UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

THIAGO DE AGUIAR CALOTI

IDENTIFICAÇÃO DE PACIENTES COM DIABETES BASEADA NA VARIABILIDADE DA FREQUÊNCIA CARDÍACA

VITÓRIA
2013

THIAGO DE AGUIAR CALOTI

IDENTIFICAÇÃO DE PACIENTES COM DIABETES BASEADA NA VARIABILIDADE DA FREQUÊNCIA CARDÍACA

Dissertação 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 Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Rodrigo Varejão Andreão.

Co-orientador: Prof. Dr. Mário Sarcinelli Filho.

VITÓRIA
2013

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

Caloti, Thiago de Aguiar, 1988 -

C165i Identificação de pacientes com diabetes baseada na variabilidade da frequência cardíaca / Thiago de Aguiar Caloti. - 2013.

154 f.: il.

Orientador: Rodrigo Varejão Andreão.

Coorientador: Mario Sarcinelli Filho.

Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Variabilidade do batimento cardíaco. 2. Diabetes. 3. Processamento de sinais - Técnicas digitais. 4. Máquina de suporte vetorial. I. Andreão, Rodrigo Varejão. II. Sarcinelli Filho, Mário. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

THIAGO DE AGUIAR CALOTI

IDENTIFICAÇÃO DE PACIENTES COM DIABETES BASEADA NA VARIABILIDADE DA FREQUÊNCIA CARDÍACA

Dissertação 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 Mestre em Engenharia Elétrica.

Aprovada em 13 de maio de 2013.

COMISSÃO EXAMINADORA

Prof. Dr. Rodrigo Varejão Andreão
Instituo Federal de Educação, Ciência e Tecnologia do Espírito Santo
Orientador

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

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

Prof. Dr. Eduardo Miranda Dantas
Universidade Federal do Vale do São Francisco

Dedico este trabalho a Deus, à Ivone de Aguiar Caloti e João Amancio Caloti.

Agradecimentos

Agradeço a Deus por ter me dado força e capacidade para desenvolver este trabalho, pois até aqui o Senhor me ajudou. E, expresso meus sinceros agradecimentos

- A minha família, que me apoiou durante as etapas desta jornada. Em especial, minha mãe *Ivone de Aguiar Caloti* e meu pai *João Amâncio Caloti*;
- Ao meu orientador e professor Rodrigo Varejão Andreão, pela oportunidade de desenvolver este trabalho, por toda ajuda e conhecimento transmitidos;
- Aos professores do Programa de Pós-Graduação em Engenharia Elétrica da UFES dispostos a ajudar, em especial o professor Dr. Evandro Ottoni Teatini Salles;
- Aos amigos Kaio, Hugo, Thiago, Guilherme e Gabriel que são parte inerente de muitos momentos fundamentais de minha história e são pessoas de valor inestimável;
- A todos os amigos da IFES/UFES, que me apoiaram e pelos momentos de descontração durante os períodos de tempo livres;
- A todos que de alguma forma contribuíram para a concretização desta Dissertação.

Sumário

Lista de Tabelas

Lista de Figuras

Resumo

O diabetes *mellitus* (DM), usualmente referido como diabetes, é uma doença crônica caracterizada por hiperglicemia e leva a complicações específicas a longo prazo: retinopatia, neuropatia, nefropatia e cardiopatia. A análise da variabilidade da frequência cardíaca (VFC), sendo uma ferramenta não invasiva, tornou-se um método amplamente empregado em pesquisas para avaliar a atividade do sistema nervoso autônomo (SNA). A frequência cardíaca (FC) é sinal biológico que está em constante mudança. Essas mudanças podem ser um indício de doença ou servir como um indicativo de iminentes doenças cardiovasculares.

Neste trabalho, analisam-se sinais de VFC de 360 indivíduos saudáveis e 360 indivíduos diabéticos, usando métodos no domínio do tempo, no domínio da frequência e técnicas não lineares. Os resultados mostram que os índices no domínio do tempo (RR_{mean}, SDNN, RMSSD, pNN50 e Δ *index*), no domínio da frequência (VLF, LF, HF, HF_{norm} e LF/HF) e os índices não lineares (ApEn, SampEn, SD1, SD2, s, α_1, FD, REC, DET, L_{mean}, L_{max} e ShanEn) são clinicamente significativos na identificação de pacientes com diabetes. O sistema de diagnóstico proposto classifica indivíduos saudáveis e com DM, com acurácia de 75.69%, especificidade de 80.56% e sensibilidade de 70.83%.

Abstract

Diabetes mellitus (DM), usually referred to as diabetes, is a chronic disease characterized by hyperglycaemia and leads to specific long-term complications: retinopathy, neuropathy, nephropathy and cardiomyopathy. Analysis of heart rate variation (HRV), being a noninvasive tool, has become a popular method to assess the activitie of the autonomic nervous system (ANS). Heart rate (HR) are bio-signals that are in constantly changing. These changes may be an indication of current disease or serve as a pre-warning to imminent cardiovascular diseases.

In this work, we analyse HRV signals from 360 normal and 360 diabetic subjects, using time domain, frequency domain and nonlinear techniques. Our results show that the indexes in the time domain (RR_{mean}, SDNN, RMSSD, pNN50 and Δ *index*), in the frequency domain (VLF, LF, HF, HF_{norm} and LF/HF) and the nonlinear indexes (ApEn, SampEn, SD1, SD2, s, α_1, FD, REC, DET, L_{mean}, L_{max} and ShanEn) are clinically meaningful in the identification of patients with diabetes. The proposed diagnostic system classifies, DM patients and normal subjects, with an accuracy of 75.69%, specificity of 80.56% and sensitivity of 70.83%.

Capítulo 1

Introdução

Diabetes *mellitus* (DM) é uma doença crônica caracterizada por hiperglicemia (Association, 2011). A Organização Mundial da Saúde (OMS) reconhece três formas principais de diabetes: tipo I, tipo II e diabetes gestacional, as quais possuem sinais, sintomas e consequências similares, mas diferentes causas e diferentes distribuições na população. Atualmente, DM representa um grave problema de saúde pública pela alta prevalência no mundo, em especial nos países em desenvolvimento, pela morbidade e por ser um dos principais fatores de risco cardiovascular e cerebrovascular.

A história natural do diabetes é marcada pelo aparecimento de complicações crônicas, geralmente classificadas como, microvasculares: retinopatia (Harman-Boehm et al., 2007), (Faust et al., 2010), neuropatia (Ewing e Clarke, 1982), nefropatia (Guthrie, 1998) e cardiomiopatia (Grundy et al., 1999), e macrovasculares: doença arterial coronariana, doença cerebrovascular e vascular periférica. Todas as complicações são responsáveis por: expressiva morbimortalidade, com taxas de mortalidade cardiovascular e renal, cegueira, amputação de membros e perda de função e qualidade de vida superior a indivíduos sem diabetes.

1.1 Motivação

O diabetes *mellitus* tipo 2 (DM2) é considerado uma das grandes epidemias mundiais do século XXI e problema de saúde pública, tanto nos países desenvolvidos como em desenvolvimento. As crescentes incidência e prevalência são atribuídas ao envelhecimento populaci-

onal e ao estilo de vida atual, caracterizado por inatividade física e hábitos alimentares que predispõem ao acúmulo de gordura corporal.

A maior sobrevida de indivíduos diabéticos aumenta as chances de desenvolvimento das complicações crônicas da doença que estão associadas ao tempo de exposição à hiperglicemia. Tais complicações podem ser muito debilitantes ao indivíduo e são muito onerosas ao sistema de saúde. Neste contexto,

- a doença cardiovascular é a primeira causa de mortalidade de indivíduos com DM2;
- a retinopatia representa a principal causa de cegueira adquirida;
- a nefropatia é uma das maiores responsáveis pelo ingresso a programas de diálise e transplante;
- o pé diabético se constitui em importante causa de amputações de membros inferiores.

Como consequências das complicações da doença podem-se citar procedimentos diagnósticos e terapêuticos (cateterismo, bypass coronariano, fotocoagulação retiniana, transplante renal e outros), hospitalizações, absenteísmo, invalidez e morte prematura, as quais elevam substancialmente os custos diretos e indiretos da assistência à saúde da população diabética. Este quadro evidencia a viabilidade da prevenção, tanto da doença como de suas complicações crônicas.

O número de indivíduos com DM dá uma ideia da magnitude do problema, e estimativas têm sido publicadas para diferentes regiões do mundo, incluindo o Brasil. Em termos mundiais, 135 milhões apresentavam a doença em 1995, 240 milhões em 2005, e há projeção para atingir 366 milhões em 2030, sendo que dois terços habitarão países em desenvolvimento (WHO, 2011; Wild et al., 2004; Barceló et al., 2003).

Dados sobre prevalência de DM representativos da população residente em 9 capitais de estados brasileiros datam do final da década de 80 (Malerbi e Franco, 1992). Nesta época, estimou-se que, em média, $7,6\%$ dos brasileiros entre 30 e 69 anos de idade apresentavam DM, que incidia igualmente nos dois sexos, mas que aumentava com a idade e a adiposidade corporal. As maiores taxas foram observadas em cidades como São Paulo e Porto Alegre, sugerindo o papel da urbanização e industrialização na patogênese do DM2. Um achado interessante foi que 46% dos indivíduos (cerca da metade destes) diagnosticados diabéticos desconhecia sua condição.

De acordo com o VIGITEL 2007 (Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico), a ocorrência média de diabetes no Brasil na população adulta (acima de 18 anos) é de $5,2\%$, mas a prevalência do diabetes atinge $18,6\%$ da

população com idade superior a 65 anos, sem diferença entre os sexos. Em 2008, a prevalência observada entre idosos na mesma faixa etária foi de $20,7\%$ (BVS, 2002).

Segundo dados do Ministério da Saúde, no Brasil existem cerca de 11 milhões de portadores. Aproximadamente $6,3\%$ da população brasileira com 18 anos ou mais possuem diabetes, o equivalente a $8,3$ milhões de pessoas. Deste grupo, 3 milhões de brasileiros não sabem que têm a doença (BVS, 2002).

Apesar de o diabetes estar aumentando, poucos estudos têm investigado a epidemiologia do diabetes em nível populacional, e uma parcela elevada dos pacientes acometidos apresenta desconhecimento sobre o diagnóstico da doença, e de suas consequências. Isso significa que os serviços de saúde têm diagnosticado casos de DM tardiamente, dificultando o sucesso do tratamento, em termos de prevenção das complicações crônicas.

1.2 Justificativa

No ambiente clínico, o diagnóstico de diabetes usualmente é baseado na glicemia de jejum (GJ), no teste oral de tolerância à glicose (TOTG) e na hemoglobina glicada (A1C). Estudos recentes vêm investigando alterações da variabilidade da frequência cardíaca (VFC) em portadores de diabetes *mellitus* (DM) (Faust et al., 2011). Não existem estudos, entretanto, nos quais a VFC pudesse funcionar como preditor de diabetes em indivíduos com glicemia ainda normal ou com TOTG já alterado. Em outras palavras, uma questão importante nessa área seria se as alterações da VFC precedem o desenvolvimento da doença, clinicamente manifesta, em indivíduos ainda normoglicêmicos.

A análise da VFC, por meio da série de intervalos entre ondas R de batimentos consecutivos no eletrocardiograma (intervalos RR), tem encontrado grande aplicabilidade clínica na avaliação da modulação simpática e parassimpática dos batimentos cardíacos, e tem sido usada como um preditor de riscos de mortalidade e de outros eventos em pacientes com distúrbios metabólicos e cardiovasculares, tais como diabetes e infarto do miocárdio (Faust et al., 2011; Lombardi, 2000; Singh et al., 2000).

Há indícios que baixa VFC esteja associada ao aumento da morbidade cardiovascular em indivíduos com hiperglicemia (Singh et al., 2000). Assim, a detecção precoce de disfunção autonômica em indivíduos diabéticos é importante para a estratificação e gestão de risco de morbidade cardiovascular e mortalidade. No caso, a detecção precoce de disfunção autonômica permite a gestão por meio de intervenções farmacológicas e/ou alterações do estilo de vida dos indivíduos.

1.3 Objetivo

O objetivo geral desta dissertação é desenvolver um sistema automatizado de diagnóstico de diabetes baseado na variabilidade da frequência cardíaca. Uma etapa fundamental do sistema consiste na extração de características da série de intervalos RR, facilitando a tarefa do classificador na discriminação entre indivíduos saudáveis e diabéticos. Para tal, métodos lineares, no domínio do tempo e da frequência, e métodos não lineares são avaliados e empregados para extração de características (Task Force, 1996; Acharya et al., 2004). Após esta etapa, avaliam-se os efeitos do diabetes nestes índices (características) visando identificar aqueles que discriminam entre a existência ou não da doença. Na sequência, diferentes classificadores são treinados e testados a partir de uma base composta por registros de séries RR de indivíduos saudáveis e diabéticos.

Como objetivos específicos, podem ser citados:

- filtrar as séries de intervalos RR contidas na base de dados do projeto ELSA;
- analisar a estacionariedade das séries RR;
- estudar conceitos da literatura de dinâmica não linear e teoria de caos;
- implementar índices de variabilidade da frequência cardíaca (VFC);
- analisar a discriminância estatística entre grupos de indivíduos (normais e diabéticos) por meio dos índices de VFC;
- desenvolver classificador para identificar indivíduos diabéticos;
- aplicar modelo de classificador na análise de outros grupos de indivíduos.

1.4 Metodologia Proposta

Realizar um estudo exploratório sobre a base de dados do projeto ELSA, com o intuito de testar a adequação de um sistema capaz de separar indivíduos diabéticos e não diabéticos com base na glicemia de jejum (GJ). Inicialmente serão incluídos apenas indivíduos sem uso de medicação para diabetes. A referência em relação ao diagnóstico prévio de diabetes será obtida a partir da história médica pregressa (HMPA04). O grupo controle será definido pelo relato de ausência de diagnóstico prévio de diabetes e presença de GJ menor que 100 mg/dL enquanto que o grupo "diabético" será definido pelo relato da presença da doença na história médica pregressa e por GJ superior a 140 mg/dL. Caso o algoritmo desenvolvido mostre

desempenho satisfatório (adotando como critério subjetivo a discriminação entre grupos superior a 75%), pretende-se realizar os seguintes testes:

1. Testar o modelo do classificador em um terceiro grupo de participantes, sob uso de medicação antidiabética e com glicemia normal (≤ 125 mg/dL);
2. Verificar a eficiência de discriminação de diabetes em indivíduos com GJ entre 126 e 139 mg/dL e entre 100 e 125 mg/dL;
3. Confrontar o diagnóstico automatizado pelo algoritmo com outro exame laboratorial também utilizado para diagnóstico de diabetes: glicemia pós-prandial (> 200 mg/dL). Para tal, serão gerados índices lineares, nos domínios tempo/frequência, e não lineares, conforme apresentado no Capítulo 3.

A comparação estatística entre os grupos será feita via teste-t e análise do p-valor para cada índice. Os índices estatisticamente significantes serão a entrada do classificador (k vizinhos mais próximos, *naive* Bayes, redes neurais artificiais e máquinas de vetor suporte), cujo treinamento utilizará um conjunto selecionado aleatoriamente (80% dos participantes de cada grupo) para ajustar as características de sua estrutura. O desempenho deste será avaliado por medidas de sensibilidade, especificidade, e valores preditivos positivo e negativo.

1.5 Estrutura da Dissertação

Na sequência, o Capítulo 2 apresenta conceitos básicos sobre o coração, descreve a série de intervalos RR, explica o diabetes *mellitus*, juntamente com a sua epidemiologia e classificação, bem como os métodos e critérios para o seu diagnóstico. No Capítulo 3 são descritos os índices de VFC lineares e não lineares calculados sobre séries de intervalos RR. No Capítulo 4, os índices são avaliados quanto à capacidade de discriminação dos dois grupos de indivíduos através de teste estatístico e medidas de correlação. Os resultados de classificação são apresentados juntamente com as demais análises e resultados. Finalmente, no Capítulo 5 são apresentadas as conclusões e perspectivas de trabalhos futuros.

Capítulo 2

Variabilidade da Frequência Cardíaca e Diabetes

Este capítulo explica a importância do coração para o corpo humano, juntamente com alguns detalhes do seu funcionamento, do ponto de vista fisiológico e elétrico. Em seguida descreve como é gerada a série de intervalos RR abordando detalhes da variabilidade da frequência cardíaca e sua relação com o controle metabólico. Explica-se o diabetes *mellitus*, juntamente com a sua epidemiologia e classificação, concluindo com os métodos e critérios para o seu diagnóstico.

2.1 O Eletrocardiograma (ECG)

O coração é semelhante a uma bomba. Para que ele possa bombear o sangue através da contração e do relaxamento, é necessário que as células miocárdicas sejam inicialmente ativadas por um estímulo elétrico que atua sobre a membrana celular. Este estímulo elétrico, espontâneo e contínuo a um ritmo específico, é gerado no nodo sinusal, que é a estrutura cardíaca mais excitável e a que possui a maior capacidade de automatismo, por isso é denominado marca-passo natural do coração (Burton, 1977).

Em situação de repouso ou de inatividade, a membrana celular de todas as células do coração encontra-se eletricamente polarizada, isto é, possui um potencial elétrico negativo de -60 mV a -80 mV no caso do tecido excito-condutor, e de -90 mV no caso do miocárdio comum, o que significa dizer que o interior da célula é negativo em relação ao seu exterior. Este potencial elétrico de repouso é chamado potencial de membrana, ou potencial de repouso, e está associado à maior permeabilidade de íons potássio na célula, e maior acúmulo de íons sódio, cloreto e cálcio fora da célula (Guyton e Hall, 2002).

Devido a propriedades eletrofisiológicas da membrana celular das células do nodo sinusal e das demais estruturas do tecido condutor, o potencial de repouso automaticamente se inverte, recuperando-se alguns milissegundos depois, de maneira cíclica e ritmada. Este processo de despolarização da membrana celular é representado por um novo potencial elétrico através das células, chamado potencial de ação (Figura 2.1), que agora é positivo em relação ao exterior da célula. Nestas células, a inversão do potencial elétrico, que gera o potencial de ação, resulta da entrada intracelular de íons sódio e, principalmente, de cálcio.

Em alguns casos, a membrana excitável não se repolariza imediatamente após a despolarização, mas, pelo contrário, o potencial permanece em *plateau* ou fase onde o período da despolarização é prolongado, com valor próximo ao do potencial em ponta, durante muitos milissegundos antes do começo da repolarização. A causa do *plateau* é uma combinação de diversos fatores distintos. O principal é o processo de ativação lenta associada à difusão de íons cálcio. A recuperação do potencial de repouso, ou repolarização, se faz pela progressiva atenuação do potencial de ação, resultado da saída de íons de potássio e cloro para o exterior das células. Estes movimentos iônicos através da membrana celular decorrem do gradiente elétrico existente e da diferença de concentração dos íons em cada lado da membrana (Guyton e Hall, 2002).

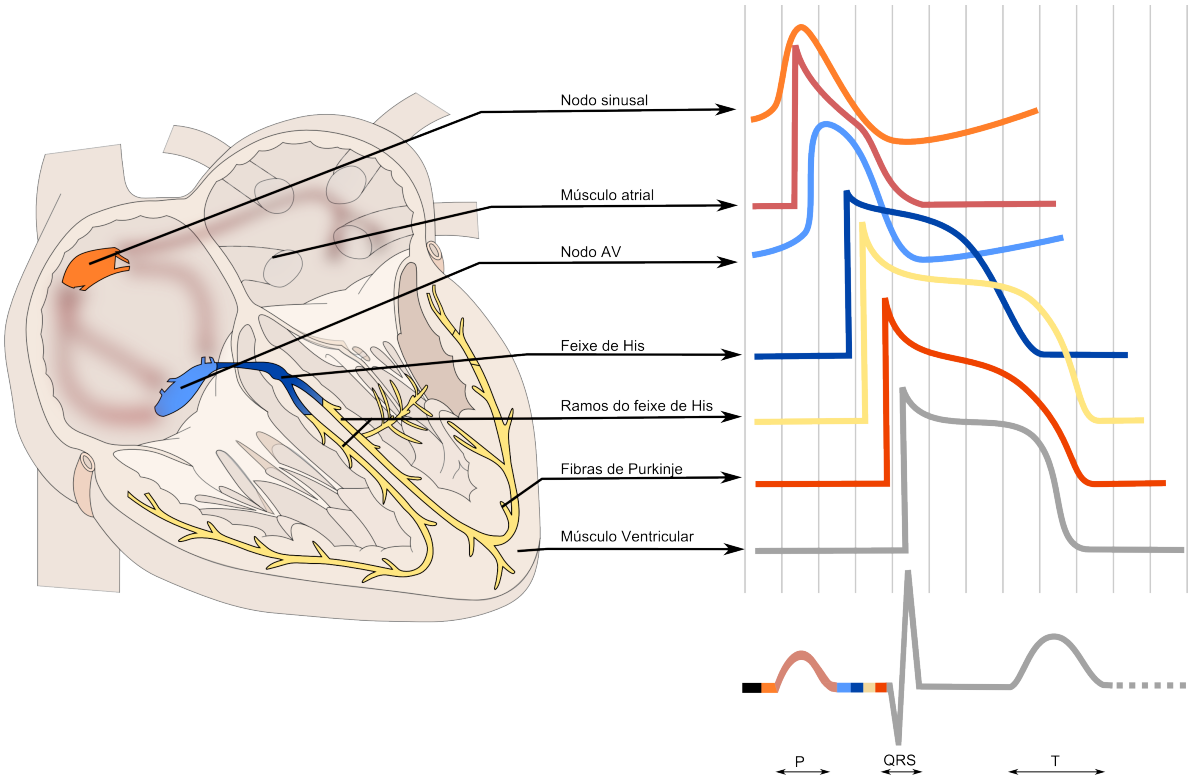


Figura 2.1: Potenciais de ação do coração.

A propagação sequencial do potencial de ação célula a célula, ao longo das suas mem-

branas, a partir do nodo sinusal, constitui-se no impulso ou estímulo elétrico do coração, que se espalha rapidamente por todo o órgão por meio dos ramos e sub-ramos do tecido de condução.

No caso das células miocárdicas comuns atriais e ventriculares, quando estas são atingidas pelo estímulo elétrico proveniente do nodo sinusal, abrem-se canais específicos para os íons sódio na membrana celular, que entram em grande quantidade e rapidamente nas células, obedecendo ao gradiente elétrico e químico presente, o que provoca a inversão da polaridade da membrana celular, ficando o interior da célula carregado positivamente em relação ao seu exterior. Esta despolarização inicia o potencial de ação que é conduzido por todo o miocárdio contrátil atrial e ventricular. Nestas células, a manutenção da despolarização, que também é dependente da entrada de íons cálcio para o interior celular se faz por tempo mais prolongado que nas células do tecido excito-condutor, o que resulta em um potencial de ação com *plateau* mais longo, conforme observado na Figura 2.1. O processo de repolarização da membrana das células miocárdicas também decorre da saída de íons potássio para o exterior celular (Guyton e Hall, 2002).

Para que a célula esteja novamente apta a se ativar, logo após a repolarização, os íons sódio que se dirigiram para o seu interior, e aí ficaram aprisionados, devem ser repostos para o exterior, e os íons potássio que saíram da célula devem retornar para o seu interior. Este processo de recuperação do estado iônico de repouso é feito por meio da chamada "bomba de sódio e potássio", que nada mais é que um sistema bioquímico enzimático existente na membrana celular, que funciona consumindo energia para tornar esta membrana permeável a esses íons, nessa fase do fenômeno elétrico celular. Portanto, o potencial de ação do coração constitui-se, de maneira geral, de três componentes:

1. um componente inicial, de curtíssima duração, dependente principalmente da entrada intracelular de íons sódio, no caso do miocárdio comum (componente inicial rápido), ou de íons cálcio, no caso do tecido excito-condutor (componente inicial lento), que inverte o potencial de membrana, e é traduzido pela despolarização da membrana celular, do que resulta o início do fenômeno da contração sistólica do coração;

2. um componente intermediário, de maior duração, que segue o anterior, e é dependente da manutenção da entrada intracelular de íons cálcio previamente iniciada, o qual é traduzido pela persistência da despolarização, dando ao potencial de ação a configuração de um *plateau*. A manutenção da despolarização constitui-se na base eletrofisiológica do prolongado processo de ativação ventricular do qual decorre a continuidade da contração sistólica. Portanto, a etapa de *plateau* é mais visível no processo de despolarização ventricular;

3. um componente final, dependente da saída extracelular de íons potássio, traduzido pela repolarização ou recuperação elétrica da membrana celular, que resulta no restabelecimento do potencial de membrana, do qual decorre o fenômeno mecânico do relaxamento diastólico do coração (Guyton e Hall, 2002).

Quanto às diferenças entre o potencial de ação dos nodos sinusal e atrioventricular, e o potencial de ação do tecido condutor intraventricular e do miocárdio comum, no tecido nodal o limiar de disparo da despolarização é mais baixo (o potencial de membrana é menos negativo), a despolarização inicial é mais lenta e dependente do íon cálcio, o *plateau* é acentuadamente mais curto, e existe peculiarmente o pré-potencial. Estas são as características eletrofisiológicas do tecido nodal que lhe conferem a propriedade do automatismo e, em decorrência, a capacidade de comandar a atividade elétrica do coração.

Assim, o potencial de ação do coração, ou o seu estímulo elétrico, origina-se automaticamente no nodo sinusal e, a partir desta estrutura, propaga-se pelo miocárdio atrial, atingindo o nodo atrioventricular, de onde ganha o tecido especializado condutor dos ventrículos, representado pelo feixe de *His* e seus ramos e sub-ramos direito e esquerdo, terminando no sistema de *Purkinje*, e ativando sequencialmente toda a musculatura ventricular numa direção e sentido bem definidos.

Em relação à propagação sequencial do potencial de ação, observa-se que quando as células do nódulo sinoatrial se contraem, o impulso elétrico da despolarização é conduzido de tal forma que o átrio direito é o primeiro a se contrair, seguido pelo átrio esquerdo. Dessa forma, o sangue contido nos átrios é bombeado para os ventrículos. A seguir, a onda de despolarização é conduzida rapidamente para os ventrículos através do feixe de *His*, tal que os ventrículos também se contraem. Ao mesmo tempo em que as células do ventrículo se despolarizam, as células dos átrios estão se repolarizando e, estes voltam a relaxar. A seguir, o mesmo acontece com os ventrículos, e o coração permanece relaxado até que as células do nódulo sinoatrial voltem a se despolarizar (Burton, 1977).

Pode acontecer de uma célula que não faz parte do nódulo sinoatrial se despolarizar antes das outras. Nesse caso acontece o que se denomina de um batimento ectópico (caracterizado como arritmia ou extrassístole). Contudo, normalmente o ritmo cardíaco é determinado pelo nódulo sinoatrial.

A atividade elétrica gerada no coração pode ser captada na superfície corporal por meio de eletrodos colocados em determinadas posições padrão. Esta atividade elétrica, representada pelas diferenças de potencial elétrico criadas em cada ponto do coração, que nada mais são que o potencial de membrana e o potencial de ação alternando-se ciclicamente, expressa o *eletrocardiograma* (ECG) que, assim, pode ser definido como o registro gráfico da atividade elétrica do coração, captada ao longo do tempo na superfície corporal. Diferentes

ondas, intervalos e segmentos são observados no eletrocardiograma (vide Figura 2.2), e traduzem as atividades elétricas das diferentes regiões do coração nas distintas fases do seu funcionamento. Assim,

- a onda "P", que é a primeira a surgir, representa a despolarização dos átrios;
- as ondas intermediárias "Q", "R" e "S", que formam o complexo "QRS", representam a despolarização das diferentes partes dos ventrículos;
- a onda "T", que é a última observada, traduz a repolarização dos ventrículos.

Por meio da análise da morfologia, da amplitude, da duração e da polaridade dos diferentes acidentes eletrocardiográficos (ondas, intervalos e segmentos), dentre outros aspectos, pode-se estabelecer o diagnóstico da condição de normalidade ou de diversas condições patológicas do coração (Guyton e Hall, 2002).

2.1.1 ECG Normal

O sinal de ECG de um coração normal é composto por alguns traços característicos, tais como o complexo QRS, a onda T e a onda P, que ocorrem de maneira cíclica, onde cada ciclo, ou período completo, corresponde a um batimento cardíaco.

A Figura 2.2 apresenta um ciclo cardíaco típico completo, com a indicação das ondas características. Esta dissertação usará a palavra batimento quando se referir a ciclo cardíaco.

