisão do modelo e o custo de aquisição do conhecimento. Os autores testaram esta metodologia em muitos exemplos de problemas de diagnóstico, sistemas de comunicação de satélites, locomotivas a diesel e equipamentos de testes de satélites. Muitos sistemas complexos foram modelados, sendo possível construir Redes Bayesianas em tempo razoável e com o mínimo custo para domínios do especialista. O desempenho das ferramentas de diagnóstico baseadas em Redes Bayesianas foi muito bom, em todos os casos.

Motivado pelos resultados obtidos usando a abordagem estatística, e tendo em conta que ainda há muito o que se explorar em tal abordagem, dado que os primeiros resultados são bem recentes, este trabalho adota o paradigma de Redes Bayesianas para analisar a ocorrência de extrassístole ventricular (ESV) em um registro de ECG, para levar em conta a incerteza embutida no problema.

Basicamente, os motivos que levaram à escolha da técnica de Redes Bayesianas para atuar como sistema de auxílio à decisão médica são os seguintes:

- a construção de um sistema baseado na opinião de especialistas pode ser feita utilizando-se as estruturas causais da RB;
- facilidade de apresentação do funcionamento da RB aos médicos. Esse fator possibilita aos especialistas envolvidos terem uma visão mais clara do funcionamento do sistema, pois a representação gráfica é muito intuitiva e de fácil assimilação;
- a possibilidade de trabalhar com incertezas (através das variáveis aleatórias propagadas na RB), porque em problemas de diagnóstico médico realizados por especialista muitas vezes existem condições em que não se tem certeza sobre o diagnóstico correto, e, assim, o tratamento da incerteza é de fundamental importância.

Antes de se discutir sobre Redes Bayesianas profundamente, serão revisados conceitos básicos de probabilidade, a seguir.

2.2. Probabilidade: Uma Visão Geral

O estudo da probabilidade pode ser iniciado com a idéia primitiva de uma experiência aleatória¹. Associados à experiência aleatória estão os seus resultados possíveis. Por exemplo, associados à experiência ‘arremesso de um dado’ estão os resultados possíveis: como um ponto, dois pontos (Clarke e Disney, 1970).

Quando não se pode ter qualquer informação a respeito dos resultados possíveis, pode-se descrevê-los como resultados equiprováveis, ou seja, igualmente prováveis. Caso seja realizada alguma experiência com elementos equilibrados que gerarão n resultados, a probabilidade de cada resultado será de $1/n$. A *probabilidade a priori*² é estimada pela razão do número de resultados equiprováveis favoráveis a uma determinada situação pelo número total de resultados (Lopes, 1999).

Quando se deseja ter um resultado sem recorrer a experimentação, deve-se proceder da seguinte maneira: seja S um espaço amostral associado ao experimento E . A cada evento A , um subconjunto de S , associa-se um número real $P(A)$, ao qual se denomina probabilidade de A , que satisfaça as seguintes propriedades:

- $0 \leq P(A) \leq 1$ para todo evento A ;
- se A é um evento certo, $A = S$, então $P(A)=1$;
- se A_1, A_2, A_3, \dots , formam uma sequência (finita ou infinita) de eventos mutuamente exclusivos ($A_i \cap A_j = \emptyset$ para $i \neq j$), e se $B = A_1 \cup A_2 \cup A_3 \cup \dots$ é o evento ‘pelo menos um dos A_{i_s} ocorre’, então

$$P(B) = P(A_1) + P(A_2) + \dots = \sum_i P(A_i) \quad (2.1)$$

Muitas propriedades importantes podem ser extraídas dos axiomas acima. Por exemplo (Clarke e Disney, 1970):

- $P(\emptyset) = 0$;
- se $A \subseteq B \subseteq S$, então $P(A) \leq P(B)$;
- se $A, B \subseteq S$, então $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- se $A \subseteq S$, então $0 \leq P(A) \leq 1$.

¹ A teoria da Probabilidade tem suas raízes em situações da vida real, quando se realiza uma experiência cujo resultado não pode ser previsto com certeza. Uma experiência desse tipo é denominada de *experiência aleatória* (Clarke e Disney, 1970).

² Dada uma experiência aleatória uniforme, definida em um espaço de amostragem S , define-se a probabilidade de ocorrer um evento A , contido em S , como sendo o quociente entre o número de elementos do evento A , n_A , e o número de elementos do espaço amostral S , n_S , isto é: $P(A) \cong n_A/n_S$.

Sendo A e B dois eventos pertencentes ao experimento E , então existe $P(A/B)$, que define a probabilidade condicional ou *probabilidade a posteriori*³ do evento A , quando B tiver ocorrido primeiro, que é determinada pela fórmula

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.2)$$

A probabilidade condicional (*probabilidade a posteriori*) $P(A/B) = X$ pode ser interpretada como: ‘Dado o evento B , a probabilidade do evento A ocorrer é X ’.

Independência Condicional ocorre quando dois eventos A e B forem independentes, ou seja, $P(A/B)=P(A)$ (neste caso, o evento A não depende dos resultados do evento B) e $P(B/A)=P(B)$ (o evento B não depende dos resultados do evento A). Nestes dois casos pode-se dizer que $P(A \cap B) = P(A) P(B)$.

Em geral, se o interesse está no cálculo da proposição A , e o conhecimento da evidência i está disponível, deve-se calcular $P(A/B)$. Em alguns casos este valor não está disponível na base de conhecimento e, portanto, deve-se utilizar algum método de inferência para obtê-lo.

A partir de resultados obtidos em um evento, pode-se descrever a inferência para caracterizar a população como um todo. Generalizar aspectos importantes de uma população a partir dos resultados obtidos na ocorrência de um evento (amostra) caracteriza a inferência (Lopes, 1999).

Para que as inferências sejam válidas, é preciso que a ocorrência de um evento qualquer seja representativa na população. Escolhendo-se uma ocorrência, geralmente opta-se pela mais conveniente, podendo, desta forma, provocar um resultado de inferência tendencioso. Para resolver esse problema deve ser feita a escolha aleatória da ocorrência, para que o resultado não seja subestimado ou superestimado.

Inferências probabilísticas são utilizadas em algoritmos de propagação de crenças em Redes Bayesianas, podendo ser do tipo causal (parte das causas para os efeitos), diagnósticos (dos efeitos para as causas), intercausal (discrimina entre causas de um efeito comum) ou misto (combinação de dois ou mais tipos acima) (Ladeira, 2000).

Cada variável que é representada por *nós* na topologia da RB pode assumir valores discretos e contínuos. Definimos a distribuição de probabilidades ou função de densidade de probabilidade (*pdf* - *probability density function*) sobre pontos da reta de Borel. No caso de

³ Dada uma experiência aleatória uniforme, definida em um espaço de amostragem S , define-se a probabilidade de ocorrer um evento A , contido em S , dado que foi observado B , que se lê como $P(A|B)$.

variáveis discretas, o valor da função de densidade de probabilidade corresponde à frequência relativa de que o resultado de um experimento seja igual ao argumento da função. Exemplo: $P(X=5) = f(5)$. No caso de variáveis contínuas, o valor da densidade de probabilidade é tal que a integral da função sobre um intervalo corresponda à frequência relativa do resultado de um experimento cair dentro do intervalo, ou

$$P(a \leq X \leq b) = \int_a^b f(x)dx \quad (2.3)$$

2.3 Inferência Bayesiana

Atualmente, a comunidade acadêmica tem se interessado muito pela utilização de inferência probabilística no desenvolvimento de sistemas inteligentes. Desenvolver técnicas de inferência eficientes é uma das metas mais importantes para os pesquisadores. Entretanto, a utilização destas técnicas pressupõe a disponibilidade de modelos de conhecimento válidos.

No enquadramento deste tema, modelos gráficos apresentam um formalismo unificado para muitos modelos probabilísticos, os quais são amplamente utilizados em estatística, aprendizagem de máquina e engenharia (Jordan, 1999).

Apresenta-se, então, a tecnologia de Redes Bayesianas, que objetiva a descoberta de conhecimento a partir de uma base de dados previamente disponível. De acordo com Pearl (1991), as Redes Bayesianas foram desenvolvidas nos anos 70, com o objetivo de modelar a leitura e compreensão em sistemas distribuídos, onde as expectativas semântica⁴ e percentual⁵ de evidência devem estar combinadas para que, desta forma, seja apresentada uma interpretação coerente. A partir dos anos 80, as Redes Bayesianas emergiram como uma ferramenta importante para a representação de incertezas.

A utilização da abordagem de Redes Bayesianas é cada vez mais comum para a representação gráfica de conhecimento incerto nas mais diversas áreas. Tal abordagem também oferece uma estrutura unificada e intuitiva para representar modelos de seus dados, demonstrando com clareza formas de expressão e de interpretação do conhecimento codificado nas Redes Bayesianas. Segundo Hruschka (1996) a probabilidade Bayesiana é uma teoria consistente, que permite a representação de conhecimentos certos e incertos via

⁴ Representação qualitativa da RB

⁵ Representação quantitativa da RB

distribuição de probabilidade conjunta, que pode ser representada por um produto de distribuições condicionadas.

Thomas Bayes, em 1762, apresentou uma regra bastante importante onde se calcula a probabilidade de um evento ocorrer dado que um outro tenha ocorrido. O teorema de Bayes tem importância fundamental, pois fornece a base para uma abordagem da inferência estatística. Este teorema possibilita representar numericamente o grau de certeza sobre um evento, pois a teoria Bayesiana está fundamentada na teoria da probabilidade.

As Redes Bayesianas são baseadas neste teorema, que é utilizado para calcular a probabilidade de um evento ocorrer, dado que foi observada a ocorrência de outro evento. Esta probabilidade é calculada através da expressão

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.4)$$

a qual caracteriza o teorema de Bayes, para o qual se tem que $P(A|B)$ é a *probabilidade a posteriori* de A , $P(B|A)$ é a *probabilidade a posteriori* de B , $P(A)$ é a *probabilidade a priori* da hipótese A , e $P(B)$ é a *probabilidade a priori* de B (Pearl, 1991).

Esta fórmula simples possui grande importância prática num domínio como o de diagnóstico. Geralmente é mais fácil deduzir a probabilidade, por exemplo, observando um sintoma dada a doença do que a doença dado o sintoma. Porém, o que geralmente acontece é o segundo caso, isto é, o mais complicado. Este teorema pode ser usado para manipular o caso onde a hipótese é uma proposição no domínio do conhecimento (por exemplo, uma doença específica) e a evidência é a observação de algumas condições (um sintoma, neste exemplo).

Com base no teorema de Bayes foram criadas as Redes Bayesianas, que trabalham com relações causais quantificadas por valores de probabilidade condicional. Matematicamente, uma Rede Bayesiana é uma representação compacta⁶ da probabilidade conjunta do universo do problema $U = (A_1, A_2, \dots, A_n)$, e também um modelo gráfico que representa de forma simples as relações de causalidade das variáveis do domínio modelado. Desta forma, se houver um arco indo de um nó⁷ A para outro nó B , admite-se que o nó A representa a causa de B , e diz-se que o nó A é o nó pai do nó B . Da mesma forma, o nó B é o nó filho do nó A (Pearl, 1991).

Dizer que a Rede Bayesiana é uma representação compacta significa que ela representa a probabilidade conjunta de um domínio de forma compacta. Esta característica

⁶ Compacta, pois são definidas somente as probabilidades condicionais de cada nó em relação aos seus pais.

⁷ Utiliza-se a denominação nó ou variável, tendo ambas o mesmo significado.

decorre do fato de que ela possui uma propriedade denominada localidade estrutural. Esta propriedade define que cada subcomponente do sistema (*nó* da árvore) interage diretamente apenas com uma parte do sistema, os *nós* que estão a ele conectados, e não com todos os *nós* modelados na rede. Assim, em tais redes cada *nó* está diretamente relacionado com no máximo i outros *nós*, dentre os n ($n > i$ *nós* da rede). Pressupondo, para simplificar, que os *nós* possam assumir apenas valores booleanos, isto é, 0 ou 1 (*nós* discretos), para se criar um conjunto de probabilidades condicionais qualquer a quantidade de informação necessária será 2^i números, diferentemente de uma tabela de conjunção de probabilidades, onde a quantidade de informação necessária será 2^n . Apenas para a melhor visualização do poder de compactação, seja uma rede com 15 *nós* ($n = 15$), cada um com no máximo 4 pais ($i = 4$). Uma Rede Bayesiana necessitaria de 240 itens de informação, enquanto uma tabela de conjunção de probabilidades seria formada por 491.520 itens de informação. No entanto, existem sistemas onde pode ser necessária a conexão de cada um dos *nós* da árvore, gerando um grafo fortemente conexo. Neste caso, a especificação de uma tabela (ou *pdf* no caso de variáveis contínuas) da rede que representa sua probabilidade condicional necessita da mesma quantidade de informação que uma Rede Bayesiana. Entretanto, a maioria dos problemas podem ser simplificados, garantido-se, desta maneira que a rede não seja fortemente conexa, sem perda de informação útil para isto.

As Redes Bayesianas são representadas por diagramas que organizam o conhecimento numa dada área através de um mapeamento entre *causas e efeitos*. Os sistemas baseados nesta rede são capazes de gerar automaticamente predições ou decisões, mesmo na situação de inexistência de algumas peças de informação.

A rede não possui *nós* correspondendo a todas as causas de um evento, sendo os fatores irrelevantes resumidos na incerteza associada às probabilidades de algumas variáveis (fugindo-se de grafos fortemente conexos). Com poucas variáveis pode-se lidar com um universo muito grande de causas, e introduzindo-se mais informação relevante é possível melhorar o grau de aproximação.

Nestas redes, as representações devem ser feitas levando-se em conta fatores que realmente têm influência direta nos resultados. Isto se deve ao fato de que representar todas as probabilidades implícitas pode tornar a construção da rede muito onerosa ou mesmo impossível.

Em síntese, as RB são grafos acíclicos direcionados, onde os *nós* representam as variáveis com medida de incerteza associada, e os *arcos* representam a existência causal direta entre os *nós* conectados (parte qualitativa), sendo a força dessa influência quantificada por probabilidades condicionais (Pearl, 1991).

Uma Rede Bayesiana consiste, portanto, de:

- um conjunto de variáveis $U = \{A_1, A_2, \dots, A_n\}$;
- um conjunto de arestas direcionadas entre variáveis, não podendo haver ciclos direcionados (DAG - *directed acyclic graph*);
- cada variável possui um conjunto limitado de estados mutuamente exclusivos;
- para cada variável A que possui como pais B_1, \dots, B_n , existe uma tabela $P(A|B_1, \dots, B_n)$. Se A não tem pais, deve-se ter uma tabela de *probabilidade a priori* $P(A)$.

A condição de Markov, que enuncia a não existência de uma relação de dependência direta entre dois *nós* a não ser que exista um arco conectando-os na rede, é respeitada pela teoria de Redes Bayesianas. A distribuição de probabilidade na rede é calculada a partir das probabilidades condicionais, ou seja,

$$P(U) = P(A_1, A_2, \dots, A_n) = \prod_{i=1}^n P(A_i | pa(A_i)) \quad (2.5)$$

sendo:

- $P(U)$ = probabilidade conjunta da rede;
- $pa(A_i)$ = pais do *nó* A_i ;
- $P(A_i | pa(A_i))$ = probabilidade condicional de A_i em relação aos seus pais.

Cada variável possui um número finito, maior ou igual a dois, de estados, que representam os valores possíveis da variável representada. Cada *nó* pode ser denominado *observável* ou *não observável*. A característica do *nó observável* é que há conhecimento sobre o estado de cada variável que define o *nó*. Estes possuem grande importância na inferência da rede, pois determinam as probabilidades dos *nós não observáveis*, juntamente com as probabilidades condicionais especificadas para a rede,

$$P(U) = P(A_i | O), i \in U, \quad (2.6)$$

que se lê: probabilidades de cada *nó* A_i , dado o conjunto de *nós* observados (O).

Existem diversos tipos de Redes Bayesianas, isto é, redes com variáveis discretas (distribuição multinomial), redes com variáveis contínuas (distribuição Gaussiana ou exponencial) e redes mistas (variáveis discretas e contínuas) (Buntine, 1991; Lauritzen e Spiegelhalter, 1988; Shachter, 1996). No caso deste trabalho, são empregadas redes mistas.

Como já discutido anteriormente, cada *nó* da rede Bayesiana retrata uma variável aleatória (discreta ou contínua, representada por tabela ou *pdf*), que representam as probabilidades condicionais. Associada ao grafo existe uma representação destas probabilidades, como mostra a Figura 2.1.

As Redes Bayesianas (Figura 2.2) são compostas de duas partes complementares, uma parte **qualitativa** e a outra **quantitativa** (Gaag, 1996). A parte qualitativa é retratada no modelo gráfico (grafo acíclico direcionado), onde as variáveis são os *nós*, com suas relações de dependência condicional representadas pelos arcos direcionados. Resumidamente, a parte qualitativa representa a dependência entre os *nós*.

Já a parte quantitativa é retratada pelo conjunto de probabilidades condicionais associadas aos arcos existentes no modelo gráfico, juntamente com as probabilidades estimadas à *priori* do *nó* de saída. Segundo Pearl (1991), a parte quantitativa avalia as dependências representadas na parte qualitativa, por meio das tabelas (ou *pdf* no caso de variáveis contínuas) que representam probabilidades condicionais.

Assim que a topologia da Rede Bayesiana tenha sido definida, é necessário que se especifiquem as probabilidades condicionais para cada *nó* descrito na rede, utilizando estes valores para calcular qualquer outro valor de probabilidade. O conhecimento prévio das informações é necessário para que se realize o cálculo das probabilidades condicionais. Estas informações são obtidas dos dados e contidas nas tabelas (ou *pdf*'s), conforme exemplo

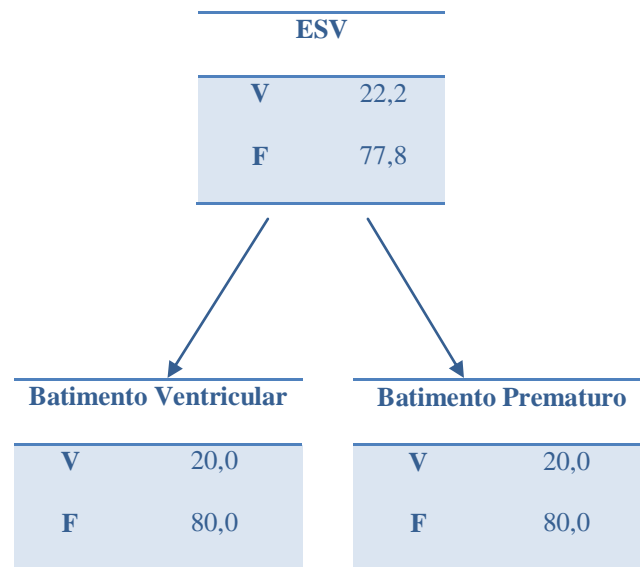


Figura 2.1: Exemplo gráfico de Rede Bayesiana com nós discretos e tabelas de probabilidades associadas a cada nó.

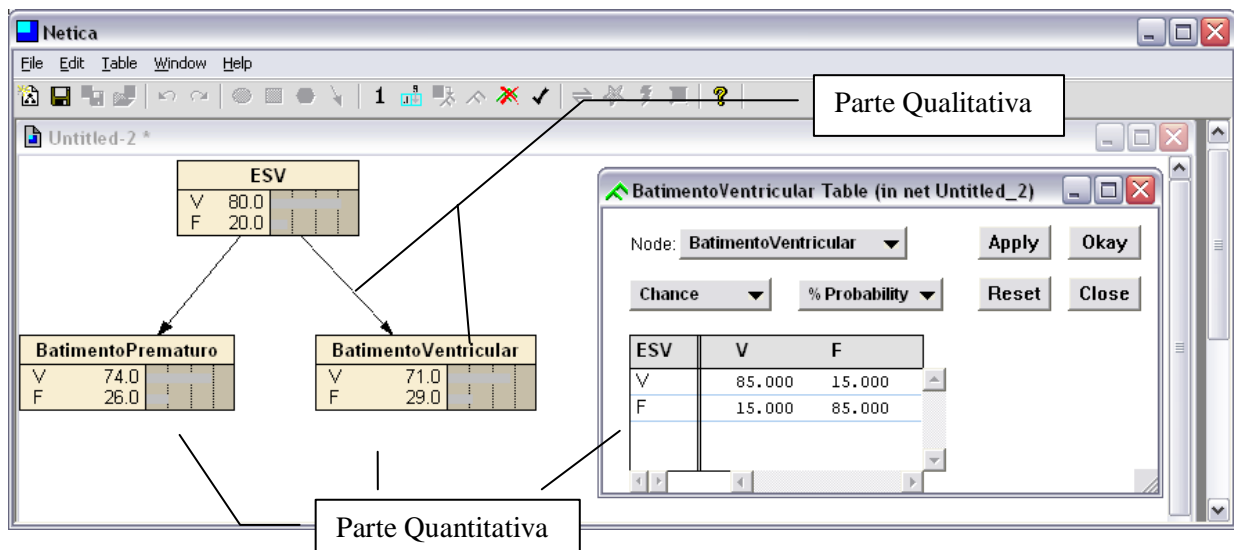


Figura 2.2: Exemplo de uma Rede Bayesiana demonstrando seus valores quantitativos e qualitativos.

ilustrado na Figura 2.3, que, a partir dos nós 'Batimento Prematuro' e 'Batimento Ventricular', permitiu o cálculo da probabilidade do nó 'ESV'.

A aplicação de Redes Bayesianas é ideal para o tratamento de incerteza, muito comum na área médica, e, além disso, modela o conhecimento do especialista de uma forma intuitiva. Este modelo é muito atraente pela simplicidade de sua estrutura, facilitando a compreensão de sua modelagem pelo especialista, provendo, desta forma, um maior entendimento e mutabilidade na adaptação da mesma, para que o especialista possua a autonomia de inserção ou exclusão de variáveis de acordo com a necessidade de cada paciente. Sua complexidade computacional também é bastante reduzida, conforme discutido anteriormente.

Redes Bayesianas constituem uma área em que grandes empresas mundiais e a comunidade científica têm muito interesse. Os mais importantes produtos comerciais nesta área são o Netica (Norsys Software Corporation, Noruega), que foi utilizado para ilustrar as Figuras 2.2 e 2.3, e o sistema HUGIN (Hugin Expert A/S, Dinamarca). Este último é utilizado pela GM, NASA e Ford, entre outras grandes corporações mundiais. Na NASA, as Redes Bayesianas são aplicadas para interpolação de dados espaciais para aquisição de conhecimentos (Stutz, Tatyler e Cheeseman, 1998). A Microsoft, por sua vez, também possui software nesta área. Primeiramente ela criou o grupo DTAS (Decision Theory and Adaptive Systems), que desenvolveu o MSBN (Microsoft Bayesian Network), software para avaliação deste tipo de rede. A equipe do DTAS incorporou ao sistema operacional da Microsoft diversas aplicações fundamentadas em Redes Bayesianas (pode-se encontrar esta tecnologia a

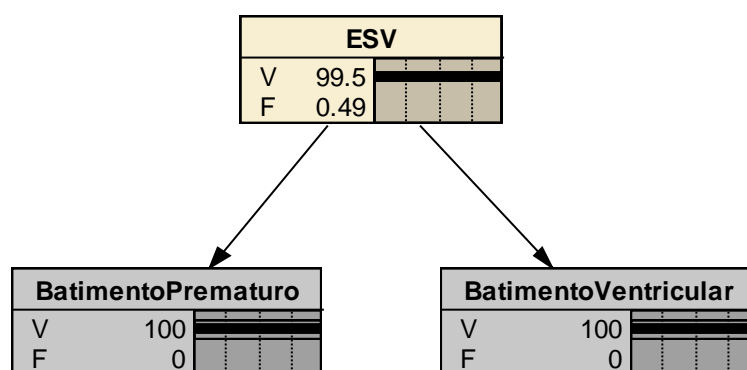


Figura 2.3: Exemplo de uma Bayesiana que calcula a probabilidade de ocorrência do nó ‘ESV’, dado que a probabilidade dos nós ‘BatimentoPrematuro’ e ‘BatimentoVentricular’ é de 100% V.

partir Windows 95, no *troublesholver* do Microsoft Word). Elas são utilizadas também pela Intel, realizando diagnósticos de defeitos em processadores, entre outros.

2.4 Aprendizado da Rede Bayesiana

Logo após a criação da estrutura da RB, definindo sua parte qualitativa, surge a necessidade da especificação de suas probabilidades (parte quantitativa). Estas probabilidades podem ser obtidas por meio de especialistas ou pelo aprendizado a partir de uma base de dados pré-existentes.

Utilizando o conhecimento do especialista, o método utilizado consiste em especificar os valores das probabilidades, listando valores numéricos entre 0% e 100% nas tabelas (ou escolhendo a *pdf* adequada à representação do problema). Este tipo de expressão tem dois problemas. O primeiro se encontra na dificuldade de especificar variáveis contínuas, ressaltando que este tipo de variável será utilizado neste trabalho. O segundo é devido à dificuldade de encontrar um especialista disponível para esta missão, problema este que foi encontrado no decorrer de desenvolvimento desta Tese. Por conseguinte, para definição das probabilidades, optou-se por utilizar o aprendizado a partir de uma base de dados.

Construir uma RB com auxílio de especialistas é um problema que envolve grande consumo de tempo, e, em alguns casos, o próprio domínio não é bem conhecido. Neste sentido, observa-se também que especialistas humanos são bons em julgar relações qualitativas entre as variáveis de um domínio, mas não tão bons em quantificar as probabilidades associadas. Este problema foi encontrado no decorrer deste trabalho: quando

se questiona a um cardiologista a probabilidade de acerto na detecção de ESV, por exemplo, não se encontra uma resposta precisa. Daí, processos de geração automática da parte quantitativa das redes serem implementados para auxiliar este processo de aquisição do conhecimento pela exploração, entre outras, das informações constantes em bancos de casos.

Existe também a alternativa de combinar o conhecimento de especialistas com técnicas de aprendizado, utilizando o conhecimento do especialista para modelar a topologia da rede, por exemplo.

O Aprendizado em Redes Bayesianas tem como objetivo a obtenção das probabilidades de cada *nó* da rede (Heckerman, Geiger e Chickering, 1995). Para este aprendizado, é necessário um algoritmo e uma base de dados.

No caso de auxílio ao diagnóstico médico, a base de dados corresponde, geralmente, a dados clínicos de casos anteriores, isto é, o domínio do problema modelado pela rede.

O algoritmo de aprendizado⁸ da Rede Bayesiana empregado nesta Tese é o algoritmo denominado de Máxima Verossimilhança (*Maximum Likelihood (ML)*) (Duda et al, 2001; Theodoridis, Koutroumbas, 2006 e Han, Kamber, 2000).

Para uma breve explicação sobre este algoritmo, considere um conjunto X de variáveis aleatórias e um modelo probabilístico paramétrico totalmente descrito pelo conjunto de parâmetros θ . Assim, pode-se calcular a distribuição conjunta $P(X|\theta)$ (probabilidade de ocorrência de X nesse modelo). Dois conceitos básicos em modelos probabilísticos devem ser discutidos: a classificação e o treinamento. A técnica empregada na classificação consiste em decidir dentre os vários modelos θ_i qual gerou o dado de entrada, a observação X . Já na técnica empregada no treinamento deve-se estimar os parâmetros θ a partir da observação X .

A Classificação (fase de teste) da RB baseada neste algoritmo consiste em decidir dentre os vários modelos θ_i qual tem maior probabilidade de ocorrência da observação X , isto é, escolhe-se o modelo θ_i tal que

$$\theta^* = \underset{i}{\operatorname{argmax}} P(X|\theta_i). \quad (2.7)$$

No treinamento (fase de aprendizado) do ML estimam-se os parâmetros θ do modelo como aqueles que fornecem o máximo da função de verossimilhança $L(\theta) = P(X|\theta)$, isto é

$$\theta^* = \underset{\theta}{\operatorname{argmax}} L(\theta) = \underset{\theta}{\operatorname{argmax}} P(X|\theta), \quad (2.8)$$

⁸ Algoritmo de aprendizado também pode ser referenciado com algoritmo de treinamento da rede.

onde, θ^* é chamado de estimador de máxima verossimilhança (maximum likelihood estimator) de θ .

A função da aprendizagem em uma Rede Bayesiana é encontrar a distribuição de probabilidades mais próxima possível da distribuição de probabilidades representada pelos dados.

É importante que se evite o problema de superaprendizagem, que significa a cópia exata da distribuição representada pelos dados do exemplo. Para que isto não ocorra, é necessário a divisão da base de dados em duas partes, treinamento e teste. Para se estimar as probabilidades de cada *nó* da rede são selecionados uma parte dos dados da base de dados, que são denominados dados de treinamento (ou bases de treinamento). Se a base de treinamento é rotulada, isto é, os dados de treinamento são rotulados por um especialista, é dito que há um *treinamento supervisionado*. Caso contrário, isto é, quando os dados não são rotulados por especialista, é dito que há um *treinamento não supervisionado*.

O algoritmo de propagação de evidência utilizado na fase de teste neste trabalho é também de grande importância na construção da RB. Conforme Russell e Norvig (2004), os algoritmos de propagação de evidência em Redes Bayesianas são divididos em dois grupos, a saber,

- exatos: resultam em probabilidades posteriores exatas. Exemplos: *junction tree*, *clustering*, *polytree*, *cutset conditioning method* e *relevance-based method*;
- aproximados: resultam em probabilidades posteriores aproximadas. Exemplos: *forward sampling (logic sampling)*, *likelihood weighting*, *gibbs sampling*, algoritmo *Metropolis-Hasting*, *self importance sampling*, *backward sampling* e *adaptive importance sampling for Bayesian Networks (AIS-BN)*.

O método que foi utilizado neste trabalho é o *junction-tree* (Jordan, 1999) (observe-se também um comentário do toolbox BNT no Anexo 3). O conceito de *junction-tree* surgiu na teoria dos grafos, possuindo um papel importante na resolução de problemas que envolvam inferência probabilística. É também conhecido como *Clique Trees*, *Tree Decomposition* ou *Join Trees*. Esse método é usado para realizar inferência (marginalização) exata em grafos em geral. Muito utilizado em aprendizado de máquina, ele obtém uma sequência ótima de eliminação de variáveis e cria uma estrutura para propagar as multiplicações e marginalizações das tabelas (Martins, 2008).

O algoritmo *junction-tree* propõem uma metodologia que utiliza a Regra da Cadeia⁹ em conjunto com a distribuição conjunta universal, dada na equação

$$P(U, e) = \prod P_{A_i|pais(A_i)} \prod e_i, \quad (2.9)$$

onde tem-se a equação para propagação de evidências (crenças) e_1, \dots, e_n , sobre o universo U de variáveis, e A_i representa um *nó* da rede, onde $i = 1, \dots, n$.

Este algoritmo propõe uma metodologia que faz uso da lei distributiva, que tem por objetivo a intenção de reduzir e otimizar o tamanho das tabelas (Jensen, 2001).

De acordo com Russell e Norvig (2004) a idéia básica do algoritmo *junction-tree* é ‘unir *nós* individuais da rede para formar *nós* de agrupamento, de tal modo que a rede resultante seja uma poliárvore¹⁰. No entanto, é necessário verificar antes de construir a *junction-tree* se a rede causal (grafo) do problema é triangulado, pois para a construção desta árvore é necessário que o grafo seja deste tipo. Caso o grafo não seja triangulado, realiza-se o processo de triangulação. Para determinar a triangularidade de grafos, inicialmente é necessário transformar o grafo direcionado em um grafo não direcionado. Para isso são substituídas as arestas direcionadas por arestas não direcionadas e são incluídas as conexões morais¹¹ (Jensen, 2001).

O objetivo da etapa de teste é calcular a probabilidade condicional de um *nó*, ou de um conjunto de *nós*, dado os valores dos *nós* observados. O algoritmo *junction-tree* decompõe o cálculo global das probabilidades conjuntas em um conjunto de cálculos locais interligados. Este algoritmo é descrito, resumidamente, da seguinte forma (Paskin, 2003; Xu, Zhu, 2004 e Huang, Darwiche, 1994):

⁹ A regra da cadeia é uma metodologia para realizar a propagação de evidências em Redes Bayesianas, quando é conhecida a probabilidade conjunta universal de todas as variáveis.

¹⁰ Redes deste tipo são chamadas de **singularmente conectadas** (singly connected networks) ou **poliárvores** (polytrees), e possuem uma propriedade particular: a complexidade espacial e temporal da inferência exata nelas é linear no tamanho da rede, e contém apenas um caminho não direcionado entre quaisquer dois nós da rede.

¹¹ Conexões morais, segundo Jensen (2001), são arestas não direcionadas que conectam dois ou mais nós pais que possuem um filho em comum.

1. moralização: converte um grafo direto em um grafo indireto. Este item consiste em conectar os *nós* que possuem filhos em comum e tornar os arcos não direcionados;
2. triangulação e construção da árvore de junção: determina a ordem de escolha dos *nós* do grafo para a construção da árvore de junção;
3. propagação das probabilidades: que tem como objetivo a propagação das probabilidades (inferência).

2.5 Redes Bayesianas Dinâmicas

A Rede Bayesiana é um método utilizado para representar o conhecimento incerto, o que permite a argumentação baseada em teoria de probabilidade (Castillo, Gutiérrez e Hadi, 1997). Em ambientes incertos, porém, é necessário que se mantenham atualizações sobre o estado do ambiente.

Cada *nó* em uma Rede Bayesiana está associado a uma variável aleatória, que pode assumir valores dentro de uma escala discreta ou contínua (observe-se que as redes representadas nas Figuras 2.1, 2.2 e 2.3 são representações discretas). Em geral, as relações de dependência definidas entre as variáveis da rede são estáticas, embora determinadas aplicações necessitem de uma relação temporal entre os *nós* da rede. Isto se faz necessário quando se deseja saber o estado atual de um ambiente que muda e recebe novas evidências ao longo do tempo. Um problema que pode ser apontado na abordagem Bayesiana Clássica¹² é que esta não possui um mecanismo para retratar o tempo.

Redes Bayesianas Clássicas não provêm um mecanismo direto para representar dependências temporais. Por exemplo, é difícil representar uma situação como a variabilidade de quando um empregado chega ao trabalho e as relações causais entre o tempo de chegada e eventos mais recentes (Young e Santos, 1996).

Até o momento, a definição de uma Rede Bayesiana não explora relacionamentos temporais entre as variáveis. Portanto, para que se possam captar mudanças ocorridas nas variáveis ao longo do tempo e, por conseguinte, a sequência de relações causais, é necessária a utilização de Redes Bayesianas Dinâmicas¹³ (RBD). Neste contexto, se faz necessário inserir um fator tempo na identificação das variáveis.

¹² Uma Rede Bayesiana Clássica é aqui representada por modelos estáticos, isto é, a variável tempo não intervém nas mesmas.

¹³ A literatura utiliza as palavras ‘dinâmico’ e ‘tempo’ de maneira permutável. Dinâmico é o oposto de estático.

Redes Bayesianas Dinâmicas têm recebido atenção crescente como uma ferramenta para modelar processos estocásticos complexos (Darwiche, 2001). Um processo estocástico é uma variável que se comporta, durante o tempo, de uma maneira em que pelo menos parte é considerada randômica, ou seja, se X é um processo estocástico, então $X(t)$ é uma variável aleatória para cada valor de t pertencente ao conjunto índice T . De maneira mais formal, é definida pela probabilidade da evolução x_t da variável x durante o tempo t . Para instantes t_1 , t_2 e t_3 é calculada a probabilidade dos valores correspondentes, x_1 , x_2 e x_3 , estarem numa faixa específica de valores. Desta forma, é necessário condicionar a probabilidade dos eventos futuros com base em informações passadas.

Para modelar um processo estocástico, entre outras abordagens, Redes Bayesianas Dinâmicas são uma opção, pois elas conseguem capturar o desenvolvimento do processo ao longo do tempo. Uma Rede Bayesiana Dinâmica é uma cadeia de repetição da mesma Rede Bayesiana Clássica quantas vezes forem necessárias. A utilização de RBD nesta tese é ideal, já que o batimento cardíaco é um processo estocástico (Belot, 2002).

Para tratar esta idéia, é preciso especificar uma variável aleatória para cada aspecto representado a cada instante de tempo. Neste sentido, um processo estocástico pode ser um processo em tempo discreto ou em tempo contínuo. O primeiro é aquele cujo valor da variável só varia em determinados instantes de tempo, e esta variável é classificada como variável discreta. Já para o segundo caso, a variável assume valores a qualquer instante de tempo, e por isso é denominada variável contínua.

Nesta circunstância, não é interessante a abordagem do processo em tempo contínuo, pois o interesse nesta Tese é unicamente por processos estocásticos em tempo discreto.

Desta forma, uma Rede Bayesiana Dinâmica é um modelo especial da Rede Bayesiana Clássica, utilizada para modelar processos estocásticos dinâmicos. Este tipo de rede modela a distribuição de probabilidades sobre um conjunto de variáveis aleatórias V_1, V_2, \dots , as quais podem ser subdivididas em variáveis observáveis e variáveis não observáveis (seção 2.3), geralmente denotadas por E_t . Cada variável em uma RBD é associada a um instante de tempo t , e denotada por V_t .

A característica chave desta abordagem é definir o número de instantes necessários para modelar um problema qualquer.

A principal utilidade de uma RBD é a predição, isto é, calcular a probabilidade de alguns eventos futuros ocorrerem, analisando evidências passadas.

Normalmente, a distribuição de algumas variáveis V_t^i no tempo t depende de uma variável no tempo $t-1$, V_{t-1}^i , definida por alguma distribuição condicional $P(V_t^i | V_{t-1}^i)$.

Uma RBD é um par $D = (B_I, B_{\rightarrow})$ onde B_I é a Rede Bayesiana que define um estado inicial $P(V_I)$, e B_{\rightarrow} é uma Rede Bayesiana temporal que representa dois instantes de tempo (2TRB), que define $P(V_t/V_{t-1})$ por uma representação gráfica acíclica da forma (Belot, 2002)

$$P(V_t | V_{t-1}) = \prod_{i=1}^N P(V_t^i | pa(V_t^i)), \quad (2.10)$$

onde V_t^i é o i -ésimo nó no instante t e $pa V_t^i$ são os pais de V_t^i no grafo.

As variáveis possuem diferentes formas (como tabelas, Gaussianas). Os pais de um nó V_t^i são os mesmos, tanto no primeiro instante quanto nos instantes posteriores. Os arcos que conectam os nós entre os diferentes instantes de tempo vão da esquerda para a direita, representando, desta maneira, o passar do tempo. Se existe um arco conectando um nó V_{t-1} a um nó V_t , então denomina-se esta variável como um nó ‘persistente’. Consequentemente, pode-se dizer que quando arcos conectam dois nós em instantes de tempo distintos, isto pode ser considerado como a persistência de um fenômeno durante o tempo. No entanto, os arcos de um mesmo instante de tempo são vistos como um efeito causal imediato.

A semântica da Rede Bayesiana Dinâmica pode ser definida estendendo a definição de 2TRB sobre um comprimento de T instantes de tempos. A distribuição de probabilidade conjunta, então, seria (Belot, 2002)

$$P(V_{i:T}) = \prod_{t=1}^T \prod_{i=1}^N P(V_t^i | pa(V_t^i)), \quad (2.11)$$

Uma Rede Bayesiana Dinâmica, no primeiro instante de tempo, é também considerada como uma Rede Bayesiana Clássica (estática). Por exemplo, a Figura 2.4 representa uma Rede Bayesiana Clássica.

A Figura 2.5 mostra um exemplo de uma Rede Bayesiana Dinâmica, que é uma rede representada pelo modelo 2TRB, possuindo como primeiro instante de tempo a rede apresentada na Figura 2.4. Assim, uma RBD possui um conjunto de variáveis aleatórias para cada instante de tempo t , e a cada instante t é representada como uma RB estática, nunca deixando de fixar a algumas variáveis sua respectiva dependência temporal. A Figura 2.5 também demonstra o fenômeno de persistência, unicamente entre os nós A_t , C_t e F_t . Esta rede

é mais detalhada que a rede da Figura 2.4, pois o resultado, tanto das dependências temporais como das dependências causais, pode ser calculado num mesmo instante de tempo.

Assim, em vez de utilizar apenas dois instantes de tempo ($t-1$ e t), podem-se ter quantos instantes de tempo forem necessários, reconhecendo que os arcos que conectam dois *nós* em tempos diferentes representam dependências temporais a cada instante de tempo (Figura 2.6). O objetivo é que se possa formalizar a representação da rede tanto estática quanto dinâmica.

Em síntese, a representação das RBD tem como base geral as seguintes suposições (Gallego, 2010):

1. suposição Markoviana: o futuro é condicionalmente independente do passado, dado o presente. (Esta propriedade não deve ser confundida com a condição de Markov que se aplica na definição formal de uma Rede Bayesiana, embora ambas estejam muito relacionadas);
2. processo estacionário no tempo: as probabilidades condicionais do modelo não se alteram ao longo do tempo.

Estas duas suposições implicam que se pode construir uma RBD a partir da representação da mesma RB estática que se repete ao longo de certo intervalo de tempo predefinido, com base em uma rede de transição entre as fases consecutivas.

Modelos de Redes Bayesianas Dinâmicas são uma escolha atrativa, pois combinam representações em grafos intuitivos com algoritmos eficientes de inferência e aprendizado (Pavlovic *et al.*, 2000). Nesta Tese, a utilização de Redes Bayesianas Dinâmicas para detecção de episódios de extrassístoles ventriculares em registros de ECG é discutida em detalhes, e é mostrado que esta é uma abordagem adequada para tratar este problema.

Modelos ocultos de Markov (*hidden Markov models* - HMM) (Rabiner; Juang, 1986) e

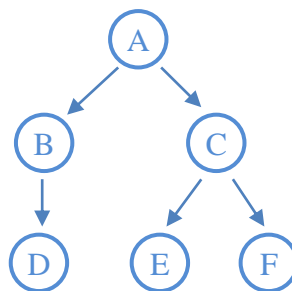


Figura 2.4: Exemplo de uma Rede Bayesiana Clássica.

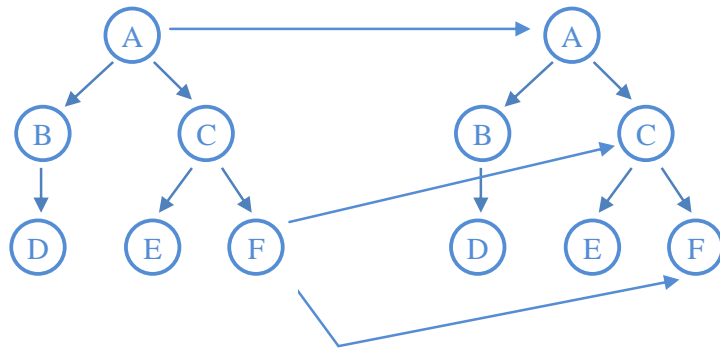


Figura 2.5: Representação gráfica de uma Rede Bayesiana Dinâmica em dois instantes de tempo. (2TRB).

filtros de Kalman (*Kalman filters*) (Kalman, 1960) podem ser considerados como casos particulares de RBD (Nilsson, 1998). Porém, não serão discutidos neste trabalho.

2.5.1 Inferência em RBD

Cálculos comuns em Redes Bayesianas, como o cálculo da distribuição conjunta, a inferência e o treinamento, também são os problemas básicos numa Rede Bayesiana dinâmica. Todos esses problemas em Redes Bayesianas dinâmicas podem ser resolvidos usando algoritmos para Redes Bayesianas comuns, levando em conta que a estrutura repetida de uma Rede Bayesiana Dinâmica pode simplificar a notação e estrutura dos algoritmos.

Em princípio, a inferência em RBD é a mesma para as Redes Bayesianas, porque se aplicam os mesmos métodos (Gallego, 2010). No entanto, ao lidar com RBD a complexidade da estrutura desta rede é maior, porque os métodos mais comuns são aqueles baseados em simulações estocásticas, como, por exemplo, o método de Monte Carlo, que é também conhecido como filtro de partículas.

O problema geral da inferência em RBD está em calcular $P(V_{i:t_0}/E_{t_1:t_2})$, onde $V_{i:t_0}$ (t_0 é o tempo atual) representa a i -ésima variável não observável no tempo t e $E_{t_1:t_2}$ representa todas

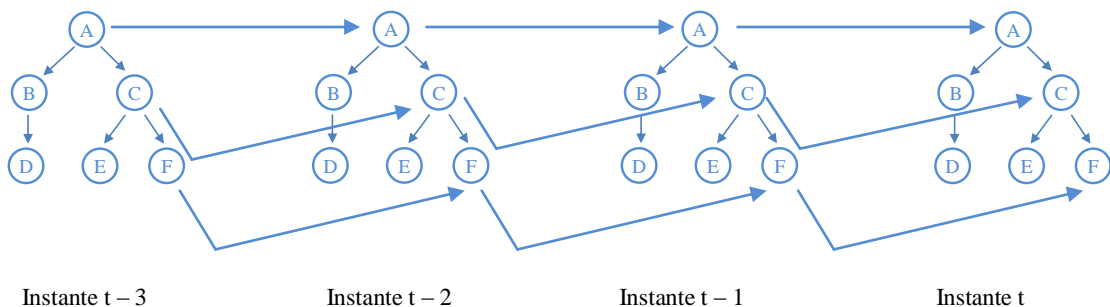


Figura 2.6: Rede Bayesiana Dinâmica representada em 4 instantes de tempo.

as evidências entre os tempos t_1 e t_2 . A observação no instante t é $E_t = e_t$ para algum conjunto de valores e_t .

Existem vários casos especiais de inferência básica que podem ser considerados (Murphy, 2002). Três tipos importantes de inferência em uma RBD são a filtragem (*filtering*), a predição (*prediction*) e a suavização (*smoothing*).

Na filtragem computa-se $\mathbf{P}(V_t/E_{1:t})$, onde é calculada a distribuição a *posteriori* do estado atual, dada toda evidência observada até o momento. Na predição, $\mathbf{P}(V_{t+k}/E_{1:t})$ para $K > 0$, onde é calculada a distribuição a *posteriori* do estado futuro, dada toda evidência observada até o momento. E, na suavização, $\mathbf{P}(V_k/E_{1:t})$ para $1 \leq k < t$, tarefa esta que consiste em calcular a distribuição a *posteriori* do estado passado, dado toda a evidência observada até o momento. Normalmente assume-se que o modelo é invariante no tempo (estacionário), ou seja, $\mathbf{P}(V_t/V_{t-1})$ e $\mathbf{P}(E_t/V_t)$ são os mesmos para todo t .

No primeiro instante de tempo a representação de uma RBD é o reflexo de uma RB estática. É como se realizasse um corte no primeiro instante de tempo. Neste caso, o processo de atualização nas RBD se pode inferir utilizando algoritmos já aplicados nas RB. Isto significa que, dada uma sequência de observações, pode-se construir a representação da RB completa duplicando cortes até a que a rede seja suficientemente grande para representar todas as observações. Uma vez completa a RBD (duplicados todos os cortes), pode-se aplicar qualquer algoritmo de inferência (claro que deve estar de acordo com a estrutura da rede). Esta técnica é conhecida como desdobramento.

Neste sentido, é importante ressaltar que uma Rede Bayesiana Dinâmica é uma cadeia de repetição da mesma Rede Bayesiana Clássica quantas vezes forem necessárias. Assim, pode-se afirmar que as RBD's estendem as RB, introduzindo-se no segundo modelo a noção de fatias de tempo (ou instantes de tempo), e especificando-se probabilidades de transição entre fatias consecutivas, ou seja, $\mathbf{P}(V_t/V_{t-1})$, onde $V_t = \{V_t^1, \dots, V_t^i\}$. Como consequência, RBD podem ser representadas como duas RB interconectadas: a RB que representa a dependência dos estados do sistema no tempo t (RB_t , intradependência) e a RB que representa a dependência entre RB_t e RB_{t-1} (interdependência). Nesta proposta a interdependência foi modelada assumindo-se que o nó V_i em RB_t é unicamente dependente do nó V_i em RB_{t-1} (esta proposição somente será válida e aplicada em variáveis persistentes).

Para um conjunto de observações, pode-se construir uma representação completa da RBD usando uma RB (desdobrando da RBD até que a rede fique grande o suficiente para

acomodar as observações). Uma vez desdobrada, pode-se usar qualquer algoritmo de inferência em Redes Bayesianas. Entretanto, as aplicações de tais algoritmos não são tão eficientes quanto a aplicação de algoritmos próprios para o processamento dinâmico, principalmente quando se tem uma sequência elevada de observações e a técnica de desdobramento não é suficiente. Neste caso, se precisará de um espaço $O(t)$ (Russell e Norvig, 2004), e deste modo a complexidade crescerá sem limites, conforme se adicionassem mais informações. Além disso, cada vez que se acrescenta uma nova observação o algoritmo de inferência é executado novamente. O tempo de processamento de inferência por atualização também aumentará em $O(t)$. Geralmente, para conseguir a otimização da complexidade $O(t)$ usa-se programação dinâmica, evitando, desta forma, o cálculo repetido da mensagem propagada.

Como RBD's são RB's que representam um modelo probabilístico temporal, os mesmos algoritmos para inferência probabilística em RB podem ser empregados. Por questões de eficiência, geralmente emprega-se o algoritmo aproximado para inferência conhecido como *particle filtering* (Russel e Norvig, 2000), que é uma modificação do algoritmo *likelihood weighting*¹⁴.

Segundo Gallego (2010), para lidar com os casos em que as RBD's são muito grandes, isto é, RBD's com muitas fatias de tempo, e ainda aplicar algoritmos comuns à RB, utiliza-se um processo que consiste na implantação de uma 'janela móvel', de tamanho fixo, que avança com o tempo. Quando esta janela é movida para um próximo instante de tempo, um corte de uma fatia de tempo é efetuado e excluído, com uma nova fatia de tempo sendo adicionada. Com a aplicação da técnica de 'janela móvel', cada vez que se avança no tempo as observações anteriormente recebidas não estarão diretamente disponíveis. Em seu lugar se terá a crença atual dos nós, que passarão posteriormente a serem as probabilidades *a priori* seguintes.

A Figura 2.7 mostra o processo anteriormente apresentado de uma RBD que utiliza 'janela móvel' com 2TRB (dois cortes de tempo ou duas fatias de tempo). Esta estrutura pode ser considerada como uma RBD genérica.

¹⁴ Gera amostras dos nós de estado da rede em ordem topológica, dando pesos a cada amostras de acordo com a probabilidade de ocorrência, dadas as variáveis de evidência observadas. Aplicar *Likelihood Weightening* diretamente na RBD desdobrada apresentaria os mesmos problemas de tempo e espaço.

O algoritmo baseado em Gallego (2010) que realiza o processo de atualização de uma RBD que utiliza ‘janela móvel’ é descrito no Anexo 2. Os passos do processo são similares aos da técnica utilizada no filtro de Kalman. Cabe aqui ratificar a possibilidade de utilização de inferências aplicadas na RB aplicando-se as mesmas nas RBD, porém criando a estrutura topológica que leva em consideração vários instantes de tempo.

Neste trabalho foram utilizados para a RBD os mesmos algoritmos aplicados na RB. Porém, foram criadas algumas regras para que a complexidade não seja excessiva. Tais regras são denominadas Três Itens de Restrição (TIR)¹⁵, e correspondem a:

1. deve-se considerar um espaço de amostragem (*sample*) razoavelmente pequeno (por exemplo de 2TRB a 4TRB). Observa-se que nesta concepção o espaço amostral é relacionado aos instantes de tempo, isto é, definindo-se o tamanho da ‘janela móvel’;

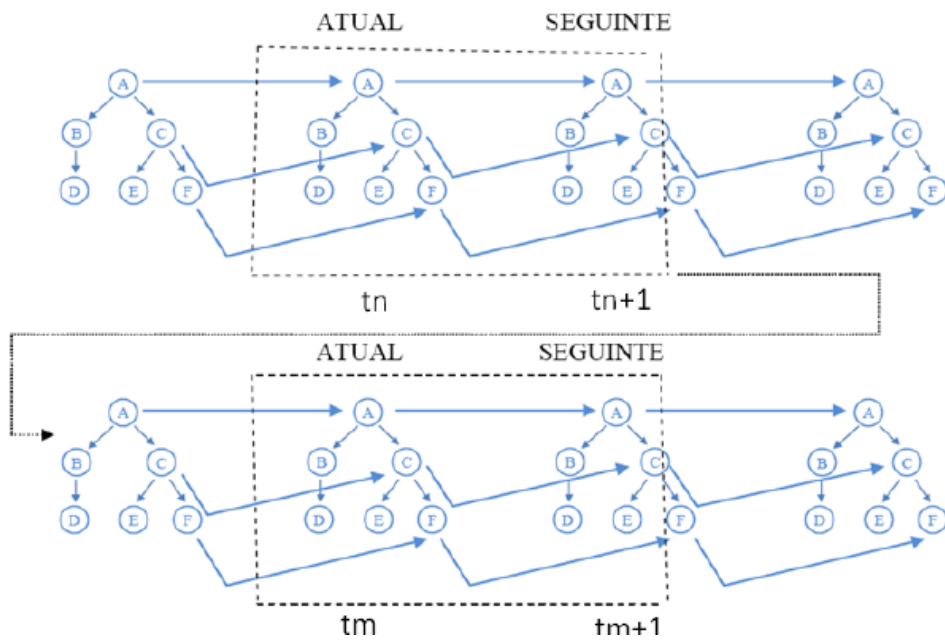


Figura 2.7: Exemplo de uma ‘janela móvel’ com duas fatias de tempo em uma RBD. A seta pontilhada indica o decorrer do tempo e a direção em que a janela se move. O segundo retângulo pontilhado da figura representa a janela após um certo tempo.

¹⁵ Visando que o cálculo repetido da inferência propagada seja considerado pequeno, isto é, que a repetição do cálculo dos valores de probabilidade propagados na rede ocorra em tempo hábil.

2. o tempo gasto para a compilação do algoritmo de treinamento e teste desta rede deve ser pequeno¹⁶;
3. a quantidade de variáveis da estrutura da rede representada em cada instante de tempo. Esta última restrição influencia diretamente no tamanho da amostragem e na complexidade de compilação do algoritmo, e por este motivo esta restrição deve ser tratada com cuidado.

Normalmente as distribuições $\mathbf{P}(V_t/V_{t-1})$ e $\mathbf{P}(V_t/E_t)$ de uma RBD devem ser aprendidas a partir dos dados disponíveis. Da mesma forma que em uma Rede Bayesiana, conhecendo-se a estrutura da rede e sendo todas as variáveis observáveis, o aprendizado é feito pela estimação ML (*Maximum Likelihood*). Se a estrutura é conhecida, mas existem variáveis não observáveis, utilizam-se métodos de gradiente ou EM (*Estimation-Maximisation*) para o aprendizado (Friedman et al, 1998; Dempster et al, 1997).

Para o aprendizado da RB, este trabalho utilizou o algoritmo ML disponível na biblioteca do MATLAB denominada BNT (2007) (ver também o Anexos 1), porque no aprendizado foram implementadas variáveis observáveis. A RB implementada inicialmente é considerada dentro dos padrões do TIR, pois:

1. quanto ao tamanho da amostragem: foram implementadas várias topologias da RBD consideradas no intervalo de 2TRB a 4TRB, onde a melhor solução é apresentada em 3TRB, isto é, a ‘janela móvel’ é de 3 fatias de tempo;
2. quanto ao tempo de compilação do algoritmo: na fase de treinamento da rede, o tempo é de menos de 7 minutos para um conjunto de batimentos de 70.832 batimentos. Na fase de teste, o tempo gasto para realização da classificação de 4.000 batimentos foi de 2,15 minutos. Sabendo-se que, em média, um indivíduo possui uma frequência de aproximadamente 75 batimentos por minuto, fica claro que a RBD pode realizar a classificação em tempo real¹⁷;

¹⁶ A velocidade de execução de um algoritmo, de uma maneira geral, pode ser considerada rápida ou não de acordo com o problema abordado, levando-se em consideração o tempo de execução e o objetivo. Por exemplo, no caso de detecção de ESV, é preciso que o tempo de resposta na fase de teste seja praticamente em tempo real, mesmo que ocorra um atraso imperceptível. (Em outras palavras, o resultado da classificação de um evento deve ocorrer antes que o evento seguinte seja gerado, ou seja, o tempo de processamento deve ser menor que o tempo de amostragem dos dados.)

¹⁷ Destaca-se neste momento que o computador utilizado é um Core2Duo com 1Gb de RAM.

3. quanto à quantidade de nós da rede: no exemplo apresentado neste trabalho, foi propagada uma rede estática no tempo (3TRB), de 5 nós, sendo esta rede aquela que apresentou os resultados mostrados no item 2.

É importante salientar que as RBD's obedecem a proposição de Markov, de que o estado atual depende de estado anterior, pois, de outra forma, novas tabelas de probabilidade teriam que ser montadas a cada fatia de tempo. Sendo assim, mais uma vez se ratifica a possibilidade de utilização da tecnologia descrita anteriormente como TIR.

Capítulo 3

Redes Bayesianas Estáticas para a Detecção de Extrassístoles Ventriculares

Este capítulo apresenta um sistema para apoio à decisão médica, utilizando Rede Bayesiana estática implementada usando o *toolbox* BNT (*Bayesian Net toolbox*) do MATLAB[®]. O problema abordado é a identificação de extrassístoles ventriculares (ESV) em sinais de ECG (Eletrocardiograma). A motivação para o uso da tecnologia de Redes Bayesianas, como já mencionado, é o fato de ela ser adequada para o tratamento de incerteza, presente no raciocínio clínico. Mais ainda, a inferência probabilística é a priori adequada para modelar este tipo de problema, pois usa variáveis aleatórias para propagar a incerteza pertinente ao problema.

Várias topologias de Redes Bayesianas são implementadas e testadas neste capítulo, procurando uma estrutura mais adequada ao problema proposto, conforme é detalhado a seguir. Neste sentido, este capítulo engloba todo estudo da RB estática para detecção de ESV, detalhando todos os passos que foram executados para se decidir pela topologia da rede, além da base de dados e do algoritmo de segmentação empregados no treinamento e nos testes das redes.