Cada ciclo de contração e relaxamento cardíacos é iniciado pela despolarização espontânea do nodo sinusal (evento não observado no ECG). A onda P é gerada por correntes oriundas da despolarização dos átrios, a qual precede sua contração. A primeira parte dessa onda corresponde à despolarização do átrio direito (pois o nodo sinusal está localizado no átrio direito) e a segunda parte corresponde à despolarização do átrio esquerdo. Ela é arredondada, monofásica e tem amplitude entre $0,25$ e $0,3$ mV (Norman, 1991).

Após ser completada a ativação atrial, o ECG torna-se novamente "silencioso", não havendo registro de nenhuma onda até o início do complexo QRS. Este retardo fisiológico ocorre no nodo átrio-ventricular, para permitir o esvaziamento atrial. A onda de despolarização espalha-se, então, ao longo do sistema de condução ventricular (feixe de *His*, ramos do feixe e fibras de *Purkinje*) e para dentro do miocárdio ventricular. A primeira parte dos ventrículos a ser despolarizada é o septo interventricular. A despolarização ventricular gera o complexo QRS.

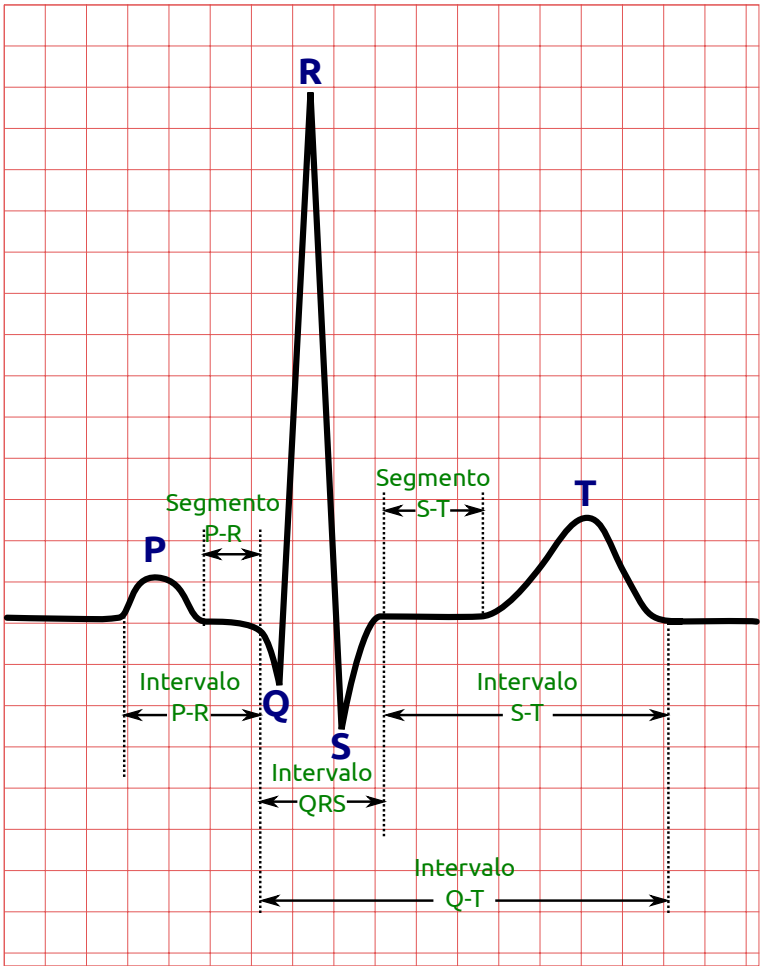


Figura 2.2: Eletrocardiograma com informações características.

O complexo QRS é gerado pela despolarização total dos ventrículos antes da contração. O tempo de ativação ventricular, que representa o momento da despolarização, é medido do início do complexo QRS ao final da deflexão negativa após a onda R.

A onda T é a onda de repolarização ventricular, que é causada por correntes geradas enquanto os ventrículos se recuperam da fase de despolarização. Ela é a primeira onda positiva ou negativa que surge após o complexo QRS. A repolarização ventricular possui amplitude menor que o complexo QRS.

O segmento ST é o intervalo entre o final do complexo QRS e o início da onda T. Portanto, ele registra o tempo do fim da despolarização ventricular ao início da repolarização ventricular. Ele é normalmente isoelétrico, e sua duração geralmente não é determinada, pois é avaliado englobado ao intervalo QT (mede o tempo do início da despolarização ventricular ao fim da repolarização ventricular).

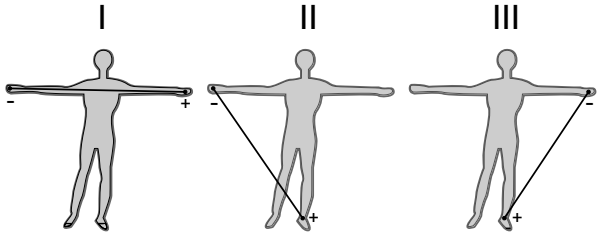
Existe também a onda U, que aparece raras vezes no ECG, e segue-se à onda T. A onda

U não é constante, e quando normal é sempre positiva. Sua gênese ainda é discutida, mas poderia representar um pós-potencial, ou seja, a repolarização dos músculos papilares (Khan, 2007).

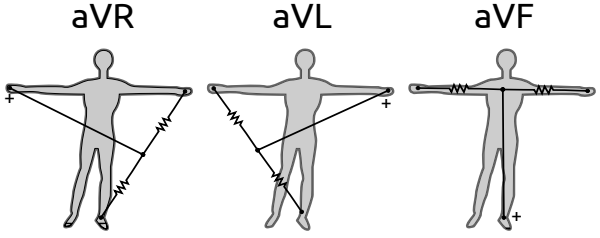
2.1.2 Derivações do ECG

Em toda superfície do corpo existem diferenças de potencial decorrentes dos fenômenos elétricos gerados durante a excitação cardíaca. Estas diferenças podem ser medidas e registradas. Assim, os pontos do corpo a serem explorados são ligados ao aparelho de registro por meio de eletrodos, obtendo-se as chamadas derivações, que podem ser definidas de acordo com a posição dos eletrodos sobre o corpo.

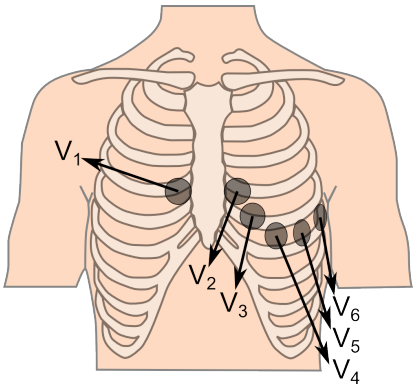
O ECG padrão é composto de 12 derivações principais: seis derivações de membros (periféricas) e seis derivações precordiais. As derivações dos membros são I, II, III, aVR, aVL e aVF. As derivações precordiais são V1, V2, V3, V4, V5 e V6, de acordo com a Figura 2.3. Eventualmente, são utilizadas derivações precordiais adicionais para uma melhor visualização da parede posterior do coração (V7 e V8) e do ventrículo direito (V3R e V4R) (Khan, 2007).



(a) Bipolar.



(b) Unipolar aumentada.



(c) Precordial.

Figura 2.3: Derivações de ECG.

As derivações podem ser divididas em três subgrupos diferentes: bipolar ou de *Einthoven*, unipolares ou de *Goldberger*, e precordiais de *Wilson*.

2.1.3 Derivações Bipolares

As derivações de membros observam o coração em um plano vertical, chamado de plano frontal, que pode ser imaginado, de forma simplificada, como sendo uma circunferência sobreposta ao corpo do ser humano. Esta circunferência é, então, separada em ângulos. As derivações de membros consideram as ondas de despolarização e repolarização (potenciais elétricos) como se movendo sobre este plano. Cada derivação possui sua perspectiva específica do coração, também chamada de ângulo de orientação (ângulo associado à circunferência no sentido horário). As três derivações padronizadas de membros são definidas pelos critérios abaixo:

- a derivação I é criada pela conversão do braço esquerdo em referencial positivo (lado referente ao coração) e o braço direito em referencial negativo. Seu ângulo de orientação é 0°.
- a derivação II é criada pela transformação das pernas em referencial positivo e do braço direito em referencial negativo. Seu ângulo de orientação é 60°.
- a derivação III é criada pela conversão das pernas em referencial positivo e do braço esquerdo em referencial negativo. Seu ângulo de orientação é 120°.

Em 1913, *Einthoven* desenvolveu um método de estudo da atividade elétrica do coração representando-a graficamente numa simples figura geométrica bidimensional, no caso um triângulo equilátero (Figura 2.4) (Hansen, 2011). As derivações bipolares de membros (I, II e III) são as derivações originais escolhidas por *Einthoven* para registrar os potenciais elétricos no plano frontal.

No triângulo de *Einthoven*, o coração está localizado no centro do triângulo equilátero e os vértices do triângulo estão posicionados no ombro esquerdo, no ombro direito e na região púbica. As derivações bipolares representam uma diferença de potencial entre dois locais selecionados:

- I equivale à diferença de potencial entre o braço esquerdo e o braço direito (aVL − aVR);
- II equivale à diferença de potencial entre a perna esquerda e o braço direito (aVF − aVR);
- III equivale à diferença de potencial entre a perna esquerda e o braço esquerdo (aVF − aVL).

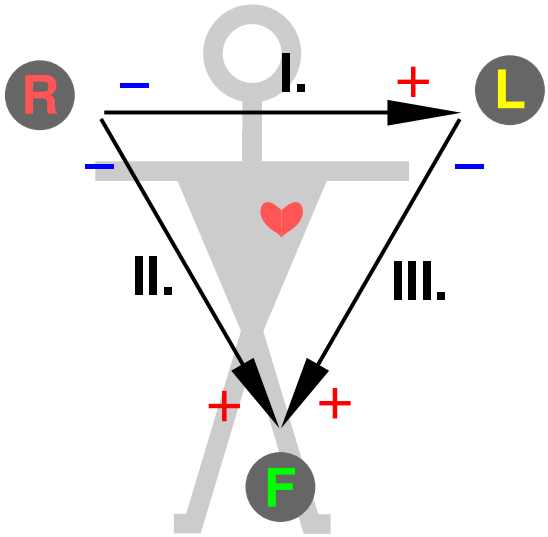


Figura 2.4: Triângulo de *Einthoven*.

Segundo a lei de *Einthoven*, baseada nas leis de *Kirchoff*, se o potencial elétrico de duas derivações bipolares quaisquer for conhecido num dado instante, a terceira pode ser calculada por

$$\mathrm{II} = \mathrm{I} + \mathrm{III}.$$

Os braços são considerados apenas como extensões dos ombros, e a perna esquerda é considerada a extensão do púbis, por convenção. Desta forma, os eletrodos são posicionados logo acima dos pulsos e do tornozelo.

2.1.4 Derivações Unipolares

As derivações unipolares foram introduzidas por *Wilson* em 1932 (Norman, 1991). Elas medem a diferença de potencial entre um eletrodo indiferente e um eletrodo explorador. O eletrodo indiferente é formado por três fios elétricos que estão ligados entre si a um terminal central. As extremidades livres destes fios são ligadas aos eletrodos do braço esquerdo (LA), braço direito (RA) e perna esquerda (LL). O eletrodo indiferente é ligado ao pólo negativo do eletrocardiógrafo, e o eletrodo explorador é ligado ao pólo positivo.

Considera-se que a soma dos três potenciais $\mathrm{LA} + \mathrm{RA} + \mathrm{LL}$ é igual à zero, ou seja, o potencial do eletrodo indiferente é nulo. A princípio, as derivações unipolares tentam medir potenciais locais, e não diferenças de potencial. *Goldberg* modificou o sistema de derivações unipolares de *Wilson* para obter três derivações unipolares aumentadas, chamadas aVL, aVR e aVF, amplificando a variação de potencial por um fator de 1.5. Por exemplo, usando o eletrodo indiferente ligado à perna esquerda e ao braço direito, e um eletrodo explorador

ligado ao braço esquerdo, é obtido o potencial do braço esquerdo amplificado (aVL) (Figura 2.3).

Portanto, as três derivações aumentadas são criadas de modo diferente. Um eletrodo é escolhido como positivo, e todos os demais são transformados em negativos, com sua média servindo como eletrodo negativo, tal que

1. a derivação aVL é criada pela transformação do braço esquerdo em pólo positivo e os demais membros em pólo negativo. Seu ângulo de orientação é 330°;
2. a derivação aVR é criada pela conversão do braço direito em pólo positivo e dos outros membros em pólo negativo. Seu ângulo de orientação é 210°;
3. a derivação aVF é criada pela transformação das pernas em pólo positivo e dos outros membros em pólo negativo. Seu ângulo de orientação é 90°.

2.1.5 Derivações Precordiais

As seis derivações precordiais, ou derivações torácicas, são dispostas através do tórax em um plano horizontal. Enquanto as derivações do plano frontal observam as forças elétricas movendo-se para cima e para baixo e da esquerda para a direita, as derivações precordiais registram as forças movendo-se anterior e posteriormente (de forma simplificada, frente - trás).

Portanto, as derivações precordiais permitem o mapeamento elétrico do coração no plano horizontal. O eletrodo indiferente permanece ligado às três extremidades, enquanto o eletrodo explorador varia de posição ao longo da parede torácica. Uma derivação unipolar feita por este método é denominada pelo prefixo V seguido de um número, que indica a sua posição correspondente.

As derivações precordiais não registram apenas os potenciais elétricos da pequena área de miocárdio que está subjacente, mas os eventos elétricos de todo o coração, tal como são vistos no eixo elétrico da sua posição específica.

A seguir estão descritas as posições dos eletrodos precordiais (vide Figura 2.3(c)):

1. V1: quarto espaço intercostal direito junto ao esterno;
2. V2: quarto espaço intercostal esquerdo junto ao esterno;
3. V3: equidistante de V2 e V4;

4. V4: quinto espaço intercostal esquerdo na linha médio-clavicular (linha hemiclavicular);

5. V5: linha axilar anterior (mesmo plano horizontal de V4);

6. V6: linha axilar média (mesmo plano horizontal de V4), no quinto espaço intercostal.

2.1.6 Formato do ECG nas Diferentes Derivações

Cada uma das 12 derivações principais mede os potenciais elétricos do coração em diferentes pontos de observação. Logo, cada uma delas irá produzir uma forma de onda diferente. Isso pode ser muito útil para o diagnóstico de determinadas anomalias do coração, que podem ser facilmente percebidas em determinada derivação ao passo que em outras derivações o ECG não apresentaria alterações significativas. A Figura 2.5 apresenta o sinal de ECG nas 12 derivações.

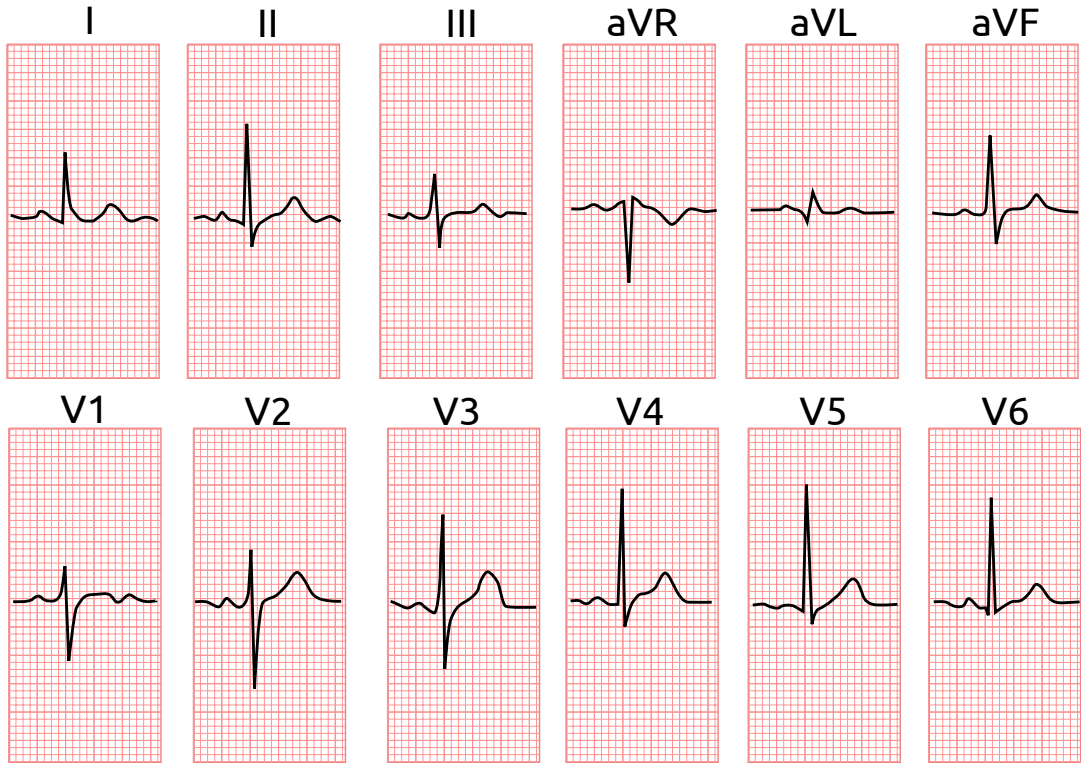


Figura 2.5: Eletrocardiogramas em todas as 12 derivações.

2.1.7 Série RR

O intervalo RR é definido pela distância entre as ondas R de dois batimentos consecutivos, de acordo com a Figura 2.6. A frequência cardíaca, dada pelo inverso do intervalo RR, é a velocidade dos batimentos cardíacos. Em condições normais a frequência cardíaca é de aproximadamente 70 bpm, e em esforço físico cerca de 120 bpm.

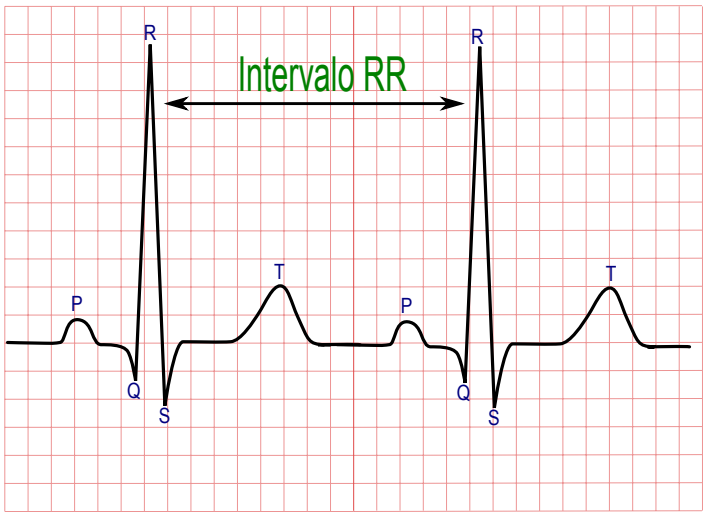


Figura 2.6: Intervalo RR.

A série RR é definida como a variação temporal dos intervalos RR consecutivos (vide Figura 2.7).

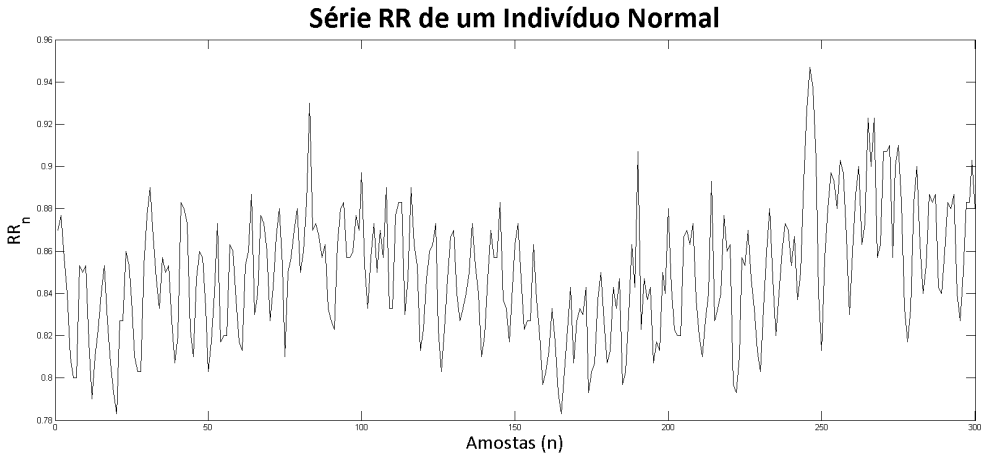


Figura 2.7: Série RR.

Como a onda P indica o início da contração cardíaca, pode-se acreditar que seria melhor utilizar os intervalos PP, ao invés dos intervalos RR, para o estudo da VFC. Porém, a onda

P é de baixa amplitude, sendo muitas vezes difícil de distinguir no ECG. Já o complexo QRS, por sua vez, quase sempre tem grande amplitude, além de uma característica espectral peculiar, sendo fácil de ser isolado através da filtragem do ECG. Assim, no estudo de VFC utilizam-se os intervalos RR como representação do período decorrido entre dois batimentos do coração.

Em específico, a detecção da onda R, e consequentemente do complexo QRS, pode ser feita através da combinação de estágios de filtragem do sinal (filtragem passa-faixa para eliminar componentes de frequência que não pertencem ao complexo QRS, filtro derivada para reforçar componentes de alta frequência do sinal e filtragem média móvel para construir o envoltório final do sinal filtrado) com regras heurísticas para identificar o complexo QRS, que seria a observação do pico dominante numa janela temporal específica do sinal de ECG.

2.2 Regulação do Sistema Cardiovascular

A regulação cardiovascular tende a manter a pressão arterial limitada a uma pequena faixa de variação, de tal forma que seja mantido o volume de sangue adequado às demandas dos diferentes tecidos (Hall, 2010).

Sua atuação engloba diversos sistemas inter-relacionados de controle, cada um executando uma função específica, podendo ser classificada em três grupos, de acordo com o tempo de reação às variações da pressão arterial (Ursino e Magosso, 2003):

1. resposta rápida (regulação de curto prazo): age dentro de segundos ou minutos, consistindo quase que totalmente de reflexos nervosos agudos ou outras respostas nervosas regidas pelo sistema nervoso autônomo (SNA);

2. resposta intermediária: age de minutos a horas, formada pelos mecanismos vasoconstritores da renina-angiotensina, distensão contínua dos vasos ou relaxamento por estresse da vasculatura e a redução do volume sanguíneo pelo desvio de líquido capilar;

3. resposta de longo prazo: age dentro de dias, meses e anos. Neste caso, o controle da pressão é realizado por mecanismo de regulação renal, através do ajuste do volume dos líquidos corporais.

Esse trabalho restringe-se à análise da regulação cardiovascular de curto prazo, realizada através do SNA.

2.2.1 Sistema Nervoso Autônomo

O organismo humano é constituído por um conjunto de órgãos e tecidos com estruturas e funções diversas. Para que todos esses órgãos e tecidos funcionem em harmonia com os demais, o sistema nervoso exerce um papel fundamental nos mecanismos de regulação dos sistemas biológicos. O sistema nervoso pode ser dividido em duas partes principais: o sistema nervoso central (SNC), representado pelo encéfalo e medula espinhal, e o sistema nervoso periférico, que consiste de todas as outras estruturas nervosas. O sistema nervoso periférico é ainda subdividido em duas outras partes: sistema nervoso somático (ou voluntário) e sistema nervoso autônomo (SNA). O SNA é responsável pelas funções involuntárias das estruturas do organismo humano, e, juntamente com o SNC, atua na regulação das diversas funções fisiológicas autonômicas, como a pressão arterial e temperatura corporal.

O sistema nervoso autônomo (SNA) é um sistema predominantemente eferente que transmite impulsos do sistema nervoso central (SNC) para os órgãos periféricos. Seus efeitos na regulação cardiovascular incluem o controle da frequência e força de contração cardíaca (ou bombeamento cardíaco), a constrição e dilatação dos vasos sanguíneos, a contração e relaxamento dos músculos e secreções de glândulas em diferentes órgãos. Os nervos autônomos constituem-se de todas as fibras eferentes, ou motoras, que deixam o SNC, exceto aquelas que inervam o músculo esquelético. Existem algumas fibras autônomas aferentes, ou sensoras (que partem dos sistemas periféricos para o SNC), que inervam os receptores de estiramento (barorreceptores) e receptores químicos (quimiorreceptores) no seio carotídeo e no arco aórtico, sendo importantes no controle da frequência cardíaca, da pressão arterial, e da atividade respiratória (Freeman et al., 2006). Existem ainda receptores de calor (termorreceptores) e outros receptores presentes nos músculos esqueléticos e no fluido extracelular dos tecidos capazes de ativar fibras aferentes, atuando no controle cardiovascular (Li, 2004). A regulação autônoma da função cardiovascular assegura o suprimento do sangue a regiões específicas do organismo, em resposta a estímulos fisiológicos (alimentação e sono) e comportamentais, como exercício físico e estresse (Robertson et al., 2011).

Com base em diferenças funcionais e anatômicas o SNA pode ser dividido em dois ramos (Hall, 2010): sistema nervoso simpático (SNS) e sistema nervoso parassimpático (SNP). Tais ramos apresentam as inervações simpática e parassimpática, respectivamente, utilizadas na regulação da circulação.

O SNS pode ser considerado como o maior responsável pela regulação cardiovascular. Ele age diretamente no nodo sinoatrial, reduzindo o limiar do potencial de membrana. Assim, ocorre um aumento na frequência de despolarização espontânea, elevando, por consequência, a frequência cardíaca. Ele aumenta também a contratilidade cardíaca e o relaxamento do ventrículo esquerdo durante a diástole (Robertson et al., 2011), produzindo um

aumento do débito cardíaco e da pressão arterial. Por fim, o SNS atua na constrição vascular, proporcionando uma rápida redistribuição do fluxo sanguíneo em dadas ocasiões, tais como ao se levantar ou durante exercício (Robertson et al., 2011).

No coração, o SNP age em oposição ao SNS. Enervado pelo nervo vago (por isso também denominado de sistema nervoso vagal), aumenta o potencial de membrana, reduzindo a frequência de despolarização espontânea e, portanto, a frequência cardíaca. O SNP diminui a contratilidade cardíaca apenas dos átrios e, portanto, não reduz o débito cardíaco por meio da atuação sobre os ventrículos. Assim, a redução de débito cardíaco pelo SNP ocorre através da redução da frequência cardíaca.

A atuação combinada destes dois ramos autônomos determina a frequência cardíaca e a pressão arterial média, bem como suas variações (Akselrod et al., 2006), sendo essencial para a homeostase do sistema cardiovascular. Assim, a permanente influência exercida pelo sistema nervoso autônomo sobre o funcionamento dos diversos órgãos, aparelhos e sistemas que compõem o organismo humano é essencial para a preservação das condições do equilíbrio fisiológico interno, permitindo que o mesmo exerça, adequadamente, a sua interação com o meio ambiente em sua volta. Qualquer fator que provoque a tendência ao desequilíbrio promove, prontamente, respostas orgânicas automáticas e involuntárias que têm por finalidade reverter o processo em andamento e restabelecer o equilíbrio funcional (Berntson et al., 1997).

2.2.2 Barorreflexo Cardiovascular

Na regulação do sistema cardiovascular, o barorreflexo é um importante mecanismo de realimentação negativa que atua na modulação da frequência cardíaca através da estimulação ou inibição das atividades do SNS e do SNP (Li, 2004). Localizados nas paredes do arco aórtico e no seio carotídeo, os barorreceptores são sensores extremamente importantes na rápida regulação da pressão arterial. O aumento ou diminuição da pressão arterial provocam, respectivamente, um estiramento ou uma retração nas paredes do seio carotídeo, aumentando ou diminuindo a frequência de disparo gerada pelos barorreceptores (Li, 2004).

Para exemplificar a atuação do mecanismo de barorreflexo no controle da pressão arterial, consideram-se as seguintes situações:

- uma queda da pressão arterial é detectada por uma diminuição na frequência de disparo dos receptores de estiramento. Isto implica na redução da quantidade dos impulsos do nervo vago (inibição parassimpática) e um aumento da quantidade dos impulsos do nervo simpático, aumentando a contratilidade, a frequência cardíaca e a constrição dos

vasos sanguíneos. A resposta conjunta do SNA produz elevação da pressão arterial, retornando-a para valor próximo do ideal (Li, 2004; Xiao et al., 2004);

- de forma correspondente, um aumento da pressão arterial é detectado pelos barorreceptores, produzindo um aumento na sua frequência de disparo. Isto implica no aumento da quantidade dos impulsos do nervo vago e uma redução da quantidade dos impulsos do nervo simpático (inibição simpática). A inibição simpática reduz a contratilidade ventricular, a frequência cardíaca e diminui a vasoconstrição nos vasos sanguíneos, fazendo o valor da pressão arterial diminuir, com o intuito de retornar ao seu valor ideal (Li, 2004).

O sistema regulatório cardiovascular conta ainda com quimiorreceptores periféricos (localizados nos seios carotídeo e aórtico) e com quimiorreceptores centrais (localizados nos núcleos respiratórios do tronco encefálico). Estes têm a função de responder rapidamente (dentro de alguns segundos) a aumento na frequência cardíaca, a aumento na concentração de dióxido de carbono (CO_2), a redução na concentração de oxigênio (O_2) ou a queda do pH do sangue (Li, 2004). Existem, ainda, receptores de volume localizados no interior dos átrios e receptores de estiramento presentes no pulmão, completando esse complexo sistema de receptores envolvidos na regulação do sistema cardiovascular. Entretanto, estes mecanismos de controle não são discutidos nesse trabalho.

2.2.3 Arritmia Sinusal Respiratória

A respiração modula a atuação dos mecanismos de controle cardiovascular, tal que uma relação entre o volume pulmonar e a frequência cardíaca é observada. Em indivíduos saudáveis, durante a expiração a frequência cardíaca diminui, ocorrendo aumento no intervalo entre batimentos (intervalos RR). Na fase da inspiração, observa-se um aumento da frequência cardíaca, e consequente redução dos intervalos RR. Essa variação do ritmo cardíaco em sincronia com a respiração é denominada de arritmia sinusal respiratória (do inglês, *Respiratory Sinus Arrhythmia*, ou RSA).

Apesar dos mecanismos exatos que ocasionam a RSA não serem ainda conhecidos, tem-se como consenso que a RSA é gerada, predominantemente, pelo controle autonômico parassimpático do nodo sinoatrial (Eckberg, 2003; Patwardhan, 2006). Observa-se que variações no volume sanguíneo central induzidas mecanicamente pela atividade respiratória contribuem pouco para modulação cardíaca (2 a 8% (Cohen e Taylor, 2002)), logo estas não são consideradas aqui.

2.3 Variabilidade da Frequência Cardíaca

Quando se está relaxado, o coração bate mais lentamente. Porém, quando uma pessoa sofre um estímulo emocional, ou começa a realizar uma atividade de maior esforço físico, o organismo se ajusta à sua nova realidade, de modo a corrigir deficiências metabólicas que possam surgir, ou a poder oferecer recursos para uma reação à nova situação. Assim, a respiração fica mais forte, e a frequência cardíaca aumenta. Quando o estímulo emocional cessa ou a exercício físico acaba, o organismo volta a relaxar, e a respiração e o ritmo cardíaco diminuem. Quem controla essa variação na respiração e na frequência cardíaca, entre outras coisas, é o sistema nervoso autônomo. Nesse exemplo, o sistema nervoso simpático atua aumentando o ritmo cardíaco e respiratório, e o sistema nervoso parassimpático atua no sentido contrário, diminuindo esses ritmos (Berntson et al., 1997), de tal forma que os ramos simpático e parassimpático controlam o ritmo dos batimentos do coração, atuando diretamente no nódulo sinoatrial.

O coração, apesar de ter a sua enervação intrínseca e, portanto, ser capaz de regular o seu ritmo, promover a condução dos estímulos intracardíacos e ter contratilidade, tem todas essas funções moduladas pelo sistema nervoso autônomo. De fato, o sistema nervoso autônomo é o responsável pela regulação do ritmo e da função do bombeamento cardíaco, adequando essas funções às necessidades metabólicas e teciduais, às quais estão expostos os seres humanos em suas atividades diárias. Conforme visto na Seção 2.2.1, a integração entre a modulação rápida (parassimpática) e lenta (simpática) determinam a variabilidade da frequência cardíaca (VFC).

A partir da série RR, obtida do registro do canal de ritmo do ECG (derivação II), é efetuada a análise da variabilidade da frequência cardíaca (VFC) do indivíduo. Uma vez que as alterações nos intervalos RR permitem avaliar a atuação do sistema nervoso (Anderson, 1992). Daí, indicadores sobre a atuação dos ramos simpático e parassimpático do sistema nervoso sobre o nódulo sinoatrial podem ser gerados, com o intuito de diagnosticar doenças no sistema nervoso, mesmo que estas estejam relacionadas a outros órgãos do corpo humano.

2.4 Variabilidade da Frequência Cardíaca e Controle Metabólico

Fatores de risco cardiovasculares como hiperglicemia, hipersinsulinemia, assim como hipertensão arterial sistólica e diastólica são associados com a redução da variabilidade da frequência cardíaca (diminuição de índices de avaliação da VFC), sobretudo em obesos e

indivíduos diabéticos (Novak et al., 1994; Singh et al., 1998; Laederach-Hofmann et al., 2000; Singh et al., 2000; Rissanen et al., 2001; Schroeder et al., 2005).

Pacientes diabéticos comumente apresentam comprometimento silencioso dos ramos do sistema nervoso autônomo (SNS e SNP), devido à degeneração neurológica que afeta as pequenas fibras do sistema nervoso central (SNC) e sistema nervoso periférico (SNP), conhecido como neuropatia autonômica, que pode ser caracterizada pela redução da VFC, e é causa frequente de morbi-mortalidade cardiovascular nestes indivíduos (Schroeder et al., 2005), (Howorka et al., 1997; Liao et al., 1998; Ribeiro e Filho, 2005).

Diabéticos apresentam mais casos de morte súbita após infarto do miocárdio comparado aos não diabéticos, e a neuropatia autonômica pode ser a responsável (Khandoker et al., 2009). Com isso, a detecção precoce da disfunção autonômica nestes indivíduos é importante para estratificar o risco e posteriormente intervir farmacologicamente e no hábito de vida (Schroeder et al., 2005).

A VFC é correlacionada à glicemia de jejum, visto que diabéticos e indivíduos com glicemia alterada apresentam atividade parassimpática do SNA reduzida (Singh et al., 2000). A baixa atividade do sistema nervoso parassimpático é um fator de risco para doença arterial coronariana, assim como é um fator que predispõe arritmia e morte súbita em obesos (Rissanen et al., 2001).

Em (Schroeder et al., 2005) observou-se uma correlação inversa entre glicemia de jejum e VFC nos indivíduos não diabéticos. Diabéticos e indivíduos com hiperinsulinemia apresentaram menor VFC que não diabéticos e os indivíduos sem hiperinsulinemia, respectivamente.

Indivíduos hipertensos tendem a ter menor VFC e maior frequência cardíaca (FC) de repouso que normotensos (Novak et al., 1994; Singh et al., 1998). Normotensos com VFC reduzida tendem a apresentar maior risco de desenvolver hipertensão (Singh et al., 1998).

A contribuição dos diferentes componentes da síndrome metabólica (diabetes *mellitus*, hipertensão e dislipidemia) sobre os padrões de VFC é complexa. Em (Liao et al., 1998) verificou-se associações negativas entre os níveis de insulina em jejum e VFC, o que sugere que a resistência à insulina pode explicar parcialmente a reduzida VFC nos indivíduos com múltiplas desordens. Igualmente, observou-se redução significativa da VFC com o aumento do número de desordens associadas. A presença de hipertensão ou diabetes *mellitus* isolados foi associada com um decréscimo na VFC, ao contrário da dislipidemia que não apresentou associação isolada (Liao et al., 1998).

Há evidências de que o treinamento físico pode amenizar a neuropatia cardiovascular autonômica. Em diabéticos com diferentes graus de neuropatia, submetidos a algumas se-

manas de exercícios controlados, a VFC aumentou nos indivíduos sem neuropatia ou com neuropatia cardiovascular autonômica branda, ao contrário dos indivíduos com neuropatia severa (Howorka et al., 1997).

2.5 Diabetes *Mellitus*

O diabetes *mellitus* (DM) é um grupo de doenças metabólicas caracterizadas por hiperglicemia e associadas a complicações, disfunções e insuficiência de vários órgãos, especialmente olhos (Harman-Boehm et al., 2007), rins (Guthrie, 1998), nervos (Ewing e Clarke, 1982), cérebro, coração (Grundy et al., 1999) e vasos sanguíneos. Pode resultar de defeitos de secreção e/ou ação da insulina envolvendo processos patogênicos específicos, por exemplo destruição das células beta do pâncreas (produtoras de insulina), resistência à ação da insulina, distúrbios da secreção da insulina, entre outros.

2.5.1 Epidemiologia do Diabetes

O diabetes é comum e de incidência crescente. Estima-se que, em 1995, atingia 4% da população adulta mundial, e que em 2025 alcançará $5,4\%$. A maior parte desse aumento se dará em países em desenvolvimento, acentuando-se, nesses países, o padrão atual de concentração de casos na faixa etária de 45 a 64 anos (Association, 2011).

No Brasil, no final da década de 1980 estimou-se que o diabetes ocorria em cerca de 8% da população, de 30 a 69 anos de idade, residente em áreas metropolitanas brasileiras. Essa prevalência variava de 3% a 17% entre as faixas de 30 a 99 e de 60 a 69 anos. A prevalência da tolerância à glicose diminuída era igualmente de 8%, variando de 6 a 11% entre as mesmas faixas etárias (BVS, 2002).

Hoje estima-se que 11% da população igual ou superior a 40 anos, aproximadamente 6 milhões, são portadores de diabetes (população estimada IBGE 2005).

O diabetes apresenta alta morbi-mortalidade, com perda significante na qualidade de vida. É uma das principais causas de insuficiência renal, amputação de membros inferiores, cegueira e doença cardiovascular. A Organização Mundial da Saúde (OMS) estimou em 1997 que, após 15 anos de doença, 2% dos indivíduos acometidos estarão cegos e 10% terão deficiência visual grave. Além disso, estimou que no mesmo período de doença 30 a 45% terão algum grau de retinopatia, 10 a 20%, de nefropatia, 20 a 35%, de neuropatia e 10 a 25% terão desenvolvido doença cardiovascular (WHO, 2011).

Mundialmente, os custos diretos para o atendimento ao diabetes variam de $2,5\%$ a 15% dos gastos nacionais em saúde, dependendo da prevalência local de diabetes e da complexidade do tratamento disponível. Além dos custos financeiros, o diabetes acarreta também outros custos associados à dor, ansiedade, inconveniência e menor qualidade de vida, que afeta doentes e suas famílias. O diabetes representa também carga adicional à sociedade, em decorrência da perda de produtividade no trabalho, aposentadoria precoce e mortalidade prematura.

2.5.2 Classificação do Diabetes

Há duas formas para classificar o diabetes, a classificação em tipos de diabetes (etiológica), definidos de acordo com defeitos ou processos específicos, e a classificação em estágios de desenvolvimento, incluindo estágios pré-clínicos e clínicos, este último incluindo estágios avançados em que a insulina é necessária para controle ou sobrevivência.

Os tipos de diabetes mais frequentes são o diabetes tipo 1, anteriormente conhecido como diabetes juvenil, que compreende cerca de 10% do total de casos, e o diabetes tipo 2, anteriormente conhecido como diabetes do adulto, que compreende cerca de 90% do total de casos. Outro tipo de diabetes encontrado com maior frequência, e cuja etiologia ainda não está esclarecida, é o diabetes gestacional, que, em geral, é um estágio pré-clínico de diabetes, detectado no rastreamento pré-natal. Ainda existem duas categorias, referidas como pré-diabetes, que são a glicemia de jejum alterada e a tolerância à glicose diminuída. Essas categorias não são entidades clínicas, mas fatores de risco para o desenvolvimento do diabetes *mellitus* e de doenças cardiovasculares.

Outros tipos específicos de diabetes menos frequentes podem resultar de defeitos genéticos da função das células beta, defeitos genéticos da ação da insulina, doenças do pâncreas exócrino, endocrinopatias, efeito colateral de medicamentos, infecções e outras síndromes genéticas associadas ao diabetes.

Pré-diabetes

Refere-se a um estado intermediário entre a homeostase normal da glicose e o DM. A categoria glicemia de jejum alterada refere-se às concentrações de glicemia de jejum que são inferiores ao critério diagnóstico para o DM, porém mais elevadas do que o valor de referência normal. A tolerância à glicose diminuída representa uma anormalidade na regulação da glicose no estado pós-sobrecarga, que é diagnosticada através do teste oral de tolerân-

cia à glicose (TOTG), que inclui a determinação da glicemia de jejum e de 2 horas após a sobrecarga com 75 g de glicose.

Diabetes *Mellitus* Tipo 1

O diabetes *mellitus* tipo 1 (DM1), forma presente em 5% a 10% dos casos, é o resultado de uma destruição das células beta pancreáticas, com a consequente deficiência de insulina. Na maioria dos casos essa destruição das células beta é mediada por auto-imunidade, porém existem casos em que não há evidências de processo auto-imune, sendo, portanto, referida como forma idiopática do DM1. Os marcadores de auto-imunidade são os auto-anticorpos: antiinsulina, antidescarboxilase do ácido glutâmico (GAD 65) e antitirosina-fosfatases (IA2 e IA2B) (Palmer et al., 1983; Baekkeskov et al., 1990; Rabin et al., 1994). Esses anticorpos podem estar presentes meses ou anos antes do diagnóstico clínico, ou seja, na fase pré-clínica da doença, e em até 90% dos indivíduos quando a hiperglicemia é detectada. Além do componente auto-imune, o DM1 apresenta forte associação com determinados genes do sistema antígeno leucocitário humano (HLA) que podem ser predisponentes ou protetores para o desenvolvimento da doença (Todd et al., 1987).

A taxa de destruição das células beta é variável, sendo, em geral, mais rápida entre as crianças. A forma lentamente progressiva ocorre geralmente em adultos, e é referida como *latent autoimmune diabetes in adults* (LADA).

O DM1 idiopático corresponde a uma minoria dos casos. Caracteriza-se pela ausência de marcadores de auto-imunidade contra as células beta, e não-associação com haplótipos do sistema HLA. Os indivíduos com essa forma de DM podem desenvolver cetoacidose (a cetoacidose diabética consiste em uma tríade bioquímica de hiperglicemia, cetonemia e acidemia) e apresentam graus variáveis de deficiência de insulina.

Como a avaliação dos auto-anticorpos não é disponível em todos os centros, a classificação etiológica do DM1 nas subcategorias auto-imune e idiopático pode não ser sempre possível.

Diabetes *Mellitus* Tipo 2

O diabetes *mellitus* tipo 2 (DM2) é a forma presente em 90% a 95% dos casos e se caracteriza por defeitos na ação e na secreção da insulina. Em geral ambos os defeitos estão presentes quando a hiperglicemia se manifesta, porém pode haver predomínio de um deles. A maioria dos pacientes com essa forma de DM apresenta sobrepeso ou obesidade, e cetoacidose raramente se desenvolve espontaneamente, ocorrendo apenas quando associada a

outras condições, como infecções. O DM2 pode ocorrer em qualquer idade, mas é geralmente diagnosticado após os 40 anos. Os pacientes não são dependentes de insulina exógena para sobrevivência, porém podem necessitar de tratamento com insulina para a obtenção de um controle metabólico adequado.

Diferentemente do DM1 auto-imune não há indicadores específicos para o DM2. Existem provavelmente diferentes mecanismos que resultam nessa forma de DM, e com a identificação futura de processos patogênicos específicos ou defeitos genéticos, o número de pessoas com essa forma de DM irá diminuir à custa de uma mudança para uma classificação mais definitiva em outros tipos específicos de DM.

Diabetes *Mellitus* Gestacional

É qualquer intolerância à glicose, de magnitude variável, com inicio ou diagnóstico durante a gestação. Não exclui a possibilidade de a condição existir antes da gravidez, desde que não seja diagnosticada nesse período. Similar ao DM2, o DM gestacional é associado tanto à resistência a insulina quanto à diminuição da função das células beta (Kuhl, 1991).

O DM gestacional ocorre, geralmente, em 1% a 14% de todas as gestações, e é associado a aumento de morbidade e mortalidade perinatal (Coustan, 1995). Pacientes com DM gestacional devem ser reavaliadas quatro a seis semanas após o parto, e reclassificadas como apresentando DM, glicemia de jejum alterada, tolerância à glicose diminuída ou normoglicemia. Na maioria dos casos há reversão para a tolerância normal após a gravidez, porém existe um risco de 17% a 63% de desenvolvimento de DM2 dentro de 5 a 16 anos após o parto (Hanna e Peters, 2002).

2.5.3 Métodos e Critérios para o Diagnóstico de Diabetes *Mellitus*

Conforme descrito anteriormente, o diabetes *mellitus* (DM) é um grupo heterogêneo de distúrbios metabólicos caracterizados por hiperglicemia crônica com alterações do metabolismo de carboidratos, proteínas e lipídios, resultante de defeitos na secreção ou ação da insulina, ou ambos. Independente de sua etiologia, o DM passa por vários estágios clínicos durante sua evolução natural.

Atualmente, em todo o mundo ocorre uma pandemia de obesidade e diabetes *mellitus* (DM) do tipo 2. O aumento do número de diabéticos e pré diabéticos (termo utilizado para enquadrar aqueles indivíduos cujos níveis glicêmicos encontram-se acima dos valores normais da população não diabética) se deve ao estilo vida contemporâneo, que induz sobrepeso

e obesidade. Essas alterações, acompanhadas de predisposição genética e resistência insulínica, resultam no aumento dos níveis glicêmicos.

A doença pode ser reconhecida nos estágios iniciais a que chamamos de intolerância à glicose. O DM pode se apresentar com sintomas característicos, como sede, polúria, visão turva, perda de peso acompanhado de fome excessiva, e em suas formas mais graves, com cetoacidose (acúmulo dos chamados corpos cetônicos, substâncias que deixam o pH sanguíneo mais baixo do que o normal) ou estado hiperosmolar não cetótico (desidratação extrema). Estes últimos, na ausência de tratamento adequado, podem levar ao coma, e até à morte. Frequentemente, os sintomas não são evidentes ou estão ausentes, principalmente no estágio de pré-diabetes. Desta forma, hiperglicemia pode já estar presente muito tempo antes do diagnóstico de DM. Consequentemente, o diagnóstico de DM ou pré-diabetes é frequentemente descoberto em decorrência de resultados anormais de exames de sangue ou de urina realizados em avaliação laboratorial, ou quando da descoberta de complicação relacionada ao DM. Estima-se que o número de casos não diagnosticados seja igual ao dos diagnosticados. Existem evidências sugerindo que as complicações relacionadas ao DM começam precocemente, ainda na fase de mínimas alterações na glicemia, progredindo nos estágios de pré-diabetes e, posteriormente, DM. Por esse motivo se torna extremamente importante diagnosticar alterações na glicemia precocemente. Níveis glicêmicos elevados em jejum e, principalmente, pós-prandiais (após refeição) implicam em maior risco cardiovascular.

A evolução para o diabetes *mellitus* tipo 2 (DM2) ocorre ao longo de um período de tempo variável, passando por estágios intermediários (pré-diabetes) que recebem a denominação de glicemia de jejum alterada (GJA) e tolerância à glicose diminuída (TGD). Tais estágios seriam decorrentes de uma combinação de resistência à ação insulínica e disfunção de célula beta. Já no diabetes *mellitus* tipo 1 (DM1), o início geralmente é abrupto, com sintomas indicando de maneira sólida a presença da enfermidade (Association, 2003; Engelgau et al., 1997).

Define-se como glicemia de jejum alterada valores de glicemia em jejum mais elevados do que o valor de referência normal, porém inferiores aos níveis diagnósticos de DM: GJ entre 100 e 125 mg/dL. Embora a Organização Mundial de Saúde ainda não tenha adotado esse critério, tanto a Sociedade Brasileira de Diabetes quanto a Academia Americana de Diabetes já utilizam tal ponto de corte (GJ normal até 99 mg/dL). Já a tolerância à glicose diminuída é caracterizada por uma alteração na regulação da glicose no estado pós-sobrecarga (teste oral de tolerância à glicose com 75 g de dextrosol, ou TOTG). Níveis glicêmicos 2 horas após o TOTG entre 140 e 199 mg/dL definem a TGD.

O critério diagnóstico foi modificado, em 1997, pela *American Diabetes Association* (ADA), posteriormente aceito pela Organização Mundial da Saúde (OMS) e pela Sociedade

Brasileira de Diabetes (SBD) (Association, 2003; Engelgau et al., 1997).

As modificações foram realizadas com a finalidade de prevenir de maneira eficaz as complicações micro e macrovasculares do DM (Fuller et al., 1980; Charles et al., 1991; Group, 1999).

Atualmente são três os critérios aceitos para o diagnóstico de DM:

- sintomas de poliúria, polidipsia e perda ponderal acrescidos de glicemia casual acima de 200 mg/dL. Compreende-se por glicemia casual aquela realizada a qualquer hora do dia, independentemente do horário das refeições (Association, 2003);
- glicemia de jejum $= 126$ mg/dL (7 milimois). Em caso de valores de glicemia pouco maiores que 126 mg/dL, o diagnóstico deve ser confirmado pela repetição do teste em outro dia (Association, 2003);
- glicemia de 2 horas pós-sobrecarga de 75 g de glicose acima de 200 mg/dL (Engelgau et al., 1997).

O teste de tolerância à glicose deve ser efetuado com os cuidados preconizados pela OMS, com colheita para diferenciação de glicemia em jejum e 120 minutos após a ingestão de glicose.

É reconhecido um grupo intermediário de indivíduos em que os níveis de glicemia não preenchem os critérios para o diagnóstico de DM. São, entretanto, muito elevados para serem considerados normais ("hiperglicemia intermediária") (Association, 1997). Nesses casos foram consideradas as categorias de glicemia de jejum alterada e tolerância à glicose diminuída, cujos critérios são apresentados na Tabela 2.1.

Tabela 2.1: Valores de glicose plasmática (em mg/dL) para diagnóstico de diabetes *mellitus* e seus estágios pré-clínicos.

Categoria	**Jejum**	**2h após 75g de glicose**	**Casual**
Glicemia Normal	<100	<140	-
Tolerância a Glicose Diminuída	> 100 a < 126	≥ 140 a < 200	-
Diabetes *mellitus*	≥ 126	≥ 200	≥ 200 (com sintomas clássicos)

Em tal tabela,

- o jejum é definido como a falta de ingestão calórica por no mínimo 8 horas;
- casual refere-se à glicemia plasmática casual, ou seja, aquela realizada a qualquer hora do dia, sem se observar o intervalo desde a última refeição;

- os sintomas clássicos de DM incluem poliúria, polidipsia e perda não-explicada de peso.

Deve-se lembrar que o diagnóstico de DM deve sempre ser confirmado pela repetição do teste em outro dia, a menos que haja hiperglicemia inequívoca com descompensação metabólica aguda ou sintomas óbvios de DM.

Glicemia de Jejum Alterada

A glicemia de jejum alterada é caracterizada por:

- glicemia de jejum acima de 100 mg/dL e abaixo de 126 mg/dL. Há discussões para alteração do ponto de corte para 100 mg/dL.
- tolerância à glicose diminuída, em outras palavras, quando, após uma sobrecarga de 75 g de glicose, o valor de glicemia de 2 horas se situar entre 140 e 199 mg/dL (Engelgau et al., 1997; Fuller et al., 1980; Charles et al., 1991; Group, 1999; on the Diagnosis e of Diabetes Mellitus, 2003)

Observa-se que indivíduos com hiperglicemia intermediária apresentam alto risco para o desenvolvimento do diabetes. São também fatores de risco para doenças cardiovasculares, fazendo parte da assim chamada síndrome metabólica, um conjunto de fatores de risco para diabetes e doença cardiovascular. Um momento do ciclo vital em que a investigação da regulação glicêmica alterada está bem padronizada é na gravidez, em que a tolerância à glicose diminuída é considerada uma entidade clínica denominada diabetes gestacional. O emprego do termo diabetes nessa situação transitória da gravidez é justificado pelos efeitos adversos à mãe e ao feto, que podem ser prevenidos/atenuados com tratamento imediato, às vezes insulínicos.

O método preferencial para determinação da glicemia é sua aferição no plasma. O sangue deve ser coletado em um tubo com fluoreto de sódio, centrifugado, com separação do plasma, que deverá ser congelado para posterior utilização. Caso não se disponha desse reagente, a determinação da glicemia deverá ser imediata ou o tubo mantido a 4°C por, no máximo, 2 horas (Bennet, 1994).

Teste Oral de Tolerância à Glicose com 75 g de dextrosol

Para a realização do teste de tolerância à glicose oral algumas considerações devem ser levadas em conta:

- período de jejum entre 10 e 16 horas;
- ingestão de pelo menos 150 g de glicídios nos três dias anteriores à realização do teste;
- atividade física normal;
- comunicar a presença de infecções, ingestão de medicamentos ou inatividade;
- utilizar $1,75$ g de glicose por quilograma de peso, até o máximo de 75 g (Bennet, 1994).

Aos indivíduos com GJA e/ou TGD, aplica-se, então, a expressão pré-diabetes, em virtude do alto risco de que venham a desenvolver DM no futuro. Tais condições representam um estado intermediário de alteração do metabolismo da glicose, não devendo ser encaradas como uma condição benigna, uma vez que aumentam em até 2 vezes a mortalidade cardiovascular. Cerca de metade dos pacientes portadores de TGD preenchem os critérios de síndrome metabólica. A progressão para DM nos pacientes com GJA é de $6-10\%$ por ano, enquanto que a incidência cumulativa de DM nos portadores de GJA e TGD é da ordem de 60% em 6 anos. No entanto, tais condições não devem ser encaradas como entidades clínicas isoladas e distintas, mas como fatores de risco para DM, assim como para doença cardiovascular. Com base nisso, recentemente a Academia Americana de Diabetes definiu as chamadas "Categorias de Risco Aumentado para Diabetes", nomenclatura vista por vários autores como mais adequada do que o termo pré-diabetes, uma vez que nem todos os indivíduos com esta condição evoluirão para DM. Dentro destas categorias de risco aumentado encontram-se, além da GJA e TGD, aqueles com níveis de hemoglobina glicada (A1C) entre $5,7$ e $6,4\%$.

Nos últimos anos, o interesse no estudo desta fase que antecede o DM vem aumentando exponencialmente. Ensaios clínicos aleatórios mostraram que aos indivíduos de alto risco de progressão para DM (GJA, TGD ou ambos) podem ser oferecidas intervenções que diminuam tal taxa de progressão. Estas medidas incluem:

- modificação do estilo de vida, a qual se mostrou ser muito eficaz com redução significativa do risco;
- uso de medicações (metformina, acarbose, orlistat, tiazolidinedionas e outros), os quais reduzem em graus variados tais taxas de progressão da doença.

O *Finish Diabetes Prevention Study* (DPS) e o *Diabetes Prevention Study* (DPP) mostraram que mudanças no padrão alimentar e na atividade física implicaram numa redução do risco

de progressão para DM de até 58%. O DPP, o qual testou a metformina (MTF), e o STOP-NIDDM, o qual testou acarbose, identificaram uma redução no risco de progressão para DM de 31% e 32%, respectivamente. O estudo XENDOS, o qual utilizou orlistat por 4 anos em indivíduos obesos e portadores de pré-diabetes, mostrou uma redução de 37% na progressão para DM nestes indivíduos. O ACT-NOW, o qual encontra-se em andamento, avaliará o impacto da pioglitazona neste contexto. O estudo NAVIGATOR, o qual avaliou o papel na nateglinida e do valsartan sobre a progressão para DM, no entanto, não encontrou redução de risco alguma.

A ADA, em sua mais recente diretriz (2011) recomenda, de modo consensual, a metformina (MTF) como única droga a ser considerada no estado de pré-diabetes, em virtude do baixo custo, segurança e persistência de seu efeito a longo prazo. É válido, no entanto, registrar que foi significativamente menos eficaz do que modificação do estilo de vida e atividade física, as quais indubitavelmente devem ser sempre tentadas ao máximo. Ela deve, portanto, ser considerada para aqueles pacientes de muito alto risco (vários fatores de risco para DM e/ou hiperglicemia progressiva e de grande magnitude). Ressalta-se, ainda, que no estudo DPP ela foi mais eficaz até do que a modificação do estilo de vida nos indivíduos com índice de massa corporal maior que 35 $\mathrm{kg/m^2}$ e não foi mais eficaz do que o placebo naqueles com idade superior a 60 anos.