Inicialmente são apresentadas informações preliminares sobre a RB, e, em seguida, o capítulo é dividido em quatro momentos. Primeiramente é feita uma comparação entre os resultados das Redes Bayesianas, Redes Heurísticas e Redes Empíricas, objetivando demonstrar a maior eficiência da RB. Em um segundo momento, a RB é testada com duas bases de dados diferentes, objetivando-se escolher a base de dados a ser efetivamente utilizada bem como comprovar a eficiência da fusão de canais, proposta neste capítulo. No terceiro momento, é apresentada a comparação do funcionamento da RB embasada em dois algoritmos de segmentação, visando assim demonstrar que a RB gera resultados adequados independentemente do algoritmo de segmentação do ECG (supondo-se algoritmos que geram resultados de segmentação adequados). Na sequência, no quarto momento é discutido o efeito da utilização da onda P como um *nó* da rede e também como critério de desempate na Fusão de resultados entre RB com diferentes bases de aprendizado (também proposta neste

capítulo). Por fim, são apresentadas as conclusões obtidas durante o estudo da RB e o sistema proposto para continuação do trabalho, apresentado no capítulo quatro.

3.1 Informações Preliminares à Implementação da Rede Bayesiana Estática

Quando se quer desenvolver um sistema de apoio ao diagnóstico médico, como no caso presente, o primeiro passo é identificar claramente os seguintes pontos: ‘o que se deseja classificar?’, ‘como o especialista age?’, ‘como ele classifica cada doença?’, ‘quais as informações relevantes para um diagnóstico correto?’. A partir destas informações ou dados relevantes resta, então, perguntar ‘qual a metodologia que ele utiliza para concatenar estes dados e gerar um resultado correto?’.

Focalizando estes questionamentos é que se faz a proposta de um sistema de auxílio ao diagnóstico médico para classificação de episódios de ESV, usando como dados de entrada os dados relevantes, neste caso o sinal de ECG devidamente rotulado.

O primeiro passo foi utilizar Redes Bayesianas Estáticas para a classificação automática de batimentos ESV presentes no sinal de ECG. Para isto, foram implementados vários modelos (estruturas) de Rede Bayesiana, visando encontrar aquele com melhor desempenho para esta aplicação específica. Para todos os casos, as simulações foram realizadas utilizando o MATLAB[®].

Foram seguidos os passos mostrados na Figura 3.1 para elaboração da Rede Bayesiana, cuja descrição está contemplada nos itens A, B, C e D a seguir.

A. Identificar Variáveis Aleatórias e Evidências

As variáveis aleatórias foram descritas através dos seguintes *nós*, os quais modelam a Rede Bayesiana:

- ‘*ESV*’, que representa a possibilidade de ocorrer ou não esta arritmia (ESV);
- ‘*Batimento Prematuro*’, que indica a probabilidade de acontecer ou não um complexo QRS prematuro;
- ‘*Batimento Ventricular*’, *nó* que tem a função de indicar a probabilidade de ocorrência de uma deformação do complexo QRS, o qual fica mais largo quando deformado (ver Figura 1.7);

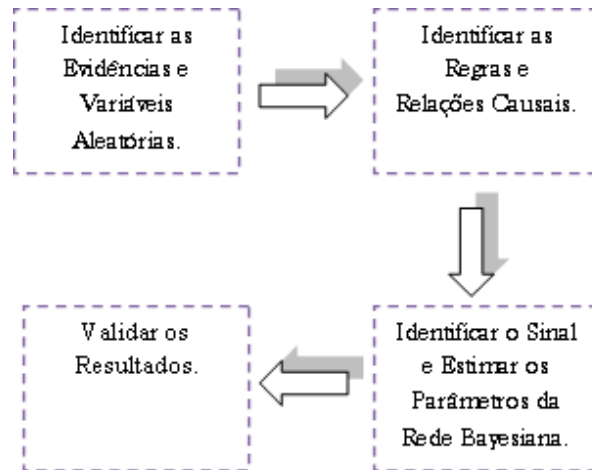


Figure 3.1: Diagrama que representa os passos seguidos para a criação da Rede Bayesiana.

- ‘*Batimento Prematuro_BA*’, um vetor que possui um atraso de um batimento, isto é, o vetor ‘*Batimento Prematuro_BA(i)*’ armazena o valor do vetor ‘*Batimento Prematuro(i-1)*’;
- ‘*Batimento Ventricular_BA*’, um vetor que possui um atraso de um batimento, isto é, o vetor ‘*Batimento Ventricular_BA(i)*’ armazena o valor do vetor ‘*Batimento Ventricular(i-1)*’;
- ‘*LLC1*’ e ‘*LLC2*’, que são os *nós* que representam exatamente o valor numérico da verossimilhança do complexo QRS, utilizados para determinar o valor do *nó* ‘*Batimento Ventricular*’. O valor quantificado deste *nó* foi calculado através de um algoritmo que indica a verossimilhança do complexo QRS, desenvolvido por Andreão, Dorizzi e Boudy (2006). A diferença entre estes dois *nós* está na terminação, C1 ou C2, que representam, respectivamente, canal 1 e canal 2. Ou seja, pode-se lidar com dois registros de ECG tomados simultaneamente para o mesmo indivíduo, como é o caso dos registros gravados na base de dados utilizada no treinamento e no teste da rede;
- ‘*RRC1*’ e ‘*RRC2*’, correspondente aos *nós* que representam exatamente o valor numérico do tempo entre dois picos R (intervalo RR). Eles são utilizados para calcular a probabilidade do *nó* ‘*Batimento Prematuro*’. O que diferencia estes dois *nós* é a terminação, que significa canal 1 (C1) ou canal 2 (C2);
- ‘*LLC1BA*’ e ‘*LLC2BA*’, correspondentes aos valores de ‘*LLC1*’ e ‘*LLC2*’, descritos anteriormente, porém a terminação ‘*BA*’ significa que é um vetor que armazena os valores dos batimentos anteriores, ou seja, este vetor possui um atraso

de um batimento¹, isto é, o vetor ' $LLC1BA(i)$ ' armazena o valor do vetor ' $LLC1(i-1)$ ';

- ' $RRC1BA$ ' e ' $RRC2BA$ ', os quais correspondem aos valores de ' $RRC1$ ' e ' $RRC2$ ', descritos anteriormente, porém a terminação ' BA ' significa que é um vetor que armazena os valores dos batimentos anteriores, ou seja, este vetor possui um atraso de um batimento, isto é, o vetor ' $RRC1BA(i)$ ' armazena o valor do vetor ' $RRC1(i-1)$ ';
- ' $LLC1BP$ ' e ' $LLC2BP$ ', os quais correspondem aos valores de ' $LLC1$ ' e ' $LLC2$ ', descritos anteriormente, porém a terminação ' BP ' significa que é um vetor que armazena os valores dos batimentos posteriores, isto é, o vetor ' $LLC1BP(i)$ ' armazena o valor do vetor ' $LLC1(i+1)$ ';
- ' $RRC1BP$ ' e ' $RRC2BP$ ', os quais correspondem aos valores de ' $RRC1$ ' e ' $RRC2$ ', descritos anteriormente, porém a terminação ' BP ' significa que é um vetor que armazena os valores dos batimentos anteriores, isto é, o vetor ' $RRC1BP(i)$ ' armazena o valor do vetor ' $RRC1(i+1)$ '.

Em relação aos nós *ESV*, *Batimento Prematuro* e *Batimento Ventricular*, é preciso evidenciar que sua representação quantitativa é descrita através de uma tabela, a qual possui alternativas binárias V (Verdadeiro) e F (Falso). Quanto aos outros nós ($LLC1$, $LLC2$, $RRC1$, $RRC2$, $LLC1BA$, $LLC2BA$, $RRC1BP$, $RRC2BP$), eles são representados quantitativamente por distribuições Gaussianas, as quais são estimadas através de algoritmos que geram histogramas obtidos de exemplos da base de treinamento, sendo que estes algoritmos são explicados de forma detalhada no item C.

A partir dos histogramas, foram calculados os valores das médias e variâncias de cada Gaussiana², os quais foram normalizados, para uma melhor distinção entre os batimentos descritos como 'Outros'³, Prematuros e Ventriculares. Os valores da média e da variância utilizados nos nós $LLC1$ e $LLC2$ foram gerados, a partir do sinal rotulado, pelo sistema

¹ Utiliza-se a palavra 'batimento' para identificar cada batimento cardíaco observado no sinal ECG, que compreende a sequência de ondas P-QRS-T-U.

² Foi escolhida a distribuição Gaussiana, pois é a que mais se aproxima dos valores representados nos histogramas.

³ A palavra 'Outros', neste contexto significa que a classificação do batimento cardíaco não é *ESV*, isto é, quando a rede realiza a classificação de cada batimento, ela identifica a probabilidade do batimento em questão ser *ESV* ou 'Outros', este último representa todos os outros batimentos da rede, os Normais e o que possuem outras arritmias e outras deformações na estrutura do sinal.

proposto por Andreão, Dorizzi e Boudy (2006) (Sistema HMM-ECG), o qual calcula a verossimilhança da onda. Da mesma maneira, para o cálculo dos nós gaussianos *RRC1* e *RRC2*, foram utilizados os rótulos da base de dados adotada.

B. Identificar as Regras e Relações Causais

Para que um médico possa diagnosticar corretamente uma arritmia ventricular como o ESV, ele deve observar o sinal de ECG e analisar algumas alterações no mesmo. Isto é, o especialista faz o processamento do sinal. Para isto, ele precisa saber que sempre quando ocorre um episódio ESV ocorre também um batimento prematuro (representado pelo nó Batimento Prematuro), no qual não ocorre a onda P e ocorre um adiantamento do complexo QRS (ver Figura 1.7). A arritmia em questão causa também uma deformação na onda QRS, e esta informação é representada pelo nó Batimento Ventricular. Portanto, quem se propuser a desenvolver um classificador para simular um especialista deve se preocupar com estas características.

A estrutura da Rede Bayesiana especifica alguns tipos de variáveis (ocultas versus observáveis, discretas versus contínuas), o nome da variável e um tipo de distribuição de probabilidade da variável (Gaussiana, discreta, analítica) (Buntine, 1991; Lauritzen e Spiegelhalter, 1988; Shachter, 1986). Neste trabalho, para que fosse feito o treinamento da rede, todos os nós foram considerados observáveis. Ressalte-se, porém, que apenas na etapa de treinamento todos os nós foram considerados observáveis, o que não ocorre na etapa de teste. Para poder assumir que todos os nós são observáveis, no treinamento, foram realizadas várias simulações, e o melhor resultado é apresentado desta forma.

As relações causais, quantificadas por valores de probabilidade condicional, estão ilustradas no grafo direcionado representado na Figura 3.2, que corresponde à Rede Bayesiana implementada. Destaque-se, porém, que tal rede é apenas um modelo geral, o qual foi alterado no decorrer de desenvolvimento de cada simulação.

Este trabalho também testa várias topologias de redes, visando encontrar a melhor estrutura para a classificação de ESVs. Para isto, foram implementadas redes com um único canal, que é representada pela estrutura da Figura 3.2, utilizando apenas *LLC1* e *RRC1* ou *LLC2* e *RRC2*, e com 2 canais, isto é, realizando uma fusão de canais, empregando neste caso os nós *LLC1* e *LLC2* e *RRC1* e *RRC2*. Também foram implementadas RB estáticas combinando os valores de ‘RR’ e ‘LL’ de dois nós consecutivos. Na próxima seção são mostradas em mais detalhes as estruturas das redes implementadas.

C. Processar o Sinal e Estimar Parâmetros da Rede Bayesiana

Neste trabalho foram utilizadas duas bases de dados. A primeira é a MIT-BIH *Arrhythmia Database* (1997), base de dados que possui 48 registros de gravação do sinal ECG, com um pouco mais de 30 minutos cada um. Toda a base de dados possui aproximadamente 109.000 batimentos cardíacos rotulados, dos quais foram utilizados nesta pesquisa 85.169 batimentos cardíacos rotulados, por causa de ruídos presentes em alguns registros que dificultaram e atrapalharam o processamento do sinal. A segunda base de dados utilizada é a *QT Database* (Laguna *et al.*, 1997), a qual corresponde a um conjunto representativo de batimentos rotulados, os quais são encontrados no banco de dados Physionet (Moody *et al.*, 2001 e *QT Database*, 1997). A base de dados *QT Database* possui registros que foram escolhidos principalmente a partir de ECG existente entre diversas bases de dados, incluindo o MIT-BIH *Arrhythmia Database* (1997) e da *European ST-T Database* (Taddei *et al.*, 1991 e *QT Database*, 1997). Esta base de dados possui 105 registros, cada um contendo entre 30 e 100 batimentos cardíacos, todos rotulados manualmente por cardiologistas. Desta base de dados também não foram utilizados todos os registros, por causa de ruídos. É proposta uma Rede Bayesiana para fazer a concatenação dos dados relevantes. Ela é mostrada na Figura 3.2, e é composta tanto de variáveis discretas como contínuas. As variáveis discretas são os nós *ESV*, *Batimento Prematuro* (*BA*, *BP*) e *Batimento Ventricular* (*BA*, *BP*), enquanto os nós denominados *RRC1*, *RRC2*, *LLC1*, *LLC2*, *RRC1BA(BP)*, *RRC2BA(BP)*, *LLC1BA(BP)* e *LLC2PA(BP)* são variáveis contínuas. As variáveis discretas são representadas por valores binários (Verdadeiro ou Falso), enquanto as variáveis contínuas são representadas por distribuições Gaussianas.

A estimação dos parâmetros da Rede Bayesiana é obtida juntamente com a fase de aprendizado, utilizando um conjunto de exemplos rotulados por um médico. Tanto para o

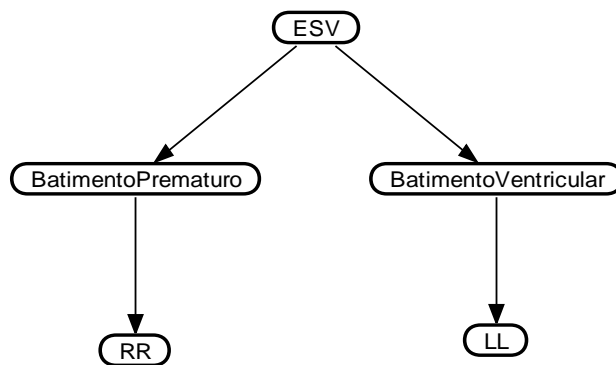


Figura 3.2: Modelo geral da Rede Bayesiana.

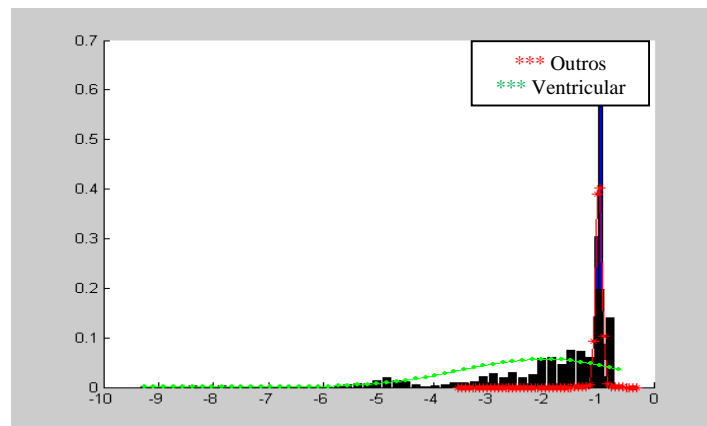
treinamento quanto para o teste da rede foram utilizados sinais obtidos de duas bases de dados. A primeira é a MIT-BIH *Arrhythmia Database* (1997), da qual foram utilizados 95.257 batimentos cardíacos para teste e treinamento da rede. Deste total de batimentos, porém, apenas 6,6% correspondem a episódios de extrassístole ventricular. Foram utilizados para o treinamento 64.074 batimentos cardíacos, enquanto os outros 31.183 foram utilizados para teste. A segunda base de dados utilizada é a *QT Database* (Laguna *et al.*, 1997), a qual corresponde a um conjunto representativo de batimentos rotulados, os quais são encontrados no banco de dados Physionet (Moody *et al.*, 2001). Desta base de dados, foram utilizados 75% do total de batimentos (58.744 batimentos) para a etapa de treinamento, enquanto os outros 25% (23.765 batimentos) foram utilizados para a etapa de teste. Ressalte-se, neste caso, que tanto no teste quanto no treinamento da rede, apenas 5,5% de batimentos correspondem a extrassístoles ventriculares (que é o percentual de batimentos com episódios de ESV em tal base de dados).

Os comprimentos dos intervalos R-R e a verossimilhança foram estimados, utilizando um algoritmo que calcula o valor numérico do tempo entre os picos R-R e o valor numérico da verossimilhança do complexo QRS. Estes dois valores foram normalizados, para melhorar a separação entre os tipos de batimentos de interesse (Batimento Normal, Ventricular e Prematuro). Esta normalização é necessária porque existe uma grande diversidade de morfologias no sinal ECG relacionado a cada pessoa. Por exemplo, uma pessoa pode ter sua frequência cardíaca normal em 70 batimentos por minuto, mas pode ter sua frequência cardíaca elevada para 85 batimentos por minuto quando estiver realizando algum tipo de esforço físico. Neste caso, pode-se observar que o tempo entre os picos R na primeira situação é maior que na segunda. Este exemplo mostra que pode ocorrer uma grande variação dos valores apresentados no sinal ECG considerado como normal para cada pessoa. Porém, esta variedade pode ser minimizada utilizando normalização, e para isto são levados em consideração os dois últimos batimentos classificados como ‘Outros’ (janela = 2). O mesmo procedimento é utilizado para o cálculo da verossimilhança do complexo QRS.

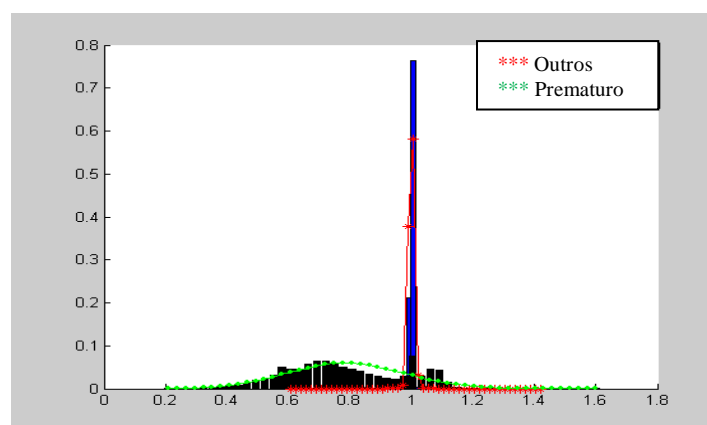
Esta janela foi implementada tanto na fase de teste quanto na fase de treinamento da rede. Um exemplo da normalização na fase de treinamento está representado na Figura 3.3. Porém, é importante ressaltar que a normalização realizada nesta fase é *offline*, pois inicialmente é realizada toda a normalização dos valores relacionados aos nós ‘RR’ e ‘LL’ e somente após encontrar estes valores ocorre a inferência dos mesmos como nós observáveis da RB. Já a normalização realizada na etapa de teste é considerada *online*, porque cada vez que o algoritmo recebe um valor de um dos nós (RR e LL) ele realiza suas respectivas

normalizações antes de realizar a inferência na rede. Isto implica uma idéia dinâmica no cálculo da normalização e, por conseguinte, no valor assumido por estes *nós* serão alterados de acordo com os valores dos respectivos *nós* anteriores juntamente com o resultado da classificação dos mesmos, já que o valor de cada *nó* é calculado mediante o seu passado próximo (janela). O trecho do algoritmo descrito em linguagem natural mostrado na Figura 3.4 ilustra esta idéia.

A Figura 3.3 ilustra um exemplo do resultado das Gaussianas formadas com os registros da base de dados MIT-BIH *Arrhythmia Database* (1997). Por meio desta figura, fica fácil identificar que os valores foram normalizados, já que os valores “*Outros*” estão incidindo no valor 1 ou -1, para batimentos prematuro e ventricular, respectivamente.



(a)



(b)

Figura 3.3: (a) Gaussianas normalizadas dos batimentos classificados como ‘Outros’ e Ventricular, representando o valor da verossimilhança da onda QRS. (b) Gaussianas normalizadas dos batimentos classificados como ‘Outros’ e Prematuro, representando a distância entre dois picos R sucessivos.

```

% Calcular a média para a realizar a normalização:
% Somente compilado quando detectado um novo batimento normal
mprem = marginal_nodes(engine,BatimPrem);

if mprem.T(1,:) > mprem.T(2,:)
    % O batimento não é prematuro
    RRmediamovel = (1-2/(janela+1))*RRmediamovel + 2/(janela+1)*viRRGlobTst(ee);
end

mventr = marginal_nodes(engine,BatimVentr);

if mventr.T(1,:) > mventr.T(2,:)
    % O batimento não é ventricular
    LLmediamovel = (1-2/(janela+1))*LLmediamovel + 2/(janela+1)*mfVraissGlobTst(ee);
end
%

```

Figura 3.4: Algoritmo em pseudocódigo que representa o cálculo da normalização na fase de Teste.

D. Validar os Resultados

Os resultados foram avaliados utilizando Tabela de Contingência (ver estrutura mostrada na Tabela 3.1), a qual favorece a visualização dos erros e oferece uma medida efetiva do modelo de classificação ao mostrar o número de classificações corretas e incorretas (falsos positivos e falsos negativos).

Tabela 3.1: Representação de uma Tabela de Contingência para classificação de batimentos ‘Outros’ e ‘ESV’.

TABELA DE CONTINGÊNCIA			
DIAGNOSTICO DA REDE BAYESIANA	DIAGNOSTICO DA BASE DE DADOS		
	ESV	OUTROS	TOTAL
ESV	A (VERDADEIRO POSITIVO)	B (FALSO NEGATIVO)	A+B (total de batimentos classificados como ESV)
OUTROS	C (FALSO POSITIVO)	D (VERDADEIRO NEGATIVO)	C+D (total de batimentos classificados como Outros)
	A+C (total de batimentos rotulados como ESV)	A+D (total de batimentos rotulados como Outros)	A+B+C+D

Com base nesta tabela pode-se avaliar o desempenho do sistema calculando as seguintes medidas de desempenho (Nassar, 2007):

- Sensibilidade, definida como

$$Sensibilidade (Se) = \frac{a}{a + c} \quad 3.1$$

que é a capacidade de o sistema reconhecer os verdadeiros positivos em relação ao total de batimentos ESV. No caso presente, este índice indica a probabilidade da classificação ser “ESV”, dado que o rótulo do batimento também é “ESV”;

- Especificidade, definida como

$$\text{Especificidade (Es)} = \frac{d}{b + d} \quad 3.2$$

que é a probabilidade de o sistema classificar os verdadeiros negativos. No caso presente, este índice indica a probabilidade da classificação ser “Outros”, dado que o rótulo do batimento também é “Outros”;

- Valor Preditivo Positivo, definido como

$$\text{Valor Preditivo Positivo (VPP)} = \frac{a}{a + b} \quad 3.3$$

No caso presente, este índice indica o grau de certeza da hipótese “Batimento ESV” quando o diagnóstico oferecido pelo sistema é negativo, neste caso, batimento “ESV”;

- Valor Preditivo Negativo, definido como

$$\text{Valor Preditivo Negativo (VPN)} = \frac{d}{c + d} \quad 3.4$$

No caso presente, este índice indica o grau de certeza da hipótese “Batimento Outros” quando o diagnóstico oferecido pelo sistema é positivo, neste caso, batimento “Outros”.

Quando um sistema possui uma alta *sensibilidade*, indica que a hipótese diagnóstica em questão é boa. Neste trabalho, tal hipótese é a detecção de ESV's. Do mesmo modo, quando o resultado aponta uma alta *especificidade*, significa que o sistema detectou muito bem episódios identificados como ‘Outros’.

O trabalho desenvolvido nesta Tese pode ser representado pelo diagrama apresentado na Figura 3.5. Ali, a camada zero corresponde ao sistema proposto por Andreão, Dorizzi e Boudy (2006), que segmenta os batimentos e fornece os valores de entrada para a camada um, a qual é efetivamente a tarefa aqui tratada, correspondendo à identificação de extrassístoles ventriculares. Na camada zero também será utilizado um outro segmentador proposto em (Aguiar, 2008), para efeito de análise da dependência do desempenho da RB com o segmentador utilizado.

Como o objetivo proposto é fazer uma comparação entre diferentes bases de dados e, conseqüentemente, um número distinto de batimentos classificados, é necessário que se

aplique um diferencial levando em consideração a quantidade de batimentos utilizada em cada problema abordado. Neste contexto, o intervalo de confiança é a metodologia adotada para validar as comparações entre bases de dados. Ele está associado a um grau de confiança, que é uma medida da certeza de que o intervalo contém o parâmetro populacional. Para cada resultado de desempenho obtido, será acrescentando um intervalo de confiança de 95%⁴ em torno do valor de Se e VPP. Para construir um intervalo de confiança de 95%, o valor do parâmetro da distribuição normal (Gaussiana) é $z_{\alpha/2} = 1,96$ (ver Anexo 3). O erro máximo da estimativa, \mathcal{E} , com confiança de $100 \cdot 1 - \alpha$ %, é definido pela equação (Larson e Fabber, 2004)

$$\mathcal{E} = \pm z_{\alpha/2} \sqrt{\frac{p \cdot q}{N}} \quad 3.5$$

onde p é a proporção de sucessos em uma amostra e q é a proporção de fracassos na mesma amostra. Considerando a medida de sensibilidade, $p = Se$, $q = 1 - Se$ ($p + q = 1$), sendo N a quantidade de batimentos utilizados no teste. Assim, o intervalo de confiança da medida Se é

$$\mathcal{E} = \pm 1,96 \sqrt{\frac{Se(100 - Se)}{N}} \quad 3.6$$

O mesmo se aplica para VPP.

3.2 Primeiro Momento: Comparação entre RB e Outras Metodologias de Classificação

Diferentes estruturas de Redes Bayesianas foram implementadas e avaliadas, a partir da Rede Bayesiana básica descrita na Figura 3.2, com a inclusão de novos *nós* na rede, dentre aqueles descritos na seção anterior. Assim, a partir da rede básica foram criadas outras configurações, para fins de comparação e identificação da melhor estrutura para o problema a ser tratado.

A base de dados MIT-BIH *Arrhythmia Database* (1997) empregada para a classificação de ESV foi dividida em duas porções, para treinamento e teste. A primeira

⁴ Um fator de 1,96 permite a obtenção de um intervalo de confiança de 95% para a distribuição quase normal típica da maioria dos resultados das medições.

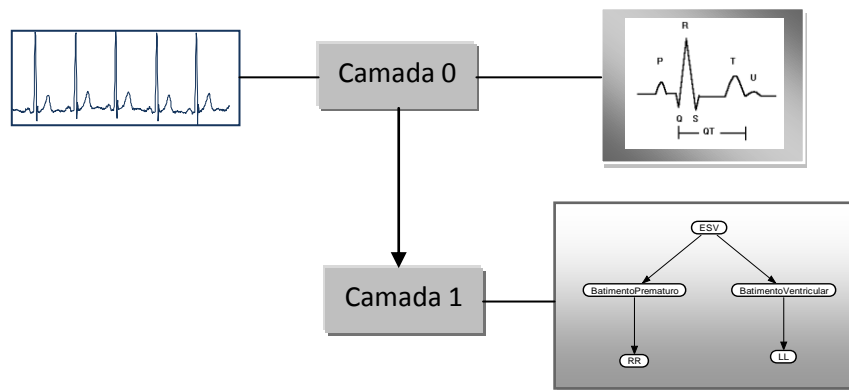


Figura 3.5: Diagrama que representa o trabalho realizado nesta Tese. A Camada 0 segmenta o sinal identificando as ondas, enquanto a Camada 1 classifica o batimento de acordo com a abordagem Bayesiana.

corresponde ao ajustamento da parte quantitativa da rede, isto é, das tabelas de probabilidade. Já o teste foi utilizado para verificar a credibilidade da classificação realizada. Em todos os casos, a rede foi implementada utilizando uma ferramenta disponibilizada pelo MATLAB[®], denominada BNT (2007).

3.2.1 Classificação Heurística de ESV (utilizando limiares)

Nesta classificação não é utilizado, em nenhum momento, a metodologia de Redes Bayesianas (Figura 3.6). É denominada como Classificação Heurística a classificação da arritmia cardíaca em questão pela utilização de limiares. Objetivando-se a comparação da metodologia de Redes Bayesianas com a metodologia heurística, como discutido no capítulo 2. Nesta fase pretende-se ressaltar a viabilidade na utilização de RB, comprando com outros métodos de classificação empregados para este fim.

A partir do algoritmo utilizado para determinar o valor normalizado do intervalo R-R e o valor numérico da verossimilhança do complexo QRS, foram geradas as funções Gaussianas apresentadas na Figura 3.6. Estes valores são gerados da mesma forma como descrito no item C do tópico anterior (Figura 3.4), isto é, utilizando a partir dos valores relacionados ao tempo do intervalo R-R e a verossimilhança do complexo QRS utilizando o algoritmo de segmentação apresentado por Andreão (2004).

Para realizar a classificação proposta, foram colocados dois limiares em cada uma destas funções (LimiarEq, representado o limiar mais a esquerda da Figura 3.6 e o LimiarDir, representando o limiar mais a direita da Figura 3.6), com o objetivo de separar os batimentos classificados como ‘Outros’ dos prematuros e ventriculares.