Há décadas o diagnóstico de DM vem se baseando na GJ e no TOTG, utilizando os níveis de GJ e sua associação com retinopatia para se definir o ponto de corte acima do qual o risco de comprometimento da retina aumenta. Com base nisso, chegou-se aos pontos de corte de 126 mg/dL em jejum e 200 mg/dL após a sobrecarga de glicose anidra (Association, 1997; Association, 2011).

Hemoglobina Glicada

A hemoglobina glicada, também conhecida como glicohemoglobina ou HbA1C, embora seja utilizada desde 1958 como ferramenta na avaliação do controle glicêmico de diabéticos, passou a ser cada vez mais empregada e aceita pela comunidade científica após 1993, quando foi validada pelos estudos DCCT (*Diabetes Control and Complications Trial*) e UKPDS (*United Kingdom Proscpective Diabetes Study*). A A1C é sabidamente um marcador de hiperglicemia crônica, refletindo a média dos níveis glicêmicos nos últimos 2 a 3 meses. Tem impacto crucial no acompanhamento dos diabéticos, uma vez que possui uma boa correlação com lesão microvascular e, em menor proporção, com lesão macrovascular. Até pouco tempo sua utilidade era apenas para acompanhamento do controle glicêmico, e não para fins diagnósticos, uma vez que não havia padronização adequada do método. Atualmente já existe padronização do teste, que deve ser realizado pelo método de cromatografia líquida

de alto desempenho (HPLC). O HPLC foi validado em diferentes populações com uma boa reprodutibilidade entre elas e permanece estável após a coleta, o que não ocorre quando se afere a glicose diretamente. É válido lembrar que, mesmo quando se realiza a dosagem da glicemia nas condições ideais, há chance de erro pré-analítico, de modo que reduções na ordem de 3 a 10 mg/dL na glicemia plasmática podem ocorrer mesmo em não diabéticos, determinando erro de até 12% dos indivíduos. A determinação da A1C, além de não requerer jejum, tem as seguintes vantagens:

- maior estabilidade pré-analítica;
- menor interferência de outras condições agudas que possam interferir com a glicemia, como infecções e outros estresses metabólicos.

Recomenda-se que os laboratórios clínicos usem preferencialmente os métodos de ensaio certificados pelo *National Glycohemoglobin Standardization Program* (NGSP) com rastreabilidade de desempenho analítico ao método utilizado no DCCT (HPLC).

Com base nisso, em 2009, após publicação em seu compêndio oficial, a ADA passou a adotar a hemoglobina glicada como mais uma ferramenta diagnóstica para o DM. Valores de A1C maiores ou iguais a 6,5% indicam o diagnóstico de DM (Vide Tabela 2.2). O ponto de corte de 6,5% não é arbitrário, e representa o ponto de inflexão da curva de prevalência de retinopatia, assim como ocorre com os valores diagnósticos da GJ e TOTG (Association, 2011). Os já consagrados e conhecidos critérios diagnósticos de DM baseados na GJ e no TOTG permanecem válidos e inalterados. E, conforme apresentado na Tabela 2.2 deve-se observar que na ausência de hiperglicemia inequívoca os critérios diagnósticos de DM (1 a 3) devem ser repetidos e confirmados.

Tabela 2.2: Critérios Diagnósticos de Diabetes *Mellitus*.

Critério 1	A1C maior ou igual a 6,5%
Critério 2	GJ maior ou igual a 126 mg/dL
Critério 3	TOTG maior ou igual a 200 mg/dL
Critério 4	Glicemia randômica (casual) maior ou igual a 200 mg/dL em pacientes com sintomas clássicos de hiperglicemia ou em crise hiperglicêmica

Capítulo 3

Métodos de Avaliação da Variabilidade da Frequência Cardíaca

Nos últimos anos, observa-se interesse crescente no desenvolvimento de métodos que possam descrever o comportamento das oscilações cardiovasculares. Entre os vários métodos disponíveis para avaliar a variabilidade da frequência cardíaca destacam-se os métodos no domínio do tempo e no domínio da frequência. Além disso, métodos de abordagem não linear também têm sido desenvolvidos (Task Force, 1996; Acharya et al., 2006).

3.1 Índices Lineares

3.1.1 Métodos no Domínio do Tempo

Em um registro eletrocardiográfico (ECG) contínuo, cada complexo QRS é detectado, e os intervalos denominados normal-normal (NN) também chamados de intervalos RR, ou a frequência cardíaca instantânea, são determinados. A representação gráfica dos intervalos RR normais em função do tempo é denominada de tacograma. Variáveis simples no domínio do tempo podem ser calculadas. Entre elas, encontram-se a média dos intervalos RR, a média da frequência cardíaca, a diferença entre o intervalo RR mais longo e o mais curto, a diferença da frequência cardíaca entre o período diurno e noturno, entre outros.

Ewing et al. (1985) padronizaram um conjunto de testes para avaliar a integridade do SNA, baseados nas respostas da frequência cardíaca e da pressão arterial a estímulos padronizados. Na década de 1970, eles aplicaram este conjunto de testes sobre as diferenças de intervalos RR curtos para detectar neuropatia autonômica em pacientes diabéticos. De uma

forma geral, os testes, propostos por Ewing et al. (1985), que avaliam a variação da frequência cardíaca referem-se à integridade do sistema nervoso parassimpático e os que avaliam a variação da pressão arterial, ao sistema nervoso simpático. Apesar de serem muito usados na prática clínica e em laboratório, ainda há dúvidas quanto à escolha do melhor teste. Além disto, eles apresentam problemas de padronização, baixa sensibilidade e reprodutibilidade, necessitam de cooperação por parte do paciente e fornecem informação exclusivamente sobre um período restrito de tempo (Ewing et al., 1985).

Com o passar do tempo, inúmeros índices começaram a ser calculados. Os diferentes índices, em sua maioria, são calculados utilizando apenas os intervalos RR normais, desprezando-se os artefatos e batimentos ectópicos. Os índices mais populares atualmente, com as suas abreviações conhecidas internacionalmente, são os seguintes (Task Force, 1996): desvio padrão de todos intervalos RR normais (SDNN), média do desvio padrão dos intervalos RR normais calculados em intervalos de cinco minutos (SDNNi), desvio padrão das médias dos intervalos RR normais calculadas em intervalos de cinco minutos (SDANN$_i$), raiz quadrada da média das diferenças sucessivas entre intervalos RR normais adjacentes (RMSSD), e percentagem das diferenças entre intervalos RR normais adjacentes que excedem a 50 milissegundos (pNN50).

Métodos Estatísticos

A partir de séries RR, medidas estatísticas podem ser calculadas (Task Force, 1996). Estas podem ser divididas em duas classes,

- as derivadas a partir de medidas diretas dos intervalos RR ou frequência cardíaca instantânea;
- as derivadas a partir das diferenças de intervalos RR.

Todas estas medidas podem ser aplicadas na análise de toda a série RR ou sobre pequenos segmentos da mesma. Alguns métodos são:

1. RR$_{\text{mean}}$: Média de todos intervalos RR, em ms,

$$\text{RR}_{\text{mean}} \equiv \overline{\text{RR}} = \frac{1}{N}\sum_{i=1}^{N}(\text{RR}_i)\,; \tag{3.1}$$

2. SDNN: Desvio padrão de todos os intervalos RR, em ms,

$$\text{SDNN} = \sqrt[2]{\frac{1}{N-1}\sum_{i=1}^{N}\left(\text{RR}_i - \overline{\text{RR}}\right)^2}; \tag{3.2}$$

3. SDANN: Desvio padrão das médias dos intervalos RR em todos os segmentos de 5 minutos do registro, em ms,

$$\text{SDANN} = \sqrt[2]{\frac{1}{N-n-1}\sum_{i=1}^{N-n}\left(\frac{1}{n}\sum_{i}^{i+n}\text{RR}_i - \overline{\text{RR}}\right)^2}, \tag{3.3}$$

onde, $n \equiv$ número de intervalos de 5 min e $N =$ tamanho da série;

4. RMSSD: Raiz quadrada da média da soma dos quadrados das diferenças entre intervalos RR adjacentes, em ms,

$$\text{RMSSD} = \sqrt[2]{\frac{1}{N-1}\sum_{i=1}^{N-1}(\text{RR}_{i+1} - \text{RR}_i)^2}; \tag{3.4}$$

5. SDNN_i: Média do desvio padrão dos intervalos RR em todos os segmentos de 5 min do registro, em ms,

$$\text{SDNN}_i = \frac{1}{N-n-1}\sum_{i=1}^{N-n}\left(\sqrt[2]{\frac{1}{n}\sum_{i=i}^{i+n}\left(\text{RR}_i - \overline{\text{RR}}_n\right)^2}\right), \tag{3.5}$$

onde, $n \equiv$ número de intervalos de 5 min e $\overline{\text{RR}}_n =$ média em cada segmento de 5 min;

6. NN50: Número de pares de intervalos RR adjacentes com diferença de duração maior que 50 ms em toda a gravação.

$$\text{NN50} = \sum_{i=1}^{N} cont, \text{ tal que: } cont = \begin{cases} 1 & \text{se} \quad |\text{RR}_{i+1} - \text{RR}_i| > 50\text{ms} \\ 0 & \text{se} \quad |\text{RR}_{i+1} - \text{RR}_i| \leq 50\text{ms} \end{cases}; \tag{3.6}$$

7. pNN50: NN50 dividido pelo número total de intervalos RR, ou porcentagem de NN50 em relação ao registro, em %, a saber

$$\text{pNN50} = \frac{\text{NN50}}{N-1} \times 100\%. \tag{3.7}$$

Observações:

- Como a variância é matematicamente igual à potência total da análise espectral, SDNN reflete todas as componentes cíclicas responsáveis pela variabilidade dentro do período de gravação. SDNN é dependente do tamanho do período de gravação;
- SDANN e SDNN_{index} (SDNN_i) medem a variabilidade da frequência devido aos ciclos maiores e menores que 5 min, respectivamente;
- Três variações no cálculo do NN50 são possíveis, visto que se pode contar todos os intervalos RR que atendem o método ou somente pares nos quais o primeiro ou o segundo intervalo são maiores que 50 ms.

Métodos Geométricos

As séries de intervalos RR também podem ser convertidas em um padrão geométrico, como a distribuição de densidade amostral das durações dos intervalos RR, a distribuição de densidade amostral das diferenças entre intervalos RR adjacentes, entre outros. Uma simples fórmula é utilizada para analisar a variabilidade com base em padrão geométrico e/ou gráfico resultante. Geralmente, três abordagens são utilizadas nos métodos geométricos:

- uma medida básica do padrão geométrico (por exemplo, a largura do histograma da distribuição) é convertida em uma medida da VFC;
- o padrão geométrico é interpolado por uma forma geométrica definida (por exemplo, o histograma da distribuição é aproximado por um triângulo ou por uma curva exponencial) e, então, os parâmetros da forma geométrica são utilizados;
- a forma geométrica é classificada em várias categorias baseadas nos padrões representando diferentes classes de VFC.

A maioria dos métodos geométricos requer que a sequência de intervalo RR seja medida sobre ou convertida para uma escala discreta que permita a construção de histogramas suaves. Normalmente os histogramas são gerados com intervalos de aproximadamente 8 ms de comprimento (precisamente $7,8125$ ms $= 1/128$ s) que corresponde à precisão dos equipamentos comerciais atuais. Alguns métodos geométricos são:

1. *Triangular index* (Δ *index*): Número total de intervalos RR dividido pelo número de intervalos RR com frequência modal. Calculado por meio da integral da função de densidade da distribuição dividida pelo máximo da distribuição. Utilizando uma escala discreta, o valor do índice é dado pelo número total de intervalos RR dividido pelo número de intervalos RR no intervalo modal (maior "frequência" do histograma) o qual é dependente do comprimento do intervalo. Sendo assim, se a frequência de amostragem for diferente de 128 Hz, o tamanho dos intervalos deve ser normalizado.
2. TINN: Largura de linha de base da distribuição medida como a base de um triângulo que aproxima a distribuição dos intervalos RR (interpolação triangular utilizando o mínimo das diferenças quadradas), em ms;
3. *Differential index*: Coeficiente ϕ da curva exponencial negativa $k \cdot e^{-\phi t}$, que representa a melhor aproximação do histograma das diferenças absolutas entre intervalos RR adjacentes.

Observa-se que o *triangular index* tem alta correlação com o desvio padrão de todos os intervalos RR (SDNN). O TINN não sofre a influência de batimentos ectópicos e artefatos, pois na geração do índice os intervalos RR representativos de batimentos ectópicos e artefatos fazem parte dos extremos da distribuição de densidade de intervalos RR e, consequentemente, ficam fora do triângulo.

Como várias medidas no domínio do tempo se correlacionam fortemente uma com a outra, os índices estatísticos e geométricos utilizados foram: média dos intervalos RR, SDNN, RMSSD, pNN50 e o Δ *index*.

3.1.2 Métodos no Domínio da Frequência

Os métodos no domínio da frequência permitem analisar individualmente o sistema nervoso autônomo, simpático e parassimpático, em variadas situações fisiológicas e patológicas e sua relação com outros sistemas que também interferem na VFC. Portanto, permitem identificar quais oscilações são responsáveis pela variabilidade.

Os métodos no domínio da frequência conseguem individualizar e quantificar os diferentes componentes de frequência de uma oscilação complexa. A análise espectral é o método no domínio da frequência mais utilizado, tendo se tornado valioso instrumento na avaliação do sistema cardiovascular. No estudo da variabilidade da frequência cardíaca, usualmente esse método identifica oscilações em três bandas de frequência. Essas oscilações podem ser classificadas como sendo de alta frequência, entre $0,2$ a $0,4$ Hz, de média frequência, ao redor de $0,1$ Hz, e de baixa frequência, entre $0,02$ a $0,07$ Hz.

Nos métodos no domínio da frequência, a estimativa da densidade espectral de potência (do inglês, *Power Spectrum Density*, ou PSD) é calculada para a série de intervalos RR. Assim, a PSD nos fornece informações básicas de como a potência (por exemplo, a variância) se distribui em função da frequência.

Na análise da VFC, a estimativa PSD é geralmente obtida utilizando métodos não paramétricos baseados na transformada discreta de Fourier (do inglês, *Discrete Fourier Transform*, ou DFT) por meio do algoritmo FFT (do inglês, *Fast Fourier Transform* ou FFT) ou por meio do método paramétrico baseado no modelo autorregressivo (AR).

As técnicas de estimação da PSD, baseadas na modelagem não paramétrica, dependem da amostragem uniforme dos dados, visto que os intervalos de tempo entre as ondas R não são constantes. Assim, a série de intervalos RR é convertida em uma série uniformemente amostrada por meio de métodos de interpolação, antes de se obter a PSD. Entre as técnicas

de interpolação, utiliza-se a *cubic spline*, uma das mais populares, que determina polinômios cúbicos que mais se aproximam da curva entre cada par de pontos dados.

Transformada Rápida de Fourier (FFT)

O cálculo da DFT pode ser computacionalmente intenso para sinais longos. Se o sinal tiver um comprimento total de N amostras, o número de operações complexas necessárias para o cálculo da DFT é de $2N^2$. Note-se que a complexidade computacional da DFT aumenta exponencialmente com o comprimento do sinal. No entanto, se N for uma potência de 2 $(256, 512, 1.024, \ldots, 2^n)$, é possível usar uma implementação que reduz o número de operações complexas para $2N \times \log_2 N$. Este algoritmo para o cálculo rápido da DFT é chamado de transformada rápida de Fourier (FFT). Esta implementação explora simetrias e redundâncias presentes no cálculo da DFT para torná-lo mais rápido.

Note-se que a FFT e a DFT produzem resultados idênticos, ou seja, são a mesma transformada. A única diferença diz respeito ao número de operações necessárias para seu cálculo. Enquanto a DFT necessita de 2.097.152 operações para o cálculo da transformada de um sinal de 1.024 amostras, a FFT utiliza apenas 20.480. A diferença é ainda mais significativa para sinal de maior comprimento. Caso o comprimento do sinal não seja uma potência de 2, pode-se completar o sinal com zeros para que se possa usar a FFT para calcular o espectro de frequência de maneira mais eficiente. Por exemplo, se o sinal tem 400 amostras, a DFT gastaria 320.000 operações. No entanto, ao se inserir 112 zeros ao final do sinal, completando o número de amostras para 512, pode-se utilizar a FFT, que gastará apenas 9.216 operações. O processo de inserir zeros ao final do sinal, utilizado para aumentar o número de amostras a uma potência de 2, é chamado de *zero-padding*.

Recentemente, pesquisadores do *Massachusetts Institute of Technology* (MIT) desenvolveram outro algoritmo rápido para o cálculo da DFT. Este algoritmo, denominado de *Fastest Fourier Transform in the West* (FFTW), é capaz de acelerar o cálculo da DFT de tamanhos arbitrários sem a necessidade de realizar o *zero-padding* (Frigo e Johnson, 1997; Frigo e Johnson, 1998; Frigo et al., 1999; Frigo e Johnson, 2005; Johnson e Frigo, 2007; Johnson e Frigo, 2008).

Modelo Autorregressivo (AR)

A análise espectral usando a DFT (ou a FFT) é muito útil para a observação de certas propriedades do sinal biológico, bem como extrair informações que possam não estar

evidentes no domínio do tempo. Por exemplo, ao se calcular a DFT de um sinal de variabilidade da frequência cardíaca (VFC), observam-se duas componentes principais no domínio da frequência. A de frequência mais baixa diz respeito à modulação conjunta simpática e parassimpática do sistema nervoso. Já a componente de frequência mais alta está relacionada à influência parassimpática da variabilidade cardíaca. Avaliando a relação entre a amplitude dessas componentes, é possível, por exemplo, diagnosticar deficiências no controle do sistema nervoso sobre a frequência cardíaca, e avaliar o risco de morte súbita, entre outros.

No entanto, em muitos casos, a análise espectral usando a DFT não oferece resultados tão claros como se poderia esperar. Isso se deve a vários fatores, como curtos intervalos de observação e variações de comportamento do sinal durante o período de aquisição. Por isso, alguns pesquisadores utilizam outros modelos para representar sinais de VFC no domínio da frequência. O mais popular destes modelos é a modelagem autorregressiva (AR). Esta modelagem consiste em tentar aproximar a envoltória do espectro de frequência por meio de uma equação só de pólos. Pólos são bons para modelar picos de energia, mas não são tão bons para modelar vales. Todavia, usando um maior número de coeficientes, consegue-se aproximar bem o resultado ao obtido com a DFT. A vantagem é que o espectro de potência calculado com o modelo AR é, em geral, mais claro que o espectro de Fourier, pois quando há muitos picos em frequências próximas a modelagem AR representa a energia desses picos com um único pico mais largo, tornando o gráfico mais limpo (Figura 3.1). A desvantagem do modelo AR é a complexidade na seleção da ordem do modelo. Na Figura 3.1, os vários picos próximos às faixas de $0,18$ Hz e $0,23$ Hz são representados por um único pico mais largo em cada frequência central, tornando o gráfico mais claro e a interpretação imediata.

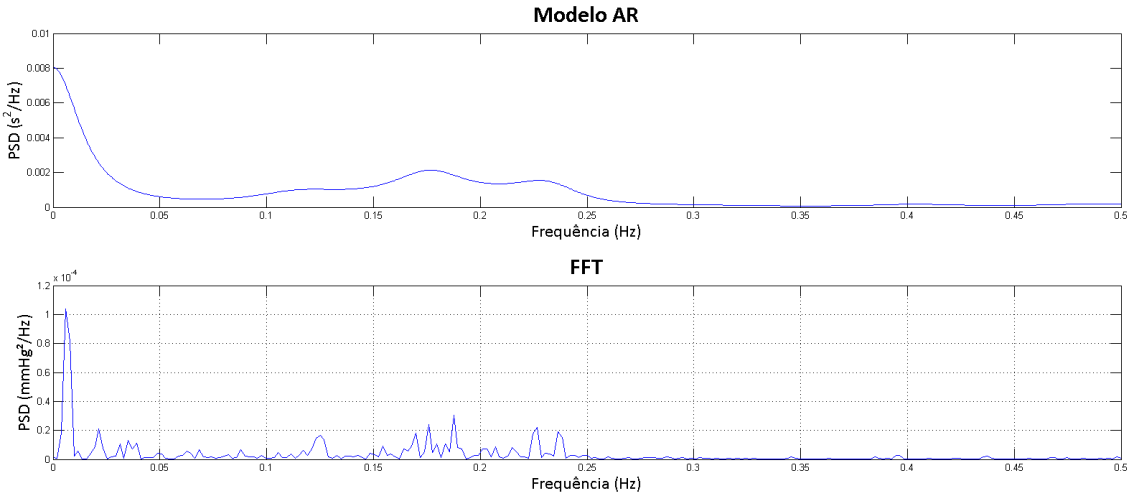


Figura 3.1: Espectro de potência de Fourier e espectro de potência calculado com modelagem autorregressiva, para um sinal de VFC.

Conforme apresentado, o espectro da VFC pode ser calculado com base no método do

periodograma de *Welch* analisando a FFT (média do espectro dos segmentos, o que implica em menor variância da estimativa de PSD) e utilizando a modelagem paramétrica autorregressiva (AR).

Nos métodos no domínio da frequência, a estimativa da densidade espectral de potência (PSD) foi calculada por meio do modelo autorregressivo (AR) com ordem de modelo fixa em 16 (Boardman et al., 2002; Carvalho et al., 2003). Os valores de potência absoluta para cada banda de frequência foram calculados por meio do método dos resíduos, porém pode-se calcular por meio de integração da área sob a curva gerada pelo espectro de potência do sinal ao longo dos limites de cada banda (Vide Apêndice A).

Nos registros curtos de intervalos RR, identificam-se componentes espectrais (Task Force, 1996) de muito baixa frequência (*Very Low Frequency*, VLF), de baixa frequência (*Low Frequency*, LF) e alta frequência (*High Frequency*, HF).

A análise espectral aplicada ao estudo da variabilidade cardíaca emprega índices associados aos mecanismos regulatórios com base em duas bandas: baixa frequência (LF) e alta frequência (HF). A medida da energia da componente espectral HF ($0,15$ a $0,4$ Hz) está relacionada à modulação parassimpática da variabilidade cardíaca (Task Force, 1996; Akselrod et al., 2006), enquanto que a energia da componente espectral LF ($0,04$ a $0,15$ Hz) está relacionada à modulação conjunta simpática e parassimpática (Malliani et al., 1994). A razão LF/HF, denominada de balanço simpatovagal, é sugerida como uma medida de atividade relativa entre as duas componentes (Task Force, 1996). Os índices LF, HF e LF/HF foram extensivamente estudados e associados a uma grande variedade de condições clínicas (Berntson et al., 1997; Task Force, 1996; Akselrod et al., 2006). Entre elas temos a alteração destes índices em função de doenças como a neuropatia diabética autonômica (Pagani et al., 1988; Freeman et al., 1991).

A medida das componentes de potência VLF, LF, e HF é, normalmente, feita em valores absolutos de potência (ms^2), mas LF e HF podem também ser medidas em unidades normalizadas (n.u.) que representam o valor relativo de cada componente de potência em relação à potência total menos a componente VLF, enfatizando o comportamento controlado e balanceado dos dois ramos do SNA (SNS e SNP).

A análise espectral também deve ser utilizada para analisar séries de intervalos RR em períodos de 24 h (registros longos). O resultado então inclui a componente de ultra baixa frequência (*Ultra-Low Frequency*, ULF), além das componentes VLF, LF e HF.

Os índices no domínio da frequência são descritos abaixo:

1. ULF: Potência na faixa de ultra baixa frequência ($f \leq 0,003$ Hz), em ms^2;

2. VLF: Potência na faixa de muito baixa frequência ($f \leq 0,04$ Hz), em ms^2;

3. LF: Potência na faixa de baixa frequência ($0,04 < f \leq 0,15$ Hz), em ms^2;

4. LF$_{\text{norm}}$: Potência na faixa de baixa frequência em unidades normalizadas, em n.u.

$$\text{LF}_{\text{norm}} = 100 \times \frac{\text{LF}}{\text{Potência Total} - \text{VLF}}. \tag{3.8}$$

5. HF: Potência na faixa de alta frequência ($0,15 < f \leq 0,4$ Hz), em ms^2;

6. HF$_{\text{norm}}$: Potência na faixa de alta frequência em unidades normalizadas, em n.u.

$$\text{HF}_{\text{norm}} = 100 \times \frac{\text{HF}}{\text{Potência Total} - \text{VLF}}. \tag{3.9}$$

7. LF/HF: Razão $\frac{\text{LF}[\text{ms}^2]}{\text{HF}[\text{ms}^2]}$;

8. Potência Total: Energia de todos os intervalos RR ($f \leq 0,4$ Hz), em ms^2.

3.2 Índices Não Lineares

Nos últimos anos, o crescente interesse pelo comportamento dos sistemas dinâmicos não lineares, em diversas áreas da ciência, começou a influenciar também o estudo da regulação cardiovascular. A não linearidade está presente em todos os sistemas vivos, produzindo comportamentos irregulares, que não são identificados corretamente pelas técnicas estatísticas convencionais. Diversas ferramentas específicas para o estudo dos sistemas dinâmicos não lineares já foram utilizadas para estudar as oscilações do sistema cardiovascular.

Assim, em função da complexidade do sistema cardiovascular é adequado supor que mecanismos não lineares estejam envolvidos na geração da VFC (Task Force, 1996). Portanto, além das técnicas lineares convencionais, utilizam-se índices não lineares na análise da VFC, a saber: Gráfico de Poincaré, Entropia Aproximada, Entropia Amostral, Dimensão Fractal, Análise das Flutuações Destendenciadas e os Gráficos de Recorrência.

3.2.1 Gráfico de Poincaré

Uma análise geométrica da VFC pode ser feita com o uso do gráfico de Poincaré. O gráfico de Poincaré dos intervalos RR é dado pelo conjunto de pontos $(\mathbf{RR}_i, \mathbf{RR}_{i+1})$. Desta forma, o resultado consiste em uma nuvem de pontos, a partir da qual são determinados

os parâmetros quantitativos. Como as coordenadas dos pontos são definidas por um par de intervalos RR consecutivos, podemos observar que há uma certa quantidade de intervalos consecutivos que são aproximadamente iguais, e outros que apresentam certa variação nas coordenadas x e y, implicando em certa dispersão em torno da linha identidade (pontos com coordenadas iguais). A linha identidade divide os pontos que caracterizam a atuação do sistema nervoso simpático (intervalo RR diminuindo ou, de forma equivalente, a frequência cardíaca está aumentando) dos pontos que caracterizam a atuação do sistema nervoso parassimpático (intervalo RR aumentando).

Implementação do Gráfico de Poincaré

Considere o vetor de dados, $\mathbf{x} = (\mathrm{RR}_1, \mathrm{RR}_2, ..., \mathrm{RR}_N)$, correspondente à série RR de comprimento N. Sejam:

$$\mathbf{x}^+ = (\mathrm{RR}_1, \mathrm{RR}_2, \ldots, \mathrm{RR}_{N-1}) \text{ e } \mathbf{x}^- = (\mathrm{RR}_2, \mathrm{RR}_3, \ldots, \mathrm{RR}_N). \tag{3.10}$$

Normalmente, estes vetores são descritos como $\mathbf{RR}_n(i)$ e $\mathbf{RR}_{n+1}(i)$.

O gráfico de Poincaré consiste de todos os pares ordenados:

$$(\mathbf{x}^+, \mathbf{x}^-),\ i = 1, 2, \ldots, N-1, \tag{3.11}$$

sendo caracterizado pelos índices SD1, SD2 e SD1/SD2, alguns dos quais são apresentados na Figura 3.2.

SD2 é definido como o desvio padrão da projeção do gráfico sobre a linha identidade ($y = x$), SD1 é o desvio padrão da projeção sobre a linha perpendicular a linha identidade ($y = -x + 2 \times \overline{\mathrm{RR}}$) (Brennan et al., 2001). Define-se $\mathrm{SD1} = \sqrt{\mathrm{var}(\mathbf{x}_1)}$ e $\mathrm{SD2} = \sqrt{\mathrm{var}(\mathbf{x}_2)}$, onde $\mathbf{x}_1 = \frac{\mathbf{x}^+ - \mathbf{x}^-}{\sqrt{2}}$ e $\mathbf{x}_2 = \frac{\mathbf{x}^+ + \mathbf{x}^-}{\sqrt{2}}$. Em outras palavras, $\mathbf{x}_1$ e $\mathbf{x}_2$ correspondem à rotação de $\mathbf{x}^+$ e $\mathbf{x}^-$ de $\theta = \pi/4$, ou seja,

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \cos\pi/4 & -\sin\pi/4 \\ \sin\pi/4 & \cos\pi/4 \end{bmatrix} \times \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \end{bmatrix}. \tag{3.12}$$

Se definirmos SDRR como a raiz quadrada da variância de toda a série temporal, temos

$$\mathrm{SDRR} = \sqrt{\mathrm{var}(\mathbf{x})} \equiv \mathrm{std}(\mathbf{x}), \tag{3.13}$$

que é aproximadamente igual a

$$\mathrm{SDRR} \approx \frac{1}{\sqrt{2}} \sqrt{\mathrm{SD1}^2 + \mathrm{SD2}^2}. \tag{3.14}$$

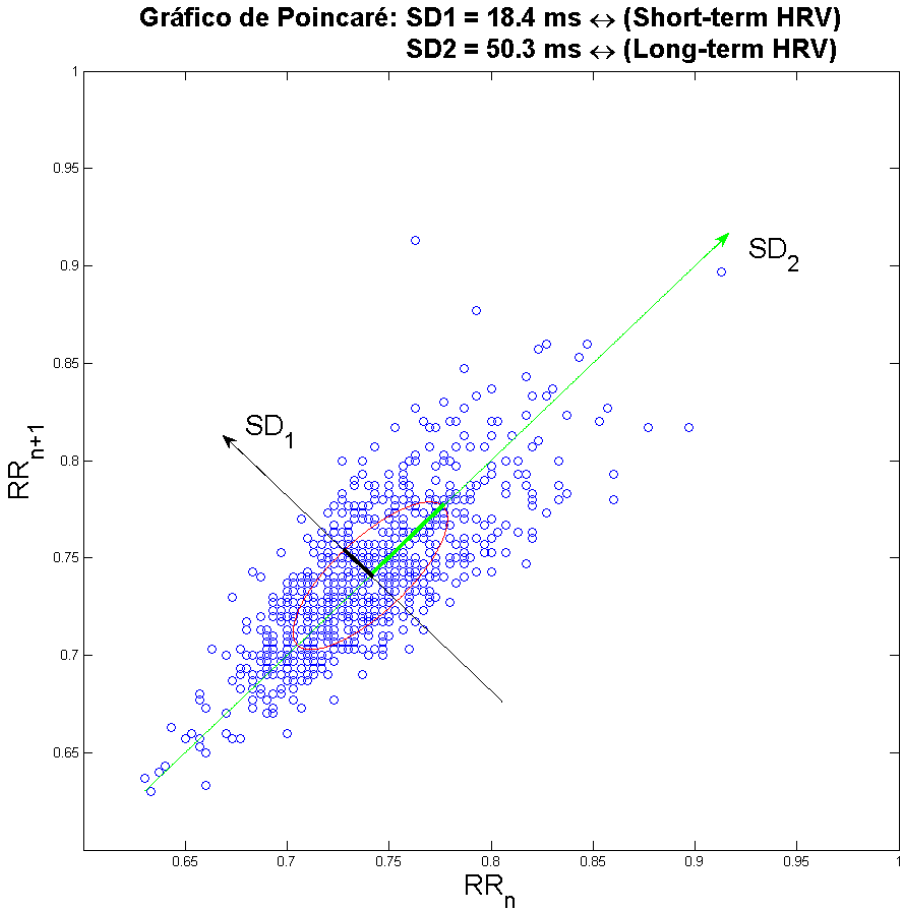


Figura 3.2: O gráfico de Poincaré obtido de uma de série RR.

Verifica-se que a relação anterior não é exatamente igual em função do fato de que os índices SDRR, SD1 e SD2 são calculados a partir de conjunto de dados ligeiramente diferentes, como descrito anteriormente.

SD1 é o índice que reflete a variabilidade instantânea da série, ou ainda, a variabilidade batimento a batimento (do inglês, *short-term variability*), SD2 representa a variabilidade em registros de longa duração (do inglês, *long-term variability*), e a relação SD1/SD2 a razão entre as variações curta e longa dos intervalos RR (Acharya et al., 2006). Normalmente, desenha-se sobre o gráfico de Poincaré uma elipse com eixos $(\mathrm{SD1}, \mathrm{SD2})$ centrada sobre os valores médios de $\mathbf{x}^+$ e $\mathbf{x}^-$ $(\bar{\mathbf{x}}^+, \bar{\mathbf{x}}^-)$. Em diversos artigos isto é descrito erroneamente como ajustando uma elipse (do inglês, *fitting an elipse*) sobre o gráfico, visto que nenhum ajuste é feito (Brennan et al., 2001). De fato, a elipse é desenhada no gráfico e seu objetivo é auxiliar a visualização dos índices de VFC envolvidos por meio de uma figura geométrica.

A análise qualitativa do gráfico de Poincaré é baseada na análise das figuras geradas. Algumas são classificadas como (Tulppo et al., 1988):

1. figura com característica de um cometa, na qual um aumento na dispersão dos intervalos RR batimento a batimento é observado com aumento nos intervalos RR. Característica de um gráfico normal;

2. figura com característica de um torpedo, com pequena dispersão global batimento a batimento (SD1) e sem aumento da dispersão dos intervalos RR de longa duração (SD2);

3. figura complexa ou parabólica, na qual duas ou mais extremidades distintas são separadas do corpo principal do gráfico, com pelo menos três pontos incluídos em cada extremidade.

3.2.2 Entropia - ApEn e SampEn

A análise da dinâmica não linear é uma abordagem poderosa para a compreensão de sistemas biológicos. Os cálculos, no entanto, geralmente requerem conjuntos grandes de dados, o que pode ser difícil, ou impossível, de se obter. Em Pincus (1991) encontra-se o desenvolvimento da teoria e método para medir a regularidade associada à entropia de *Kolmogorov* (Grassberger e Procaccia, 1983) (taxa de geração de informação nova) que pode ser aplicada a séries temporais de dados clínicos, tipicamente curtas e ruidosas.

O método analisa a correspondência de janelas de uma dada série temporal, tal que janelas mais frequentes e similares implicam em menores valores de entropia . Assim, um valor baixo de ApEn reflete um elevado grau de regularidade.

Nesse trabalho, são estudados e implementados os algoritmos para o cálculo da entropia aproximada (do inglês, *approximate entropy* ou ApEn) e entropia amostral (do inglês, *sample entropy* ou SampEn).

ApEn

ApEn é uma medida estatística de regularidade/similaridade que quantifica a imprevisibilidade das flutuações em uma série temporal, como as séries de intervalos RR. Intuitivamente, pode-se pensar que a presença de padrões de flutuação repetitivos em uma série temporal a torna mais previsível do que uma série temporal em que tais padrões estão ausentes. Uma série temporal contendo muitos padrões repetitivos tem um valor de ApEn relativamente pequeno, ao passo que um processo menos previsível (ou seja, mais complexo) tem um maior valor de ApEn.

Descrição do Algoritmo

O parâmetros N, m e r devem ser fixados para cada cálculo. N é o comprimento da série temporal $\mathbf{x}$, m é o comprimento das sequências (subséries da série original) que serão comparadas, e r é a tolerância para aceitação de correspondência das sequências. É conveniente definir a tolerância em função do desvio padrão da série de dados ($r = f(\mathrm{std}(\mathbf{x}))$, tal que $f(\cdot)$ representa uma função de seu argumento), de tal forma que sequências no conjunto de dados com diferentes amplitudes possam ser comparadas.

Para uma série temporal de N pontos, $\mathrm{RR}(j) : 1 \leq j \leq N$ forma os $N-m+1$ vetores $\mathbf{x}_m(i)$ para $\{i | 1 \leq i \leq N-m+1\}$, onde $\mathbf{x}_m(i) = \{\mathrm{RR}(i+k) : 0 \leq k \leq m-1\}$ é o vetor de tamanho m de $\mathrm{RR}(i)$ a $\mathrm{RR}(i+m-1)$. A distância entre dois destes vetores é definida como a máxima diferença das componentes escalares correspondentes, de acordo com

$$\mathrm{d}[\mathbf{x}(i), \mathbf{x}(j)] = \max\{|\mathrm{RR}(i+k) - \mathrm{RR}(j+k)| : 0 \leq k \leq m-1\}. \tag{3.15}$$

Seja B_i o número de vetores $\mathbf{x}_m(j)$ com distância menor ou igual a r de $\mathbf{x}_m(i)$ e seja A_i o número de vetores $\mathbf{x}_{m+1}(j)$ com distância menor ou igual a r de $\mathbf{x}_{m+1}(i)$. Considere a função

$$C_i^m(r) = \frac{B_i}{N-m+1}. \tag{3.16}$$

No cálculo de $C_i^m(r)$, o vetor $\mathbf{x}_m(i)$ é denominado de modelo. Quando o vetor $\mathbf{x}_m(j)$ está distante a menos de r do vetor $\mathbf{x}_m(i)$ dizemos que há correspondência de modelo e os vetores atendem ao critério de similaridade. Assim, $C_i^m(r)$ é a probabilidade que qualquer vetor $\mathbf{x}_m(j)$ esteja a uma distância menor ou igual a r de $\mathbf{x}_m(i)$.

De forma análoga, podemos definir $C_i^{m+1}(r)$ como a probabilidade que qualquer vetor $\mathbf{x}_{m+1}(j)$ esteja a uma distância menor ou igual a r de $\mathbf{x}_{m+1}(i)$, tal que

$$C_i^{m+1}(r) = \frac{A_i}{N-m}. \tag{3.17}$$

A estimativa da entropia aproximada pode ser definida como (Pincus, 1991)

$$\mathrm{ApEn}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r), \tag{3.18}$$

$$\Phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \ln[C_i^m(r)], \tag{3.19}$$

onde, $\Phi^m(r)$ é a média dos logaritmos naturais das funções $C_i^m(r)$, de acordo com a Equação 3.19. Manipulações algébricas levam a

$$\mathrm{ApEn}(m, r, N) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \ln[C_i^m(r)] - (N-m)^{-1} \sum_{i=1}^{N-m} \ln[C_i^{m+1}(r)]. \tag{3.20}$$

Quando N é grande, ApEn(m,r,N) é aproximadamente igual a $(N-m)^{-1}\sum_{i=1}^{N-m} -\ln\frac{A_i}{B_i}$, a média sobre i do negativo do logaritmo natural da probabilidade condicional de que duas sequências que são semelhantes para o comprimento m permaneçam semelhantes (a uma tolerância de r) em $m+1$, ou seja, de que $\mathrm{d}[\mathbf{x}_{m+1}(i),\mathbf{x}_{m+1}(j)] \leq r$ tal que $\mathrm{d}[\mathbf{x}_m(i),\mathbf{x}_m(j)] \leq r$.

Retomando a descrição anterior da entropia aproximada, vemos que o algoritmo ApEn é tendencioso (o valor esperado não é igual ao parâmetro estimado) e sugere maior similaridade do que a realidade, pois considera $\mathrm{d}[\mathbf{x}_m(i),\mathbf{x}_m(i)] = 0$ nos cálculos. É importante notar que o ApEn utiliza a abordagem de comparação de modelo para calcular uma probabilidade logarítmica média, primeiramente calculando a probabilidade para cada modelo. Então, o algoritmo ApEn requer que cada modelo contribua com uma probabilidade definida, diferente de zero. Ou seja, que $C_i^m(r) \neq 0$ tal que não se tenha $\ln(0)$ na Equação 3.19. Esta limitação é contornada permitindo que cada modelo seja similar a si. Formalmente, em função de $\mathrm{d}[\mathbf{x}_m(i),\mathbf{x}_m(i)] = 0 \leq r$, o algoritmo conta cada modelo como correspondente a si, fenômeno conhecido como *self-matching*. Isto garante que as funções $C_i^m(r)$ sejam maiores que zero para todo i, consequentemente evitando a ocorrência de $\ln(0)$ nos cálculos. A eliminação das auto correspondências não é trivial, visto que a remoção faria ApEn altamente sensível a *outliers*, e se houvesse um único modelo tal que não houvesse outro modelo corresponde, ApEn não poderia ser calculada devido à ocorrência de $\ln(0)$.

SampEn

Simplificadamente, os cálculos de SampEn podem ser vistos como um processo de amostragem de informações sobre a regularidade da série temporal e utilizam estatísticas amostrais para nos informar sobre a confiabilidade do resultado obtido. Existem duas grandes diferenças entre SampEn e ApEn. Primeiro, SampEn não conta os *self-matches*. Esta característica é justificada visto que a entropia é concebida como uma medida da taxa de produção de informação (Eckmann e Ruelle, 1985), e, neste contexto, comparar dados com eles próprios não tem significado. Segundo, SampEn não utiliza a abordagem de comparação de modelo ao estimar probabilidades condicionais. Para ser definido, SampEn requer apenas que um modelo encontre uma correspondência de sequências de comprimento $m+1$. Em Grasberger e Procaccia (1983) tem-se a definição de $C^m(r)$ como a média de $C_i^m(r)$, que por sua vez é a probabilidade que qualquer vetor $\mathbf{x}_m(j)$ esteja a uma distância inferior a r de $\mathbf{x}_m(i)$, ou seja,

$$C^m(r) = (N-m+1)^{-1}\sum_{i=1}^{N-m+1} C_i^m(r). \tag{3.21}$$

Isto difere da Equação 3.19 somente pelo fato de que $\Phi^m(r)$ é a média de logaritmos naturais de $C_i^m(r)$. Grasberger e Procaccia (1983) sugerem aproximar a entropia de *Kolmogorov* de

um processo, representado por uma série temporal, por

$$\lim_{r\to 0}\lim_{m\to\infty}\lim_{N\to\infty}\ln\left[\frac{C^{m+1}(r)}{C^{m}(r)}\right], \tag{3.22}$$

$$\frac{C^{m+1}(r)}{C^{m}(r)}=\frac{(N-m)^{-1}\sum_{i=1}^{N-m}A_i}{(N-m+1)^{-1}\sum_{i=1}^{N-m+1}B_i}\therefore\frac{C^{m+1}(r)}{C^{m}(r)}=\frac{(N-m+1)\sum_{i=1}^{N-m}A_i}{(N-m)\sum_{i=1}^{N-m+1}B_i}. \tag{3.23}$$

Nesta forma, no entanto, os limites tornam o algoritmo inadequado para a análise de séries temporais finitas com o ruído e os *self-matches* são contados. Por isso, duas alterações são feitas ao algoritmo para adaptá-lo. Em primeiro lugar, não se considerou *self-matches* ao calcular $C^m(r)$. Em segundo lugar, considerou-se somente os primeiros $N-m$ vetores de comprimento m, garantindo que, para $1\leq i\leq N-m$, $\mathbf{x}_m(i)$ e $\mathbf{x}_{m+1}(i)$ sejam definidos.

Define-se $B_i^m(r)$ como $(N-m-1)^{-1}$ vezes o número de vetores $\mathbf{x}_m(i)$ distantes a menos de r de $\mathbf{x}_m(j)$, tal que $1\leq i\leq N-m$, e $i\neq j$ para eliminar *self-matches*. Assim,

$$B^m(r)=(N-m)^{-1}\sum_{i=1,i\neq j}^{N-m}B_i^m(r). \tag{3.24}$$

De forma semelhante, define-se $A_i^m(r)$ como $(N-m-1)^{-1}$ vezes o número de vetores $\mathrm{x}_{m+1}(i)$ distantes a menos de r de $\mathrm{x}_{m+1}(j)$, com $1\leq i\leq N-m$ e $i\neq j$. Assim,

$$A^m(r)=(N-m)^{-1}\sum_{i=1,i\neq j}^{N-m}A_i^m(r). \tag{3.25}$$

Então, $B^m(r)$ é a probabilidade que duas sequências de tamanho m sejam correspondentes, ao passo que $A^m(r)$ é a probabilidade que duas sequências de tamanho $m+1$ sejam correspondentes. Finalmente, o parâmetro

$$\mathrm{SampEn}(m,r)=\lim_{N\to\infty}\left\{-\ln\left[\frac{A^m(r)}{B^m(r)}\right]\right\}, \tag{3.26}$$

o qual é estimado pela estatística

$$\mathrm{SampEn}(m,r,N)=-\ln\left[\frac{A^m(r)}{B^m(r)}\right]. \tag{3.27}$$

Para evitar confusão quanto aos parâmetros r e o comprimento m do vetor modelo, considera-se

$$\begin{aligned}B&=\{[(N-m-1)(N-m)]/2\}B^m(r)\\A&=\{[(N-m-1)(N-m)]/2\}A^m(r)\end{aligned} \tag{3.28}$$

onde, B é o número total de correspondências de modelo de comprimento m e A o número total de correspondências de modelo de comprimento $m+1$. Observa-se que $A/B = [A^m(r)/B^m(r)]$. Logo, da Equação 3.27, tem-se

$$\text{SampEn}(m,r,N) = -\ln\left(\frac{A}{B}\right). \tag{3.29}$$

No cálculo de SampEn, a razão $\frac{A}{B}$ é a probabilidade condicional de que duas sequências, dentro da tolerância r para o comprimento m, permaneçam dentro de r uma da outra para o comprimento $m+1$.

Normalmente, tanto ApEn quanto SampEn apresentam menores valores para indivíduos com redução da variabilidade da frequência cardíaca (Acharya et al., 2006). Embora a medida de complexidade do sinal (entropia) seja semelhante para os índices ApEn e SampEn, vê-se que ambos dependem da escolha dos parâmetros m e r, de forma diferenciada. Tipicamente, o parâmetro m possui valores na faixa de 1 a 5 e r na faixa 5% a 20% do desvio padrão da série RR.

3.2.3 Dimensão Fractal - FD

Geometria Fractal

Se um segmento de reta for medido usando algum elemento medidor que tenha comprimento r equivalente a $1/3$ daquele segmento de reta, obviamente serão necessários três elementos para sobrepor completamente o referido segmento, e isso será válido para qualquer fração escolhida, infinitamente. Matematicamente, a quantidade de elementos necessários para sobrepor esse segmento de reta é dada por $1/r$. Se agora tentarmos sobrepor totalmente um quadrado de lado 1 com pequenos quadrados com comprimento de lado r, se cada pequeno quadrado tiver como medida de lado r igual a $0,5$ então serão necessários quatro quadrados, e matematicamente essa quantidade pode ser obtida pela formulação $1/r^2$. Por fim, se usarmos pequenos cubos para ocupar totalmente o volume de um cubo de $1 \times 1 \times 1$, e se cada lado r dos pequenos cubos, por exemplo, medir $0,5$ serão necessárias oito unidades. Alternativamente, se cada lado r medir $1/3$ (como no cubo mágico) então serão necessárias 27 unidades, ou seja, independentemente do valor de r, no caso de preenchimento de um cubo sempre serão necessárias $1/r^3$ unidades.

Relembrando que, no caso da superposição de um quadrado com vários pequenos quadrados, a quantidade esperada era de $1/r^2$ e para sobrepor um segmento de reta com vários pequenos segmentos era de $1/r$. Então, fica claro que o expoente, especificamente para essas

estruturas consideradas, é sempre o mesmo que a dimensão do objeto que estamos tentando sobrepor ou preencher, ou seja, um no caso da linha reta, dois no caso do quadrado e três no caso do cubo. Generalizando, tem-se

$$N_r = r^{-D}, \tag{3.30}$$

onde N_r é a quantidade de elementos iguais necessários para sobrepor ou preencher o objeto original, sendo r a escala aplicada ao objeto e D a dimensão da referida estrutura ou objeto.

Até o momento analisamos apenas estruturas ditas regulares e que obedecem à geometria euclidiana. Nessas estruturas, o tamanho do objeto medido é sempre o mesmo, independentemente do tamanho da escala utilizada. A grande maioria das estruturas da natureza, porém, são ditas "fractais".

O termo fractal é derivado do latim *fractus*, que significa irregular ou quebrado, indicando que o objeto ou estrutura observada não tem um número inteiro como dimensão (dimensão não-inteira), podendo assumir valores como $0,25$ ou $3,1415926535897932\ldots$. Neste contexto, abre-se espaço para estudo e aplicação da geometria fractal.

A "geometria fractal" surgiu com o intuito de reproduzir as formas da natureza. Benoit Mandelbrot (Mandelbrot e Freeman, 1983) desenvolveu estudos sobre a geometria da natureza visando representar o contorno de uma nuvem, as costas marítimas, o contorno de uma folha ou um floco de neve.

Da abordagem determinística da geometria fractal origina-se a dimensão fractal (do inglês, *fractal dimension*, ou FD), definida como

$$\mathrm{FD} = \lim_{r \to 0} \frac{\log N_r}{\log r^{-1}}. \tag{3.31}$$

A dimensão fractal mede o grau de complexidade de uma estrutura, avaliando o quão rápido as medidas aumentam ou diminuem em função da escala. Além disso, a FD de um sinal representa uma ferramenta poderosa para a detecção de transitório. Tal característica tem sido utilizada nas análises de sinais de EEG e ECG para identificar e distinguir estados específicos de funções fisiológicas.

Existem diversos algoritmos para a determinação da FD de um sinal. Neste trabalho são implementados os algoritmos de Higuchi e Katz para análise das séries de variabilidade de frequência cardíaca (Acharya et al., 2006; Polychronaki et al., 2010).

Algoritmo de Higuchi

Seja a série temporal $\mathbf{x} = \mathrm{x}(1), \mathrm{x}(2), \ldots, \mathrm{x}(N)$. Construa k novas séries temporais $\mathbf{x}_m^k$ como $\mathbf{x}_m^k = \{\mathrm{x}(m), \mathrm{x}(m+k), \mathrm{x}(m+2k), \ldots, \mathrm{x}\left(m + \lfloor (N-m)/k \rfloor k\right)\}$, para $m = 1, 2, \ldots, k$, onde

m é o valor inicial de tempo, k é o intervalo discreto de tempo entre pontos e $\lfloor \cdot \rfloor$ representa a parte inteira do argumento. Para calcular a dimensão fractal pelo algoritmo de Higuchi, as etapas a seguir devem ser observadas:

1. para cada uma das k séries temporais $\mathbf{x}_m^k$, o comprimento $L_m(k)$ é calculado por

$$L_m(k) = \frac{\sum_{i=1}^{\lfloor a \rfloor} |\mathrm{x}(m+ik) - \mathrm{x}(m+(i-1)k)| \, (N-1)}{\lfloor a \rfloor k}, \tag{3.32}$$

   onde $\lfloor a \rfloor$ é a parte inteira de $a = (N-m)/k$ obtida pelo arredondamento para baixo, N é o comprimento da série temporal $\mathbf{x}$, e $\frac{N-1}{\lfloor a \rfloor k}$ é um fator de normalização;

2. obtém-se o comprimento médio de $L_m(k)$, em k amostras (para $m = 1, 2, \ldots, k$), tal que

$$L(k) = \frac{1}{k} \times \sum_{m=1}^{k} L_m(k); \tag{3.33}$$

3. esse processo é repetido para cada k na faixa de 1 a k_{max};

4. plota-se a curva $\ln(L(k))$ versus $\ln\left(\frac{1}{k}\right)$.

5. obtém-se o coeficiente angular (α) ou inclinação do polinômio de primeira ordem, uma reta, que melhor se aproxima da curva (analisando o erro quadrático médio). Esse coeficiente é a estimativa da dimensão fractal FD pelo método de Higuchi ($D^{Higuchi}$) (Higuchi, 1988). Assim,

$$D^{Higuchi} \models \mathrm{FD} = \left\{ \frac{\delta[\ln(L(k))]}{\delta\left[\ln\left(\frac{1}{k}\right)\right]} \right\}_{k=1 \to k_{max}}. \tag{3.34}$$

Algoritmo de Katz

Utilizando o método de Katz (1988), a dimensão fractal de uma curva pode ser definida como

$$\mathrm{FD} = \frac{\log_{10}(L)}{\log_{10}(d)}, \tag{3.35}$$

onde L é o comprimento total da curva ou a soma das distâncias entre pontos sucessivos, e d é o diâmetro estimado como a distância do primeiro ponto da sequência ao ponto que provê a maior distância. Matematicamente, $d = \max \|x(1), x(i)\|$, para $i \leq N$.

Considerando a distância entre cada um dos pontos da sequência e o primeiro ponto, o ponto i é o que maximiza a distância em relação ao primeiro ponto. A dimensão fractal

(FD) compara o número atual de unidades que compõem a curva com o número mínimo de unidades necessárias para reproduzir um padrão de mesma extensão espacial. Dimensões fractais calculadas desta forma dependem das unidades de medida utilizadas. Se as unidades são diferentes, então tem-se distintos valores de dimensão fractal. O método de Katz resolve este problema criando uma unidade geral (ou critério): neste caso, cria-se o passo médio ou distância média entre pontos sucessivos a. Portanto, normalizando as distâncias, a dimensão fractal pelo método de Katz (D^{Katz}) (Katz, 1988) é dada por

$$D^{Katz} \models \text{FD} = \frac{\log_{10}(L/a)}{\log_{10}(d/a)}. \tag{3.36}$$

3.2.4 Análise das Flutuações Destendenciadas - DFA

Há uma tendência em se acreditar que a atividade normal de um coração sadio é caracterizada por um ritmo sinusal regular e que, para isso, o sistema nervoso atua sobre o coração de forma a minimizar a variabilidade da frequência cardíaca, objetivando atingir um ponto de equilíbrio no sistema circulatório. Porém, a variabilidade é um fator saudável.

A análise das flutuações destendenciadas (do inglês, *Detrended Fluctuation Analysis*, ou DFA) foi desenvolvida com o objetivo de permitir a distinção entre as complexas flutuações intrínsecas ao sistema nervoso no comando das ações vitais do corpo humano, daquelas advindas do meio e que também exercem influência sobre a frequência cardíaca. Observa-se que as flutuações do sinal intrínsecas ao sistema são caracterizadas por ocorrerem ao longo de todo o sinal, independentemente da escala utilizada na divisão do mesmo, o que impõe sobre elas uma característica de comportamento fractal (conforme descrito na Seção 3.2.3). Ao passo que as flutuações extrínsecas, ou do meio, caracterizam-se por apresentar efeitos locais e de curto prazo, implicando no surgimento de diferentes tendências ao longo do sinal.

Como as interações intrínsecas do sinal (interações entre o SNS e o SNP no controle do sistema circulatório) são de origem não linear, surge o problema de utilizar métodos no domínio do tempo e da frequência para a caracterização de sinais não estacionários. O uso da DFA na análise de sinais não estacionários foi inicialmente proposto por Peng et al. (1995). Essa análise gera dois coeficientes, α_1 e α_2, que representam a correlação de curto e longo prazo, respectivamente, do sinal destendenciado. Isso foi validado pelos autores ao mostrar que essa análise distinguia sinais com características patológicas de sinais de indivíduos saudáveis.

Algoritmo da DFA

O cálculo da DFA baseia-se em encontrar os coeficientes α_1 e α_2 que irão caracterizar o sinal de VFC e as influências do sistema nervoso sobre o ritmo cardíaco. O cálculo tradicional desses coeficientes, para uma série de intervalos RR, baseia-se nos passos (Peng et al., 1995; Leite et al., 2010):

1. integrar o sinal temporal sem o seu valor médio, o que equivale a obter um sinal de tendências e não um sinal de diferenças temporais. Logo,

$$y(n) = \sum_{i=0}^{n} \mathrm{RR}(i) - \overline{\mathrm{RR}}, \text{ para } 1 \leq n \leq N, \tag{3.37}$$

   onde $\overline{\mathrm{RR}}$ é o valor médio da série de intervalos RR;

2. segmentar o sinal $\mathbf{y}$ $(y(n))$, de tamanho N, em múltiplas janelas de comprimento l. Para o trecho de sinal presente em cada uma dessas janelas é calculada uma reta de regressão, por meio do método dos mínimos quadrados, gerando-se um sinal de tendências do intervalo. Esse sinal de tendências, formado por segmentos de reta, é denotado por $y_{trend}(n)$;

3. obter o sinal destendenciado subtraindo-se o sinal $y_{trend}(n)$ do sinal integrado $y(n)$;

4. calcular o erro de aproximação, $e(n)$, denominado de sinal destendenciado, como $e(n) = y(n) - y_{trend}(n)$;

5. calcular a raiz da média quadrática do erro de aproximação, de acordo com

$$E = \sqrt{\frac{1}{N} \times \sum_{n=0}^{N-1} e^2(n)}, \tag{3.38}$$

   onde N é o tamanho da série RR;

6. repetir os passos 2 a 4 para múltiplos valores de l;

7. obtém-se a curva 2D $f(x) \times x$, ou ainda $\log_{10} E \times \log_{10}(l)$. Como E tem relação exponencial com l, observa-se uma relação linear entre as variáveis na escala *log-log*;

8. assim, faz-se um ajuste linear de $f(x)$ pelo método dos mínimos quadrados, para $4 \leq l \leq 16$. O coeficiente angular obtido é denominado de α_1;

9. de forma análoga, faz-se o ajuste linear de $f(x)$ para tamanhos de janela variando de $16 \leq l \leq N$. O coeficiente angular obtido é denominado de α_2.