Em suma, os valores da média e da variância de cada variável (prematura e ventricular) que estiverem compreendidas entre os dois limiares são consideradas ‘Outros’

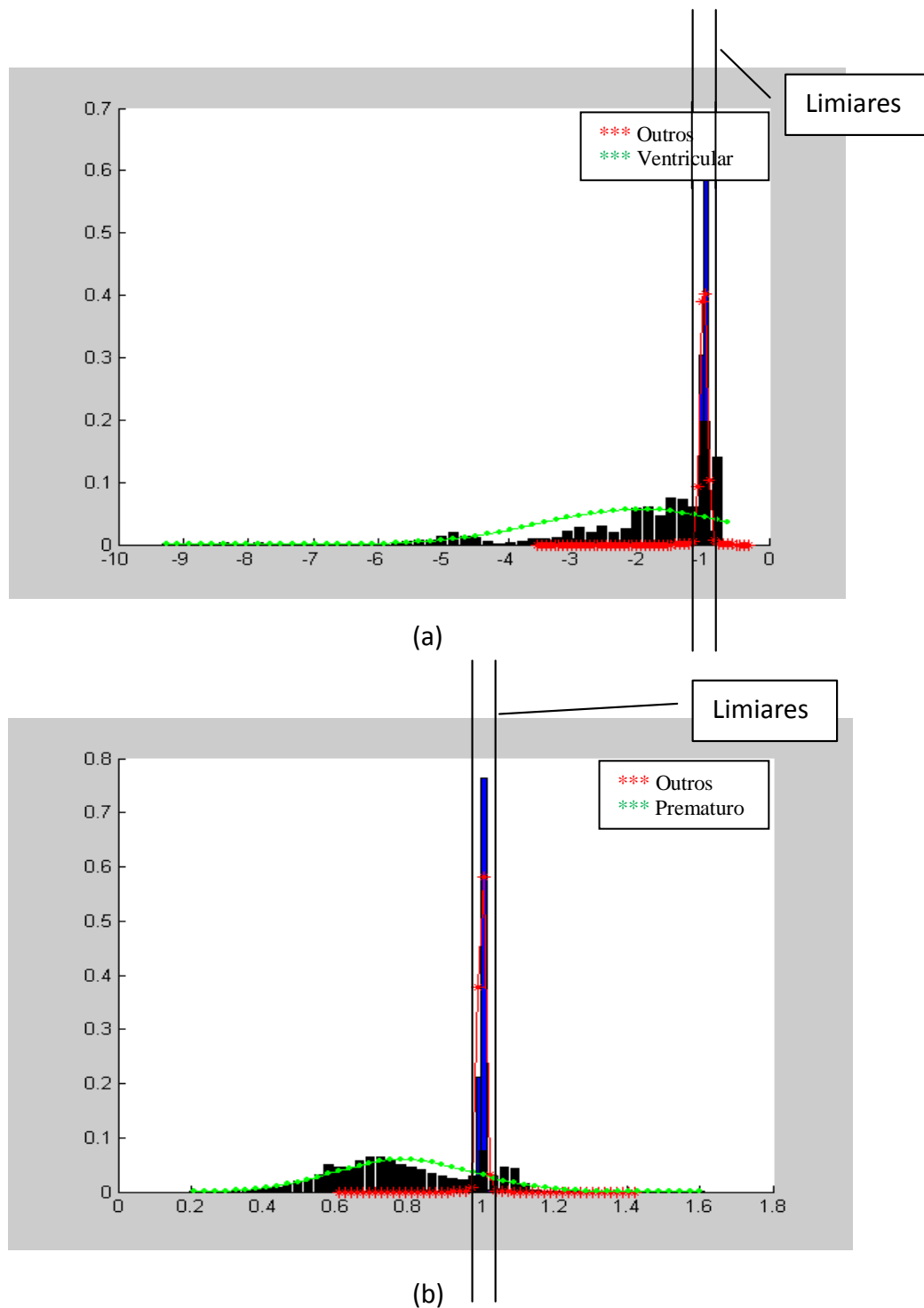


Figura 3.6: (a) Gaussianas do batimento ‘Outros’ e Ventricular (valor da verossimilhança da onda QRS); e (b) Gaussianas do batimento ‘Outros’ e Prematuro, todas com seus respectivos limiares.

(LimiarEsq \geq ‘Outros’ \leq LimiarDir), por conseguinte, os valores que não pertencem a este conjunto são considerados ESV’s (ESV $<$ LimiarEsq e ESV $>$ LimiarDir). O resultado obtido com esta classificação está na Tabela 3.2.

3.2.2 Rede Bayesiana com estimação empírica das tabelas de probabilidade

Realizou-se, inicialmente, uma estimação empírica das tabelas de probabilidade de cada *nó* da rede descrita na Figura 3.2. Ressalte-se que os valores dos *nós* contínuos

Tabela 3.2: Resultado da Classificação Heurística.

Sensibilidade	73,98%
Especificidade	98,32%
Valor Preditivo Positivo	67,65%
Valor Preditivo Negativo	97,73%
Intervalo de Confiança - Se (IC - Se)	$\pm 2,0$
Intervalo de Confiança - VPP (IC - VPP)	$\pm 2,2$

(representados por funções Gaussianas) foram estimados por um algoritmo que calcula o valor dos *nós* 'RR' e 'LL'. Já para os *nós* discretos, a estimação foi feita empiricamente (tentativa e erro), gerando o resultado apresentado na Tabela 3.3. Estes valores foram estimados utilizando o mesmo conjunto de batimentos separados para o treinamento da Rede Bayesiana. Após a estimação empírica dos valores probabilísticos de cada *nó*, inicia-se a fase de teste, utilizando o conjunto de batimentos que foram separados para esta finalidade.

Assim, não foi utilizado o algoritmo de aprendizado da RB, somente uma estimação empírica baseada em tentativa e erro, como descrito na Tabela 3.3. E após esta estimação, realizou-se o teste, isto é, a verificação da confiabilidade desta metodologia. O resultado obtido do teste é apresentado na Tabela 3.4.

Como pode ser observado nos resultados apresentados na Tabela 3.4, o VPP do sistema não gerou um bom resultando, significando que este algoritmo não é um bom classificador (o VPP é muito baixo, evidenciando que a detecção de ESV's não foi boa).

O próximo passo foi criar uma rede que utilize a inferência Bayesiana tanto para a fase de treinamento quanto para a fase de teste, reajustando automaticamente seus próprios parâmetros.

3.2.3 Rede Bayesiana com aprendizado bayesiano

Tabela 3.3: Valores dos *nós* discretos, estimação empírica.

ESV	Probabilidade	ESV	Batimento Prematuro	%	ESV	Batimento Ventricular	%
V	40%	F	F	95%	F	F	95%
F	60%	V	F	5%	V	F	5%
		F	V	5%	F	V	5%
		V	V	95%	V	V	95%

Tabela 3.4: Resultados obtidos com a Rede Bayesiana com estimação empírica.

Sensibilidade (Se)	82,16%
Especificidade (Es)	89,78%
Valor Preditivo Positivo (VPP)	32,34%
Valor Preditivo Negativo (VPN)	98,83%
Intervalo de Confiança - Se (IC - Se)	$\pm 1,8$
Intervalo de Confiança - VPP (IC - VPP)	$\pm 2,2$

Considerando a mesma estrutura de rede da Figura 3.2, esta nova abordagem objetiva que a própria rede aprenda e altere seus parâmetros, isto ocorrendo na fase de treinamento. Desta forma, foi realizada a fase de aprendizagem bayesiana, com a finalidade de melhorar o desempenho da rede. O procedimento de aprendizagem considera todos os *nós* como *nós* observáveis, isto é, para cada *nó* da rede foi atribuído um valor exato.

Desta forma, os parâmetros da rede foram ajustados, com base no mesmo conjunto de treinamento, e os resultados da classificação (utilizando o mesmo conjunto de dados separados para a etapa de teste) são apresentados na Tabela 3.5.

Pode-se observar, através da comparação dos resultados da Rede Bayesiana apresentados nas Tabelas 3.4 e 3.5, que o VPP da Tabela 3.5 é bem maior que aquele da Tabela 3.4, evidenciando que ao utilizar a etapa de aprendizado bayesiano o desempenho do sistema melhorou significativamente. Isto demonstra que a estratégia de aprendizagem bayesiana, para estimar o valor de cada variável, resulta em uma melhor classificação.

Estas comparações confirmam que a metodologia de Redes Bayesianas é uma ferramenta poderosa de aprendizado e classificação. Visto os resultados obtidos, dispõe-se de um classificador com um bom desempenho, pois a avaliação de desempenho deste sistema, apresentado via comparação entre a resposta oferecida pelo classificador heurístico e empírico, apresentou uma percentagem de acerto muito representativa.

A utilização de Redes Bayesianas oferece a possibilidade de uma abordagem que trata a incerteza pela sua representação probabilística, trabalhando com dados incompletos e aleatórios em sua máquina de inferência.

Nesta seção, a capacidade de aprendizagem e classificação da rede foi verificada, sendo o melhor resultado aquele correspondente à utilização de *nós* observáveis (Tabela 3.4).

O próximo passo deste trabalho foi utilizar uma estratégia de fusão de canais. Para esta finalidade, dois canais foram utilizados, bem como informações de batimentos cardíacos anteriores, isto é, batimentos já classificados, visando, desta maneira, melhorar o desempenho do classificador. Para isto foi necessária a utilização de outra base de dados, a *QT Database*,

Tabela 3.5: Resultados obtidos para a Rede Bayesiana com a fase de aprendizado bayesiano.

Sensibilidade (Se)	79,53%
Especificidade (Es)	98,79%
Valor Preditivo Positivo (VPP)	93,92%
Valor Preditivo Negativo (VPN)	99,69%
Intervalo de Confiança - Se (IC - Se)	$\pm 1,9$
Intervalo de Confiança - VPP (IC - VPP)	$\pm 1,1$

pois a base de dados MIT-BIH *Arrhythmia Database* apresentou problema no processamento do sinal correspondente ao canal 2 na Camada 0 do sistema.

Pretendia-se fazer todos os testes com a base MIT-BIH *Arrhythmia Database*. Porém, devido ao problema da baixa qualidade do canal 2 da base, partiu-se para estudar o efeito da fusão com a base de dados QT *Database*.

3.3 Segundo Momento: Fusão de Canais

Para que fosse possível a Fusão de Canais, foi necessário que se utilizasse uma nova base de dados para a aplicação da RB, pois a base de dados MIT-BIH *Arrhythmia Database* não apresentou resultados satisfatórios para a aplicabilidade da utilização do canal 2, um dos motivos principais foi o ruído apresentado neste segundo canal.

Então, neste tópico a base de dados QT *Database*⁵ (1997), foi implementada para utilização da Fusão de Canais, já que nesta base de dados o canal 2 foi gerado sem maiores problemas para utilização na RB.

Da QT *Database*, foram utilizados 82.509 batimentos cardíacos rotulados, obtidos a partir do sinal ECG. Desta base de dados, 75% de batimentos foram selecionados para a etapa de treinamento (58.744 batimentos), enquanto os outros 25% foram utilizados para a etapa de teste (23.765 batimentos), ressaltando-se que, tanto no teste quanto no treinamento, apenas 5,5% do total de batimentos utilizados correspondiam a episódios de extrassístole ventricular.

⁵ Será utilizado somente o termo QT para designar a base de dados QT *database* (1997).

3.3.1 Rede Bayesiana com Fusão de Canais e Nós que Armazenam Informações Anteriores

É importante ressaltar que diferentes estruturas de Redes Bayesianas foram implementadas e avaliadas, utilizando dois canais⁶ e os valores de classificação do batimento cardíaco anterior. O modelo da rede que gerou o melhor resultado é aquele apresentado na Figura 3.7. Deve-se ressaltar que na etapa de treinamento todos os nós foram considerados observáveis, mas na etapa de teste foram utilizados nós observáveis (evidências, ou seja, existem de forma explícita nos dados) e não observáveis (variáveis que se deseja inferir). O resultado de classificação desta rede é apresentado na Tabela 3.6.

Tabela 3.6: Resultados para a classificação modelada pela Figura 3.7.

Sensibilidade (Se)	78,15%
Especificidade (Es)	99,98%
Valor Preditivo Positivo (VPP)	98,33%
Valor Preditivo Negativo (VPN)	99,72%
Intervalo de Confiança - Se (IC - Se)	$\pm 4,6$
Intervalo de Confiança - VPP (IC - VPP)	$\pm 1,4$

Desta feita, foram utilizados 2 Canais (nó com terminação C1 e C2) e os valores de classificação do batimento cardíaco anterior, isto é, os nós com a terminação ‘BA’ possuem um vetor que armazena os valores do batimento cardíaco precedente.

Foram geradas diferentes estruturas de redes, mas será apresentado apenas o modelo da rede que gerou o melhor resultado, aquele apresentado na Figura 3.7. Deve-se evidenciar que, mais uma vez, na etapa de treinamento todos os nós foram considerados observáveis.

O próximo passo foi excluir da rede os nós com informações de batimentos anteriores,

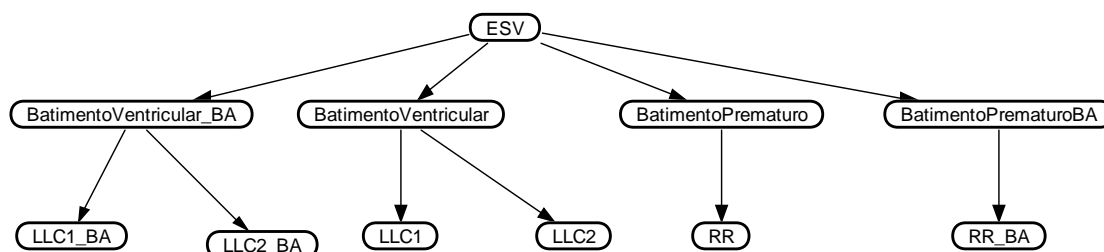


Figura 3.7: Rede Bayesiana com fusão de canais, isto é, utilizando o canal 1 e o canal 2, bem como valores correspondentes ao batimento cardíaco anterior.

com o objetivo de verificar se a fusão de canais melhora realmente o desempenho da rede, já

⁶ 2 Canais, significando duas derivações, as quais são representadas pelo Canal 1 e o Canal 2, respectivamente. Estas derivação estão acessíveis na base de dados QT.

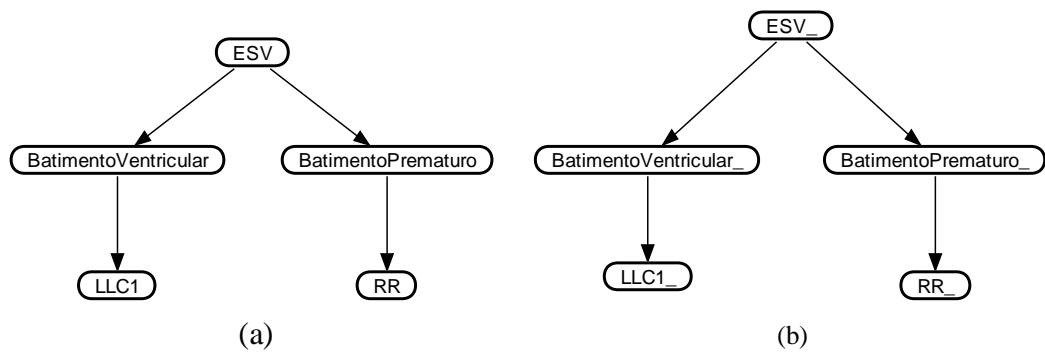


Figura 3.8: Rede Bayesiana sem fusão de canais, (a) utilizando o canal 1 e (b) utilizando o canal 2.

que não se pode comparar com os resultados obtidos a partir da base de dados MIT-BIH *Arrhythmia Database*, por causa da diferença na quantidade de batimentos rotulados como ESV's.

3.3.2 Rede Bayesiana com Apenas um Canal

Neste momento, foram implementadas duas estruturas de Redes Bayesianas, a primeira utilizando apenas o Canal 1(Figura 3.8 (a)), e a segunda utilizando apenas o Canal 2 (Figura 3.8 (b)). Como implementado nas simulações anteriores bem como nas simulações que serão apresentadas posteriormente, durante a etapa de treinamento todos os *nós* foram considerados observáveis.

O objetivo desta seção é a comparação entre o aprendizado das redes em cada canal (C1 e C2), para que posteriormente se possa realizar a comparação entre a fusão de canais.

O resultado destas classificações é apresentado nas Tabelas 3.7 e 3.8, respectivamente.

Pode-se concluir que a rede que é representada pela Figura 3.8 (a) é melhor que a rede que é representada pela Figura 3.8 (b), principalmente no valor da 'Se'. Com relação ao valor da 'Se' houve um aumento significativo maior que 20% bem como um aumento no intervalo de confiança de 1% . Isto ocorre porque o sinal do canal 2 é de pior qualidade que o sinal do canal 1. De fato, isto também ocorre com relação a base de dados MIT-BIH *Arrhythmia Database*, que por esta razão não foi utilizado a fusão de canais.

Comparando-se o resultado da rede representada pela Figura 3.8 (a) e o resultado da rede representada pela Figura 3.7 (com a fusão de canais e com os *nós* relacionados ao batimento anterior), pode-se verificar que a fusão de canais melhorou o resultado final. Observe que o valor da 'Se' apresentado no resultado referente à Figura 3.7 obteve um aumento de aproximadamente 3%, isto pode ser afirmado baseando-se no 'IC-Se'. Observa-se

Tabela 3.7: Resultado da classificação modelada pela Figura 3.8 (a).

Sensibilidade (Se)	75,50%
Especificidade (Es)	99,85%
Valor Preditivo Positivo (VPP)	86,69%
Valor Preditivo Negativo (VPN)	99,69%
Intervalo de Confiança - Se (IC - Se)	$\pm 4,8$
Intervalo de Confiança – VPP (IC - VPP)	$\pm 3,8$

também que o ‘VPP’ aumentou aproximadamente 10%. A observação dos intervalos de confiança serve para permitir esta comparação.

Para ressaltar que a fusão de canais melhora o desempenho da rede, foi implementada uma rede com fusão de canais, porém sem os batimentos anteriores. Com a criação desta nova topologia é possível também verificar se os batimentos anteriores melhoram o desempenho de uma Rede Bayesiana estática, isto é, se realmente a inclusão destes *nós* na rede aumentam a confiabilidade dos resultados.

3.3.3 Rede Bayesiana com Fusão de Canais

Variadas topologias de Redes Bayesianas com fusão de canais foram implementadas e avaliadas, porém a que gerou o melhor resultado é a apresentada na Figura 3.9. A razão para se excluir o *nó* representando o valor *RR* do canal 2 é porque o *nó RR* representa o valor numérico do tempo entre dois picos R, que deve ser o mesmo para ambos canais (ressaltando que cada canal é segmentado separadamente, e o resultado pode ser influenciado pela qualidade do sinal dos mesmos). Então, como o intervalo de tempo entre dois picos R’s deve ser o mesmo, e como já foi demonstrado anteriormente que o canal 2 tem qualidade inferior ao canal 1, optou-se por utilizar o valor do *nó* ‘RR’ referente ao canal 1.

Os resultados de classificação para as redes da Figura 3.9 são apresentados na Tabela 3.9.

Comparando inicialmente a RB com fusão de canais (Figura 3.8 (a)) e a mesma topologia sem a fusão de canais (Figura 3.9) observa-se um aumento no desempenho da rede com a fusão de resultados, isto pode ser visualizado pelo aumento do valor da ‘Se’ que é de

Tabela 3.8: Resultado da classificação modelada pela Figura 3.8 (b).

Sensibilidade (Se)	53,31%
Especificidade (Es)	99,98%
Valor Preditivo Positivo (VPP)	96,99%
Valor Preditivo Negativo (VPN)	99,40%
Intervalo de Confiança - Se (IC - Se)	$\pm 5,6$
Intervalo de Confiança – VPP (IC - VPP)	$\pm 1,9$

Tabela 3.9: Resultado da classificação modelada pela Figura 3.9, com fusão de canais.

Sensibilidade (Se)	78,15%
Especificidade (Es)	99,96%
Valor Preditivo Positivo (VPP)	96,33%
Valor Preditivo Negativo (VPN)	99,72%
Intervalo de Confiança - Se (IC - Se)	$\pm 4,6$
Intervalo de Confiança - VPP (IC - VPP)	$\pm 2,1$

aproximadamente 3% e também pelo valor apresentado por ‘VPP’ que adquire um acréscimo de 10%. Estas afirmações podem ser ratificadas observando os respectivos intervalos de confiança, os quais permitem a legitimidade desta afirmação.

É importante comentar que quando se compara a Rede Bayesiana com fusão de canais e batimentos cardíacos anterior (Figura 3.7) e a Rede Bayesiana apenas com fusão de canais, representada pela Figura 3.9, verifica-se que a diferença entre os valores da ‘Se’ de ambos os casos é muito baixa, e o ‘VPP’ referente a rede da Figura 3.7 obteve um aumento de aproximadamente 2%. Conclui-se, a partir desta observação, que implementar uma Rede Bayesiana estática com um nó representado uma informação referente ao passado (batimento anterior), é um pouco melhor para ‘VPP’. Desta forma, é possível visualizar a implementação de RBD objetivando-se um aumento mais significativo.

Por fim, pode-se concluir que as melhores topologias são aquelas cujos resultados são obtidos a partir da fusão dos canais.

3.3.4 Conclusões Referentes à Rede Bayesiana e à Base de Dados QT Database

Até este momento foram apresentadas Redes Bayesianas estáticas para o problema particular de classificação de batimentos cardíacos, de acordo com o diagnóstico médico para a classificação de extrassístole ventricular (ESV). Foram implementadas diferentes estruturas de Redes Bayesianas, cujo desempenho foi avaliado usando a base de dados QT Database,

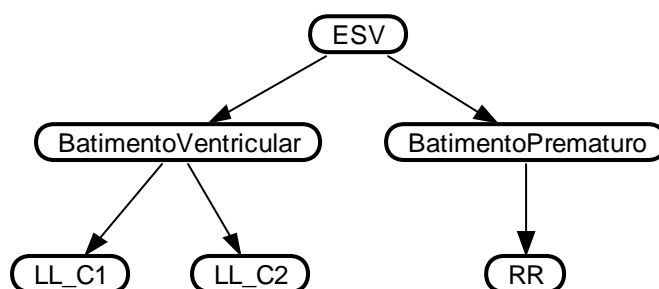


Figura 3.9: Rede Bayesiana com fusão de canais.

um banco de dados rotulado que contém um conjunto representativo de registros do sinal de ECG com os batimentos devidamente classificados por especialistas. A Rede Bayesiana que apresentou os melhores resultados, em termos de ‘Se’ e ‘VPP’, foi a que combinou a informação de dois canais do sinal ECG.

Embora os melhores resultados obtidos apontam um resultado muito significativo para a detecção de ‘Outros’ (referenciados por ‘Es’ e ‘VPN’), os valores de ‘Se’ e ‘VPP’ ainda precisam ser melhorados, já que se tem a pretensão de que esta rede seja utilizada para monitoração remota de pacientes. O que ocorre aqui é que o número de batimentos ESV disponíveis na base de dados usada é muito pequeno (cerca de 5,5 %), o que restringe bastante o treinamento da rede para classificar exatamente este tipo de batimento. Isto, porém, não invalida a proposta de utilizar as Redes Bayesianas com a utilização da base de dados QT *Database*.

3.3.5 Conclusões e Perspectivas

Os resultados obtidos confirmam que a combinação de diferentes canais do sinal ECG melhora o desempenho do sistema de classificação, assim como demonstram a viabilidade de usar Redes Bayesianas como uma ferramenta para classificação deste sinal. De fato, as Redes Bayesianas representam um modelo eficaz, pois elas permitem uma representação, no mesmo modelo, de conhecimento qualitativo e quantitativo.

Afirmar que o treinamento da rede ficou restrito pelo fato do número de batimentos ESV’s disponíveis na base de dados QT *Database* ser muito pequeno é uma verdade. Esta afirmação é claramente demonstrada quando se compara seus melhores resultados com os resultados obtidos utilizando a base de dados MIT-BIH *Arrhythmia Database*, desde que a mesma possui maior número de batimentos ESV. Isto é, a base de dados MIT-BIH *Arrhythmia Database* possui aproximadamente 1%⁷ a mais de batimentos rotulados como ESV que a base de dados QT, e a primeira obteve um resultado melhor tanto no valor de ‘Se’ quanto no ‘VPP’ que são pouco mais de 4% e 7% respectivamente (Tabela 3.5 e 3.7(a)). É importante ressaltar que esta comparação é baseada na mesma topologia da rede. Esta afirmação é também ratificada pela observação dos intervalos de confianças entre as respectivas bases de dados, a MIT-BIH *Arrhythmia Database* possui uma variação em torno

⁷ Este 1% não se refere a comparação entre o número de batimentos rotulados como ESV de ambas as bases de dados. Este valor diz respeito ao total de batimentos ESV de cada base e o total de batimentos da base de dados em questão. Isto é, a base de dados MIT-BIH *Arrhythmia Database* possui aproximadamente 6,6% do total de batimentos rotulados como ESV e a base de dados QT possui aproximadamente 5,5 % do total de batimentos classificados como ESV.

de 2% e a QT possui uma variação entre 4% e 5%, isto se deve a quantidade maior do total de batimento da primeira base de dados em relação a segunda, como também o aumento de 1% no valor do total de batimentos classificados como ESV.

Assim, para melhorar o desempenho da rede, deve-se mudar a base de dados, para que se possa ter um conjunto maior de batimentos ESV's, e, desta forma, melhorar a classificação do mesmo. Porém, não existem bases com número de batimentos ESV maiores do que os utilizados neste trabalho. Isto é, não existe base de dados com número representativo de batimentos rotulados com ESV. Como exemplo, pode-se citar a *AHA Database* (2008) que possui apenas 4.600 batimentos classificados como 'ESV' e 200.000 batimentos classificados como 'Outros'.

Então, pretende-se voltar a trabalhar com a base de dados MIT-BIH *Arrhythmia Database*, já que esta possui mais batimentos classificados como ESV do que a base de dados *QT Database*, e também uma base de dados com maior número de batimentos, implicando, desta forma em uma certeza maior com relação a seus resultados, já que os intervalos de confiança da base de dados em questão deixa esta afirmação clara. Pretende-se deixar como trabalho futuro o aprimoramento do canal 2, visando o um melhor tratamento do sinal, resultando, desta forma, em uma melhor segmentação do mesmo. E quando isto for possível, utilizar-se-á a fusão de canais.

3.4 Terceiro Momento: RB Utilizando Diferentes Algoritmos de Segmentação do Sinal

Para comprovar que a Rede Bayesiana é uma boa ferramenta para classificação de ESV's, independentemente⁸ do algoritmo de segmentação do sinal ECG utilizado, (objetivando-se, assim, a comprovação que a rede não está viciada em um algoritmo de segmentação específico) foram selecionados dois algoritmos de segmentação para gerar as variáveis de entrada da rede, um deles proposto por Andreão (2004) e o outro proposto por Aguiar (2008).

O algoritmo proposto por Andreão, (2004), a partir daqui denominado Primeiro Algoritmo de Segmentação do ECG, calcula a verossimilhança da onda QRS e identifica o intervalo entre dois picos R sucessivos, enquanto que o algoritmo de segmentação proposto

⁸ É importante deixar muito claro que esta independência está condicionada a um algoritmo de segmentação satisfatório, é lógico que quando não existe um algoritmo de segmentação apropriado a rede não faz uma boa classificação. Se assim o fosse, não seria necessários tais algoritmos.

por Aguiar (2008), a partir daqui denominado Segundo Algoritmo de Segmentação do ECG, calcula a distância QS e também identifica o intervalo entre dois picos R sucessivos. Assim:

- Primeiro Algoritmo de Segmentação - PAS (Andreão, 2004): segmenta o sinal de ECG em batimentos, e em ondas (P, QRS, T), bem como gera informações a respeito do intervalo R-R e da medida estatística do complexo QRS. Ele utiliza HMM (*Hidden Markov Model*).
- Segundo Algoritmo de Segmentação - SAS (Aguiar, 2008): faz uso da informação do intervalo R-R da base de dados (*MIT-BIH Arrhythmia Database*) para obter uma informação determinística da largura do complexo QRS. Os valores das distâncias Q-S são obtidos pela diferença de morfologia dos batimentos depois de alinhados. Além da distância Q-S, o algoritmo pode também levar em consideração o tamanho do vetor, a energia do batimento, a amplitude, dentre outras. Porém somente foram empregadas informações que se referem a distância R-R e Q-S.

Portanto, de acordo com a Figura 3.5, o processo de classificação de ESV inicia-se com o processamento do sinal ECG por um dos dois algoritmos de segmentação (Andreão, 2004 e Aguiar, 2008). Após o resultado da segmentação, a RB realiza a classificação do batimento, entre 'ESV' e 'Outros'.

Deve-se destacar que como as variáveis de evidência referentes ao SAS estão incompletas, isto é, não se possui os valores do tamanho do vetor, da energia, da amplitude, entre outros, variáveis estas que são importantes para a definição da probabilidade do nó Batimento Ventriculares, optou-se em utilizar a variável da distância Q-S referentes ao canal 2, tencionado, desta maneira, suprir as deficiências que a falta de algumas informações que define o problema.

Então, no PAS, utiliza-se duas evidências, o valor da distância R-R e o valor da verossimilhança do complexo QRS, ambos referentes ao canal 1. Já no SAS são observadas três evidências, a distância R-R, QS_C1 e QS_C2. Por conseguinte, a topologia da rede para o PAS é especificada a Figura 3.2 e seus respectivos resultados são apresentados na Tabela 3.5 ('Se' = 79,53 e 'VPP' = 93,92). A topologia da rede referente ao SAS é apresentada na Figura 3.10 e seus resultados, na Tabela 3.10.