O Coeficiente Alfa

Em geral, o valor de E aumenta com o aumento do tamanho da janela l, e quando é observada uma relação linear entre as variáveis na escala *log-log* pode-se dizer que o sinal em questão apresenta características de escala como fractais. Sob essas condições, as flutuações dos sinais fisiológicos podem ser caracterizadas através do coeficiente α, o qual é o coeficiente angular da reta $\log_{10}(E)$ por $\log_{10}(l)$ (vide Figura 3.3).

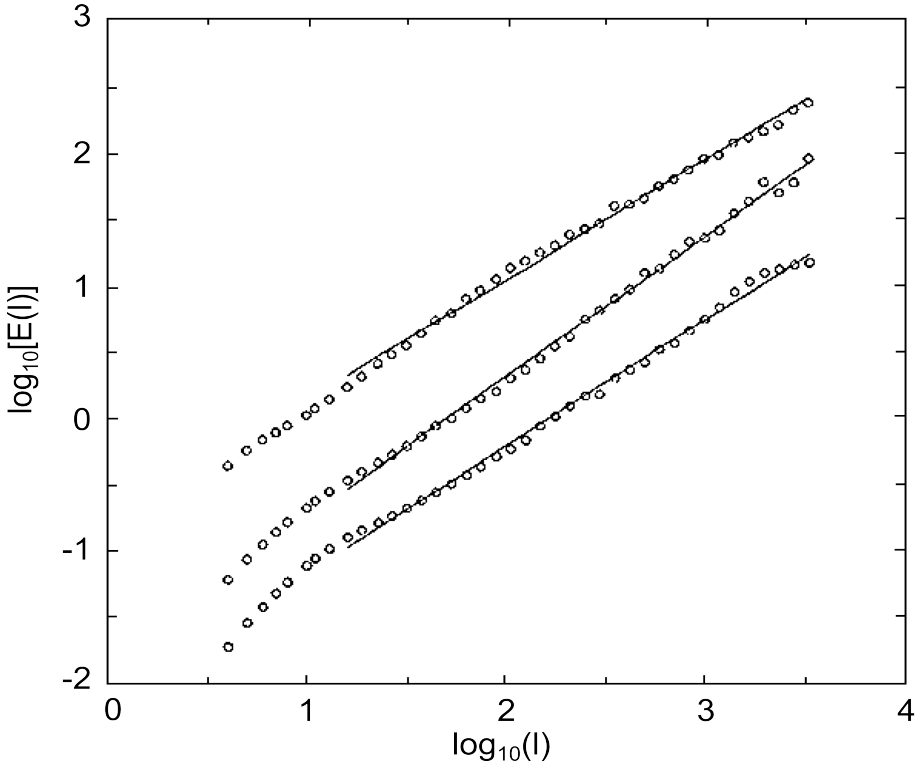


Figura 3.3: Reta característica $\log(E) \times \log(l)$ (Peng et al., 1995).

Quando o sinal possui correlações de curto prazo, o valor inicial da inclinação pode ser diferente de $0,5$, mas α se aproxima de $0,5$ para maiores tamanhos de janela. Tal valor de α é esperado para um sinal completamente descorrelacionado, ou melhor, com características de ruído branco (Peng et al., 1995). Para valores de α entre $0,5$ e 1, os sinais apresentam correlações de lei de potência (do inglês, *power-law*) de longo prazo, o que significa que um grande intervalo RR tende a ser seguido por outro grande intervalo e não por um intervalo pequeno, como é de se esperar. Quando α varia entre 0 e $0,5$, observa-se que os intervalos RR tendem a se alternar entre grandes e pequenos, o que caracteriza uma correlação de lei de potência diferente da anterior. O caso em que $\alpha = 1$ caracteriza a presença de ruído $1/f$, pois apresenta uma relação de amplitude que é o inverso da frequência em seu espectro de densidade de potência $\left(\frac{1}{f}\right)$. Para $\alpha \geq 1$ correlações existem, porém não como uma lei de

potência. Valores de $\alpha = 1,5$ indicam que o sinal comporta-se como ruído browniano, o qual corresponde à integral do ruído branco. Nesse caso, a relação entre densidade de potência e frequência é dada por $\frac{1}{f^2}$ (Peng et al., 1995).

Observa-se que quanto maior o valor de α mais suave é o sinal que está sendo analisado no domínio do tempo. Portanto, de certa forma, o coeficiente α representa a suavidade da curva. Assim, é possível interpretar o ruído $\frac{1}{f}$ $(\alpha = 1)$ como sendo o compromisso de suavidade do sinal entre a total imprevisibilidade (ruído branco, $\alpha = 0,5$), para sinais com $\alpha < 1$, e a suavidade do sinal (ruído browniano, $\alpha = 1,5$), para sinais com $\alpha > 1$.

O Fenômeno de *Crossover*

Embora o método de análise, proveniente da DFA, seja de grande utilidade na detecção e classificação de diferentes patologias fazendo uso de registros de ECG de 24 horas, é importante que haja a possibilidade de utilização desta ferramenta para sinais de curta duração, como os tipicamente utilizados na análise da VFC, com cerca de 5 minutos de duração. A este respeito, observamos que para pequenos comprimentos de janela l existe o fenômeno de *crossover*, representado por uma alteração no valor do coeficiente angular (α) (setas indicadas na Figura 3.4).

Com o intuito de verificar a possibilidade de utilização da DFA para sinais de curta duração, Peng et al. (1995) observaram que há um fenômeno de inversão dos valores de α para janelas de tamanho muito pequeno, menor que 10 batimentos. Em sinais obtidos de indivíduos normais, o valor de α para janelas pequenas é maior que o valor para janelas com mais de 10 batimentos. Talvez isto se justifique, visto que as flutuações dos intervalos em pequenas escalas (janelas com número de batimentos inferior a 10) apresentam predominantemente a influência suave da respiração, consequentemente levando a valores maiores de α (Peng et al., 1995). Para escalas maiores, as flutuações dos intervalos representam as correlações intrínsecas do sistema, tal que $\alpha \to 1$, para indivíduos normais, aproximando-se do ruído $\frac{1}{f}$, conforme discutido anteriormente.

A Figura 3.4 ilustra a diferenciação entre esses dois valores de α. Observa-se que as distintas retas apresentam distintos valores de coeficiente angular α. Para o caso normal, variando-se o tamanho da janela l, para $l \approx 10$ batimentos (o que equivale à abscissa $\log_{10}(10) = 1$) e $l \gtrapprox 10$ batimentos ($\log_{10}(l) > 1$ ou todo o restante do sinal), observamos a presença do fenômeno de *crossover* (ver as setas na Figura 3.4).

Por causa do fenômeno de *crossover*, são calculados dois valores de α, em função do tamanho das janelas utilizadas. O valor de α_1 é calculado para as janelas cujo tamanho

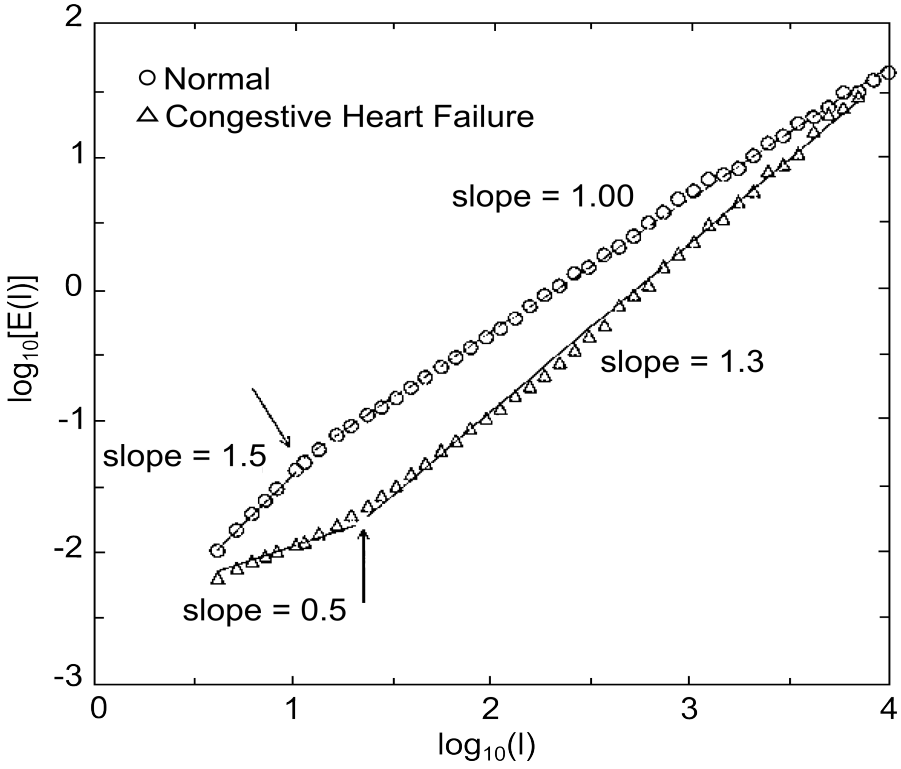


Figura 3.4: Variação do coeficiente angular de acordo com o tamanho das janelas. Adaptado de (Peng et al., 1995).

varia entre 4 e 16 batimentos, e α_2 é calculado para janelas com tamanhos superiores a 16 batimentos (Peng et al., 1995).

Melhorias no Cálculo da DFA

Três melhorias, propostas em (Almeida et al., 2012), foram implementadas no cálculo da DFA, a saber:

1. alteração no cálculo dos comprimentos de janelas. O tamanho das janelas l deve ser tal que os valores de $\log_{10}(l)$ sejam igualmente espaçados;

2. análise de intervalos por meio de janelas com sobreposição para solução de amostras faltantes ou excedentes;

3. regressão linear ponderada. Mesmo com a escolha ótima dos valores de l, ainda é observada não uniformidade no espaçamento dos valores de $x = \log_{10}(l)$. Assim, faz-se uma alteração no cálculo dos coeficientes α_1 e α_2, utilizando a interpolação pelas

distâncias entre os valores de x, com o intuito de uniformizar o espaçamento horizontal dos valores de $f(x) = \log_{10}(E)$.

Cálculo dos comprimentos de janela

Como explicado anteriormente, o sinal destendenciado é calculado para diferentes valores de janela l. Nós desejamos uma relação linear entre $f(x) = \log_{10}(E)$ e $x = \log_{10}(l)$, para que os valores dos coeficientes α sejam obtidos da regressão linear dos pontos obtidos.

Para que o espaçamento no eixo $x = \log_{10}(l)$ seja uniforme, os comprimentos de janela, l, são obtidos por

$$l = f(k) = \lfloor l'_k \rfloor, \text{tal que: } l'_k = 10^{ks} l_0, \tag{3.39}$$

onde $\lfloor l'_k \rfloor$ é o resultado do arredondamento de l'_k e l_0 é o menor comprimento de janela avaliado. Observa-se que l é uma função de k $(f(k))$ logo, a partir desse ponto iremos denominar o comprimento de janela como l_k, tendo em mente que $l = l_k$. Além disso, a equação 3.39 nos mostra que l'_k é uma função do passo s e de k.

O espaçamento horizontal dos valores de $x = \log_{10}(l_k)$ é Δx, tal que

$$\Delta x = f(k) \therefore \Delta x = \Delta x_k = \log_{10}\left(\frac{l_k}{l_{k-1}}\right). \tag{3.40}$$

A Tabela 3.1 ilustra o efeito da escolha do passo s. Com s grande, poucos valores de x (e consequentemente de $f(x)$) para $l_k \leq 16$ seriam obtidos, o que diminuiria a precisão no cálculo de α_1. No entanto, para s pequeno, são obtidos valores repetidos de l_k. Os valores repetidos poderiam ser descartados, mas nota-se que os espaçamentos para valores pequenos de k seriam maiores que os espaçamentos para valores grandes de k. O espaçamento deve ser aproximadamente igual a s para todos os valores de k, para que não seja reduzida a precisão no cálculo de α_1. Com o valor ideal de s, ou seja, valor de s que proporciona espaçamento quase uniforme nos valores de x, tem-se a melhor situação.

Na implementação da DFA o valor ideal para o passo s (s_{ideal}) é calculado em função do menor comprimento de janela avaliado (l_0). Tal valor é o menor valor de s que garanta não haver valores repetidos de l_k após o arredondamento (Tabela 3.1). Para garantir a ausência de repetições, o algoritmo calcula todos os valores de l_k entre l_0 e l_{stop} (valor limítrofe superior de l_k, por exemplo 16, 64 ou N). Havendo repetição dentro desse intervalo, o algoritmo aumenta o tamanho do passo s e repete o processo.

Isto é feito com o intuito de aumentar a precisão no cálculo de α_1 e α_2. Com base nessas observações e na Tabela 3.1, para o cálculo de α_1, considerando que o menor comprimento

Tabela 3.1: Espaçamento horizontal dos valores de x, para diferentes valores de passo s com $4 \leq l_k \leq 15$.

	$s = 0,10$ (grande)			$s = 0,05$ (pequeno)			$s = 0,0703$ (ideal)		
k	l'_k	l_k	Δx_k	l'_k	l_k	Δx_k	l'_k	l_k	Δx_k
0	4,00	4	-	4,00	4	-	4,00	4	-
1	5,04	5	0,10	4,49	4	0	4,70	5	0,10
2	6,34	6	0,08	5,04	5	0,10	5,53	6	0,08
3	7,98	8	0,12	5,65	6	0,08	6,50	7	0,07
4	10,05	10	0,10	6,34	6	0	7,64	8	0,06
5	12,65	13	0,11	7,11	7	0,07	8,98	9	0,05
6				7,98	8	0,06	10,56	11	0,09
7				8,95	9	0,05	12,42	12	0,04
8				10,05	10	0,05	14,60	15	0,10
9				10,27	11	0,04			
10				12,65	13	0,07			
11				14,19	14	0,03			

de janela é $l_0 = 4$, utilizamos o valor de s igual a $0,0703$. De forma análoga, para o cálculo de α_2, considerando $l_0 = 16$, utilizamos um menor valor de s $(s = 0,0215)$.

É importante ressaltar que a busca por repetições não é realizada de l_0 a N, pois isso resultaria em tempo desnecessário de processamento. Para cada valor de l_0, existe um valor de l_{stop}, após o qual não há possibilidade de ocorrer repetições para qualquer $s < s_{ideal}$. O valor de l_{stop} foi calculado empiricamente para $1 \leq l_0 < 5 \times 10^3$, montando-se uma tabela que é consultada antes do processo iterativo de busca do valor ideal para o passo s.

Solução para as amostras faltantes ou excedentes

De acordo com Leite et al. (2010), quando $N/l_k \notin \mathbf{Z}$, a última janela do sinal teria comprimento menor que l_k podendo ou não as amostras desta janela serem consideradas. Em ambos os casos, a precisão no cálculo de $e(n)$ e E é afetada.

Nessa implementação da DFA, todas as janelas têm l_k amostras, tal que nenhuma amostra é descartada (Almeida et al., 2012). Para isso, os ajustes lineares no cálculo do sinal das tendências foram realizados em janelas com superposição. Isto é ilustrado na Figura 3.5.

Regressão linear ponderada

Pode-se observar na Tabela 3.1 que o espaçamento Δx_k não é uniforme para todo x_k. A não uniformidade é maior para valores pequenos de k, o que prejudica o cálculo de α_1.

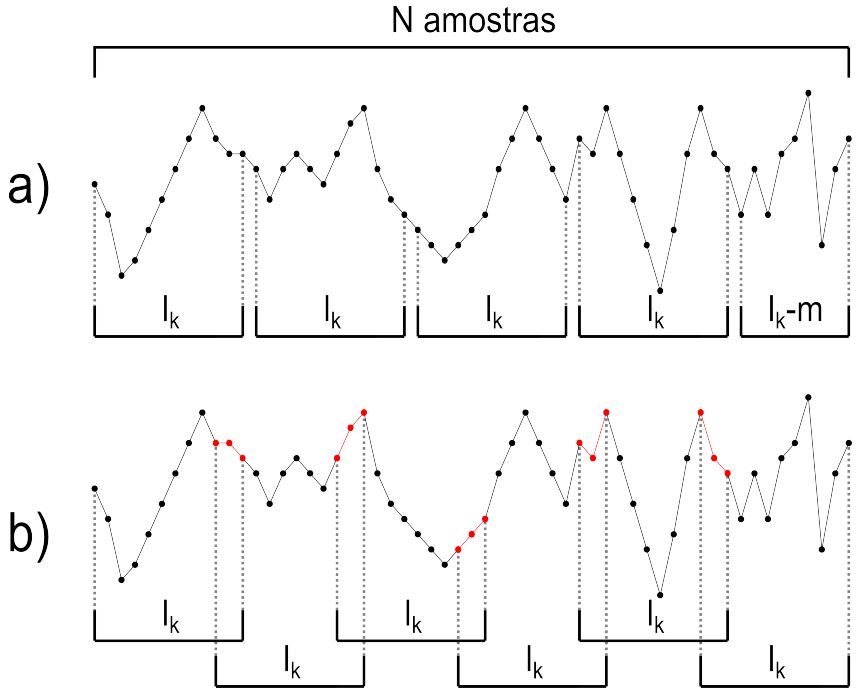


Figura 3.5: Uniformizando o tamanho das janelas para o cálculo do sinal das tendências. (a) Implementação original sobram $l_k - m$ amostras quando $N/l_k \notin \mathbf{Z}$; b) Utilização de janelas com superposição para contornar esse problema.

Isso pode ser minimizado utilizando regressão linear ponderada, onde os pesos, w_k, são calculados com base no espaçamento horizontal dos valores de $x_k = \log_{10}(l_k)$, ou

$$w_k \simeq \frac{\Delta x_k + \Delta x_{k+1}}{2}. \tag{3.41}$$

De fato, w_k é calculado com base na faixa de variação dos comprimentos das janelas l_k, visto que $x_k = \log_{10}(l_k)$. Sabendo que $\Delta x_k = \log_{10}\left(\frac{l_k}{l_{k-1}}\right)$, e obtendo o módulo da Equação 3.41, tem-se

$$w_k = \left| \frac{\log_{10}\left(\frac{l_k}{l_{k-1}}\right) + \log_{10}\left(\frac{l_{k+1}}{l_k}\right)}{2} \right| \therefore w_k = \left| \frac{\log_{10}(l_{k+1}) - \log_{10}(l_{k-1})}{2} \right|. \tag{3.42}$$

Assim,

$$w_k = \left| \frac{\log_{10}(l_{k+1}) - \log_{10}(l_{k-1})}{2} \right| \Rightarrow w_k = \left| \frac{x_{k+1} - x_{k-1}}{2} \right|. \tag{3.43}$$

3.2.5 Gráficos de Recorrência - RP

Em 1987, Eckmann et al., baseando-se no trabalho de Poincaré, introduziram uma ferramenta conhecida como gráficos de recorrência (do inglês, *Recurrence Plots*, ou RP) com

a finalidade de visualizar a dinâmica de sistemas recorrentes. De acordo com a definição original do gráfico de recorrência (Eckmann et al., 1987; Holyst et al., 2000), tem-se

"O gráfico de recorrência de uma série temporal de N pontos $x(n)$, onde n é o índice temporal, é uma "matriz $N \times N$" (matriz de recorrência) preenchida por pontos brancos e pretos. O ponto preto, chamado de ponto recorrente, é colocado na matriz de recorrência com coordenadas (i, j) somente se a distância $\mathrm{d}(i, j)$ no instante $n = i$ e $n = j$ (entre o estado corrente do sistema e o estado a ser comparado) for menor que uma certa distância (raio) ε, fixado no centro do estado corrente" (Eckmann et al., 1987). Em outras palavras, o gráfico de recorrência é uma matriz bidimensional, quadrada, na qual os seus eixos representam a evolução dos estados de um sistema dinâmico. Assim,

$$\vec{R}_{i,j}^{m,\varepsilon} = \Theta\left(\varepsilon - \left\|\vec{x}_i - \vec{x}_j\right\|\right), \vec{x}_i \in \mathbb{R}^m, \quad i, j = 1, \ldots, N, \tag{3.44}$$

onde

- N é o número de estados x_i considerados. No caso, o tamanho da série;
- ε é o raio da vizinhança (*threshold*) no ponto x_i;
- $||\cdot||$ é a norma da vizinhança, comumente a norma euclidiana;
- $\Theta(\cdot)$ é função de Heaviside (Bracewell, 1999), tal que, $\Theta(x) = 0$, se $x < 0$, e $\Theta(x) = 1$, caso contrário;
- m é a dimensão de imersão, ou ainda, a dimensão do espaço onde se observa os estados do sistema. Assim, m representa o número de componentes necessárias para caracterizar o sistema. Normalmente, tem-se $m < N$;
- $\vec{x}$ é um vetor que define uma trajetória no espaço de fase de dimensão m.

Se $\vec{R}_{i,j} = 1$, o estado é dito recorrente e, como consequência, um ponto preto é marcado no RP. Caso $\vec{R}_{i,j} = 0$, o estado é "não recorrente" e marca-se um ponto branco no RP. A Figura 3.6 ilustra a obtenção do gráfico de recorrência.

A exemplo de gráficos de recorrência, apresentam-se nas Figuras 3.7 e 3.8 os RP originados de uma senóide (sinal periódico) com uma dada frequência, ruído com distribuição normal e uma oscilação harmônica onde se altera a frequência de amostragem do sinal. Como apresentado na descrição da Figura 3.8, durante a obtenção das medidas de quantificação nos gráficos de recorrência temos que definir alguns parâmetros como: dimensão de imersão (m), tempo de retardo (τ), *threshold* (ε) e norma para o cálculo de distância entre estados do sistema (por exemplo, norma euclidiana ou norma infinito).

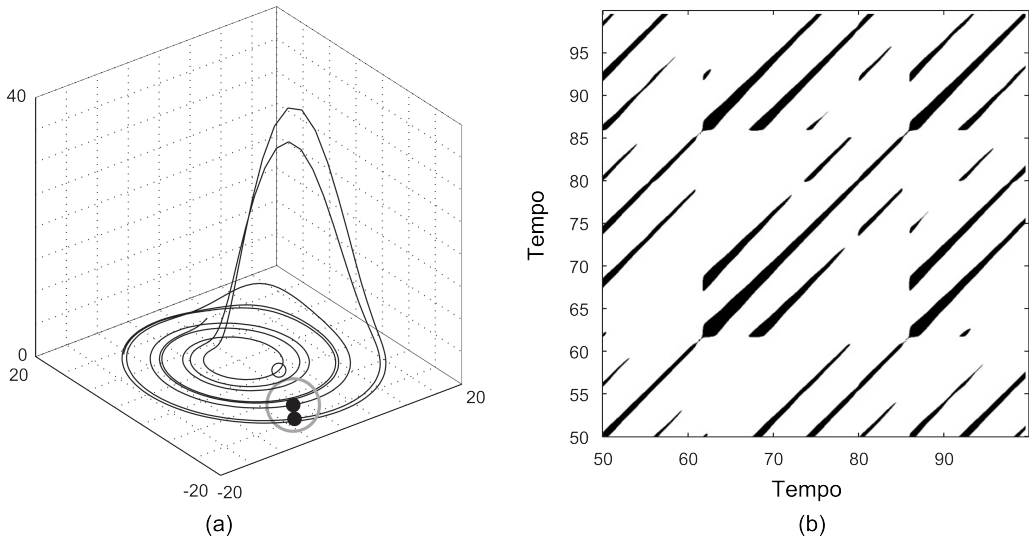


Figura 3.6: (a)Segmento da trajetória no espaço de fase de um dado sistema e (b) Gráfico de recorrência correspondente. Um vetor no espaço de fase em j que se encontra na vizinhança (círculo cinza em (a)) de um dado vetor no espaço de fase em i é considerado como um ponto de recorrência (ponto preto na trajetória em (a)). Este é marcado com um ponto preto no RP na posição (i, j). Um vetor no espaço de fase fora da vizinhança (círculo vazio em (a)) representa pontos brancos no RP. Neste exemplo, o raio da vizinhança para o RP é $\varepsilon = 5$, tal que a norma euclidiana ($L_2 - \text{norm}$) é utilizada nos cálculos de distância. Adaptado de Marwan et al (2007).

Categorias de Gráficos de Recorrência

Os gráficos de recorrência foram divididos em duas categorias (Eckmann et al., 1987): padrões de larga escala (denominados de tipológicos (do inglês, *tipology*)) e de pequena escala (*texture*). Os padrões de larga escala oferecem uma visão global do gráfico de recorrência e são divididos, de acordo com a Figura 3.9, em:

- homogêneo: apresentam pontos pequenos se comparados com o gráfico de recorrência como um todo, ou seja, o tempo de duração de cada linha diagonal ou vertical formada (estados sucessivos) é curto em relação ao tempo total de exposição do sistema;
- deriva (*drift*): acontece principalmente quando o sistema possui uma variação de parâmetros lenta. Tais variações provocam uma ausência de pontos recorrentes, tanto no canto superior esquerdo quanto no inferior direito do gráfico de recorrência;
- periódico: sistemas oscilantes sempre apresentam linhas diagonais totalmente preenchidas e paralelas à diagonal principal. Além disso, apresentam também estruturas de blocos recorrentes;

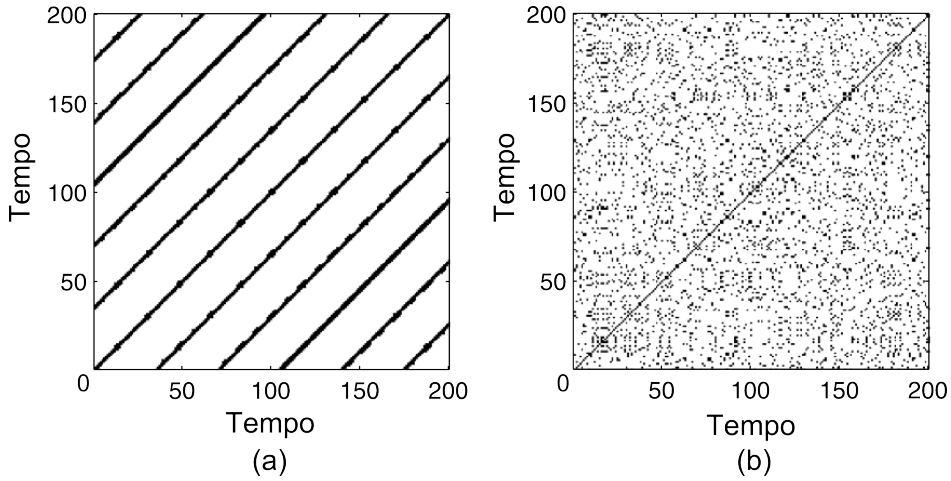


Figura 3.7: Exemplos de RPs exibindo padrões de recorrência distintos. (a) Senóide e (b) Ruído com distribuição normal de um gerador de números aleatórios. Adaptado de Marwan et al (2007).

- descontínuo (*checkboard*): a descontinuidade é causada por mudanças abruptas na dinâmica, bem como a ocorrência de eventos raros, ocasionando bandas brancas.

Os padrões de pequena escala, como o próprio nome sugere, são os pontos singulares, as linhas diagonais, verticais, horizontais e as estruturas de blocos formadas por essas linhas.

- Pontos (*pixels*): representam estados recorrentes. Se o ponto estiver isolado, significa um estado raro no sistema.
- Linhas Diagonais: ocorrem quando uma parte da trajetória evolui de forma paralela (similar) a outro segmento de trajetória, isto é, a trajetória visita a mesma região do espaço de fase em tempos diferentes. O comprimento dessas estruturas diagonais é determinado pela duração dessa evolução similar de tal segmento da trajetória (Marwan, 2003).
- Linhas Verticais (e Horizontais): a ocorrência de linhas verticais demonstra um comprimento temporal em que o estado do sistema não muda, ou seja, permanece estacionário durante a evolução temporal. Sistemas intermitentes geralmente apresentam esse comportamento.