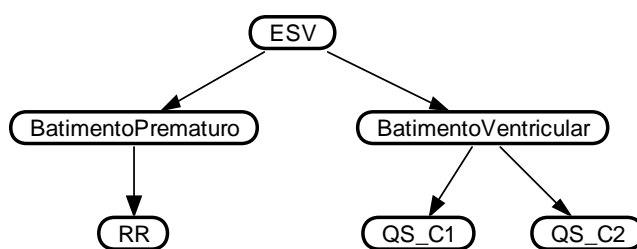


Figura 3.10: Topologia referente ao SAS.

É possível observar que a diferença não foi grande, houve um ganho de aproximadamente 3% no valor de ‘Se’ da rede baseada no SAS, porém em relação ao ‘VPP’, esta mesma rede aponta uma perda de 6%. Valores que são ratificados por seus respectivos intervalos de confiança.

Assim sendo, pode-se observar que há um equilíbrio na classificação da Rede Bayesiana, ou seja, seu desempenho é muito pouco influenciado pelo Algoritmo usado na segmentação do ECG para gerar os dados de entrada da rede, no que tange à detecção de episódios de ESV. Neste instante, é possível afirmar que a RB classifica muito bem, e gera resultados similares independentemente do algoritmo de segmentação, isto é, a RB apresentada neste trabalho não está viciada em um único algoritmo de segmentação e nem tampouco em uma mesma base de dados, conforme visto na seção anterior.

Para continuação deste trabalho, será utilizado, a partir deste momento, não somente a base de dados MIT-BIH *Arrhythmia Database* como também o algoritmo de segmentação apresentado por Andreão (2004), pois este além de apresentar um resultado satisfatório, a metodologia utilizada é baseada em técnicas probabilísticas (HMM), como a própria Rede Bayesiana. Ademais, o PAS identifica a observação da onda P, o que não ocorre no SAS.

Como descrito no Capítulo 1, para a detecção de episódios de ESV o cardiologista

Tabela 3.10: Resultados referentes a Base de Dados de MIT-BIH *Arrhythmia Database* e algoritmo de segmentação de Aguiar (2008) - SAS.

Sensibilidade (Se)	82,33%
Especificidade (Es)	99,59%
Preditivo Positivo (PP)	87,62%
Preditivo Negativo (PN)	99,37%
Intervalo de Confiança - Se (IC - Se)	$\pm 1,8$
Intervalo de Confiança - VPP (IC - VPP)	$\pm 1,5$

observa a ausência da onda P, além do tempo entre os picos R's e a forma do complexo QRS. Por esta razão, na próxima seção será discutida a utilização de mais um *nó* na rede, isto é, a observação da evidência relacionada a onda P.

3.5 Quarto Momento: RB Utilizando a Onda P e Fusão de Resultados

Discute-se neste seção a utilização da Onda P como um *nó* da rede, identificando se/como esta onda influencia de forma significativa na classificação de ESV's. Verifica-se, também, que a utilização da Fusão dos Resultados, os quais são obtidos através de uma RB com dois conjuntos de treinamento diferentes, melhora significativamente a classificação desta arritmia cardíaca.

Sendo que o diferencial nesta seção é a utilização da onda P. Esta variável é utilizada pelo especialista na detecção de ESV's, mas resta saber se esta variável faz diferença no resultado da classificação utilizando RB.

Também é abordado neste estudo o conceito de *Data Mining*, ou Mineração de Dados, que consiste em um processo analítico projetado para explorar grandes quantidades de dados. Neste contexto, o foco é a base de dados selecionada para treinamento.

Existem muitas técnicas utilizadas pelo *Data Mining* (Mitchell, 1997) para construção e segmentação de *Data Warehouse* (Inmon, 1997 e Imhoff, 2003). Porém, o objetivo nesta seção não é discutir estas técnicas, mas sim utilizar uma abordagem baseada em Mineração de Dados para separação da base de dados de treinamento. Mais informações a cerca destas tecnologias são apresentadas no Capítulo 4 deste mesmo trabalho.

Esta abordagem consiste em utilizar os dados de treinamento separado em subconjuntos diferentes. Este tipo de abordagem pode ser observado em alguns trabalhos, como em Dietterich,, (1997) e Dietterich, (2002).

3.5.1 Impacto da onda P na RB

O algoritmo proposto por Andreão (2004), denominado PAS, calcula a verossimilhança da onda QRS, identifica o intervalo entre dois picos R sucessivos (Figura 3.2), e identifica a ocorrência da onda P, dados estes que são utilizados como *nós* da RB proposta. A topologia da RB utilizando a onda P é apresentada na Figura 3.11. Verifica-se, de imediato, que nesta topologia é incluído mais um *nó*, denominado OndaP, que apresentará o

valor lógico ‘V’ ou ‘F’, para presença e ausência da onda P, respectivamente. O resultado obtido usando esta topologia é apresentado na Tabela 3.11.

Pode-se concluir que em ambas as redes (Figura 3.2 e 3.11) não houve grande alteração nos resultados (Tabela 3.5 e 3.11), lembrando que os valores referentes a ‘Se’ e ‘VPP’ apresentadas na Tabela 3.5 são 79,53% e 93,92% respectivamente. Na Tabela 3.11 é apresentado um pequeno aumento do valor de ‘Se’, isto é, um acréscimo de pouco mais de 3%. Porém observa-se também um pequeno decréscimo no ‘VPP’, isto é, uma diminuição de 6% em relação aos resultados observados na Tabela 3.5.

É importante comentar a explicação para isto, já que em um primeiro momento se deduz que a rede geraria resultados melhores, com a utilização desta nova evidência, porém, a justificativa para estes resultados está embasada no seguinte:

- A porcentagem do valor da onda P para os batimentos ESV não é suficiente para gerar informações úteis à rede (utilizando esta informação como *nó* da rede). Isto se deve ao fato da quantidade de batimentos classificados com a presença da onda P rotulados como ESV’s, na base de dados em questão, não ser muito representativo. Por exemplo, na segmentação feita por Laguna (Laguna et al, 1997) 66,28% dos batimentos rotulados como ESV’s não possuem onda P.
- Já na segmentação feita por Andreão (2004), que é utilizada neste trabalho, 64,84% de batimentos classificados como ESV’s não possuem onda P.

Outra conclusão que se admite a partir destas observações e que é ratificada no decorrer deste trabalho é que os valores das evidências descritas pelos *nós* ‘LL’ e ‘RR’ são suficientes para classificar este tipo de arritmia (ESV), pois quando se utiliza a Onda P (como um *nó* da rede) os resultados não se modificam substancialmente. Isto se pode afirmar baseado na comparação dos resultados descritos anteriormente.

É essencial a observação que em ambos os algoritmos (Laguna, 1997 e Andreão, 2004), para a identificação da onda P, apresentam resultados com confiabilidade maior que

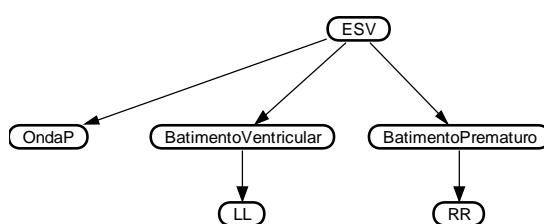


Figura 3.11: Rede Bayesiana com a onda P.

Tabela 3.11: Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, com a onda P.

Sensibilidade (Se)	82,96%
Especificidade (Es)	99,31%
Valor Preditivo Positivo (VPP)	87,78%
Valor Preditivo Negativo (VPN)	98,99%
Intervalo de Confiança - Se (IC - Se)	$\pm 1,7$
Intervalo de Confiança - VPP (IC - VPP)	$\pm 1,5$

90% de acerto. Com base nesta observação é evidente que a base de dados MIT-BIH *Arrhythmia Database* gera um número de exemplos (batimentos rotulados como ESV que não possuem onda P) insuficientes para que a rede possa realmente observar uma influência significativa deste valor, como um nó da rede. Principalmente por esta razão não houve uma melhora significativa nos resultados apresentados na Tabela 3.11.

Entretanto a utilização da onda P para melhorar o resultado da classificação da RB não está descartada. Esta será utilizada posteriormente na tecnologia de Fusão de Resultados, apresentada na próxima subseção, mas não como um nó da rede, mas sim como um critério de desempate.

3.5.2 Rede Bayesiana com Fusão de Resultados (FR)

Como continuação do trabalho, foi implementada uma fusão nos resultados da rede. Em termos mais técnicos, foram utilizadas técnicas de *Data Warehouse* para a construção de duas bases de dados para o treinamento. Isto é, o treinamento da rede foi baseado em duas bases de dados distintas, as quais são a partir deste momento baseadas em 80% dos batimentos da base de dados MIT-BIH *Arrhythmia Database* e não mais 75% como utilizado até o momento. Esta proporção foi alterada a partir de estudos realizados na base de dados em questão bem como em simulações, chegando-se a conclusão de que 80% do total de batimentos selecionados para o treinamento seria mais apropriado e não impediria a comparação com os resultados anteriormente descritos, já que os valores referentes ao intervalo de confiança também viabilizam esta idéia.

O primeiro subconjunto de dados selecionados para o treinamento utilizado para o cálculo da classificação é baseado, como dito anteriormente, em 80% dos batimentos da base de dados MIT-BIH *Arrhythmia Database* os quais foram *todos* utilizados no aprendizado. Já para a construção do segundo subconjunto de dados separados para o treinamento, realizou-se uma mineração de dados para selecionar 65% de batimentos classificados como ‘Outros’ e

Tabela 3.12: Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, para o conjunto de treinamento TT.

Sensibilidade (Se)	81,70%
Especificidade (Es)	99,52%
Valor Preditivo Positivo (VPP)	91,08%
Valor Preditivo Negativo (VPN)	98,92%
Intervalo de Confiança - Se (IC - Se)	$\pm 1,8$
Intervalo de Confiança – VPP (IC - VPP)	$\pm 1,3$

35% dos batimentos classificados como ‘ESV’. Explicando melhor esta idéia, foram selecionados dois conjuntos de treinamentos para a rede, denominados:

- Treinamento Total (TT), onde *todos* os 80% do total de batimentos selecionados para o treinamento foram utilizados, ressaltando que deste total apenas 6% dos batimentos são rotulados como ESV.
- Treinamento Parcial (TP), onde dos 80% de batimentos selecionados para o treinamento da rede foram retirados alguns batimentos rotulados como ‘Outros’, de forma a cumprir a regra que determina que somente 65% dos batimentos do treinamento poderiam ser rotulados como ‘Outros’. Ressalte-se que a seleção destes dados foi realizada de forma randômica.

Vale salientar que os batimentos selecionados para o teste da rede são os mesmos, isto é, tanto para o TT quanto para TP o conjunto de teste é o mesmo. Nele, a porcentagem de ocorrência de episódios ESV é de aproximadamente 6% do total de batimentos selecionados. Por fim, registre-se que a topologia da rede é a mesma para ambos os casos, e é representada na Figura 3.2. Os resultados da classificação para o TT e TP são apresentados nas Tabelas 3.12 e 3.13 respectivamente.

Comparando os resultados obtidos com os dois conjuntos de treinamento (Tabela 3.12 e Tabela 3.13), observa-se que a Sensibilidade do conjunto TP aumentou significativamente em relação ao conjunto TT, isto é, obteve-se um ganho de aproximadamente 5% no valor de ‘Se’. Já o ‘VPP’ do conjunto TP diminuiu, também significativamente, em comparação com

Tabela 3.13: Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, para TP.

Sensibilidade (Se)	86,45%
Especificidade (Es)	99,04%
Valor Preditivo Positivo (VPP)	84,19%
Valor Preditivo Negativo (VPN)	99,19%
Intervalo de Confiança - Se (IC - Se)	$\pm 1,6$
Intervalo de Confiança – VPP (IC - VPP)	$\pm 1,7$

TT. Isto é, obteve-se um decréscimo de 7% neste valor. Isto se deve ao fato de que quando o treinamento é realizado utilizando um conjunto que possui uma porcentagem de episódios ESV maior do que o normal é de se esperar que a classificação baseada neste treinamento identifique mais episódios ESV que o normal, pois a rede admite que exista um maior número de episódios ESV's do que os realmente apresentados. Por isto, a necessidade de se apresentar a rede um conjunto de treinamento compatível com a realidade observada. Porém, para a proposta aqui visada, este resultado servirá para a confirmação de um diagnóstico baseado nos dois resultados (TT e TP). Resumidamente, os resultados apresentados na Tabela 3.13 mostram em um aumento na Sensibilidade (já que a rede classifica um número maior de ESV's), e um decréscimo no 'VPP' (já que o valor de ESV's classificados na rede é maior em relação ao total de ESV's rotulados pela base de dados).

Com base nestes dois resultados, foi realizada uma fusão entre ambos, objetivando-se um aproveitamento significativo para a classificação de ESV. Esta fusão é denominada, neste trabalho, como Fusão de Resultados (FR).

A Fusão de Resultados tem como base identificar a classificação de cada batimento baseada no TT e compará-la com a classificação baseada no TP. Exemplificando, dado um batimento qualquer X , considere-se que a classificação utilizando o resultado do treinamento TT para X é ESV e a classificação utilizando o resultado do treinamento TP para X também é ESV. Então, a classificação baseada na FR assumirá que o batimento X é um episódio de ESV. Mas a FR tem um problema, quando a classificação usando os resultados do treinamento TP e TT são contraditórias, isto é, existe uma zona de incerteza baseada na FR. É necessário, neste momento, um critério de desempate. Este critério de desempate é baseado na Onda P, definido pelos seguintes passos:

- 1- Se Onda P = 'F' então Classificação FR = ESV.
- 2- Senão Classificação FR = 'Outros'.

Ressalte-se que este critério de desempate é utilizado apenas nos batimentos que caem nesta zona de incerteza. Então, como descrito anteriormente, a onda P será utilizada, mas não como um *nó* da rede, e sim como um critério de desempate do resultado, caso haja dúvida.

O resultado da classificação baseado em FR com critério de desempate utilizando Onda P (FR-OndaP) é apresentado na Tabela 3.14.

Comparando os resultados apresentados na Tabela 3.13 com os resultados obtidos utilizando a Fusão de Resultados (Tabela 3.14), observa-se que ocorreu um aumento significativo no valor da Sensibilidade, isto é, utilizando ‘FR-OndaP’. Neste caso, o algoritmo de classificação identifica mais precisamente episódios de ESV. Isto é, ocorreu um aumento de 3% no valor de ‘Se’ utilizando FR. Observa-se também que os valores relacionados aos intervalos de confiança permanecem os mesmos, ratificando que o aumento da classificação de ESV é válido.

3.6 Considerações Sobre o Uso de RB Estática

Quando se deseja desenvolver um sistema de apoio ao diagnóstico médico, como no caso presente, é necessário levar em consideração as incertezas inerentes a este problema, já que o especialista médico lida com incerteza em cada decisão tomada.

A metodologia baseada em Redes Bayesianas é então utilizada para representar conhecimento incerto, o que permite a argumentação baseada em teoria de probabilidade (Castillo, Gutiérrez e Hadi, 1997).

Neste capítulo foram implementadas duas topologias de Rede Bayesiana, cujo desempenho foi avaliado usando medidas de desempenho baseadas em Tabela de Contingência que gera medidas de Se, Es, VPP, VPN e Intervalo de Confiança.

É necessário apontar que a Rede Bayesiana com Fusão de Canais contribui significativamente para a melhoria na classificação de ESVs, bem como que a quantidade de ESVs contidos na base de dados a ser utilizada influi diretamente a qualidade dos resultados obtidos, por esta razão, optou-se pela utilização da base de dados MIT-BIH *Arrhythmia Database*.

Outro fator importante a ser destacado é o fato de a RB classificar de forma

Tabela 3.14: Resultado da Rede Bayesiana para a base de dados MIT-BIH ARRHYTHMIA DATABASE, ‘FR-OndaP’.

Sensibilidade (Se)	84,79%
Especificidade (Es)	99,49%
Valor Preditivo Positivo (VPP)	91,00%
Valor Preditivo Negativo (VPN)	99,10%
Intervalo de Confiança - Se (IC - Se)	$\pm 1,7$
Intervalo de Confiança – VPP (IC - VPP)	$\pm 1,3$

satisfatória, ‘independente’ do algoritmo de segmentação utilizado. Isto é, ela não está viciada em um algoritmo de segmentação do sinal ECG específico.

Outro dos resultados expressivos foi a observação de que a inclusão de um *nó* representando a Onda P na rede não gerou melhoria significativa nos resultados da classificação, concluindo-se que os *nós* representando os valores de RR e LL são suficientes para identificação de ESV’s. Esta afirmação é válida levando-se em consideração que os batimentos cardíacos da base de dados MIT-BIH *Arrhythmia Database* possuem exemplos insuficientes de batimentos rotulados como ESV que não possuem Onda P. Verifica-se esta afirmação pela porcentagem de 64,84% de batimentos classificados dentro deste padrão desejado que não apresentam Onda P. É necessário ressaltar que os algoritmos de segmentação do sinal ECG proposto por Andreão, Dorizzi e Boudy (2006) e Laguna são algoritmos de confiabilidade maior que 90% para a detecção da onda P.

Outra conclusão importante que se pode observar neste trabalho é que apesar da onda P não ter importância significativa como *nó* da rede, ela é muito útil como critério de desempate para a classificação de ESV baseada em Fusão de Resultados.

É importante destacar que os resultados obtidos pela Fusão de Resultados são de importância relevante para a classificação de ESVs, ressaltando-se desta forma que a utilização de *data mining* contribui significativamente para este tipo de classificação. Assim, uma continuação deste trabalho é prosseguir gerando diferentes bases de dados para o treinamento da rede.

As Redes Bayesianas até agora utilizadas nesta Tese são representadas por modelos estáticos, isto é, a variável tempo não intervém nas mesmas (ver diagrama da Figura 3.5, camada 1). Entretanto, para modelar um processo estocástico é mais adequado empregar Redes Bayesianas Dinâmicas, que envolvem a repetição causal de uma Rede Estática no tempo, como visto no capítulo 2. Uma Rede Bayesiana Dinâmica é uma cadeia de repetição da mesma Rede Bayesiana Estática quantas vezes forem necessárias. Utilizando Redes Bayesianas Dinâmicas pretende-se melhorar os resultados aqui obtidos, já que a sequência de batimentos que compõem um registro de ECG pode ser modelada como um processo estocástico. O desenvolvimento e os resultados da RBD são apresentados no próximo capítulo.

Capítulo 4

Rede Bayesiana Dinâmica com Definição de Limiar

Este capítulo apresenta uma nova versão do sistema para apoio à decisão médica apresentado no capítulo anterior, agora utilizando uma Rede Bayesiana Dinâmica (RBD). Mais uma vez a rede foi implementada usando o *toolbox* BNT (*Bayesian Net toolbox*) do MATLAB®.

Assim, como continuação deste trabalho, é estendido o paradigma de Redes Bayesianas e modela-se a evolução temporal da arritmia (cadeia Bayesiana Dinâmica) (Agosta e Gardos, 2004), levando-se em consideração os batimentos anteriores para que se possa deduzir a classificação do próximo batimento. A Figura 4.1 representa a arquitetura inicial construída para descrever o Sistema de Auxílio ao Diagnóstico de ESV proposto nesta Tese, utilizando a RBD como classificador. Note-se que a diferença em relação à Figura 3.5 é exatamente o uso de uma RBD na camada 1.

Inicialmente é feita uma descrição sucinta dos conceitos de *Data Warehouse*, *Data Mart* e *Data Mining*, metodologias estas que são necessárias para embasamento da avaliação e tratamento da base de dados selecionada para a aplicação da RBD (a MIT-BIH *Arrhythmia Database* (1997)). Após a introdução desses conceitos, é apresentada a proposta de *Data Mart* a ser utilizada em conexão com a MIT-BIH *Arrhythmia Database*. Posteriormente, é apresentada uma comparação entre a RB estática e RBD com aplicação de *Data Mart*. Para que esta comparação possa ser realizada, é apresentado o resultado da RB com a nova base de dados separada pelo *Data Mart* e depois os resultados da RBD são apresentados. Finalmente, é feita uma descrição da ferramenta de apoio a diagnóstico de ESV, utilizando a idéia de incerteza apresentada anteriormente, e de como utilizar esta incerteza baseada na interação do especialista médico. É necessário ressaltar que os conceitos de *Data Warehouse*, *Data Mart* e *Data Mining* são abordados de forma introdutória, pois foge ao escopo deste trabalho uma ampla discussão sobre estas metodologias.

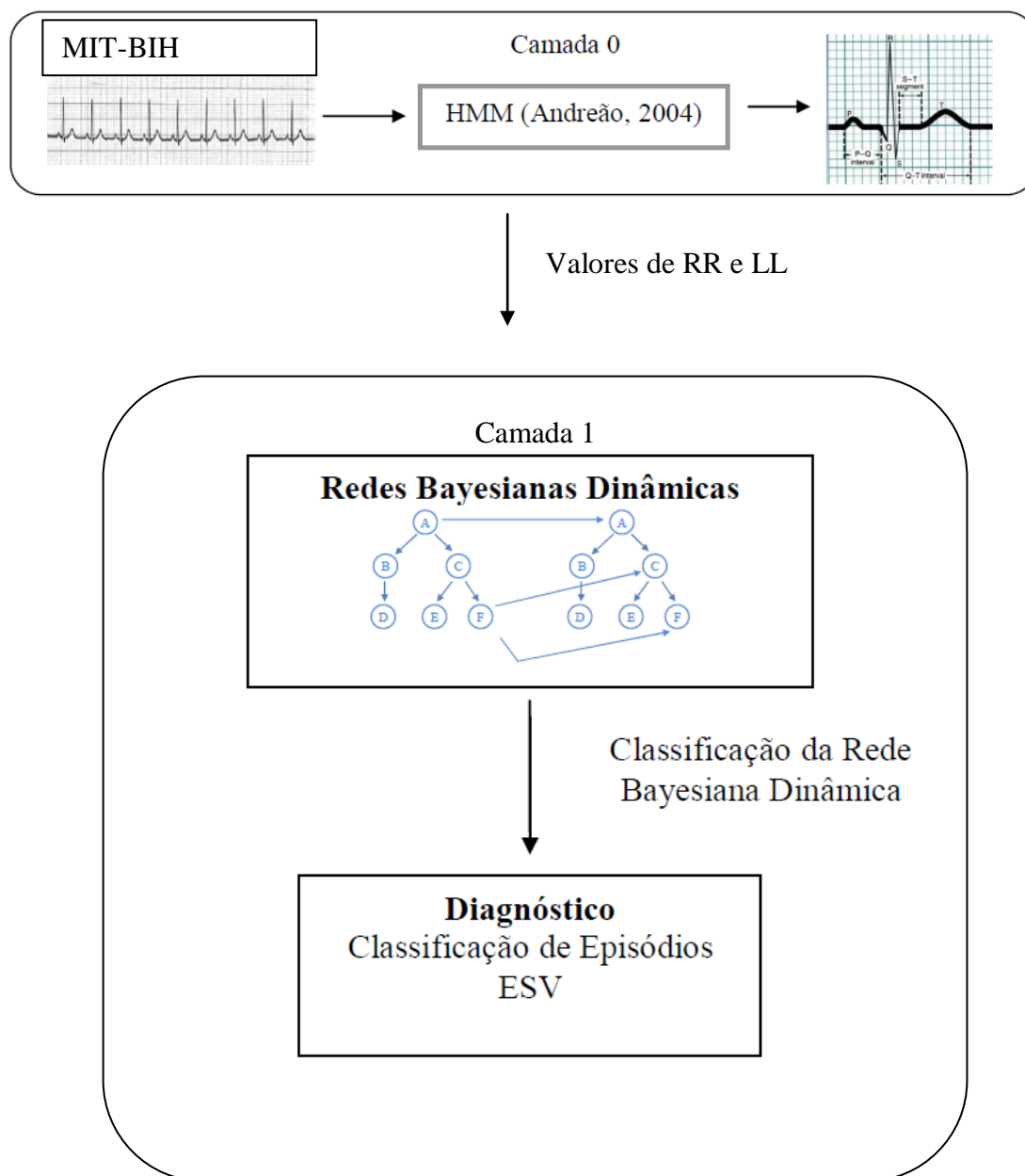


Figura 4.1: Arquitetura do Sistema de Auxílio ao Diagnóstico proposto, utilizando RBD.

4.1 *Data Warehouse (DW)*

Tecnologias aplicadas a Banco de Dados (BD) estão disponíveis há bastante tempo. As topologias iniciais de banco de dados concentravam-se em um único BD, o qual atendia os objetivos conhecidos da comunidade de informática – processamento de transações, em lote (*batch*) e analítico, entre outros. Nos primeiros sistemas de BD o

enfoque principal era o processamento operacional – geralmente transacional¹. Sistemas de informações transacionais, porém, são limitados quando utilizados como ferramenta de tomada de decisão gerencial.

Posteriormente, surgiu a necessidade de conceitos de BD mais sofisticados: um que atende às necessidades transacionais e outro que atende às necessidades informacionais ou analíticas. Suprindo as limitações de sistemas com processamentos transacionais, surgiu a tecnologia de *Data Warehouse*, que representou uma mudança fundamental nos sistemas de informação e introduziu alguns novos e importantes conceitos. O DW é o alicerce do processamento informacional ou analítico, que tem por meta apoiar a tomada de decisão no ambiente desejado.

Existem várias definições para Data Warehouse. Muitos autores que trabalham com este assunto fornecem diferentes definições:

- coleção de dados orientada por assuntos, integrados, não volátil, e variável em relação ao tempo, que tem por objetivo dar suporte aos processos de tomada de decisão (Inmon,1997);
- coleção integrada de bases de dados, orientada por assunto e otimizada, projetada para suportar a função SAD (Sistema de Apoio à Decisão), onde cada unidade de dados é relevante para algum momento do tempo (Imhoff, 2003).

Data Warehouse é uma tecnologia que reúne dados de forma organizada e eficiente, tendo por objetivo o auxílio na análise destes dados. Desta forma, as consultas para análises gerenciais de dados tornam-se mais fáceis de serem realizadas, uma vez que ele concentra, em um único local, informações específicas de um determinado contexto, provenientes de fontes de dados operacionais, que são armazenadas em estruturas multidimensionais próprias para análises de dados e consultas gerenciais. Isto se dá após os registros sofrerem uma conversão e haver uma coleta das informações gerenciais necessárias.

As técnicas de *Data Warehousing* para produção de informações gerenciais são utilizadas desde o início dos anos 90, em diversos setores da economia. Na área da saúde, por exemplo, também existem iniciativas para a adoção desta tecnologia, como

¹ Um sistema com processamento transacional é desenvolvido com o objetivo de dar suporte às atividades operacionais do banco de dados em questão. As operações atômicas são armazenadas - incluídas, alteradas e excluídas por seus usuários - constantemente, e sobre elas são executadas operações pré-definidas, as quais retratam o funcionamento diário destes bancos de dados.

programas de gerenciamento de doenças utilizando as tecnologias para gerenciamento de dados e o advento do “*managed care*”². O objetivo é aplicar programas preventivos em indivíduos doentes ou com potencial de contrair doenças, identificando estes indivíduos através de programas específicos. Neste contexto, sistemas DW encaixam-se perfeitamente. O trabalho de Ramick (2001) cita alguns casos de aplicação de DW em programas de gerenciamento de doenças.

É importante ressaltar que o aspecto físico é muito importante no projeto de DW. O particionamento dos dados diz respeito à separação destes dados em unidades físicas distintas, podendo ser tratados independentemente. *Data Mart* (DM) é um conceito que está ligado a esta questão.

4.2 *Data Mart* (DM)

De acordo com a definição exposta no trabalho de Inmon (1997) um DM é uma coleção de assuntos organizados para o suporte de decisões baseado nas necessidades de um departamento. O DM é um subconjunto lógico de um DW completo, criado para servir às necessidades particulares de um grupo específico. Essas subdivisões podem ser alocadas fisicamente em novas partições, buscando eficiência na manipulação dos dados e melhor organização.

É preciso observar que as diferenças entre DM e DW são apenas com relação ao tamanho e ao escopo do problema a ser resolvido. O *Data Mart* trata de problema local, e enquanto um *Data Warehouse* envolve todos os dados disponíveis, para que o suporte

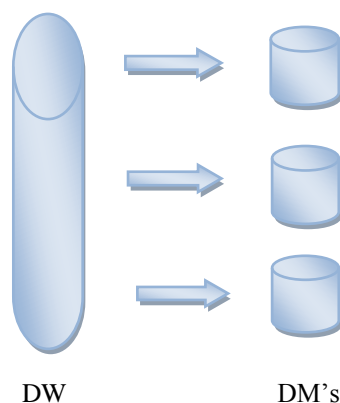


Figura 4.2: Esquema de construção de *Data Marts* a partir do *Data Warehouse*.

² O *Managed Care* nada mais é que o sistema de atendimento médico através de serviços de seguro saúde americano. Muitas de suas características foram importadas pelas empresas brasileiras.

à decisões atue em todos os níveis do problema abordado. O caminho tradicional é construir um DW, e a partir dele os *Data Marts* (DM's) (Machado, 2004), conforme ilustra a Figura 4.2.

Existe outra abordagem no desenvolvimento de DM e DW, que envolve primeiramente construir DM's e, a partir destes, o DW (inversão das setas da Figura 4.2). Não cabe aqui discutir estas técnicas (Machado, 2004), mas o objetivo é usar a técnica apresentada na Figura 4.2, isto é, a partir de um DW já criado, subdividir este DW em DM específicos de acordo com o tema abordado. Isto é, agrupar dados em um novo banco de dados, referente a um assunto em especial, dados estes que focalizam uma ou mais áreas específicas.

Em suma, o DM é um subconjunto de um *Data Warehouse*, no qual uma porção resumida ou altamente focalizada dos dados da organização é colocada em um banco de dados separado, destinado a uma população específica de usuários (Laudon, 2007).

O DM é utilizado em grande escala na área comercial, mas na área médica também existem muitos trabalhos que utilizam esta tecnologia, como mostram Isken, Littig e West (2001), que descrevem a evolução e a arquitetura de um *Data Mart* desenvolvido para atender as necessidades de modelagem e análise de operações de cuidado à saúde. Pode-se encontrar também um trabalho muito interessante em (Dallora *et al*, 2004), que destaca a importância e as oportunidades advindas da utilização de um *Data Mart* no processo de apoio à tomada de decisão gerencial do centro cirúrgico de um ambiente hospitalar.

Neste contexto, o presente trabalho visa apresentar o particionamento do *Data Warehouse* correspondente à base de dados MIT-BIH *Arrhythmia Database* (1997), para criar um *Data Mart* focalizando especificamente Extrassístoles Ventriculares. É importante deixar claro que neste trabalho iremos considerar tal base de dados como um DW já completo, pois, de acordo com as definições apresentadas em (Inmon, 1997) e em (Imhoff, 2003), a MIT-BIH *Arrhythmia Database* satisfaz as características de um DW.

4.3 *Data Mining*

Existem várias maneiras de recuperar informações de um *Data Warehouse*. Uma das ferramentas mais empregadas é o *Data Mining*.

O termo *Data Mining* se tornou bastante popular nos últimos anos, encontrando-se na literatura significativa variedade de estudos sobre o tema. Esta seção tem por

objetivo a introdução de conceitos básicos dessa tecnologia, apresentando alguns desses conceitos introdutórios que visam à descoberta de conhecimento em grandes conjuntos de dados.

A tecnologia de *Data Mining* (mineração de dados, em português) emergiu da intersecção de três áreas: inteligência artificial, estatística clássica e aprendizado de máquina. O *Data Mining* é uma das técnicas utilizadas para a realização de KDD (*Knowledge Discovery in Databases*) – em português Descoberta de Conhecimento em Bases de Dados, que permite a extração não trivial de conhecimento previamente desconhecido e potencialmente útil de um banco de dados (Addrians e Zantinge, 1996). O KDD pode ser visto como o processo da descoberta de novas correlações, padrões e tendências significativas, por meio da análise minuciosa de grandes conjuntos de dados estocados. Este processo se vale de tecnologias de reconhecimento utilizando padrões e técnicas estatísticas e matemáticas. Esse conceito é enfatizado por Fayyad *et al.* (1996), ao afirmar que é “o processo não trivial de identificação de padrões válidos, desconhecidos, potencialmente úteis e, no final das contas, compreensíveis, em dados”.

Um ambiente de apoio à tomada de decisões, integrando técnicas de *Data Mining* sobre um ambiente de *Data Warehousing*, possibilita um grande número de aplicações, que já vêm sendo implementadas em diversas áreas.

As técnicas de *Data Mining* são de caráter genérico, e podem ser implementadas por meio de ferramentas diferentes, como Redes Neurais Artificiais, Estatística ou Inteligência Artificial Simbolista, entre outras (Carvalho, 2001). Este é um campo que compreende atualmente muitas ramificações importantes, e uma destas técnicas é a descoberta de conhecimento baseado na tecnologia de Redes Bayesianas (RB) para extração de modelos de conhecimento válidos a partir de bases de dados.

Um dos fatores primordiais para o sucesso da descoberta do conhecimento é a capacidade de compreensão, entendimento ou ainda interpretação dos padrões descobertos, aplicando técnicas de mineração de dados. Neste sentido, as redes Bayesianas surgem como uma das mais proeminentes ferramentas, quando se considera a facilidade alcançada para a interpretação do conhecimento.

É importante ressaltar que o objetivo deste tópico não é discutir as diversas técnicas de *Data Mining*, mas enfatizar a utilização de RB como ferramenta de mineração de dados no contexto de *Data Mart*.

4.4 *Data Mart* para detecção de ESVs

A aplicação deste trabalho se deu na base de dados já descrita e discutida anteriormente, denominada MIT-BIH *Arrhythmia Database* (1997). Esta base de dados já está rotulada, e é considerada neste trabalho, dentro do contexto apresentado, como o *Data Warehouse*, após o pré-processamento dos registros desta base de dados pelo algoritmo de segmentação apresentado por Andreão (2004). Este DW será subdividido, gerando-se um conjunto específico para detecção de Extrassístoles Ventriculares.

De todos os registros que foram avaliados e classificados nas Redes Bayesianas Estáticas apresentadas anteriormente, foram excluídos alguns registros, para que se possa atingir o propósito de classificação de ESVs. Assim, a Figura 4.3 representa a extração de dados relevantes para detecção de ESVs da base de dados MIT-BIH *Arrhythmia Database*.

4.4.1 Método de Seleção de Registros para Classificação de ESVs

A base de dados MIT-BIH *Arrhythmia Database* é subdividida em vários registros, sendo cada registro representado por exames de ECG de pacientes diferentes. O processo de seleção de registros foi baseado no conteúdo de cada um deles. Assim, cada registro desta base de dados foi examinado e o direcionamento deste estudo visa a verificação das arritmias representativas em cada uma deles. A partir deste estudo,

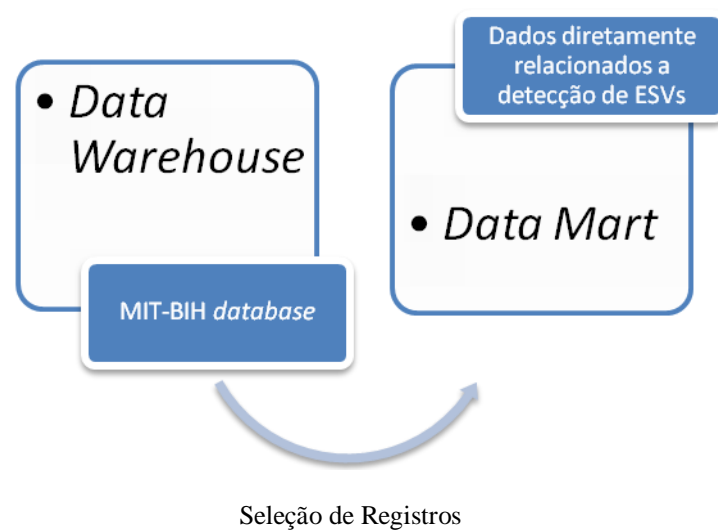


Figura 4.3: Arquitetura do *Data Mart* para a classificação de ESVs.

foram selecionados alguns registros que realmente são pertinentes ao estudo de ESVs, isto é, registros que possuem esta arritmia de forma representativa.

O questionamento que vem à tona neste momento é: porque não utilizar todos os registros? A resposta baseia-se no que foi apresentado na seção sobre *Data Mart*, pois o objetivo da utilização desta metodologia se justifica pelo aprimoramento da base de dados sobre a qual serão aplicadas técnicas de *Data Mining*, no caso específico deste trabalho as RB para classificação de ESVs.

Inicialmente, foram estudados todos os registros e selecionados os registros que poderiam, de alguma forma, não contribuir de forma significativa para a detecção de ESV's. Vejamos a seguir alguns exemplos que não serão representativos para a detecção desta arritmia cardíaca:

- registro 107, correspondente a um paciente que utiliza marca passo;
- registro 223, correspondente a um paciente que sofre de Taquicardia Ventricular;
- registro 232, correspondente a um paciente que possui outros tipos de arritmias, porém nenhuma relacionada a ESV.

No caso do registro 107 (existem mais registros na MIT-BIH *Arrhythmia Database* de pacientes com marcapasso, o registro 107 é apenas um exemplo). Quando um paciente utiliza o marcapasso o complexo QRS gerado por este aparelho causa uma deformação nesta onda, isto é, o complexo QRS fica alargado, como um batimento ventricular. Esta deformação da onda causa confusão na rede quando esta for identificar valores para o nó 'Batimento Ventricular'. Resta também ressaltar que em pacientes com marcapasso a forma de avaliação dos batimentos cardíacos não é a mesma que em um coração que gera seus próprios impulsos elétricos. Por estas razões não cabe a este estudo utilizar estes batimentos cardíacos.

No caso do registro 223, este paciente apresenta basicamente taquicardia ventricular (TV). Como foi descrito no Capítulo 1, esta arritmia cardíaca não é objetivo deste trabalho, porém na base de dados MIT-BIH *Arrhythmia Database* ela está rotulada como se fossem sucessivos episódios de ESV, o que é uma realidade, pois TV é um conjunto sucessivo de ESVs. No entanto, é necessário, neste caso, levar em consideração outro fator muito importante, a frequência cardíaca, informação que é de total relevância para a detecção de TV. Outro ponto importante é que este fator prejudica o aprendizado da rede pois como os intervalos R-R são calculados de forma dinâmica e normalizados de acordo com um batimento normal, a partir da repetição de

ESV a rede pode considerar que os intervalos R-R entre a TV (como irão tendendo para um intervalo de duração aproximada) é normal, e não prematuro. Sendo assim, além de outra variável para identificação desta arritmia específica também seria necessário a implementação de uma rotina especial no momento de normalização das variáveis ‘RR’ e ‘LL’. Por estas razões, não se julgou adequado utilizar registros que correspondem a TV.

No caso de registros que apresentam basicamente outros tipos de arritmias, como é o caso do registro 232, não se julgou necessária a utilização destas informações, uma vez que para a classificação destas arritmias seria necessário trabalhar com *nós* RB distintos dos que são utilizados neste trabalho.

É importante destacar que a não utilização destes registros não influencia de forma negativa a classificação de ESV's. De fato, quando alguma rede for construída para classificação de outras arritmias cardíacas serão utilizadas outras variáveis de entrada. Pode-se pensar que com a exclusão de alguns registros a rede possa ficar tendenciosa, ou classificar de forma incorreta outros pacientes. Na verdade, a rede não ficará tendenciosa, porque a separação do conjunto de teste e treinamento formam conjuntos totalmente disjuntos, o que significa que os pacientes utilizados para o treinamento da rede são diferentes dos pacientes utilizados no conjunto de teste. Este aspecto já foi abordado e bem discutido no Capítulo 3, onde se mostra que uma separação adequada entre o conjunto de treinamento e teste não geram vício na rede. Outro ponto importante que ratifica a aplicabilidade desta idéia é a utilização de técnicas de *Data Mart*, que particionam a base de dados em subconjuntos realmente significativos para a classificação pretendida.

Não se pode afirmar que a rede classificaria de forma incorreta outros pacientes (isto é, pacientes com outras arritmias cardíacas), pois, quando se ampliar o horizonte de doenças cardíacas detectadas pela rede, é também necessário aumentar o número de variáveis de entrada da rede (evidências ou *nós* observáveis). Por exemplo, quando se objetiva a classificação de Episódios Isquêmicos, o próprio especialista médico leva em consideração outras variáveis no ECG, isto é, ele observa o Segmento ST e o ponto J, variáveis muito significativas para a identificação desta arritmia, porém não necessárias para a detecção de ESV.

Ainda para ratificar esta idéia, é implementada uma RB que gera resultados de classificação do batimento cardíaco em ‘Normais’, ‘ESV’ e ‘Outros’. Neste ponto é criada uma nova classificação, os ‘Normais’ (Tabela 4.1). Assim, a rede não diz apenas

se o batimento é ESV. Ela diz também se o batimento é Normal ou se o batimento possui outro tipo de arritmia ('Outros'). Observe-se que até este momento a RB gera a probabilidade do batimento cardíaco ser considerado ESV ou 'Outros', esta última classificação englobando batimentos cardíacos considerados normais e arrítmicos.

Desta forma, a Tabela de Contingência desta rede é baseada na Tabela 4.1, e seus resultados são apresentados na Tabela 4.2. Os valores da Sensibilidade, Especificidade, Valor Preditivo Positivo e Valor Preditivo Negativo são apresentados na Tabela 4.3, sendo a topologia da rede a mesma mostrada na Figura 4.4.

Observa-se que o resultado obtido é inferior quando comparado às RB's

Tabela 4.1: Tabela de Contingência para representação de três níveis de classificação.

Classificação da RBE	Rótulos da Base de dados		
	ESV	Normal	Outros
ESV	A	B	C
Normal	D	E	F
Outros	G	H	I

discutidas no capítulo anterior. O valor da Sensibilidade ficou em aproximadamente 67%, enquanto os valores de Sensibilidade obtidos com as redes anteriores tiveram o valor aproximado de 80%. Outros resultados, como Especificidade e Valor Preditivo Negativo, também ficaram inferiores quando comparados às RB mostradas no capítulo anterior.

Com estes resultados se tem a confirmação de que é necessário fornecer à rede outras variáveis observáveis (evidências), para ela classificar, mais precisamente, outros tipos de arritmias. Por esta razão, quando se aumenta o número de arritmias classificadas aumenta-se também o número de variáveis observáveis, justificando-se assim a aplicação de *Data Mart* para a classificação de ESV's.

É importante ressaltar que em outros registros utilizados neste trabalho é possível encontrar Taquicardia Ventricular, batimentos Atriais Prematuros e outras

Tabela 4.2: Tabela de Contingência com os resultados da classificação da RB estática.

Classificação da RBE	Rótulos da Base de dados		
	ESV	Normal	Outros
ESV	1180	39	38
Normal	467	17123	6278
Outros	102	3820	2133

Tabela 4.3: Resultado dos valores de Se, Es, VPP e VPN, baseados na Tabela de Contingência

Sensibilidade (Se)	$A/(D+A+G)$	67,47%
Especificidade (Es)	$E/(E+B+H)$	81,60%
Valor Preditivo Positivo (VPP)	$A/(A+B+C)$	93,87%
Valor Preditivo Negativo (VPN)	$E/(B+E+F)$	71,74%

arritmias diferentes de ESV. Porém, nestes registros estes tipos distintos de arritmias ocorrem com menor frequência, isto é, não são registros destinados unicamente ao apontamento de tais arritmias. Deste modo, mesmo com a inclusão de outros tipos de arritmias é possível obter resultados representativos, isto é, não são utilizados somente registros que possuem batimentos rotulados como ESV, mas também registros que possuem outros tipos de arritmia. Porém, neste último caso, não é observado um número excessivo de tais arritmias.

Em suma, são selecionados registros representativos para a detecção de ESV, obtidos da base de dados MIT-BIH *Arrhythmia Database*, aplicando-se para este propósito técnicas de *Data Mart*. Esta seleção de subconjuntos foi justificada, e não prejudica e nem torna tendenciosa a rede para a classificação de ESV's.

4.4.2 Rede Bayesiana Estática para Classificação de ESVs

Após a utilização de técnicas de *Data Mart* para a seleção dos registros, houve a necessidade da separação da base de dados resultante em dois conjuntos disjuntos, os subconjuntos de Teste e Treinamento.

Para que se pudesse encontrar uma quantidade específica de batimentos para cada um destes conjuntos, foram realizadas diversas experiências, objetivando-se encontrar uma porcentagem ideal para o conjunto de Treinamento, sendo importante recordar que a base de Treinamento deve possuir uma parcela muito maior de dados. A solução que melhor supriu as necessidades foi a separação de 80% de batimentos cardíacos para o conjunto de Treinamento e 20% de batimentos para o conjunto de Teste, lembrando que estes conjuntos são disjuntos. Foram realizados vários testes, com vários conjuntos de Treinamento e Teste selecionados randomicamente, e esta porcentagem foi a que apresentou melhor resultado.

É importante ressaltar que tanto para o conjunto de Teste quanto para o conjunto de Treinamento é obedecida a premissa que estipula a proporção de batimentos

rotulados como ESV, a qual deve ser aproximadamente a mesma para ambos os conjuntos. Assim, o total de batimentos rotulados como ESV é em torno de 6% para qualquer um dos conjuntos.

Após a seleção dos conjuntos de Treinamento e Teste, o próximo passo é implementar o aprendizado da RB. Então, faz-se necessário a escolha da topologia da Rede Bayesiana Estática, que é a mesma apresentada na Figura 4.4. Os resultados utilizando o conjunto de Teste, isto é, resultados apresentados na Tabela de Contingência, são mostrados na Tabela 4.4.

É possível observar uma melhor classificação utilizando-se as técnicas de *Data Mart*. Basta comparar este resultado com aquele apresentado na Tabela 3.5, onde a mesma topologia da rede é implementada. É claro que nesta comparação há uma diferença entre o conjunto de batimentos selecionados para Teste e Treinamento em ambas as comparações, mas as diferenças nos valores de ‘Se’ e ‘VPP’ (13% e 5%, respectivamente) justificam a utilização de técnicas de *Data Mart*. Os valores dos respectivos intervalos de confiança permitem esta comparação, mesmo entre bases de dados com dimensões distintas. Mesmo comparando a Tabela 4.4 com a Tabela 3.23, que representa a Fusão de Resultados, melhor resultado apresentado no Capítulo 3, pode-se observar um ganho de aproximadamente 8% e 5% nos valores de ‘Se’ e ‘VPP’, respectivamente, com os intervalos de confiança ratificando esta comparação.

4.4.3 Rede Bayesiana Dinâmica (RBD) para Detecção de ESV

Como Redes Bayesianas Dinâmicas (RBD) são RB’s que representam um modelo probabilístico temporal, os mesmos algoritmos para inferência probabilística em RB’s podem ser empregados. Por questões de eficiência, geralmente emprega-se o

Tabela 4.4: Tabela de Contingência e resultados obtidos com a Rede Bayesiana estática após o *Data Mart*.

Tabela de Contingência		
	Rótulo ‘ESV’	Rótulo ‘Outros’
Classificação ‘ESV’	850	15
Classificação ‘Outros’	97	13118
Sensibilidade (Se)		92,69%
Especificidade (Es)		99,89%
Valor Preditivo Positivo (VPP)		98,26%
Valor Preditivo Negativo (VPN)		99,26%
Intervalo de Confiança - Se (IC - Se)		± 1,65
Intervalo de Confiança - VPP (IC - VPP)		± 0,8

algoritmo aproximado para inferência conhecido como *particle filtering* (Russel e Norvig, 2000), que é uma modificação do algoritmo *likelihood weighting*, como já discutido em detalhes no Capítulo 2.

Normalmente as distribuições $P(E_t|E_{t-1})$ e $P(T_t|E_t)$ de uma RBD devem ser aprendidas a partir dos dados disponíveis. Isto é feito da mesma forma que em uma Rede Bayesiana, onde, conhecendo-se a estrutura da rede e sendo todas as variáveis observáveis, o aprendizado é feito pela estimação ML (*Maximum Likelihood*). Se a estrutura é conhecida mas existem variáveis não observáveis, utilizam-se métodos de gradiente ou EM (*Expectation-Maximization*) para o aprendizado (Friedman et al, 1998; Dempster et al, 1997). Para o aprendizado da RB, este trabalho utilizou o algoritmo ML disponível na biblioteca do MATLAB denominada BNT (2007), porque no aprendizado foram implementadas variáveis observáveis, como discutido anteriormente.

Como já discutido detalhadamente no Capítulo 2, os mesmos algoritmos para inferência probabilística em RB's podem ser empregados para RBD. Utilizamos neste trabalho o mesmo algoritmo para ambos os casos, pois o que vai diferenciar a RB da RBD é sua topologia (Redes Bayesianas Dinâmicas nada mais são que Redes Bayesianas que incluem informações temporais). No Capítulo 2 também foi elaborado um conjunto de restrições (Três Itens de Restrição - TIR) que foram seguidos no desenvolvimento da RBD, restrições estas que visam o controle sobre a complexidade do algoritmo, para que a implementação da RBD a partir de algoritmos de RB seja possível.

A topologia da RB utilizada neste trabalho é apresentada na Figura 4.4, enquanto que a topologia da RBD é mostrada na Figura 4.5. Observa-se que a RBD, no primeiro instante de tempo (Tempo n-1), é a mesma Rede Bayesiana estática apresentada na

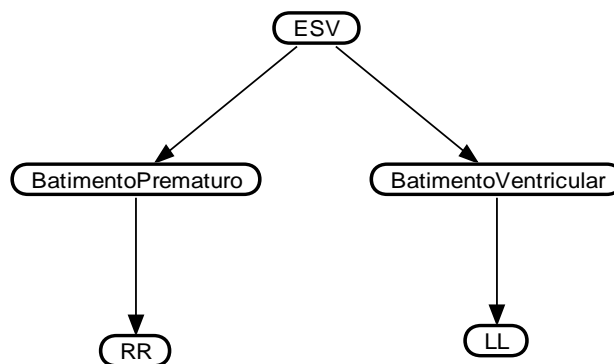


Figura 4.4: Modelo da Rede Bayesiana.

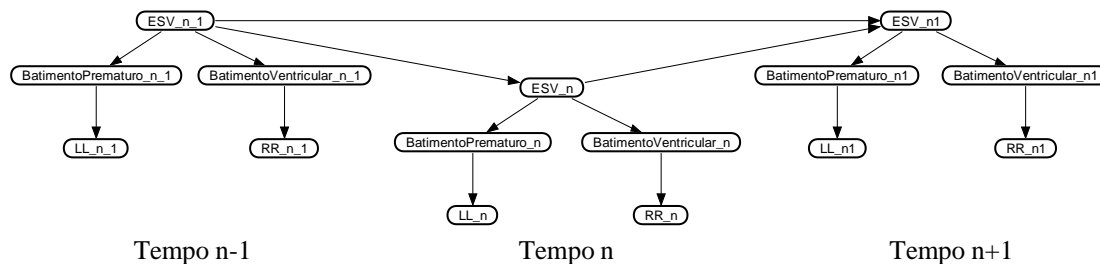


Figura 4.5: Rede Bayesiana Dinâmica com representação em três instantes de tempo.

Figura 4.4, como deve ser a representação da topologia da RBD.

Os arcos que conectam os *nós* entre os diferentes instantes de tempo vão da esquerda para a direita, representando, desta maneira, o passar do tempo.

Todos os *nós* são observáveis na etapa de treinamento. Por isto utiliza-se o algoritmo ML. Note-se que o valor do *nó* ESV_{n-1} (significa tempo anterior ou um batimento anterior, isto é, no instante de tempo $n-1$, onde n representa o instante atual) influencia o *nó* ESV_n (significa tempo atual ou batimento atual, isto é, no instante de tempo n), que, por sua vez, é influenciado pelo *nó* ESV_{n+1} (significa tempo futuro ou um batimento posterior, isto é, no instante de tempo $n+1$).

A definição do comprimento dos instantes de tempo (isto é, três instantes de tempo) foi baseada na experiência do especialista: ele verifica um instante de tempo anterior e um instante de tempo posterior ao batimento que será classificado. Porém, outras topologias também foram implementadas, visando à confirmação de que esta topologia fosse adequada, e verificou-se que o comprimento de três instantes de tempo é suficiente para uma boa classificação. Os resultados obtidos com a topologia apresentada na Figura 4.5 são mostrados na Tabela 4.5, que corresponde à classificação

Tabela 4.5: Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o *Data Mart*, como apresentada na Figura 4.5.

Tabela de Contingência		
	Rótulo 'ESV'	Rótulo 'Outros'
Classificação 'ESV'	866	15
Classificação 'Outros'	51	13118
Sensibilidade (Se)		94,44%
Especificidade (Es)		99,89%
Valor Preditivo Positivo (VPP)		98,30%
Valor Preditivo Negativo (VPN)		99,61%
Intervalo de Confiança - Se (IC - Se)		$\pm 1,45$
Intervalo de Confiança - VPP (IC - VPP)		$\pm 0,8$

de ESV no instante de tempo n , que é o instante de tempo atual.

Comparando o resultado da RB mostrado na Tabela 4.4 com o resultado da Tabela 4.5, verifica-se um ganho no valor da Sensibilidade em favor da RBD de 2%.

Como já descrito, esta topologia de RBD não foi a única a ser implementada. Foram implementadas outras RBDs com dependências temporais diferentes, como a apresentada na Figura 4.5. Esta topologia leva em consideração a dependência temporal entre os *nós* Batimento Prematuro e Batimento Ventricular, porém a dependência temporal entre os *nós* ESV's não é representada nesta rede. O resultado desta rede é mostrado na Tabela 4.6.

Na topologia da rede apresentada na Figura 4.6, os *nós* ESV no tempo $n-1$, n e $n+1$, respectivamente, acabam sendo influenciados pelos *nós* BatimentoPrematuro e

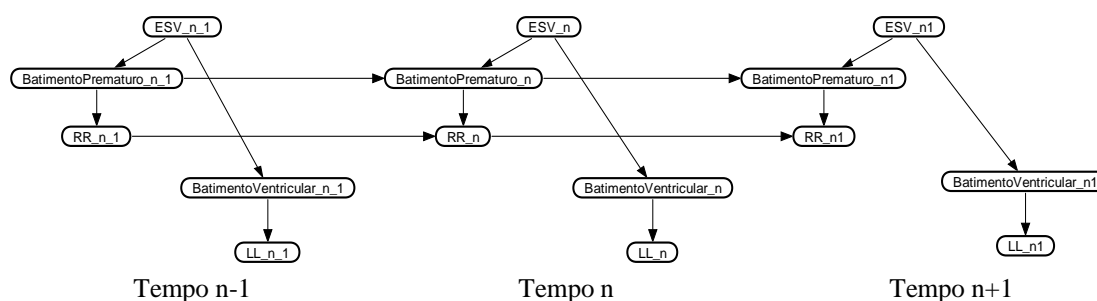


Figura 4.6: Rede Bayesiana Dinâmica com representação em 3 instantes de tempo.

BatimentoVentricular também nos tempos $n-1$, n e $n+1$. Assim, como na Figura 4.5, o *nó* ESV no tempo $n-1$ influencia indiretamente o *nó* ESV no tempo n , e assim sucessivamente.

Destaca-se o resultado apresentado na Tabela 4.6 como o melhor resultado da

Tabela 4.6: Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica apresentada na Figura 4.6, após o *Data Mart*.

Tabela de Contingência (MC)		
	Rótulo 'ESV'	Rótulo 'Outros'
Classificação 'ESV'	872	14
Classificação 'Outros'	45	13119
Sensibilidade (Se)		95,09%
Especificidade (Es)		99,86%
Valor Preditivo Positivo (VPP)		98,42%
Valor Preditivo Negativo (VPN)		99,66%
Intervalo de Confiança - Se (IC - Se)		± 1,4
Intervalo de Confiança - VPP (IC - VPP)		± 0,8

RBD. Este resultado apresenta um ganho de aproximadamente 3% em comparação ao resultado da RB estática, no valor da Sensibilidade, isto é, a RBD classifica melhor episódios de ESV. Este valor é considerado significativo, principalmente devido ao intervalo de confiança de ambas as redes serem pequenos.

Contudo, este não é o único ganho da RBD: foi observado, também, um valor de probabilidade correspondente à RBD mais discriminativo quando comparado ao valor de probabilidade correspondente à RB. Explicando melhor esta idéia, este valor de probabilidade discriminativo envolve uma nova visualização e tratamento da incerteza. Até este momento, a incerteza tratada neste trabalho envolve a incerteza aleatória³ inferida pela RBD, isto é, a incerteza propagada pela inferência do algoritmo da RBD.

Após o cálculo da probabilidade de um batimento, o algoritmo aqui implementado, até este momento, segue a seguinte regra:

1. Se a probabilidade de ESV do batimento atual for maior que 50%, então assumo que o batimento atual será classificado como ESV,
2. Senão, assumo que o batimento atual será classificado com ‘Outros’.

Observa-se que utilizando a regra descrita acima não se leva em consideração a probabilidade real de um batimento ser considerado ESV ou ‘Outros’. Por exemplo, sendo ‘B’ o conjunto de batimentos classificados pela rede e ‘ b_i ’ um batimento qualquer (tal que $b_i \subset B$), quando a rede classifica b_i como ESV ela não leva em consideração que a probabilidade de b_i ser realmente ESV é de 60% (neste exemplo). Este valor de probabilidade é importante, e deve ser levado em consideração pelo especialista. Observe-se que cada batimento classificado pela rede possui uma probabilidade de ser considerado ESV e outra de ser considerado ‘Outros’, e é este valor de probabilidade que está sendo considerado neste momento. Deste modo, seria interessante deixar a cargo do especialista decidir o grau de certeza com que a rede classifica cada batimento. Para isto, a probabilidade de cada batimento é muito importante, e deve ser ressaltada.

Também neste aspecto a RBD difere da RB estática: este valor de probabilidade é mais discriminativo com a utilização da RBD. À guisa de exemplo, a Tabela 4.7 ilustra essa diferença de capacidade discriminatória. Ela ilustra o caso de 13 batimentos classificados incorretamente tanto pela RBD quanto pela RB, na fase de Teste. Nela estão mostradas as probabilidades a partir das quais a RBD e a RB classificaram cada

³ A expressão ‘incerteza aleatória’ refere-se ao fato de que como cada nó da Rede Bayesiana retrata uma variável aleatória representada por tabelas (ou PDFs) de probabilidades condicionais e a inferência Bayesiana propaga a incerteza através destas tabelas (ou PDFs), então esta incerteza propagada na rede, neste trabalho é referenciada como incerteza aleatória.