Devido a dificuldade de visualizar riquezas na dinâmica apenas com o gráfico de recorrência, houve a necessidade de se criar quantificadores que contabilizem essas estruturas presentes no gráfico de recorrência como: linhas diagonais, verticais, etc. Isso proporciona

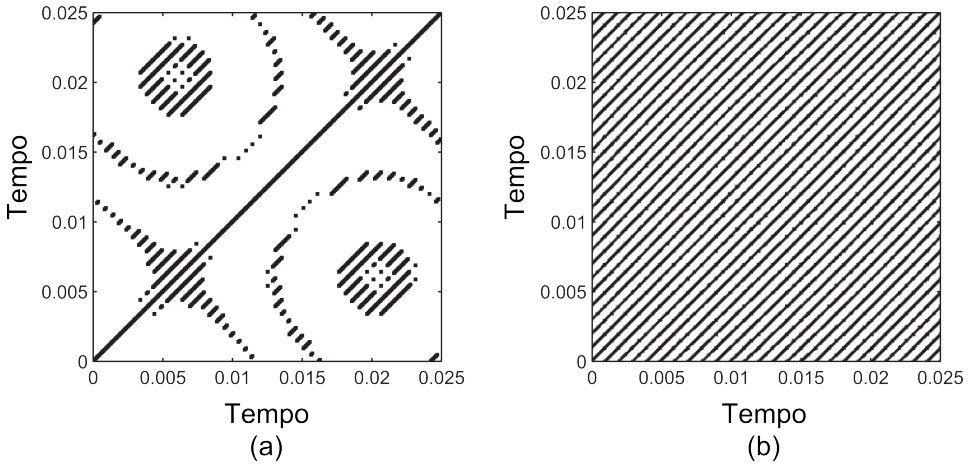


Figura 3.8: (a) Padrão de *gaps* (grandes áreas brancas) no RP da função harmônica modulada $\cos(2\pi 1000t) + 0,5\sin(2\pi 38t)$ amostrado em 1 kHz. Esses *gaps* representam ausência de recorrência devido a frequência de amostragem ser próxima a frequência harmônica do sinal. (b) Gráfico de recorrência mostrado em (a), porém com frequência de amostragem de 10 kHz. Conforme esperado, agora todo o RP consiste de estruturas periódicas de linhas devido às oscilações. Neste exemplo, $m = 3$, $\tau = 1$, $\varepsilon = 0.05$, e a norma infinito ($L_\infty - \text{norm}$) é utilizada nos cálculos de distância. Adaptado de Marwan et al (2008).

uma maior sensibilidade aos resultados, realçando principalmente transições que se desenvolvem em escalas de tempos muito curtas. Na subseção Medidas de Quantificação nos Gráficos de Recorrência, essas medidas serão introduzidas.

Escolha apropriada da dimensão de imersão (*Embedding dimension*)

De acordo com os teoremas de imersão (Souza, 2008), a preservação das estruturas topológicas da trajetória original de um sistema é garantida se, e somente se, $m \geq 2d + 1$, onde d é a dimensão do atrator e m a dimensão de imersão. Contudo, isso não impossibilita uma imersão que não satisfaça essa relação de dimensões e, caso isso ocorra, ela apenas não possuirá uma comprovação matemática rigorosa.

Por exemplo, utilizando $m = 1$, analisa-se apenas medidas de uma série temporal em um tempo t comparando-as com outras medidas dentro dessa mesma série em tempos diferentes, sem necessariamente fazer uma imersão multidimensional. Em (March et al., 2004) mostra-se, com argumentos estatísticos, que a imersão é um processo de duplicação de informação, e que muitas vezes não fornece ganhos adicionais dessa informação, apenas redundâncias.

No caso, se tentarmos imergir uma série temporal quando não é conhecida a dimensão

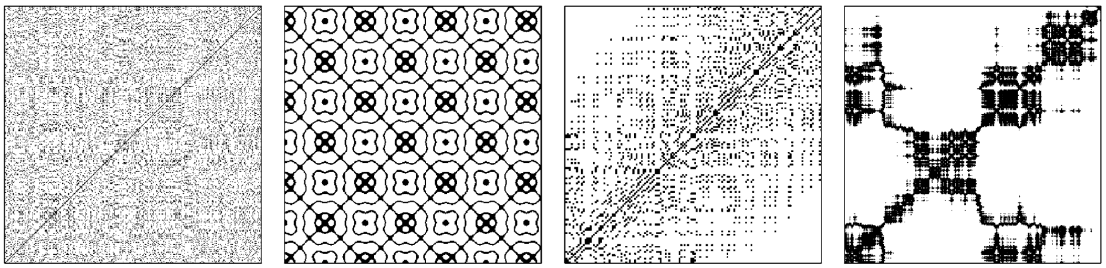


Figura 3.9: Padrões de larga escala em RPs: (a) Homogêneo (ruído branco uniformemente distribuído), (b) Periódico (superposição de oscilações harmônicas), (c) *Drift* (mapa logístico) e (d) Descontínuo (movimento Browniano). Esses exemplos ilustram o quão diferentes os gráficos de recorrência podem ser. Os parâmetros dos RPs utilizados são: $m = 1$, $\varepsilon = 0.2$ (a, c, d) e $m = 4$, $\tau = 9$, $\varepsilon = 0.2$ (b); norma euclidiana $(L_2 - \text{norm})$. Adaptado de Marwan et al. (2007).

d do atrator, torna-se impossível usufruir da relação $m \geq 2d + 1$ para a escolha de um m adequado. Quando esse for o caso, podemos recorrer a duas hipóteses:

- imersão com base em um diagnóstico conhecido. Nesse caso, começamos com uma dimensão de imersão igual à um ($m = 1$) e aumentamos essa dimensão m sucessivamente até uma imersão alta, $m = 10$ por exemplo. Se esse suposto diagnóstico conhecido permanecer inalterado com a mudança de m, usa-se a dimensão de imersão mais baixa para a qual o mesmo permanece nessa invariância. Esta técnica torna-se um pouco subjetiva, mas na prática funciona para quase todo tipo de sistema dinâmico conhecido;

- investigação de mudanças na vizinhança de pontos no espaço de fase. Se partirmos de uma imersão alta, $m = 10$ por exemplo, podemos ter uma noção de como muda a vizinhança de cada ponto ao passo que diminuímos m. Vizinhos verdadeiros sempre permanecem vizinhos, independente do m utilizado. Uma dimensão de imersão inapropriada pode causar um aumento nessa quantidade de falsos vizinhos e, por isso, induzir a falsos diagnósticos.

Existem vários métodos que utilizam a técnica de "falsos vizinhos". Nestes métodos esses vizinhos (nem sempre verdadeiros), são apenas uma consequência do decréscimo dimensional ($\downarrow m$). Assim, a dimensão de imersão ideal é aquela escolhida quando o número de falsos vizinhos cai a zero.

Escolha do tempo de retardo (*delay*)

Suponha-se que num experimento nós não possamos medir todas as componentes de um vetor $\vec{x}(t)$, dado um estado do sistema, e sim apenas uma componente deste. Por exemplo, a única medida que temos para a análise da relação da variabilidade da frequência cardíaca é o valor de intervalos RR. Neste caso, a reconstrução do atrator do sistema ocorre por meio da formação de vetores $y = (y_1, y_2, y_3, \ldots, y_m)$ em função das coordenadas de retardo. Assim,

$$\begin{aligned} y_1(t) &= g(t) \\ y_2(t) &= g(t-\tau) \\ y_3(t) &= g(t-2\tau) \\ &\vdots \\ y_m(t) &= g(t-(m-1)\tau) \end{aligned} \tag{3.45}$$

onde τ é um intervalo fixo de tempo a ser escolhido (Ott, 1993) e $g(t)$ é uma função escalar do vetor de estado $\vec{x}(t)$. Nesse caso, a dificuldade é descobrir qual é o τ adequado para cada sistema. Uma das técnicas aplicáveis ao problema é proposta por Fraser e Swinney (1986), conhecida como informação mútua média (Fraser e Swinney, 1986). Matematicamente podemos defini-la como

$$S = -\sum_{ij} p_{ij}(\tau) \ln\left(\frac{p_{ij}(\tau)}{p_i p_j}\right). \tag{3.46}$$

A informação mútua média é calculada via entropia de Shannon (Shannon, 1948), a qual procura identificar o quanto de informação podemos obter de uma medida *i*, com uma probabilidade p, em um tempo t, pertencente a uma partição (ou trecho) p_i de uma série temporal, quando se observa uma outra medida j, da mesma série, com uma probabilidade p, pertencente a uma partição p_j, em um tempo $t+\tau$.

Portanto, o valor de τ deve ser escolhido quando a informação mútua média tiver um mínimo "relevante", ou ainda, quando a probabilidade que tenhamos uma medida com informação em ambas as partições p_i e $p_j(\tau)$ seja bem pequena. Isso se justifica através do fato de que, quanto menor a informação conjunta de p_i e $p_j(\tau)$, podemos dizer que "menos importância" esse valor da série tem quando for projetado em outro eixo para servir de coordenada (o que significa que as medidas de p_i e p_j, quando mínimas, são altamente descorrelacionadas).

Para Hegger et al. (1998) o tempo de retardo não possui dependência com o número e nem com o tamanho das partições utilizadas (Hegger e Kantz, 1998).

Medidas de Quantificação nos Gráficos de Recorrência

Baseado nos padrões de pequena escala criados por Eckmann et al. (1987), Webber e Zbilut (1994) desenvolveram medidas de complexidade, as chamadas medidas de quantificação de recorrência. Essas medidas auxiliam na contabilização de pontos e diagonais, além de relacioná-los através de diagnósticos (Webber e Zbilut, 1994). A primeira medida proposta é a taxa de recorrência, REC, dada por

$$\mathrm{REC} = \frac{1}{N^2} \sum_{i,j=1}^{N} \vec{R}^{m,\varepsilon}_{i,j;\, i \neq j}. \tag{3.47}$$

REC simplesmente conta o número de pontos pretos no gráfico de recorrência. Ela é uma aplicação direta da recorrência de Poincaré, ou seja, compara cada elemento x_i de um gráfico de recorrência com N^2 termos (N^2 provém do gráfico de recorrência $N \times N$), em um tempo inicial t_o, com todos os outros elementos x_j em tempos $t = 1, \cdots, N-1$. Se porventura, x_j estiver dentro da vizinhança de x_i, x_j é recorrente à x_i. A soma de todos os x_j recorrentes à x_i é justamente o valor da taxa de recorrência.

Em relação às linhas diagonais, temos o determinismo, DET, dado por

$$\mathrm{DET} = \frac{\sum_{l=l_{min}}^{N} l \times P^{\varepsilon}(l)}{\sum_{i,j}^{N} \vec{R}^{m,\varepsilon}_{i,j}}, \tag{3.48}$$

onde l é o tamanho da estrutura diagonal, $P^{\varepsilon}(l)$ é a probabilidade dessa estrutura diagonal ocorrer dentro do gráfico de recorrência e l_{min} é o número mínimo de estruturas diagonais que se deseja contabilizar dentro do gráfico de recorrência (normalmente escolhe-se $l_{min} = 2$).

O determinismo mede o número de estruturas diagonais formadas, divididas sobre todo o gráfico de recorrência. Pode ser interpretado como a previsibilidade do sistema, a razão dos pontos recorrentes que formam estruturas diagonais pelos pontos recorrentes de todo o gráfico de recorrência. É importante observar que, apesar do nome desta medida, não há relação com o significado de um processo determinístico dado pela mecânica clássica (Marwan, 2003). As estruturas diagonais, por sua vez, mostram o intervalo em que um segmento da trajetória evolui paralelamente a outro segmento, ou seja, a trajetória de um segmento visita a mesma região do espaço de fases de outro segmento em tempos diferentes (Marwan e Kurths, 2004). O comprimento dessa linha diagonal determina o tempo em que esses estados permanecem com essa evolução similar. Baseado nisso, Webber e Zbilut (1994) propuseram uma outra medida, relacionado à média das estruturas diagonais formadas, o comprimento

médio das linhas diagonais, ou simplesmente L, dado por

$$\mathrm{L} = \frac{\sum_{l=l_{min}}^{N} l \times P^{\varepsilon}(l)}{\sum_{l=l_{min}}^{N} P^{\varepsilon}(l)}. \tag{3.49}$$

Sua interpretação fornece o tempo médio em que dois segmentos da trajetória permanecem evoluindo de forma similar em um estado do sistema. Essa medida difere um pouco do determinismo porque reflete a soma das probabilidades de se encontrar estruturas diagonais no gráfico de recorrência, dividido pela probabilidade apenas de encontrar diagonais, e não de pontos recorrentes. Ou seja, ela é mais robusta que o determinismo se aplicada a sistemas em que a riqueza da dinâmica se baseia simplesmente em estruturas diagonais (evoluções similares das trajetórias) e não em pontos recorrentes isolados ou intermitentes.

Seguindo esta ideia o comprimento máximo, L_{max}, é dado por

$$\mathrm{L}_{max} = \max(l_i;\ i = 1, \ldots, N_l). \tag{3.50}$$

De acordo com Eckmann et al. (1987), o comprimento das linhas diagonais está relacionado com o maior expoente de Lyapunov positivo do sistema, se houver algum. Baseado nisso, foi criada a medida de divergência, DIV, dada por

$$\mathrm{DIV} = \frac{1}{L_{max}}, \tag{3.51}$$

que simplesmente é o inverso do comprimento máximo de uma linha diagonal contida no gráfico de recorrência. Contudo, a relação direta entre o expoente de Lyapunov e a divergência é mais complexa do que apresentado na literatura (Thiel et al., 2004). Intuitivamente, pensa-se que se o maior comprimento diagonal de um gráfico de recorrência está associado ao maior tempo em que duas trajetórias evoluem de forma similar, o oposto disso nos fornece o tempo máximo em que duas trajetórias divergem. Não obstante, não podemos esquecer que estamos trabalhando apenas com trajetórias recorrentes e, por isso, medimos a divergência apenas dessas trajetórias recorrentes.

Outra medida que pode ser analisada é a entropia, no caso, a entropia de Shannon (Shannon, 1948) que nos fornece a frequência de distribuição das linhas diagonais

$$\mathrm{ShanEn} = -\sum_{l=l_{min}}^{N} p(l) \ln p(l), \text{tal que: } p(l) = \frac{P^{\varepsilon}(l)}{\sum_{l=l_{min}}^{N} P^{\varepsilon}(l)}. \tag{3.52}$$

Rescrevendo ShanEn em função de um único comprimento l_1 normalizado por todo o RP, tem-se

$$\text{ShanEn} = -\sum_{l=l_{min}}^{N} l_1 \ln(l_1); \quad l_1 = \frac{P^{\varepsilon}(l)}{\sum_{l=l_{min}}^{N} P^{\varepsilon}(l)}. \tag{3.53}$$

Matematicamente, a entropia sempre fornecerá a mesma informação que o comprimento médio, a menos de um fator de escala, devido à multiplicação do logaritmo neperiano. Assim, ShanEn é a soma de todos os comprimentos recorrentes reescalonados na escala logarítmica, o qual fornece um crescimento de comprimentos recorrentes em escala menor.

Em 2003, Marwan propôs novas medidas baseado em linhas verticais (ou horizontais). De forma análoga à definição de determinismo, a razão entre as linhas verticais (ou horizontais) recorrentes de todo o gráfico de recorrência, dividido pelo conjunto de pontos recorrentes contido nele, é definida como

$$\text{LAM} = \frac{\sum_{v=v_{min}}^{N} vP^{\varepsilon}(v)}{\sum_{v=1}^{N} vP^{\varepsilon}(v)}. \tag{3.54}$$

LAM é chamada de laminariedade, onde v_{min} é o tamanho mínimo com que se deseja computar uma estrutura vertical. Por exemplo, se utilizarmos $v_{min} = 2$ serão consideradas apenas estruturas verticais que tiverem um tamanho mínimo igual a dois estados recorrentes, neste caso qualquer evidência mínima de estruturas verticais será levada em consideração. A laminariedade fornece a quantidade de estruturas verticais (ou horizontais) do gráfico de recorrência, e representa a ocorrência de estados recorrentes que não mudam no tempo sem, contudo, descrever o comprimento desse estado laminar, que mede justamente o seu tempo de ocorrência.

A medida de LAM decresce se no gráfico de recorrência (RP) tiverem mais pontos singulares do que estruturas verticais (Marwan, 2003). Todavia, se quisermos medir o tempo médio dos estados laminares, utilizaremos o comprimento médio das linhas verticais, definido como

$$\text{TT} = \frac{\sum_{v=v_{min}}^{N} vP^{\varepsilon}(v)}{\sum_{v=v_{min}}^{N} P^{\varepsilon}(v)}. \tag{3.55}$$

TT é denominado de tempo de aprisionamento (*trapping time*) e mede o tempo médio que um estado permanece em um estado laminar, um estado que não muda no tempo. Da mesma forma que as estruturas diagonais, também podemos definir o tamanho máximo das estruturas verticais (*maximal length of vertical structures*) do gráfico de recorrência, como

$$\text{V}_{max} = \max(v_i;\ i = 1, \ldots, L). \tag{3.56}$$

Capítulo 4

Experimentos

Neste capítulo são descritos:

- a base de dados do projeto ELSA;
- os grupos formados a partir da base de dados;
- o processo de filtragem da base;
- a análise de estacionariedade das séries RR;
- a escolha de parâmetros associados aos índices de VFC implementados;
- os resultados obtidos.

4.1 Base de Dados

4.1.1 Projeto ELSA

A implantação do Estudo Longitudinal de Saúde do Adulto (ELSA Brasil) é uma iniciativa do Ministério da Saúde, por intermédio do Departamento de Ciência e Tecnologia (Decit), da Secretaria de Ciência, Tecnologia e Insumos Estratégicos, e do Ministério da Ciência e Tecnologia, por meio da Financiadora de Estudos e Projetos (Finep).

O objetivo do ELSA Brasil é identificar a maneira como diferentes fatores agem na predisposição para hipertensão e para o diabetes (Aquino et al., 2012). Uma amostra populacional de aproximadamente 15 mil pessoas, com idades entre 35 e 74 anos, será monitorada e

acompanhada durante pelo menos 15 anos. Esse acompanhamento demonstrará a evolução corporal dos voluntários; questionários detalhados oferecerão subsídios sobre os hábitos alimentares, o modo de vida, além de exames clínicos e laboratoriais periódicos, e checagem de medidas como o Índice de Massa Corporal.

Antes dos exames efetuados no ELSA Brasil, os participantes foram orientados a não ingerir café e não fumar. A coleta dos dados foi realizada por meio de registros eletrocardiográficos (eletrocardiógrafo digital Micromed, taxa de amostragem de 250 Hz) nos indivíduos em posição supina, em um ambiente silencioso com temperatura controlada ($\simeq 23\ ^{o}$C). Os eletrodos foram posicionados nos membros e uma série de intervalos RR de 10 minutos foi gerada (software Wincardio 4.4) a partir da derivação de maior amplitude da onda R (geralmente D_{II}) (Aquino et al., 2012). Durante a coleta, um filtro digital foi aplicado para eliminação de ruídos decorrentes de tremores musculares, oscilações periódicas da rede elétrica (frequência de 60 Hz) e para minimizar as variações da linha de base.

4.1.2 Pré-Processamento

A análise da VFC é concentrada, principalmente, nos batimentos originados no nodo sinusal. Assim, todos os outros batimentos (supraventricular, ventricular e relacionados a marcapassos) devem ser removidos. Isso também acontece quando o sinal contém artefatos. Além disso, intervalos RR anormais podem interferir na interpretação da VFC.

Realizou-se o pré-processamento dos intervalos RR da base ELSA, eliminando intervalos RR anormais. Assim,

- intervalos RR maiores que 2 s ou menores que 0.5 s foram suprimidos, visto que os intervalos RR característicos de batimentos irregulares são normalmente mais curtos ou mais longos (por exemplo, $\mathrm{RR}_i < 500$ ms ou $\mathrm{RR}_i > 2000$ ms, respectivamente) que os fisiologicamente aceitáveis para indivíduos normais;
- intervalos cuja diferença é superior a 20% da média dos dez últimos intervalos RR normais foram removidos, pois a origem de tais intervalos seriam os batimentos ectópicos. Assim, o intervalo RR que atender à condição $\left\|\overline{\mathrm{RR}} - \mathrm{RR}_{\mathrm{Normal}}\right\| \leq 20\%\overline{\mathrm{RR}}$ é considerado normal, tal que $\overline{\mathrm{RR}}$ representa a média dos dez últimos intervalos RR normais.

Os intervalos RR anormais e/ou ectópicos foram substituídos por valores sintéticos que seguem o mesmo ritmo dos valores normais adjacentes, via interpolação linear (Task Force,

1996). Nessa etapa, quando a série de intervalos RR possuía elevado percentual de interpolações (superior a 20%), o registro era classificado como inadequado para análise e, portanto, era eliminado da base de dados. Sendo assim, foram eliminados 537 registros da base.

4.1.3 Grupos

Dados foram coletados de registros de participantes do ELSA Brasil com o objetivo de verificar diferenças em índices de VFC, lineares e não lineares, entre indivíduos com diagnóstico previamente estabelecido de diabetes (HMPA04 $= 1$) com glicemia de jejum (GJ) superior a 140 mg/dL, e indivíduos sem diagnóstico de diabetes (HMPA04 $= 0$) com glicemia de jejum (GJ) inferior a 100 mg/dL. Cada indivíduo na base é identificado por um número (ID).

Para a separação dos grupos pertencentes à base de dados ELSA, as seguintes etapas de processamento foram executadas:

- eliminaram-se os IDs que possuíam informações de exame ausentes, bem como IDs com ausência de informação de nascimento e/ou sexo. Além disso, IDs com valores absurdos de idade foram eliminados. Com esta filtragem inicial, observando os exames de laboratório para a análise da glicemia (GJ, TTOG e A1C), saímos de 14885 registros para 10971 registros utilizáveis;

- fez-se a correspondência desse grupo de IDs, analisando-se a História Médica Pregressa (HMPA04). Então, 9 registros sem informação de HMPA04 foram eliminados. Nessa etapa, temos um total de 10962 registros de indivíduos da ELSA Brasil utilizáveis para análises subsequentes;

- separou-se o grupo de indivíduos diabéticos. Para tal, foram selecionados os IDs com HMP04 $= 1$ (pois o valor de HMP04 igual a 2 é indicativo de diabetes apenas em mulheres grávidas) e GJ > 140 mg/dL.

- separou-se o grupo de indivíduos saudáveis (grupo controle). Para tal, foram selecionados os IDs com HMP04 $= 0$ e GJ < 100 mg/dL;

- após a separação inicial desses dois grupos, verificaram-se os IDs para os quais existiam séries RR. Nesta filtragem, temos um total de 363 IDs de indivíduos diabéticos e 2685 IDs de indivíduos saudáveis;

- utilizaram-se apenas 360 indivíduos diabéticos nas análises pois o particionamento desse grupo em subgrupos é facilitado na etapa de classificação;

- houve escolha aleatória dos indivíduos sadios, de tal forma que obtivéssemos a mesma distribuição de homens e mulheres que o grupo de indivíduos diabéticos;
- os demais grupos são obtidos do total de 10962 registros, de acordo com a descrição apresentada na Seção 1.4.

As investigações inicias foram realizadas em dois grupos (1 e 2) da base de dados. Os demais grupos (3, 4 e 5) foram formados para realização das demais análises propostas na metodologia (Vide Seção 1.4).

Grupo 1: Indivíduos Diabéticos

A primeira amostra, denominada de Grupo 1 ou grupo de indivíduos com diagnóstico clínico de diabetes *mellitus*, contém dados de 360 indivíduos com relato da presença da doença na história médica pregressa (HMPA04) e por GJ > 140 mg/dL. Essa amostra é composta por 59% de indivíduos do sexo masculino (211 indivíduos) e 41% do sexo feminino (149 indivíduos).

O valor de GJ é superior a 140 mg/dL (GJ > 140 mg/dL), com média e desvio padrão dados por $205,9 \pm 64,3$ mg/dL. A faixa etária do grupo é de 38 a 74 anos, com média e desvio padrão dados por $58,3 \pm 8,0$ anos.

Grupo 2: Indivíduos Saudáveis

A segunda amostra, denominada de Grupo 2 ou grupo de indivíduos "saudáveis", contém dados de 360 indivíduos sem relato da presença da doença na história médica pregressa (HMPA04) e por GJ < 100 mg/dL. Esta amostra também é composta por 59% de indivíduos do sexo masculino (211 indivíduos) e 41% do sexo feminino (149 indivíduos).

O valor de GJ tem média e desvio padrão dados por $94,8 \pm 4,3$ mg/dL. A faixa etária do grupo é de 35 a 73 anos, com média e desvio padrão dados por $47,7 \pm 7,8$ anos.

Grupo 3

A terceira amostra, denominada de Grupo 3, caracteriza-se por 65 indivíduos que utilizam medicação antidiabética e possuem GJ < 125 mg/dL. Esta amostra é composta por 51% de indivíduos do sexo masculino (33 indivíduos) e 49% do sexo feminino (32 indivíduos).

O valor de GJ é inferior a 125 mg/dL, com média e desvio padrão dados por $109,4 \pm 12,6$ mg/dL). A faixa etária do grupo é de 35 a 73 anos (média e desvio padrão: $56 \pm 9,4$ anos).

Grupo 4

A quarta amostra, denominada de Grupo 4, caracteriza-se por 99 indivíduos que não utilizam medicamentos e possuem GJ entre 100 mg/dL e 125 mg/dL. Esta amostra é composta por 70% de indivíduos do sexo masculino (69 indivíduos) e 30% do sexo feminino (30 indivíduos).

O valor de GJ pertence ao intervalo $100 < \text{GJ} \leq 125$ mg/dL, com média e desvio padrão $109,5 \pm 6,2$ mg/dL. A faixa etária do grupo é de 35 a 72 anos ($49,6 \pm 7,9$ anos).

Grupo 5

A quinta amostra, denominada de Grupo 5, caracteriza-se por 99 indivíduos que não utilizam medicamentos e possuem GJ entre 126 mg/dL e 139 mg/dL. Esta amostra é composta por 81% de indivíduos do sexo masculino (80 indivíduos) e 19% do sexo feminino (19 indivíduos).

O valor de GJ pertence ao intervalo $126 \leq \text{GJ} \leq 139$ mg/dL, com média e desvio padrão $131,3 \pm 3,6$ mg/dL. A faixa etária do grupo é de 36 a 70 anos ($52,9 \pm 7,6$ anos).

4.2 Teste de Estacionariedade

A aplicação da maioria das técnicas de análise de sinais lineares e não lineares requer algum tipo de estacionariedade. Infelizmente, a detecção de estacionariedade em séries temporais não é uma tarefa trivial, principalmente quando são analisadas séries temporais experimentais. Desta forma, o teste de estacionariedade ou não estacionariedade dos sinais deve fazer parte da primeira fase da análise.

Vale ressaltar que sinais não estacionários também são importantes, pois existem situações em que a observação de alterações dinâmicas torna-se fundamental na análise de determinados processos. Por exemplo, podemos citar o uso de séries não estacionárias de VFC como ferramenta de análise da modulação autonômica da atividade cardíaca, relações com condições patológicas, entre outros.

Alguns testes de estacionariedade apresentados na literatura estimam determinados parâmetros estatísticos a partir de segmentos da série temporal, tal que se as flutuações estatísticas dos parâmetros ficam restritas a um intervalo de confiança, então a série temporal é considerada estacionária. Na análise de séries temporais, a estacionariedade é geralmente testada a partir de parâmetros estatísticos de primeira e segunda ordem, tais como média, variância e densidade espectral de potência. No caso de sinais obtidos de sistemas não lineares, a estacionariedade fraca não é muito apropriada, por isso é desejável utilizar algum quantificador não linear a fim de checar a não estacionariedade.

Antes dos cálculos dos índices de VFC, foram aplicados dois testes, a saber, *Kwiatkovski-Phillips-Schmidt-Shin test - Augmented Dickey-Fuller test* (KPSS-ADF) e *Restricted Weak Stationary* (RWS), para avaliar a estacionariedade das séries RR utilizadas. Foi aplicado um teste de estacionariedade de cada vez, tal que ambos, independentemente, mostraram que todas as séries RR de 10 minutos são não estacionárias. Depois analisaram-se séries RR de 5 minutos, constatando-se que apenas cerca de 2% eram estacionárias. Assim, de acordo com (Task Force, 1996), desconsidera-se a análise de estacionariedade e utilizam-se séries RR de 5 minutos de duração para a geração dos índices de VFC dos grupos de indivíduos da base ELSA Brasil.