Tabela 4.7: Valores de probabilidade para 13 batimentos classificados incorretamente tanto pela RBD quanto pela RB (eles foram classificados como ‘Outros’, embora sejam ‘ESV’).

Rede Dinâmica (%)	Rede Estática (%)
75,38	95,45
83,40	96,89
91,88	96,26
81,41	76,55
74,18	95,10
64,58	93,43
88,78	89,39
90,89	96,15
73,70	94,89
83,57	95,30
91,29	96,28
88,01	89,86
71,00	94,45

um dos 13 batimentos como ‘Outros’, sendo que os rótulos dados pelos especialistas apontam que eles são ‘ESV’.

Note-se que as probabilidades da RBD são aproximadamente 15% mais discriminativas que a probabilidade da RB, o que é mais uma vantagem da RBD apresentada neste trabalho. Este diferencial de aproximadamente 15% se estende ao total de batimentos classificados incorretamente por ambas as redes.

Destaque-se que tal valor de probabilidade deve ser levado em consideração para que se possa obter um resultado melhor qualificado, e com um índice de incerteza definido pelo especialista. De fato, a RBD gera este valor que implica na incerteza aleatória, e é esta incerteza que deve ser ponderada e considerada pelo especialista.

Assim, cria-se um sistema mais interativo com o usuário (no caso em questão, o médico), o qual utiliza melhor a incerteza aleatória, não somente a propagada pela rede mas também a obtida pelo resultado da classificação.

O sistema proposto permite a autonomia do especialista em avaliar o grau de certeza que a rede apresenta, decidindo qual é o limiar que a rede utiliza para decidir os resultados da classificação. Através do algoritmo a seguir pode-se visualizar com mais clareza como a definição do limiar funciona:

1. Limiar = X, então

- a. Se a probabilidade da classificação de $ESV > X$, então a classificação do batimento é *ESV*;
- b. Se a probabilidade da classificação de '*Outros*' $> X$, então a classificação do batimento é '*Outros*';
- c. Senão, a classificação do batimento é *INDEFINIDO*, isto é, a RBD não classifica estes batimentos cardíacos.

De acordo com o pseudocódigo anterior, quando o especialista define o limiar, no caso o valor de X , ele aceita que alguns batimentos irão cair em uma zona de indeterminação (batimento indefinido). Esta abordagem permite que o médico decida qual é o grau de certeza que determinada arritmia deve ter, no caso deste trabalho, a arritmia *ESV*. Assim que o especialista decide qual o grau de certeza, ele deve estar ciente que quanto maior o grau de certeza que ele estipula maior será o total de batimentos indefinidos. Porém, estes batimentos indefinidos são apresentados ao especialista, cabendo a ele a responsabilidade pela classificação. Ademais, o total de batimentos classificados como indefinidos é pequeno, se comparado ao total de batimentos classificados pela rede, como se pode verificar a seguir.

São apresentados alguns exemplos para uma melhor visualização deste método e os seus respectivos resultados.

Exemplo 1: O especialista define um limiar de 65%, isto é, ele quer apenas os resultados dos batimentos classificados com certeza maior que 65% (probabilidade $\geq 65\%$). A Tabela de Contingência, os valores de '*Se*', '*Es*', '*VPP*' e '*VPN*' e os intervalos de confiança de '*Se*' e '*VPP*' são apresentados na Tabela 4.8, bem como os intervalos de confiança da '*Se*' e '*VPP*'.

Tabela 4.8: Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o *Data Mart*, com limiar de 65%.

Tabela de Contingência (MC)		
	Rótulo ' <i>ESV</i> '	Rótulo ' <i>Outros</i> '
Classificação ' <i>ESV</i> '	863	12
Classificação ' <i>Outros</i> '	27	13113
Sensibilidade (<i>Se</i>)		96,97%
Especificidade (<i>Es</i>)		99,91%
Valor Preditivo Positivo (<i>VPP</i>)		98,63%
Valor Preditivo Negativo (<i>VPN</i>)		99,79%
Intervalo de Confiança - <i>Se</i> (IC - <i>Se</i>)		± 1
Intervalo de Confiança - <i>VPP</i> (IC - <i>VPP</i>)		$\pm 0,7$

Note-se que o valor de ‘Sensibilidade’ da RBD com definição de limiar melhora aproximadamente 2% se comparado à RBD sem definição de limiar. Vale observar, também, que esta melhora se estende para aproximadamente 5% quando se compara com a RB estática. Observa-se também o intervalo de confiança, que varia em torno de 1%, implicando, desta forma, que um ganho de 2% é realmente significativo. É importante deixar claro que esta rede, cujos resultados são apresentados na Tabela 4.8, possui batimentos classificados como indefinidos, os quais representam 0,25% do total de batimentos classificados. Isto é, a RBD com definição de limiar classificou 99,75% do total de batimentos com certeza maior ou igual a 65%. Ou seja, os batimentos classificados como indefinidos equivalem a uma porção muito pequena em relação ao total de batimentos classificados. Cabe ao médico decidir se este valor é significativo ou não para a arritmia em questão.

Exemplo 2: O especialista define uma certeza maior ou igual a 75%. Assim, a probabilidade de classificação de cada batimento tem que ser maior ou igual a 75%. Os resultados deste exemplo são apresentados na Tabela 4.9.

Observa-se que houve um aumento de mais 4% no valor da ‘Sensibilidade’ quando comparado à RBD sem definição de limiar, e de aproximadamente 7%, se comparado à RB estática. Mas, neste caso, houve um aumento no valor de batimentos indefinidos, onde a RBD com definição de limiar no valor de 75% classifica 99,43% do total de batimentos destinados ao teste. Note-se que houve um aumento de aproximadamente 0,3% do total de batimentos indefinidos, quando comparado à RBD com definição de limiar no valor de 65%.

Exemplo 3: Neste caso, é definido um limiar de 90%, ou seja, o especialista quer

Tabela 4.9: Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o *Data Mart*, com limiar de 75%.

Tabela de Contingência (MC)		
	Rótulo ‘ESV’	Rótulo ‘Outros’
Classificação ‘ESV’	861	9
Classificação ‘Outros’	7	13092
Sensibilidade (Se)		99,19%
Especificidade (Es)		99,93%
Valor Preditivo Positivo (VPP)		98,97%
Valor Preditivo Negativo (VPN)		99,95%
Intervalo de Confiança - Se (IC - Se)		± 0,6
Intervalo de Confiança - VPP (IC - VPP)		± 0,6

Tabela 4.10: Tabela de Contingência e resultados obtidos com a Rede Bayesiana Dinâmica após o *Data Mart*, com limiar de 90%.

Tabela de Contingência (MC)		
	Rótulo 'ESV'	Rótulo 'Outros'
Classificação 'ESV'	850	3
Classificação 'Outros'	3	12946
Sensibilidade (Se)		99,65%
Especificidade (Es)		99,98%
Valor Preditivo Positivo (VPP)		99,65%
Valor Preditivo Negativo (VPN)		99,98%
Intervalo de Confiança - Se (IC - Se)		± 0,4
Intervalo de Confiança - VPP (IC - VPP)		± 0,4

ter uma certeza maior ou igual a 90%. O resultado correspondente está na Tabela 4.10, onde se verifica um aumento de aproximadamente 0,5% no valor da 'Sensibilidade', quando comparado à RBD com limiar de 75%, bem como a um aumento de aproximadamente 0,7% no 'VPP'. Ocorreu, também, uma classificação de 97,71% do total de batimentos selecionados para o teste, isto é, apenas 2,29% dos batimentos foram classificados como indefinidos, número este que continua sendo pequeno em comparação ao número total de batimentos classificados. É fácil verificar, também, que os intervalos de confiança, que são muito pequenos, ratificam a classificação, assim como fazem com que um pequeno acréscimo no ganho entre as definições de limiar seja bastante representativo.

Destarte, vale ressaltar novamente que cabe ao especialista a decisão final quanto à importância da certeza inerente à classificação correta do batimento.

Outra maneira de identificar limiar é disponibilizar para o especialista a escolha de certeza tanto para a classificação de ESV's como de 'Outros'. Para isto, observa-se o seguinte pseudocódigo:

1. Limiar para definição de ESV = X e limiar para definição de 'Outros' = Y
 - a. Se a probabilidade de ESV > X, então a classificação do batimento é ESV;
 - b. Se a probabilidade de 'Outros' > Y, então a classificação do batimento é 'Outros';
 - c. Senão, a classificação do batimento é INDEFINIDO, isto é, a RBD não classifica estes batimentos.

Esta idéia é melhor ilustrada pelo seguinte exemplo:

Exemplo 4: É definido um valor de limiar de 80% para o valor de Y e 95% para o valor de X. Isto é, o especialista quer uma certeza de 80% quando a rede classificar o batimento como ‘Outros’ e uma certeza de 95% quando a rede classificar o batimento como ESV. O resultado deste exemplo é mostrado na Tabela 4.11. Ele reflete uma certeza bastante elevada quanto à classificação de ESV e uma certeza relativamente alta quanto à classificação de outros tipos de batimentos (‘Outros’). Porém este é apenas um exemplo ilustrativo: pode não ser necessário usar limiares tão elevados para este tipo de arritmia, cabendo novamente ao especialista decidir tal situação.

Analisando este exemplo pode-se verificar que os valores resultantes ficaram bem próximo aos da Tabela 4.10. Porém, é importante observar que a porcentagem de batimentos classificados é de 98,5%, isto é, a porcentagem dos batimentos indefinidos é de 1,5% do total de batimentos selecionados para o teste, ou seja, houve uma redução de aproximadamente 1% do total de batimentos indefinidos quando comparamos esta rede com a RBD apresentada no exemplo 3.

Tendo em vista estes 4 exemplos, pode-se concluir que a RBD com definição do valor de limiar é uma ótima opção para classificação de batimentos ESV, porque apresenta certeza maior que 99% na ‘Se’ e ‘VPP’ quando este limiar é maior que 75%. Do mesmo modo, a quantidade de batimentos indefinidos pode ser considerada pequena quando comparada com o total de batimentos classificados pela rede.

Outra vantagem desta rede é que ela, além de trabalhar com a incerteza aleatória propagada pelos *nós* da rede, também leva em consideração a incerteza no momento da classificação de cada batimento, deixando o especialista decidir o valor da certeza que é mais conveniente para ele.

Tabela 4.11: Tabela de Contingência e resultados obtidos com para a Rede Bayesiana Dinâmica após o *Data Mart*, com limiar de 95% e 80%.

Tabela de Contingência (MC)		
	Rótulo ‘ESV’	Rótulo ‘Outros’
Classificação ‘ESV’	843	0
Classificação ‘Outros’	4	12987
Sensibilidade (Se)		99,53%
Especificidade (Es)		100%
Valor Preditivo Positivo (VPP)		100%
Valor Preditivo Negativo (VPN)		99,97%
Intervalo de Confiança - Se (IC - Se)		± 0,4
Intervalo de Confiança - VPP (IC - VPP)		± 0

Enfim, o propósito deste capítulo é investigar a aplicabilidade das Redes Bayesianas Dinâmicas em detecção de ESVs. Para tanto, foi realizado uma comparação inicial com as Redes Bayesianas estáticas, onde foi possível observar que a RBD aumentou tanto o resultado classificatório apresentado pela Tabela de Contingência quanto os valores discriminativos de cada probabilidade resultante, este último possibilitando obter valores muito significativos quando se utiliza limiar, permitindo, desta maneira, que o especialista tenha uma interação com o sistema proposto, inserindo no mesmo o seu conhecimento e optando pelo grau de certeza desejado.

Portanto, a estrutura resultante para o sistema proposto nesta Tese, considerando o uso de RBD, fica sendo aquela da Figura 4.7. Sua operação se inicia com a segmentação do sinal ECG pelo algoritmo de Andreão (2004) (camada 0 idêntica àquela do sistema tratado no Capítulo 3). Após esta segmentação, são selecionados os registros realmente significativos para a detecção de ESV, aplicando a idéia de *Data Mart*. A partir desses registros, os dados que representam os valores das evidências utilizadas na rede são empregados como *nós* observáveis da rede. A partir deste momento, é utilizada a RBD, gerando os resultados de classificação (camada 1 do sistema proposto). Por fim, o grau de certeza que os resultados da classificação devem obedecer é definido pelo especialista, através da seleção de um limiar.

É possível encontrar alguns trabalhos que geram resultados significativos para detecção de ESV's, porém, realizar uma comparação é um pouco difícil, pois os resultados apresentados em outras bibliografias não possuem dados suficientes para uma comparação justa. Por exemplo, em Zhou et al (2003) os resultados são apresentados individualmente, isto é, é realizado o treinamento e o teste da Rede Neural no mesmo registro. Somente a nível de comparação, antes de implementar a RBD foi realizado um teste para verificação de como a RB estática reage quando o treinamento e o teste são realizados com o mesmo registro⁴, foi possível verificar que, quando se utiliza o treinamento e o teste do mesmo registro, os resultados são significativamente melhores. Tem-se, como exemplo, o registro 200, este aumentou de 78% para 92% o valor de 'Se' e 70% para 100% o 'VPP', o trabalho apresentado por Zhou et al (2003) para o registro 200 encontra uma '*accuracy*' de 96% e a '*accuracy*' encontrada pela RB estática é de 97%. Como já foi demonstrado, a RBD gera resultados mais significativos

⁴ É importante ressaltar que tanto no trabalho apresentado por Zhou et al (2003) quanto no resultado da RB estático, o conjunto de teste e treinamento são diferentes.

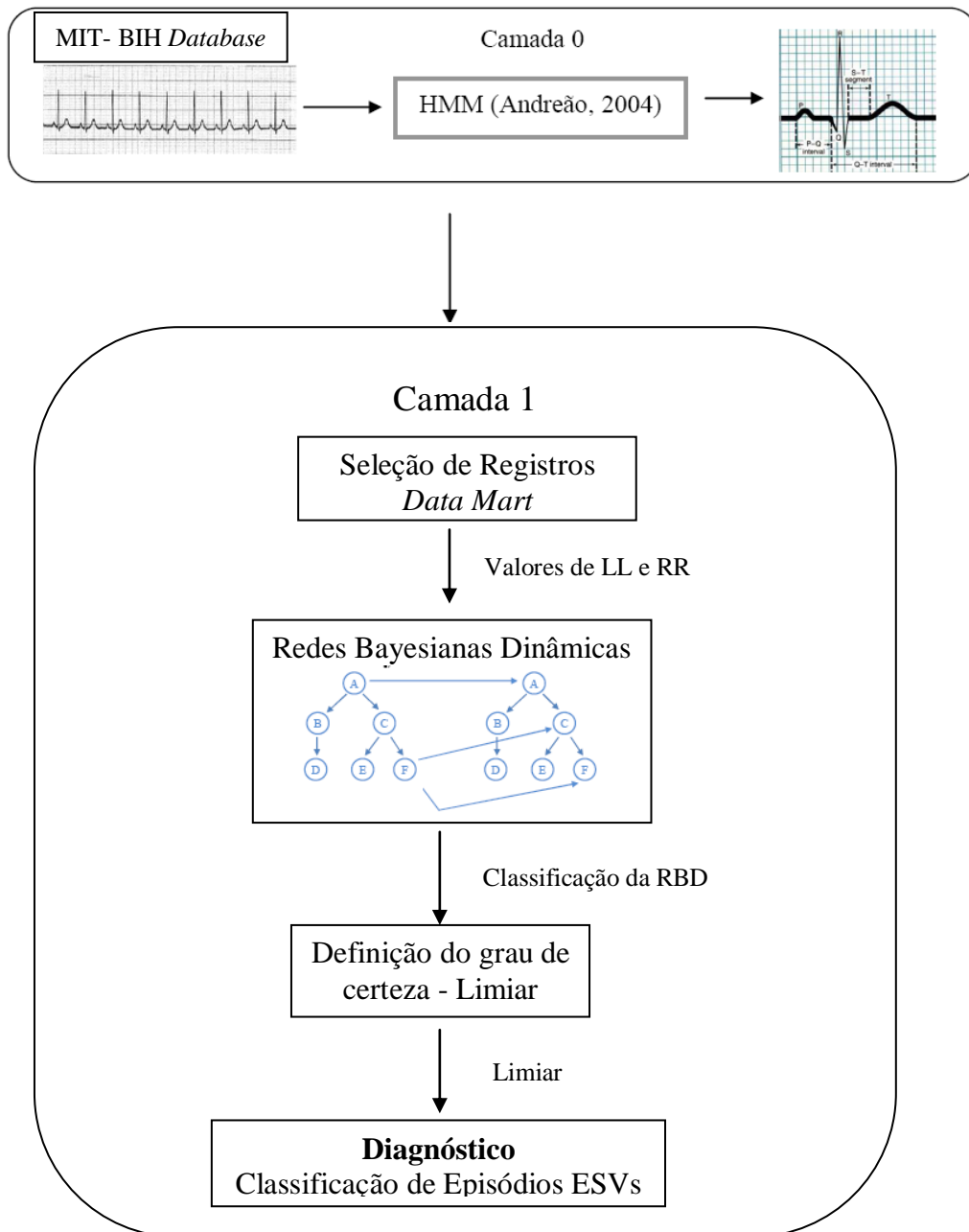


Figura 4.7: Arquitetura Final do Sistema Proposto na Tese.

que a RB estática, então se fosse aplicada a RBD somente para o registro 200, provavelmente encontraríamos melhores resultados.

Outro trabalho que classifica ESV's é apresentado por Nahar e Munir (2009) apresenta o valor de 'Se' e 'Esp' de 96,67% e 95,2%, porém este trabalho não apresenta 'VPP', informação que é muito relevante ao problema. Somente a nível e exemplificação, quando se realiza a Fusão de Resultados (ainda com a RB estática), em um dos vários teste que se realizou, foi separado para o conjunto de treinamento com 50% de ESV's e 50% de 'Outros' o valor encontrado para a 'Se' igual a 96%, e a 'Esp'

igual a 98%, porém o 'VPP' encontrado foi de 56%. É possível observar por este exemplo a necessidade de se apresentar 'VPP' e não somente a 'Se' e 'Esp', pois os falsos negativos também são importantes para distinguir um bom classificador.

E finalmente o trabalho de Zhao (2003), o qual apresenta um resultado de 'Se' de 93,8% e 'VPP' de 98,6%, porém com os dados disponíveis no artigo não é possível calcular o intervalo de confiança. Verifica-se que a RBD com definição de limiar possui valores de 'Se' e 'Esp' de 99,53% e 100% com intervalos de confiança de 0,4% e 0% respectivamente, mostrando, desta forma, uma melhor performance.

A partir destas comparações, é possível concluir que a RBD gerou resultados que até o momento não foram encontrados em outras bibliografias, demonstrando, desta forma, que a utilização desta metodologia é ideal para o tratamento de ESV's, bem como, uma ferramenta promissora para classificação de arritmias cardíacas.

Capítulo 5

Conclusão

Este trabalho teve como foco o estudo e desenvolvimento da metodologia probabilística baseada no Teorema de Bayes, isto é, Redes Bayesianas. O estudo de caso para verificação da aplicabilidade desta tecnologia é a detecção de Extrasístoles Ventriculares a partir de registros de ECG retirados da MIT-BIH *Arrhythmia Database*.

Foram realizados vários testes, inicialmente utilizando RB estáticas, com o objetivo de definir a topologia da rede, o número de canais (derivações de ECG), o algoritmo de segmentação a ser utilizado e a base de dados, entre outros aspectos. Durante o estudo da RB estática (Capítulo 3), chegou-se aos seguintes resultados:

1. a tecnologia de RB estática foi comparada com outras tecnologias, como classificação heurística e classificação empírica de tabelas de probabilidades, chegando-se à conclusão que a RB gera melhores resultados, em comparação com tais tecnologias;
2. com uma base de dados que possua mais episódios de ESV, a rede aprende melhor, podendo chegar a melhores resultados classificatórios. Este é um dos motivos pelos quais se optou por utilizar a base de dados MIT-BIH *Arrhythmia Database*, ao invés da base de dados QT *Database*, já que ela apresenta um maior número de episódios de ESV;
3. foram empregados dois algoritmos de segmentação, com a finalidade de observar se os resultados de classificação fornecidos pela rede são afetados quando se usa um outro algoritmo de segmentação do sinal de ECG. Como resultado, chegou-se à conclusão que a RB classifica bem episódios de ESV, ‘independentemente’ (isto é, há um equilíbrio no resultado da classificação) do algoritmo de segmentação utilizado, assim como não fica viciada nos resultados deste único algoritmo. Após esta análise, o algoritmo de segmentação escolhido para a continuação da Tese foi o baseado no trabalho de Andreão (2004);
4. utilizando Fusão de Canais, a RB gera melhores resultados classificatórios. Do mesmo modo que o especialista utiliza várias derivações para confirmar um diagnóstico, isto também pode ser aplicado à tecnologia de RB, fazendo com que

esta classifique mais precisamente os batimentos cardíacos. Porém, no prosseguimento dos trabalhos não foi utilizada a fusão de canais, pois optou-se pela utilização do algoritmo de segmentação desenvolvido por Andreão (2004) e a base de dados MIT-BIH *Arrhythmia Database*, escolhas estas que impossibilitaram a utilização do segundo canal, pois na base de dados escolhida este canal apresenta muito ruído, impedindo, desta forma, a boa utilização da Fusão de Canais. Todavia, é importante deixar claro que a Fusão de Canais melhora o desempenho da rede;

5. é possível concluir também, que a utilização da Onda P não influenciou de maneira significativa para o melhoramento das informações e, desta forma, fazer com que a rede classifique melhor os batimentos ESV. Porém é necessário se observar que, como discutido anteriormente, a base de dados utilizada é a MIT-BIH *Arrhythmia Database*, a qual possui poucos episódios de ESV sem onda P. De fato, aproximadamente 35% dos casos classificados por especialistas como ESV possuem onda P. Este percentual, por ser elevado, não contribui significativamente para melhorar a classificação da rede. Outra conclusão que se pode chegar é que apenas os valores da verossimilhança do complexo QRS (nó 'LL') e do tempo entre dois picos R sucessivos (nó 'RR') são suficientes para gerar uma classificação considerada adequada;
6. por fim, utilizando ainda a RB estática, foi proposta a Fusão de Resultados, abordagem esta que gerou os resultados mais significativos da rede estática. Como critério de desempate, foi empregada a onda P, que desta forma, gerou um diferencial no momento de classificação.

Então, a partir do estudo da RB estática, foi possível definir a topologia da rede, o algoritmo de segmentação e a base de dados a serem empregados em conexão com a Rede Bayesiana Dinâmica (Capítulo 4). Do mesmo modo, foi possível verificar que ela é uma ferramenta de classificação viável para realizar a classificação da arritmia cardíaca em questão. Após as definições da RB, foram realizados os testes utilizando a Rede Bayesiana Dinâmica.

Nos experimentos com a Fusão de Resultados, foi possível observar que seria interessante, antes de iniciar o desenvolvimento da RBD, trabalhar um pouco mais sobre a base de dados, pois a partir da FR é possível observar que ela, quando empregada de maneira mais objetiva, conduz a melhores resultados de classificação. Então, foram discutidas técnicas

de *Data Warehouse*, *Data Mart* e *Data Mining*, tecnologias que têm o foco no tratamento dos dados antes de se aplicar qualquer tipo de técnica de mineração de dados sobre os mesmos (no caso deste trabalho, as Redes Bayesianas). Assim, foi observado que quando se selecionam dados realmente significativos para um objetivo específico, o resultado da classificação melhora, isto é, quando se seleciona, a partir do *Data Warehouse*, subconjuntos, denominados *Data Marts*, estes últimos melhoram o desempenho da rede, como foi mostrado no Capítulo 4, ao ser implementada a RB estática utilizando este subconjunto de dados. Foi demonstrado, também, que esta seleção (registros realmente significativos a detecção de ESVs) não influencia de forma negativa no aprendizado da rede, já que para um acréscimo no desempenho, abordando várias arritmias, seria necessário trabalhar com outras variáveis, ou seja, trabalhar com outras evidências. Outro ponto que ratifica esta última afirmação é que vários registros (como, por exemplo, alguns que correspondem a pacientes com marcapasso) geram dados distorcidos, que mascaram informações relevantes à rede.

A partir destas observações foram implementadas várias topologias de RBD, das quais foi possível concluir que a RBD classifica melhor os batimentos ESV que a RB, afirmação esta que é ratificada por duas informações significativas:

1. o resultado de classificação da RBD é significativamente melhor que a RB;
2. o valor discriminante da probabilidade de cada batimento cardíaco gerado pela RBD é aproximadamente 15% mais preciso que a RB.

Principalmente a partir do item 2 descrito acima, foi possível gerar uma interação direta com o especialista, fazendo com que sejam tratadas tanto as incertezas aleatórias, propagadas pelo algoritmo de inferência da rede, quanto a incerteza visualizada a partir do valor de probabilidade resultante após a classificação de cada batimento cardíaco. Assim, o especialista pode definir o grau de certeza que ele deseja para a arritmia em questão. Para que isto fosse possível, foi definido um limiar, o qual produz resultados de acordo com o grau de certeza que se deseja alcançar.

Quando se define um limiar, é observado que alguns batimentos cardíacos não são classificados, isto é, eles são identificados como batimentos indefinidos. Isto ocorre porque a probabilidade de que se trate de um batimento de qualquer dos tipos considerados pela rede (ESV e 'Outros') não atinge um valor superior ao limiar estabelecido como mínimo para ter certeza. Porém, estes batimentos indefinidos correspondem a uma quantidade muito pequena, considerando-se o total de batimentos classificados pela rede. Assim sendo, o sistema automático de análise de ECG para identificar episódios de ESV aqui proposto pode arquivar

tais batimentos e encaminhá-los para análise de um especialista, que daria a palavra final sobre tais batimentos classificados como indefinidos.

Finalmente, pode-se afirmar que o resultado da RBD é muito bom, com mais de 99% de certeza nos valores de sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo, quando da utilização de limiar, e mais de 95% e 98% de certeza nos valores de sensibilidade e valor preditivo positivo, respectivamente. Estas afirmações podem ser ratificadas pelos respectivos intervalos de confiança. No primeiro caso, com utilização de limiar, os intervalos de confiança são de $\pm 0,4\%$, no caso da Tabela 4.10, o que significa que a certeza de um batimento ser realmente o que a rede classifica está compreendido no intervalo de 99,15% a 100%.

Como discutido durante todo o desenvolvimento do trabalho, é possível afirmar que as tecnologias de Redes Bayesianas (RB e RBD) são ferramentas adequadas para classificação de dados, como demonstrado no estudo de caso em questão (a classificação da arritmia cardíaca denominada ESV).

5.1 Perspectivas para a Continuidade do Trabalho

Este trabalho demonstrou a capacidade de aprendizado e classificação das RBs (RB e RBD), metodologias que possuem um campo muito amplo de possibilidades para desenvolvimento e aplicação. Para a realização de trabalhos futuros são sugeridos, basicamente, os seguintes itens:

1. melhorar o sinal do canal 2 da base de dados MIT-BIH *Arrhythmia Database*, ou então trabalhar com bases de dados com pelo menos dois canais com dados pouco ruidosos, para que se possa realizar a fusão de canais, e implementar uma RBD a partir desta topologia, o que não foi feito neste trabalho;
2. estender o estudo de caso para incluir outras arritmias cardíacas detectadas a partir de registros de ECG;
3. utilizar o algoritmo de inferência bayesiana para RBD denominado *particle filtering* (Russel e Norvig, 2000).

Referências Bibliográficas

Addrians, P. e Zantinge, (1996), D. *Data Mining*. Inglaterra: Addison-Wesley.

AHA *database*, American Heart Association ECG Database, *ECRI Institute*.
https://www.ecri.org/Products/Pages/AHA_ECG_DVD.aspx, visitada em 22/09/2008.

Andreão, R. V., Dorizzi B. and Boudy, J., (2006), 'ECG Signal Analysis Through Hidden Markov Models', *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, August, pp. 1541-1549.

Andreão R.V, Dorizzi B., Mota J.C.M. e Boudy J., (2004), "Ischemia Detection Based on HMM Beat Segmentation and ST Segment Analysis", *Proceedings IFMBE*, vol. 5, João Pessoa, Brasil.

Andreão R. V. (Nov 2004), *ECG Beat Segmentation through a Markovian Approach: Application to the Detection of Ischemic Episodes*, Ph. D. Thesis, Dept. EPH, National Telecommunications Institute, Evry, France, (escrita em Francês).

Aguiar, R. O. (2008). *Classificador Automático e Não-Supervisionado de Batimentos Cardíacos Baseado no Algoritmo Dynamic Time Warping*, Dissertação de Mestrado, Universidade Federal do Espírito Santo (UFES), Vitória.

Barreto, J. M. (1999), *Inteligência Artificial no Limiar do Século XXI*. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis/SC.

Belot D., (Nov 2002), *Fusion de données avec des réseaux bayésiens pour la modélisation des systèmes dynamiques et son application en telemedicine*. Ph. D. Thesis, École doctorale IAE+M Lorraine UFR STMIA. (escrita em Fancê).

- BNT, How to use the Bayes Net Toolbox (2007) viewed 20/04/2010.
<http://www.cs.ubc.ca/~murphyk/Software/BNT/usage.html>
- Buntine, W. L., (1991), Theory Refinement on Bayesian Networks. *Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence*, July 13-15, Morgan Kaufmann, Los Angeles, California, pp. 52-60.
- Carvalho, L. A.V. (2001), *DataMining: A Mineração de Dados no Marketinig, Medicina, Economia, Engenharia e Administração*, São Paulo: Érica. ISBN: 85-7194-766-X.
- Castillo E., Gutiérrez J. M. e Hadi A. S., (1997), *Expert Systems and Probabilistic Network Models*. Springer-Verlag, New York.
- Clarke A. B. e Disney R. L., (1970) Probability and Random Processes for Engineers and Scientists, *New York: John Wiley & Sons, Inc.*
- Crawford M. H. et al., (1999), "ACC/AHA Guidelines for Ambulatory Electrocardiography", *Journal of the American College of Cardiology*, vol. 34, n° 3, pp. 913-948.
- Charniak, E. (1991), "Bayesians Networks without Tears". *IA Magazine*.
- Cunha A.M. (2009), *Movimentos de Cabeça guiados pela Voz*, Tese de Doutorado, IMPA - Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro.
- Dallora M.E.L.V., Góes W.M., Rosa N.A., Carabolante F.J., Santos F.G. (2004), Proposta de Utilização de um Data Mart no Centro Cirúrgico do HCFMRP-USP. *IX Congresso Brasileiro de Informática em Saúde*, São Paulo.
- Darwiche A., (2001), Constant-Space Reasoning in Dynamic Bayesian Networks, *International Journal of Approximate Reasoning*, pp 161-178.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Royal Stat Soc*, 39:1–39.
- Dempster, A. P. (1968), A generalization of bayesian inference. *Journal of the Royal Statistical Society (Series B)*, 30:205–267.

- Dietterich, T. G., (1997). Machine Learning Research: Four Current Directions *AI Magazine*. 18 (4), 97-136.
- Dietterich, T. G. (to appear) (2002). Ensemble Learning. *In The Handbook of Brain Theory and Neural Networks, Second edition*, (M.A. Arbib, Ed.), Cambridge, MA: The MIT Press.
- Dubois, D., Prade, H. (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York.
- Duda, R. O. et al (2001), *Pattern Classification*, 2nd Edition. Wiley-Interscience.
- Farrugia S., Yee, H. and Nickolls, P. (1991), 'Neural Network Classification of Intracardiac ECGs', *Proceedings of the IEEE and INNS Int. Joint Conf. on Neural Networks*, 18-21 November, Singapore, pp. 1278-1283.
- Fayyad, U.M. et al. (1996), *Advances in Knowledge Discovery and Data Mining*. California: AAAI Press.
- Friedman N., Murphy K. e Russell S. J., (1998). Learning the structure of dynamic probabilistic networks. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 139-147.
- Gaag, L. C., (1996), *Bayesian Belief Networks: Odds and Ends*. The Computer Journal, vol 39, n° 2, p 97-113.
- Gallego D. G. (2010), *Desarrollo de un Entorno de Usuario para Aplicación de Rede Bayesianas Dinámicas a Problemas de Fusión de Información*. Proyecto Fin de Carrera Ingeniería Informática. Universidad Carlos III de Madrid Escuela Politécnica Superior.
- Gao, D., Madden, M., Chambres, D. e Lyons, G., (2005), Bayesian ANN Classifier for ECG Arrhythmia Diagnostic System: A Comparison Study, *Proceedings of the International Joint Conference on Neural Networks*, July 31 - August 4, Montreal, Canada, vol. 4, pp. 2383-2388.

- Gawande, A. (2002), *Complicações: Dilemas de um cirurgião diante de uma ciência imperfeita*. Objetiva: Rio de Janeiro.
- Gonçalves, M. A. B. (1995), *Noções Básicas de Eletrocardiograma e Arritmias*, Editora Senac, São Paulo.
- Hadjali, A., D. Dubois, et al. (2004). A possibility theory-based approach to the handling of uncertain relations between temporal points. *Proceedings - 11th International Symposium on Temporal Representation and Reasoning (TIME 2004)*, Jul 1-3 2004, Tatiou, France, Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States.
- Han, J., Kamber, M. (2000), *Data Mining Concepts and Techniques*. Morgan-Kaufmann.
- Hans-Jrgen Z., Elite Foundation, A. G. T. H. O. F. S., (1999), *Practical Applications of Fuzzy Technologies*. Kluwer Academic Publishers.
- Heat Disease Program (2010), HEART DISEASE PROGRAM. Clinical Decision Making Group MIT. Disponível em: <<http://groups.csail.mit.edu/medg/projects/hdp/>>. Acesso em: maio de 2010.
- Heckerman D., Geiger D. e Chickering D.M., (1995), *Learning Bayesian Networks: The Combination of Knowledge and Statistical Data*, Machine Learning, pp. 197-243.
- Hruschka Jr., Estevam R. e Silva, W. T., (1996), Propagação de Crença e Aprendizado em Redes Bayesianas. *Relatório de Pesquisa CIC/UnB – 03/96*, Departamento de Ciência da Computação – CIC, Universidade de Brasília. Brasília, [s.n].
- Huang, C.; Darwiche, A., (1994), Inference in Belief Networks: A Procedural Guide. *International Journal of Approximate Reasoning*, 11:1-158. *Elsevier Science Inc.* New York.
- Ibrahim, A. M. (2004), *Fuzzy Logic for Embedded Systems Applications*. Elsevier.

- Imhoff, C. Glemmo N., Geiger, J. G. (2003), *Mastering Data Warehouse Design: Relational and Dimensional Techniques*. Indianapolis: Wiley Publishing, Inc.
- Inmon, W. H. (1997), *Como construir o data warehouse*. Rio de Janeiro. Campus.
- Isken, M. W., Littig, S. J. and West (2001), M. A data mart for operations analysis. *J Healthc Inf Manag*, 15, 2, 143--153.
- Jensen, F. V. (2001), *Bayesian Networks and Decision Graphs* . 1. ed New York: Springer-Verlag.
- Jordan, M.I. (1999), *Learning in Graphical Models*. MIT Press. Visitado em maio de 2010: <http://books.google.com.br/books?hl=ptR&lr=&id=7f61BBKdJ4EC&oi=fnd&pg=P1&dq=%22junctiontree+%22&ots=PUyBL4NJjn&sig=jRLOu1cSDxKCMIjWsUdApAcMkw#v=onepage&q=%22junction-tree%20%22&f=false>.
- Kadish, A. H., Buxton, A. E., Kennedy, H. L., Knight, B. P, Mason, J. W., Schuger, C. D., Tracy, C. M., Winters, W. L., Boone, A. W., Elnicki, M., Hirshfeld, J. W., Lorell, B. H., Rodgers, G. (2001), "ACC/AHA Clinical Competence Statement on Electrocardiography and Ambulatory Electrocardiography", *JACC*, vol. 38, n° 7.
- Kalman, R. E. (1960), A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME - Journal of Basic Engineering*, Volume 82, 35-45.
- Karwowski, W.; A., M., (1986), *Appllications of Fuzzy Set Theory in Human Factors*. Elsevier Science Publishers.
- Klir, G. J.; Folger, T. A. (1988), *Fuzzy Sets, Uncertainty And Information*. Prentice Hal.
- Kuppuraj, R. N. (1993), 'A Neural Network System to Classify Simulated ECG Rhythms', *Proceedings of the IEEE Biomedical Engineering Conference*, New Orleans, Louisiana, USA.
- Ladeira, M. (2000), *Diagrama de Influências Múltiplo Secionado*. Tese de Obtenção do grau em Doutor em Ciência da Computação. Porto Alegre.

- Laguna, P., Mark, R. G., Goldberger, A. e Moody, G. B. (1997), A Database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. *Computers in Cardiology*, Lund, Sweden, vol. 24, pp. 673-676 (Piscataway, NJ: IEEE Computer Society Press). <http://www.physionet.org/physiobank/database/qtdb/doc/>. Acessado em 11/06/2010.
- Laudon, K. C. e Laudon, J. P. (2007). *Sistemas de informação gerenciais*. Pearson Prentice Hall, São Paulo.
- Lauritzen, S. L. e Spiegelhalter, D. J., (1988), Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion). *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 50, no. 2, pp. 157-224.
- Larson, R. e Farber, B. (2004),. *Estatística Aplicada*. Tradução de Cyro Patarra. 2.ed. São Paulo: Pearson prentice hall.
- Lim J.S. (2009), Finding Features for Real-Time Premature Ventricular Contraction Detection Using a Fuzzy Neural Network System, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 20, NO. 3, MARCH 2009.
- Lopes, P. A., (1999), *Probabilidades e Estatísticas*. Rio de Janeiro: Reichmann & Affonso Editores.
- Machado F. N. R. (2004), *Tecnologia e Projeto de Data Warehouse: uma visão multidimensional*. São Paulo: Érica. 318 p. ISBN 8536500123 (broch.)
- Maglaveras N., Stamkopoulos T., Diamantaras K., Pappas S C., Strintzis M., (1998), ECG pattern recognition and classification using non-linear transformations and neural networks: A review . *International Journal of Medical Informatics* , Volume 52, Issue 1 - 3, p 191 - 208 N.
- Maglaveras N. and Strintzis G. M., (July 1998), An Adaptive Backpropagation Neural Network for Real-Time Ischemia Episodes Detection: Development and Performance

- Analysis Usomg tje European ST-T Database. *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 7.
- Martins V. R., (2008), *Identificação Automática de Situações de Emergência através de técnicas de Fusão de Sinais Vitais e de Movimentos*. Dissertação de Mestrado, Vitória, UFES.
- MIT-BIH Harvard-MIT, (1997), ‘Arrhythmia DataBase Directory’, *Division of Health Sciences and Technology Biomedical Engineering Center*, viewed 11/09/2008, <http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm>.
- Mitchell T. (1997), *Machine Learning*. McGraw Hill.
- Moody, G. B., Mark, R. G. e Goldberger, A. L. (2001), PhysioNet: A webbased resource for the study of physiologic signals. *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, pp 70–75.
- Murphy, K. P. (2002), *Dynamic Bayesian Networks: Representation, Inference and Learning*. s.l. : University of California, Berkeley.
- Nahar, S.; ShahNoor bin Munir, M.; (2009), Automatic Detection of Premature Ventricular Contraction Beat Using Morphological Transformation and Cross-correlation. *Signal Processing and Communication Systems ICSIPA 2009*. Dept. of Electr. & Electron. Eng., United Int. Univ., Dhaka, Bangladesh ISBN: 978-1-4244-4473-1.
- Nassar, S. M., (2007), Tratamento de Incerteza: Sistemas Especialistas Probabilísticos, *Universidade Federal de Santa Catarina. Programa de Pós Graduação em Ciência da Computação*. Visitado em 10/09/2008. http://www.inf.ufsc.br/~silvia/disciplinas/sep/material_didatico/MaterialDidatico.pdf
- Netica TM Application, (2007), Introduction to Bayes Nets, *Norsys Software Corp*, <http://www.norsys.com/netica.html>, visitado em 10/09/2007.
- Nilsson, N. (1998). *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publishers.

- Nunes, E. M.P. (2005), *Fiabilidade de Sistemas com Processos Não-Markovianos e com Parâmetros Incertos*. Dissertação submetida à Faculdade de Engenharia da Universidade do Porto para a obtenção do grau de Doutor em Engenharia Electrotécnica e de Computadores. Portugal.
- Ono Y, Ishiyama A, Kasai N, Yamada S, On K, Watanabe S, Yamaguchi I, Miyashita T, Tsukada K., (Nov 2004), Bayesian classification of myocardial excitation abnormality using magnetocardiogram maps for mass screening. *Neurol Clin Neurophysiol.*, vol 43.
- Paskin, M. A. (2003), *A Short Course on Graphical Models*. The Junction Tree Algorithms.
- Pavlovic V., Garg A., Rehg J.M., e Huang T. S., (2000), Multimodal Speaker Detection Using Error Feedback Dynamic Bayesian Networks, *Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, June 2000, Hilton Head, SC, USA. *IEEE Computer Society*, ISBN 0-7695-0662-3.
- Pearl, J (1987), *Evidencial reasoning using stochastic simulation of causal models*. *Artificial Intelligence*, 32:247–25.
- Pearl, J. (1991), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Revised Second Printing. San Mateo, CA: Morgan Kaufmann.
- Porto C.P. (2001), *Semiologia médica*; Guanabara Koogan, 4º Ed.; pp. 422-522.
- Przytula K.W., Dash D., Thompson D., (2003), Evaluation of Bayesian Networks Used for Diagnostics. *IEEE Aerospace Conference*, pages 3177 – 3187.
- Przytula, K.W., Thompson, D., (2000), Construction of Bayesian: Networks for Diagnostics, *Aerospace Conference Proceedings IEEE*, vol.5, pp 193-200.
- QT database, (1997), The QT database, PhysioBank physiologic signal archives for biomedical research, *Computers in Cardiology*, vol. 24, pp. 673-676 (Piscataway, NJ: IEEE Computer Society Press), <http://www.physionet.org/physiobank/database/qtldb/>, viewed 11/09/2008.

- Rabiner, L. R. e Juang, B. H. (1986). An introduction to Hidden Markov Models, *IEEE ASSP Magazine* 3 (1), 4-16.
- Ramick D.C. (2001), "Data Warehousing in Disease Management Programs., *Journal of Healthcare Information Management*, Summer, v. 15, n. 2, p. 99-105.
- Rezende, S. O., (2003), *Sistemas Inteligentes: Fundamentos e Aplicações*. Monole.
- Russell S. J., Norvig P. (2000), *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995. 2nd edition draft copy, capítulo 17.
- Russell, S. J., Norvig, P. (2004), *Inteligência Artificial. Um enfoque moderno*, 2 ed. Rio de Janeiro: Elsevier.
- Sayadi O., Shamsollahi M. B., e Clifford G.D., (2010), Synthetic ECG generation and Bayesian filtering using a Gaussian wave-based dynamical model. IOP PUBLISHING PHYSIOLOGICAL MEASUREMENT. *Physiol. Meas.* **31** (2010) 1309–1329
- Shachter, R. D. (1986), Evaluating Influence Diagrams, *Operations Research*, vol. 34, no. 6, December, pp. 871-882.
- Strintzis, M., Stalidis, G., Magnisalis, X., Maglaveras N. (1992), "Use of Neural Networks for electrocardiogram feature extraction, recognition and classification.", *Neural Network World Journal*, vol. 3-4, 313-327.
- Stutz J., Taylor W., e Cheeseman P., (1998), AutoClass C, *General Information. NASA, Ames Research Center*.
- Taddei A., Biagini A., et al., (1991), "The European ST-T database: Development, distribution and use," in *Computers in Cardiology. IEEE Computer Society Press*, pp. 177-180.
- Theodoridis, S., Koutroumbas, K. (2006), *Pattern Recognition*, 3rd Edition. Academic Press.

- Toscani, L. V.; Veloso, P. A. S. (2001), *Complexidade de Algoritmos*. Porto Alegre: Instituto de Informática da UFRGS: Editora Sagra Luzzatto, 202p.
- Vieira, R.; Osório, F. S. (1999), Sistemas híbridos inteligentes. XIX Congresso S.B.C. *ENIA99 Encontro Nacional de Inteligência Artificial*, UNISINOS, Rio de Janeiro, [s.n.].
- Xu, W.; Zhu, Q. (Novembro, 2004), The Junction Tree Algorithm. *Curso CS281A/Stat241A: Statistical Learning Theory. Lecturer: Michael I. Jordan*.
- Young, J.D., e Santos, Jr., (April 1996), Introduction to Temporal Bayesian Networks, *The Seventh Midwest AI and Cognitive Science Conference*, pp 26-28.
- Yuan, B., Klir, G. J. (1995), Proceedings of the 3rd International Symposium on Uncertainty Modelling and Analysis, *IEEE Computer Society Publication*, College Park, March, pp 90-94.
- Wiggins, M., Saad, A., Litt, B. e Vachtsevanos, G., (2005), *Evolving a Bayesian classifier for ECG-based age classification in medical applications*, *Applied Soft Computing*, vol. 8, no. 1, Elsevier, pp. 599-608.
- Wilson, R. A.; C., K. F., (1999), *The MIT encyclopedia of the cognitive sciences*. The MIT Press, London.
- Zadeh, L. A., (1975), *Fuzzy Logic and Approximate Reasoning*. *Synthese*, 30: 407-428.
- Zadeh, L. A., (1995), Probability theory and fuzzy logic are complementary rather than competitive. *A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, [S.l.], v.37, Issue 3, p.271 – 276.
- Zhao L., Wiggins M., Vachtsevanos G., (2003). Premature Ventricular Contraction Beat Detection Based on Symbolic Dynamics Analysis. *IASTED International conference Circuits, Singnal and Systems*, May 19-21, Cancun, México.

Zhou J., (2003). Automatic Detection of Premature Ventricular Contraction Using Quantum Neural Networks. *IEEE Symposium on Bioinformatics and BioEngineering (BIBE'03)* 0-7695-1907-5/03. Department of Computer Science Northern Illinois University

Zipes, D. P., Camm, J. A., Borggrefe, M. et al., (2006), Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Develop Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death). *Journal of the American College of Cardiology*, Volume 48, Issue 5, 5 September, pp 1064-1108.

Anexo 1

Estimação por Máxima Verossimilhança (ML)

- Na construção do classificador, assumimos conhecida a verossimilhança $p(x|C_i)$, $i = 1 \dots M$ de como o conjunto de treinamento se distribuía.
- Supomos que essa verossimilhança assume uma forma paramétrica, dependente do vetor de parâmetros θ_i desconhecido e que desejamos estimar.
- Sejam $x_1 \dots x_N$, N observações com pdf $p(x, \theta)$.
- O conjunto $D = \{x_k\}$ é o conjunto de amostras com probabilidade conjunta $p(D, \theta)$.
- Os dados do conjunto D são independentes e identicamente distribuídos, de forma que:

$$p(D, \theta) = p(x_1, x_2, \dots, x_N) = \prod_{k=1}^N p(x_k, \theta)$$

- O estimador de máxima verossimilhança $\hat{\theta}_{ML}$ é aquele θ para qual a função de verossimilhança tem valor máximo:

$$\hat{\theta}_{ML} = \arg \max_{\theta} \prod_{k=1}^N p(x_k, \theta)$$

- Na otimização irrestrita, a condição necessária para obter o máximo é que o gradiente da função de verossimilhança em relação ao vetor θ seja nulo.

$$\frac{\partial \prod_{k=1}^N p(x_k, \theta)}{\partial \theta} = 0$$

- O logaritmo é crescente e monotônico, podemos definir a função log-verossimilhança como

$$\mathcal{L}(\theta) = \ln \prod_{k=1}^N p(x_k, \theta)$$

- Operando o gradiente do log:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \sum_{k=1}^N \frac{\partial \ln p(x_k, \theta)}{\partial \theta} = \sum_{k=1}^N \frac{1}{p(x_k, \theta)} \frac{\partial p(x_k, \theta)}{\partial \theta} = 0$$

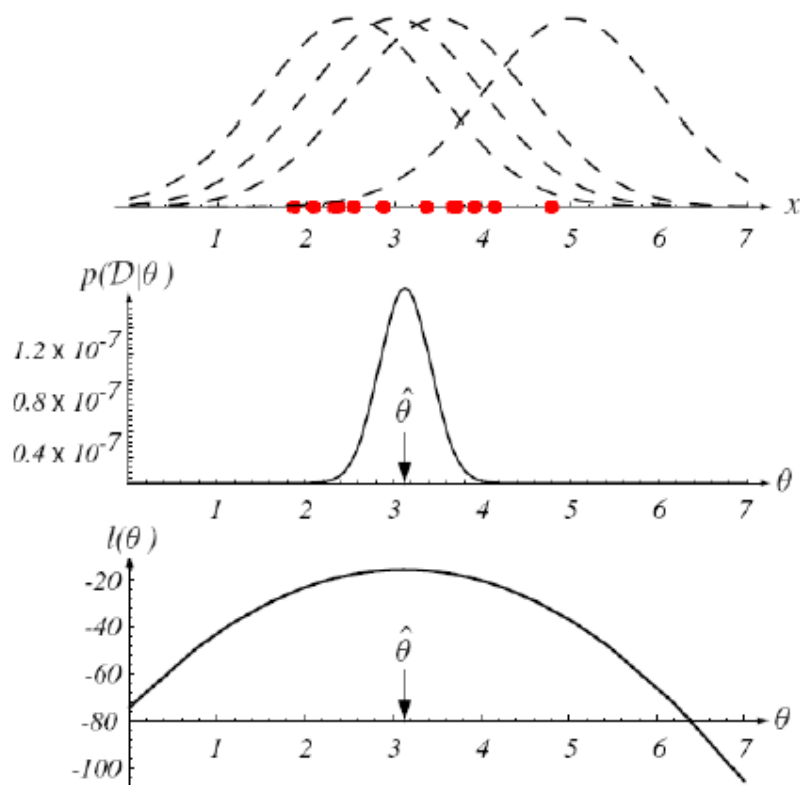


Figura 1:

The top graph shows several training points in one dimension, known or assumed to be drawn from a Gaussian of a particular variance, but unknown mean. Four of the infinite number of candidate source distributions are shown in dashed lines. The middle figure shows the likelihood $p(\mathcal{D}|\theta)$ as a function of the mean. If we had a very large number of training points, this likelihood would be very narrow. The value that maximizes the likelihood is marked $\hat{\theta}$; it also maximizes the logarithm of the likelihood—that is, the log-likelihood $l(\theta)$, shown at the bottom. Note that even though they look similar, the likelihood $p(\mathcal{D}|\theta)$ is shown as a function of θ whereas the conditional density $p(x|\theta)$ is shown as a function of x . Furthermore, as a function of θ , the likelihood $p(\mathcal{D}|\theta)$ is not a probability density function and its area has no significance. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Propriedade da ML

- Se θ_0 é o valor real, o estimador ML é assintoticamente não-viezado, isto é, para amostras grandes ($n \rightarrow \infty$) em média converge para o valor real.

$$\lim_{n \rightarrow \infty} \mathcal{E}[\hat{\theta}_{ML}] = \theta_0$$

- Em geral, o estimador ML pode ser viesado para amostras pequenas (lembrar do σ^2).
- Porém, é assintoticamente consistente: a probabilidade de estar arbitrariamente próximo do valor correto é alta quanto maior for N .

$$\lim_{n \rightarrow \infty} Pr[|\hat{\theta}_{ML} - \theta_0| < \varepsilon] = 1$$

- O estimador ML converge na média dos quadrados, isto é, para amostras grandes, a variância do estimador ML tende a zero.

$$\lim_{n \rightarrow \infty} \mathcal{E}[\|\hat{\theta}_{ML} - \theta_0\|^2] = 0$$

- O estimador de ML tende a ser o estimador de mínima variância e é aproximadamente não-tendencioso.
- A pdf do estimador ML tende a uma gaussiana.
- Princípio da invariância: Se $\hat{\theta}$ é um estimador de máxima verossimilhança para o vetor de estatísticas θ , o estimador de máxima verossimilhança das estatísticas $h(\theta)$ é $h(\hat{\theta})$ qualquer que seja h .

Exemplo

Caso Gaussiano com μ desconhecido.

Supomos Σ conhecido.

$$\ln p(x_k|\mu) = -\frac{1}{2} \ln [(2\pi)^d |\Sigma|] - \frac{1}{2} (x_k - \mu)^T \Sigma^{-1} (x_k - \mu)$$

$$\nabla_{\mu} \ln p(x_k|\mu) = \Sigma^{-1} (x_k - \mu)$$

$$\sum_{k=1}^n \Sigma^{-1} (x_k - \mu) = 0$$

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^n x_k$$

O estimador ML da média da distribuição é a média amostral.

Exemplo

Caso Gaussiano com μ e Σ desconhecidos.

Consideramos o caso unidimensional com $\theta_1 = \mu$ e $\theta_2 = \sigma^2$

$$\ln p(x_k|\theta) = \frac{-1}{2} \ln 2\pi\theta_2 - \frac{1}{2\theta_2} (x_k - \theta_1)^2$$

Com derivadas

$$\nabla_{\theta} \mathcal{L} = \nabla_{\theta} \ln p(x_k|\theta) = \begin{bmatrix} \frac{1}{\theta_2} (x_k - \theta_1) \\ -\frac{1}{2\theta_2} + \frac{(x_k - \theta_1)^2}{2\theta_2^2} \end{bmatrix}$$

Igualando a zero,

$$\sum_{k=1}^n \frac{1}{\hat{\theta}_2} (x_k - \hat{\theta}_1) = 0$$
$$- \sum_{k=1}^n \frac{1}{\hat{\theta}_2} + \sum_{k=1}^n \frac{(x_k - \hat{\theta}_1)^2}{\hat{\theta}_2} = 0$$

Assim,

$$\theta_1 = \hat{\mu} = \frac{1}{n} \sum_{k=1}^n x_k$$
$$\theta_2 = \hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \hat{\mu})^2$$

No caso multivariado:

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k$$
$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^n (x_k - \hat{\mu})(x_k - \hat{\mu})^T$$

Notar que o estimador ML para a variância é tendencioso (viesado).

$$\mathcal{E} \left[\frac{1}{n} \sum_{i=1}^n (x_k - \bar{x})^2 \right] = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

Um estimador não tendencioso seria:

$$C = \frac{1}{n-1} \sum_{k=1}^n (x_k - \hat{\mu})(x_k - \hat{\mu})^T$$

Referências

1. Duda, R. O. et al. Pattern Classification, 2nd Edition. Wiley-Interscience 2001;
2. Theodoridis, S., Koutroumbas, K. Pattern Recognition, 3rd Edition. Academic Press, 2006;
3. Han, J., Kamber, M. Data Mining Concepts and Techniques. Morgan-Kaufmann, 2000.

Anexo 2

Processo de atualização de uma RBD

Pseudocódigo:

1. Deslocamento: mover a janela
2. Predição:
 - a. Calcula-se a probabilidade de $P(X_{t-1}|E_{\{1,t-1\}})$, que é a probabilidade a posteriori sobre X_{t-1} ,
 - b. Estima-se a predição de $P(X_t|E_{\{1,t-1\}})$.
3. Desenrolar:
 - a. Eliminar o corte do tempo para $t-1$,
 - b. Utilizar as predições para o segmento de t , com a nova configuração prévia por parte de $P(X)$ para $P(X_t|E_{\{1,t-1\}})$.
4. Estimação:
 - a. Adicionar a nova estimacão E_t ,
 - b. Calcular a probabilidade de $P(X_t|E_{\{1,t\}})$, que é a probabilidade a posteriori do estado atual,
 - c. Adicionar o corte de tempo para $t+1$.

Anexo 3

- Intervalos de Confiança

A principal restrição da estimação pontual é que quando estimamos um parâmetro através de um único valor numérico toda a informação presente nos dados é resumida através deste número. É importante encontrar também um intervalo de valores plausíveis para o parâmetro. A idéia é construir um intervalo em torno da estimativa pontual de modo que ele tenha uma probabilidade conhecida de conter o verdadeiro valor do parâmetro.

O nível de confiança, c , é a probabilidade de que a estimativa intervalar contenha o parâmetro populacional em questão. Quando o tamanho amostral é superior a 30 podemos aproximar a distribuição amostral da média populacional, \bar{x} , por uma distribuição normal (pelo teorema do limite central), $N(\mu, \sigma^2)$.

Para fazer inferências sobre μ nos baseamos na média amostral \bar{x} e sabemos que

$$U = \frac{\sqrt{n} \cdot \bar{x} - \mu}{\sigma} \sim N(0,1) \quad (1)$$

A estatística U é uma função da amostra e também de μ , o parâmetro de interesse, mas sua distribuição de probabilidades não depende de μ . Usando uma tabela da distribuição normal padronizada podemos obter o valor do percentil $z_{\alpha/2}$ tal que

$$P[-z_{\alpha/2} \leq U \leq +z_{\alpha/2}] = 1 - \alpha \quad (2)$$

Assim, após isolar μ , obtemos

$$P\left[\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha \quad (3)$$

O parâmetro μ é desconhecido, mas fixo e, portanto, não é passível de descrição probabilística, ou seja, não se trata de um intervalo de probabilidade para μ . Na verdade são os limites do intervalo que são variáveis aleatórias e após a amostra ser observada dizemos que

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (4)$$

Com confiança de $100 \cdot 1 - \alpha$ % .

Para construir um intervalo de confiança de 95% o valor do parâmetro $z_{\alpha/2}$ da distribuição normal é $z_{\alpha/2} = 1,96$.

- Intervalos de Confiança de uma Proporção

A estimativa pontual de sucessos, \hat{p} , é dada pela proporção de sucessos em uma amostra e \hat{q} é a proporção de fracassos. Se $n \cdot \hat{p} \geq 5$ e $n \cdot \hat{q} \geq 5$, a distribuição amostral de \hat{p} pode ser considerada normal.

$$\hat{p} = \frac{X}{n} \quad \text{e} \quad \hat{q} = 1 - \hat{p} \quad (5)$$

O erro máximo da estimativa, E , para um intervalo de confiança, $100 \cdot 1 - \alpha$ %, é:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \quad (6)$$

O intervalo de confiança de 95% para uma proporção será: $\hat{p} - E \leq p \leq \hat{p} + E$, e o valor do parâmetro $z_{\alpha/2}$ da distribuição normal é $z_{\alpha/2} = 1,96$.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)