4.2.1 Teste KPSS-ADF

Foi implementado um teste que combina os testes complementares *Kwiatkovski-Phillips-Schmidt-Shin test* (KPSS) (Kwiatkowski et al., 1992) e *Augmented Dickey-Fuller test* (ADF) (Dickey e Fuller, 1979) para determinação da estacionariedade da série RR.

Portanto, foi utilizada uma combinação dos testes KPSS e ADF, que são testes de hipótese nula complementares. Ambos são baseados na tentativa de rejeição de uma hipótese inicial $(\mathrm{H}_0 = 1)$. Se o teste não for capaz de rejeitar a hipótese nula $(\mathrm{H}_0 = 0)$, então, nenhuma conclusão pode ser tirada.

A hipótese nula do teste KPSS é que o sinal é estacionário em torno de uma tendência determinística, contra a alternativa de que o sinal seja um processo não estacionário de raiz unitária. Já a hipótese nula do teste ADF é que o sinal é um processo não estacionário de raiz unitária, contra a alternativa de que o sinal seja estacionário em torno de uma tendência determinística.

Dessa forma, conclusões podem ser tomadas somente quando:

1. a hipótese nula de KPSS é confirmada e a hipótese nula de ADF é rejeitada, ou seja, $\mathrm{KPSS}|\mathrm{H}_0 = 0$ e $\mathrm{ADF}|\mathrm{H}_0 = 1$. Neste caso, o sinal é considerado estacionário;

2. a hipótese nula de KPSS é rejeitada e a hipótese nula de ADF é confirmada, ou seja, KPSS$|\mathrm{H}_0 = 1$ e ADF$|\mathrm{H}_0 = 0$. Neste caso, o sinal é considerado não estacionário.

Além disso, H_0 somente é rejeitada se o valor obtido no teste (p-valor) ultrapassar um valor pré-determinado com um intervalo de confiança dado por α (Kwiatkowski et al., 1992). Neste trabalho considerou-se $\alpha = 5\%$.

4.2.2 Teste RWS

Foi implementado um teste (Porta et al., 2004) que analisa a estabilidade da média e variância (do inglês, *restricted weak stationary* ou RWS) sobre séries RR de aproximadamente 5 minutos (≈ 300 amostras).

Dada a série $\mathbf{x} = \{x(i),\ i = 1,2,\ldots,N\}$ testa-se, primeiramente, a normalidade de sua distribuição por meio do *Kolmogorov-Smirnov goodness-of-fit Test* ($p < 0.05$) (Massey, 1951; Marsaglia et al., 2003). Depois, a série RR é dividia em $N-L+1$ sequências de comprimento L (ou subséries RR), como o padrão $\mathbf{x}_L(i) = [x(i), x(i+1), x(i+L-1)]$ derivado de $\mathbf{x}$. Assim, tem-se $N-L+1$ subséries da série RR original:

$$\begin{aligned} \mathbf{RR}_1 &= \mathrm{RR}(1), \mathrm{RR}(2), \ldots, \mathrm{RR}(L) \\ \mathbf{RR}_2 &= \mathrm{RR}(2), \mathrm{RR}(3), \ldots, \mathrm{RR}(L+1) \\ \mathbf{RR}_P &= \mathrm{RR}(P), \mathrm{RR}(P+1), \ldots, \mathrm{RR}(N),\ P = N-L+1 \end{aligned} \tag{4.1}$$

Depois, M subséries são selecionadas aleatoriamente, com mesma probabilidade de escolha, de $N-L+1$, tal que $M < N-L+1$.

Após a análise de normalidade da série de dados, divisão em subséries e escolha de M subséries, parte-se para a análise da média e da variância das M subséries escolhidas. No caso, a análise da média e da variância separa-se em dois casos, a saber, quando a série RR tem distribuição normal (hipótese nula verdadeira) e quando tem distribuição não normal (rejeição da hipótese nula).

1. Na análise das médias das M subséries ($p < 0.05$):

   - distribuição normal: o teste de estabilidade da média é feito por meio da análise da variância utilizando o *F-test* (Hogg e Ledolter, 1987);
   - distribuição não normal: compara a estabilidade da mediana por meio do *Kruskall-Wallis test* (Hollander e Wolfe, 1999; Gibbons e Chakraborti, 2010).

2. Na análise das variâncias das M subséries:

   - distribuição normal: o teste de estabilidade da variância é feito por meio *Barlett Test* (Bartlett, 1937);
   - distribuição não normal: aplica-se o *Levene test* (Levene, 1960) para análise da estabilidade da variância utilizando a mediana.

Nas análises, de média e variância, cada subsérie é considerada como um grupo. Então a análise consiste em comparar as distribuições de M grupos. Caso a hipótese de nulidade seja confirmada, ou seja, caso os grupos tenham a mesma distribuição na análise da média e da variância, independente da distribuição (normal ou não), considera-se a série de dados estacionária.

Além do nível de confiança ($\alpha = 5\%$), três outros parâmetros precisam ser definidos: o comprimento da série temporal (N), o comprimento das subséries (L), e o número de subséries/padrões (M). L foi escolhido suficientemente grande (50 batimentos cardíacos) para observar vários ciclos (5) do fenômeno mais lento (no caso, a oscilação LF) que foi considerado estável/estacionário na série de VFC. Por exemplo, 5 ciclos de oscilação LF em $0,1$ Hz com período cardíaco médio de 1 s (Task Force, 1996) equivalem a 50 batimentos cardíacos. M foi definido como 8 para garantir que a análise de subséries, com seleção aleatória, seja correspondente a analisar quase todas as amostras da série. N foi definido como 300 batimentos cardíacos. Os parâmetros M, N, e L, foram escolhidos de acordo com (Magagnin et al., 2011).

Modificações no Teste RWS

Foram realizadas as seguintes modificações no Teste RWS:

- alteração dinâmica do valor de N. De fato, um sinal de entrada de cinco minutos de duração não necessariamente possui 300 amostras. Sendo assim, N assume o tamanho da série de análise;
- alteração na seleção das subséries (padrões). Observou-se que em função da seleção aleatória das M subséries de tamanho L poderíamos ter a seguinte situação: $\mathbf{RR}_1, \mathbf{RR}_2, \mathbf{RR}_3, \ldots, \mathbf{RR}_M$. Nessa situação as M subséries apresentam grande sobreposição e, consequentemente, possuem distribuições aproximadamente iguais, caracterizando a série RR como estacionária. Assim, em função da escolha aleatória das posições iniciais das subséries (índices), uma série de dados poderia, em alguns momentos, ser considerado estacionária, e em outros, não estacionária. Portanto, fez-se a

alteração dos índices escolhidos aleatoriamente na faixa de $N-L+1$ de tal forma que estes fossem sempre os mesmos. No caso, escolhe-se o inicio de posições de janelas de tal forma que tenhamos M subséries de tamanho L, mesmo que N seja diferente de 300. Assim, obtêm-se janelas de posição fixa com sobreposição (quando necessário) de tal forma a abranger toda a extensão da série de tamanho N.

4.3 Escolha de Parâmetros na Geração dos Índices de VFC

Alguns dos índices implementados requerem a determinação de parâmetros de entrada para posterior geração de seu valor. Portanto, as escolhas mais relevantes no processo de geração dos índices de VFC são enumeradas abaixo:

1. no domínio da frequência, a estimativa da densidade espectral de potência foi calculada por meio de um modelo autorregressivo (AR) de ordem 16 (Boardman et al., 2002; Carvalho et al., 2003; Dantas et al., 2012). Os valores de potência absoluta para cada banda de frequência foram calculados por meio do método dos resíduos;

2. nos métodos não lineares que calculam a entropia, vemos que ambos, ApEn e SampEn, dependem da escolha dos parâmetros m e r, de forma diferenciada. A determinação dos valores ótimos dos parâmetros m e r é um problema em aberto (Pincus, 1991; Acharya et al., 2006; Faust et al., 2011). Neste trabalho, utiliza-se a pesquisa em grade sobre os valores comumente utilizados na literatura (Acharya et al., 2006; Faust et al., 2011) avaliando a discriminância estatística entre os grupos de indivíduos sadios e diabéticos (p-valor $< 0,05$). Assim, determinam-se os valores $m = 2$ e $r = 0,1$ para ApEn e $m = 2$ e $r = 0,2$ para SampEn;

3. nos cálculos da análise das flutuações destendenciadas (DFA), devemos definir os valores de comprimento de janela (l_k) para o cálculo de α_1 e α_2. O valor de α_1 é calculado para as janelas cujo tamanho varia entre 4 e 16 batimentos, tal que $(4 \leq l_k \leq 16)$. Já o valor de α_2 é calculado para janelas com tamanhos superiores a 16 batimentos $(16 \leq l_k \leq N;\ N = 64)$ (Peng et al., 1995; Acharya et al., 2006; Almeida et al., 2012);

4. nos cálculos do gráfico de recorrência (RP), os parâmetros m, τ, ε são definidos como:

   - o cálculo do valor ótimo de m ocorre através da investigação de mudanças na vizinhança de pontos no espaço de fase tal que a dimensão de imersão ideal é aquela escolhida quando o número de falsos vizinhos cai a zero (algoritmo *false nearest neighbours*) (Kennel et al., 1992). Na prática, observa-se que o número

de falsos vizinhos diminui com o aumento de m. No caso, consideraremos o valor ótimo de m igual a 10;

- o valor do *threshold* ε, ou tamanho da vizinhança, é definido como 10% do diâmetro máximo (ou médio) do espaço de fase. Outras formas de se escolher o valor de ε são: quando a taxa de recorrência (REC) for igual a 10% (Marwan et al., 2007) ou utilizando um valor fixo relativo à dimensão de imersão utilizada, como $\varepsilon = m^{1/2}$. Optou-se por utilizar $\varepsilon = m^{1/2}$ para evitar cálculos de obtenção de ε para cada série RR;
- o valor do tempo de retardo τ é definido como 1 (Acharya et al., 2006; Faust et al., 2011).

4.4 Detalhes do Algoritmo do Classificador SVM

Para utilizar o classificador SVM, deve-se adaptar o formato dos dados de entrada para o formato da biblioteca *libsvm* (Chang e Lin, 2011). Assim, as seguintes etapas são realizadas:

1. ajuste da faixa geral de variação dos índices. Realiza-se um ajuste de faixa dinâmica dos índices de VFC gerados. No caso, aplica-se apenas um fator de ponderação sobre cada índice. Isto é feito pois cada índice tem uma origem de cálculo distinta, o que implica em escalas diferenciadas. Por exemplo, a energia de uma dada banda de frequência do espectro do sinal (como VLF) é dada em ms^2, a porcentagem de intervalos RR adjacentes com diferença de duração maior que 50 ms é expresso em porcentagem, e a dimensão fractal (FD) é um número real, normalmente, na faixa $[1,2]$;

2. escalonamento. Projeta-se linearmente os dados de entrada para a faixa de $[-1,+1]$ ou $[0,1]$. A principal vantagem desta abordagem é evitar que características com elevados valores numéricos dominem as de menor valor, ou ainda evitar a saturação do classificador. Outra vantagem é evitar algum problema durante os cálculos, visto que a função núcleo, utilizada no SVM, é baseado no produto interno dos vetores de características;

3. seleção da função de núcleo. Escolheu-se o núcleo RBF (do inglês, *radial basis function*) pois é bastante utilizado para resolução de problemas de aprendizagem (não linearmente separáveis) e necessita da determinação de apenas dois parâmetros: γ (*gamma*) e C (custo), o que reduz a complexidade do classificador. No núcleo RBF o número de funções radiais e os seus respectivos centros são definidos pelos vetores suporte obtidos (Vide Anexo A);

4. escolha dos parâmetros do núcleo RBF. A priori, não são conhecidos os parâmetros C e γ ótimos do núcleo escolhido para o problema de classificação. Consequentemente, algum tipo de busca de parâmetros deve ser feita. O objetivo deste método de pesquisa é identificar os parâmetros C e γ de tal forma que o classificador SVM possa classificar com boa acurácia dados novos (por exemplo, os dados do conjunto de teste). Um método comum é separar os dados de treinamento em duas partes, sendo que uma delas é desconhecida pelo classificador (conjunto de validação). Então, a acurácia sobre esse subconjunto pode refletir mais precisamente o desempenho na classificação de dados desconhecidos. Baseado nesta ideia, surge o procedimento denominado de validação cruzada durante a fase de treinamento do classificador;

5. aplicação da validação cruzada. Realiza-se a validação cruzada para a obtenção de valores médios representativos das taxas de acerto sobre os conjuntos de treinamento / validação. Ou ainda, a estimativa de erro global é calculada como a média das v estimativas de erro de cada iteração. Para tal, o conjunto de dados de treinamento é dividido em v subconjuntos de mesmo tamanho, depois cada subconjunto é usado para validação e os demais são utilizados para treinamento. Este processo é repetido v vezes. Por isso o processo é denominado de validação cruzada por v vezes (vide Figura 4.1);

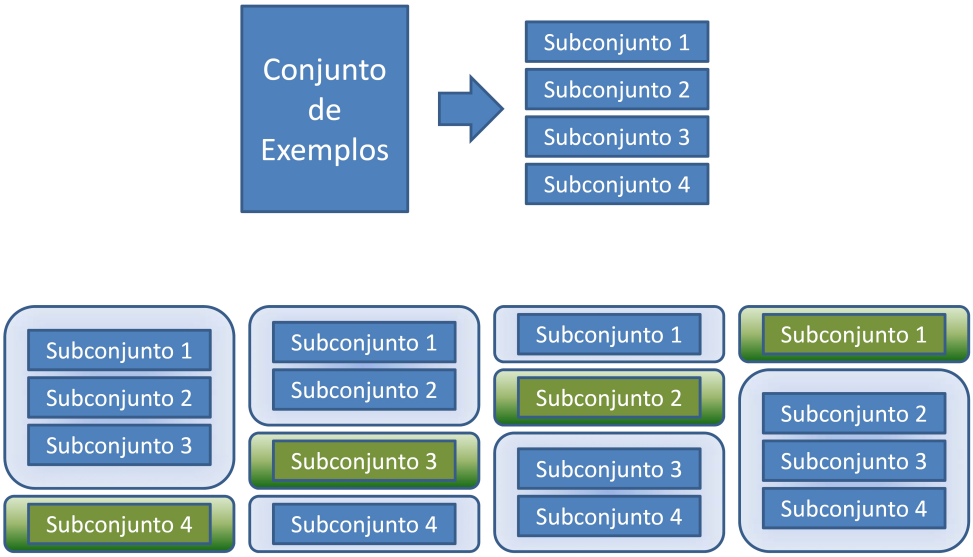


Figura 4.1: Validação cruzada com a divisão do conjunto de exemplos em 4 subconjuntos.

6. implementação do método *grid search*. O método de busca dos parâmetros utilizado é o "*grid search*", conhecido como método de pesquisa em grade. Neste método, pares

de (C, γ) são testados e o modelo que apresentar melhor acurácia, relativa à validação cruzada, é utilizado. Nos pares testados, sequências dos parâmetros com crescimento exponencial são testadas (por exemplo, $C = 2^{-5}, 2^{-3}, \ldots, 2^{15}$, e $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$). Verifica-se que o *grid search* é um método prático e simples que permite obter bons parâmetros C e γ. É importante observar que uma pesquisa em grade muito precisa implica na obtenção de parâmetros próximos dos ótimos (C e γ) sobre o conjunto de treinamento/validação, o que não necessariamente, quando aplicados ao núcleo escolhido, apresentará melhor resultado para o conjunto de teste. Em outras palavras, a determinação de parâmetros ótimos (C e γ) com vistas à otimização do modelo gerado para o classificador na fase de treinamento pode levar a um problema de generalização do mesmo, o que será expresso por 100% de acurácia no conjunto de treinamento e baixa acurácia no conjunto de teste.

4.5 Resultados

4.5.1 Teste Estatístico

As Tabelas 4.1, 4.2 e 4.3 apresentam os resultados (média e desvio padrão) dos índices associados aos grupos 1 e 2. Observa-se que apenas os índices RR_{mean} (vide Tabela 4.1), α_2 e FD (vide Tabela 4.3) possuem distribuição normal. Nestas tabelas os índices de VFC com distribuição normal são indicados com um asterisco.

A comparação foi feita utilizando o teste t (*Kolmogorov-Smirnov*) (Boneau, 1960) para os índices com distribuição normal. A hipótese nula é que os dados dos dois grupos são amostras independentes e aleatórias, de distribuições normais com médias iguais e variâncias iguais, porém desconhecidas. A hipótese não nula (alternativa) é que as médias não são iguais. O resultado indica a rejeição, ou não, da hipótese nula ao nível de significância de 5% (análise de p-valor).

Para os índices com distribuição não normal a comparação entre os grupos foi feita utilizando o *Mann-Whitney U-test* (Hollander e Wolfe, 1999; Gibbons e Chakraborti, 2010). A hipótese nula é que os dados dos dois grupos são amostras independentes a partir de distribuições, contínuas e idênticas, com medianas iguais, contra a alternativa que eles não têm medianas iguais. O resultado do teste indica a rejeição, ou não, da hipótese nula ao nível de significância de 5% (análise de p-valor).

Domínio do Tempo

A Tabela 4.1 mostra que todos os índices no domínio do tempo (estatísticos e geométricos) são estatisticamente significativos, ou seja, os índices são diferentes entre os grupos de indivíduos sadios (grupo controle) e diabéticos (p-valor $< 0,0001$). Observa-se que todos os índices no domínio do tempo, exceto o *Triangular index*, apresentam valores inferiores para o grupo de pacientes com diabetes ao compará-los com os do grupo controle.

Tabela 4.1: Métodos Estatísticos e Geométricos.

Índice	**Unidade**	**Controle** (média $\pm$ desvio padrão)	**Diabéticos** (média $\pm$ desvio padrão)	**p-valor**
RR^{*}_{mean}	ms	$939,80 \pm 123,97$	$844,43 \pm 133,57$	$< 0,0001$
SDNN	ms	$42,70 \pm 16,84$	$30,90 \pm 17,06$	$< 0,0001$
RMSSD	ms	$33,73 \pm 18,99$	$22,42 \pm 16,50$	$< 0,0001$
pNN50	%	$12,00 \pm 15,10$	$5,81 \pm 11,15$	$< 0,0002$
Δ index	-	$0,12 \pm 0,05$	$0,18 \pm 0,09$	$< 0,0001$

Obs.: Utiliza-se o *T-test* ($*$ $-$ índice com distribuição normal) ou o *Mann-Whitney U-test* para comparação entre os grupos. Onde, N $\equiv$ 360 indivíduos para cada grupo.

Domínio da Frequência

A Tabela 4.2 apresenta a análise no domínio da frequência, em que se observa que somente os índices LF_{norm} e LF/HF não apresentaram diferença estatística entre os dois grupos de indivíduos (p-valor $> 0,05$).

Tabela 4.2: Métodos no Domínio da Frequência.

Índice	**Unidade**	**Controle** (média $\pm$ desvio padrão)	**Diabéticos** (média $\pm$ desvio padrão)	**p-valor**
VLF	ms^2	$946,94 \pm 946,60$	$587,44 \pm 802,42$	$< 0,0001$
LF	ms^2	$478,79 \pm 504,42$	$261,90 \pm 391,29$	$< 0,0001$
HF	ms^2	$509,50 \pm 674,69$	$232,18 \pm 366,68$	$< 0,0001$
HF_{norm}	n.u.	$46,10 \pm 18,73$	$41,40 \pm 20,19$	$< 0,001$
LF_{norm}	n.u.	$48,82 \pm 19,83$	$45,86 \pm 23,78$	$> 0,05$
LF/HF	-	$1,64 \pm 1,93$	$2,20 \pm 3,88$	$> 0,05$

Obs.: Utiliza-se o *Mann-Whitney U-test* para comparação entre os grupos. Onde, N $\equiv$ 360 indivíduos para cada grupo.

Não Lineares

A Tabela 4.3 mostra que os índices não lineares, em sua maioria, possuem p-valor menor que $0,0001$. Além disso, somente o índice α_2 não apresentou diferença significativa entre os grupos.

Tabela 4.3: Métodos Não Lineares.

Índice	**Unidade**	**Controle** (média $\pm$ desvio padrão)	**Diabéticos** (média $\pm$ desvio padrão)	**p-valor**
SD1	ms	$23,18 \pm 13,45$	$15,87 \pm 11,69$	$< 0,0001$
SD2	ms	$55,12 \pm 21,33$	$40,19 \pm 22,07$	$< 0,0001$
SD1/SD2	-	$0,42 \pm 0,18$	$0,40 \pm 0,21$	$< 0,005$
s	-	$4600,78 \pm 4395,47$	$2586,02 \pm 4600,78$	$< 0,0001$
ApEn	-	$0,81 \pm 0,19$	$0,85 \pm 0,24$	$< 0,0001$
SampEn	-	$1,59 \pm 0,27$	$1,52 \pm 0,35$	$< 0,001$
α_1	-	$1,05 \pm 0,25$	$1,10 \pm 0,26$	$< 0,005$
α_2^*	-	$0,93 \pm 0,20$	$0,95 \pm 0,20$	$> 0,05$
FD*	-	$2,21 \pm 0,49$	$1,87 \pm 0,43$	$< 0,0001$
REC	%	$32,52 \pm 9,55$	$35,18 \pm 9,53$	$< 0,0002$
DET	%	$97,78 \pm 1,38$	$98,16 \pm 1,42$	$< 0,0005$
L_{mean}	batimentos	$11,26 \pm 4,46$	$12,30 \pm 4,54$	$< 0,005$
L_{max}	batimentos	$177,70 \pm 120,82$	$216,01 \pm 120,30$	$< 0,0001$
ShanEn	-	$3,13 \pm 0,32$	$3,23 \pm 0,34$	$< 0,0001$

Obs.: Utiliza-se o *T-test* ($*$ $-$ índice com distribuição normal) ou o *Mann-Whitney U-test* para comparação entre os grupos. Onde, N $\equiv$ 360 indivíduos para cada grupo.

4.5.2 Medidas de Desempenho

A princípio utilizou-se a matriz de confusão para se visualizar o desempenho do classificador, analisando a sensibilidade (taxa de verdadeiros positivos), especificidade, valor preditivo positivo (precisão), valor preditivo negativo, taxa de falsos positivos e acurácia. Com o intuito de exaurir dúvidas quanto à descrição e funcionalidade de cada medida de desempenho, todas estas medidas serão descritas brevemente.

A matriz de confusão é expressa por

$$\begin{bmatrix} \text{VP} & \text{FP} \\ \text{FN} & \text{VN} \end{bmatrix},$$

onde VP, FN, FP e VN, significam verdadeiro positivo, falso negativo, falso positivo e verdadeiro negativo, respectivamente. As linhas da matriz de confusão representam os valores preditos (saída do classificador) e as colunas representam os valores reais (anotação da base). O total de elementos da classe de diagnosticados como doentes é dado por $\mathrm{VP}+\mathrm{FN}$, e o total de elementos da classe de diagnosticados como sadios é dado por $\mathrm{FP}+\mathrm{VN}$.

A sensibilidade (do inglês, *sensitivity*, *recall* ou *hit rate*) é equivalente a taxa de verdadeiros positivos (do inglês, *true positive rate*). A sensibilidade, ou a razão de verdadeiros positivos, é a porcentagem dos verdadeiramente doentes, indicados como doentes. Sendo assim, é a razão entre VP e $(\mathrm{VP}+\mathrm{FN})$, ou seja,

$$\text{Sensibilidade} = \frac{\mathrm{VP}}{\mathrm{VP}+\mathrm{FN}} \times 100, \tag{4.2}$$

e

$$\mathrm{VP}_{\text{RATE}} = \text{Sensibilidade}. \tag{4.3}$$

A especificidade (do inglês, *specificity*), ou a razão de verdadeiros negativos, é a porcentagem dos verdadeiramente sadios, indicados como sadios. No caso, a sensibilidade é a porcentagem dos verdadeiramente sadios em relação a todos realmente sadios, ou seja,

$$\text{Especificidade} = \frac{\mathrm{VN}}{\mathrm{FP}+\mathrm{VN}} \times 100. \tag{4.4}$$

O valor preditivo positivo ou precisão (do inglês, *positive predictive value* ou *precision*), também chamado de seletividade positiva (VPP), é a capacidade do teste ou exame em indicar corretamente os doentes. No caso, o VPP é a porcentagem dos verdadeiramente doentes em relação a todos indicados como doentes pelo classificador, ou seja,

$$\mathrm{VPP} = \frac{\mathrm{VP}}{\mathrm{VP}+\mathrm{FP}} \times 100, \tag{4.5}$$

e

$$\text{Precisao} = \mathrm{VPP}. \tag{4.6}$$

O valor preditivo negativo (do inglês, *negative predictive value* ou VPN), também chamado de seletividade negativa, é a capacidade do teste ou exame em indicar corretamente os sadios. No caso, o VPN é a porcentagem dos verdadeiramente sadios em relação a todos indicados como sadios pelo classificador.

$$\mathrm{VPN} = \frac{\mathrm{VN}}{\mathrm{FN}+\mathrm{VN}} \times 100. \tag{4.7}$$

A taxa de falsos positivos (do inglês, *false positive rate* ou *false alarm rate*) é a razão entre FP e os verdadeiramente sadios $(\mathrm{FP}+\mathrm{VN})$, ou seja,

$$\mathrm{FP}_{\text{RATE}} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{VN}} \times 100. \tag{4.8}$$

Por último, temos a acurácia (do inglês, *accuracy*), que é a porcentagem de acerto do classificador, sendo a medida mais comumente utilizada como quantificador de desempenho de sistemas de classificação. No caso, a acurácia indica a razão entre a soma de verdadeiros positivos e verdadeiros negativos e todos os elementos de análise, ou seja,

$$\text{Acurácia} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{FP} + \text{VN}} \times 100. \tag{4.9}$$

4.5.3 Avaliação: Classificadores

Os classificadores k vizinhos mais próximos (KNN), *naive bayes* (NB), e máquinas de vetor suporte (SVM) apresentaram os resultados descritos na Tabela 4.4. O número de vizinhos mais próximos utilizados no KNN foi 13 ($k = 13$, obtido através da análise da acurácia do classificador para diferentes valores de k) e os parâmetros do núcleo RBF da SVM foram $C = 16.384$ e $\gamma = 9,7656 \times 10^{-4}$.

Analisando a acurácia, verifica-se que o classificador SVM apresenta melhores resultados,e por isso é escolhido como padrão nas análises subsequentes dos demais grupos.

Tabela 4.4: Medidas de Desempenho.

Indicador/Classificador	KNN	NB	SVM
Especificidade	87,50	63,89	80,56
Sensibilidade	52,78	77,78	70,83
VPN	64,95	74,19	73,42
VPP	80,85	68,29	78,46
Acurácia	70,14	70,83	75,69

Primeiramente vamos analisar apenas o conjunto de teste apresentado ao classificador SVM. Esse conjunto é composto por 144 indivíduos igualmente distribuídos nos grupos 1 e 2 (72 indivíduos cada). De acordo com o classificador, o grupo 1 possui 21 indivíduos diabéticos indicados como sadios, dentre os quais 8 são realmente sadios pelo exame de glicemia pós-prandial (TOTG). Já o grupo 2 possui 14 indivíduos sadios indicados como diabéticos, dentre os quais apenas 1 é diabético pelo TOTG. Assim, apenas analisando o elementos do conjunto de teste, temos que o classificador erra a identificação de 13 indivíduos de cada grupo. Portanto, a acurácia apresenta para os grupos 1 e 2 é de $81,94\%$.

Antes de analisar os resultados de classificação sobre os grupos intermediários (3, 4 e 5) foram feitos gráficos que comparam GJ com as duas primeiras principais componentes obtidas de uma transformação linear sobre os índices de variabilidade de frequência cardíaca

em direções ortogonais de máxima variância, no caso, da PCA (do inglês, *Principal Component Analysis*), ilustrando a relação entre índices de VFC e GJ, para GJ entre 100 e 140 $(100 < \text{GJ} < 140\,\text{mg/dL})$.

A ideia é reforçar que não é possível inferir uma relação explícita do valor de GJ com as componentes da PCA. De fato, observa-se uma diminuição do valor da componente principal em função do valor de glicemia. Isto é ilustrado no gráfico da Figura 4.2 uma vez que há uma sutil tendência entre GJ e a primeiras componentes da PCA, tal que maiores valores de glicemia implicam em menores valores da primeira componente da PCA na faixa de glicemia intermediária. Ainda na Figura 4.2, observa-se que os diabéticos identificados pelo

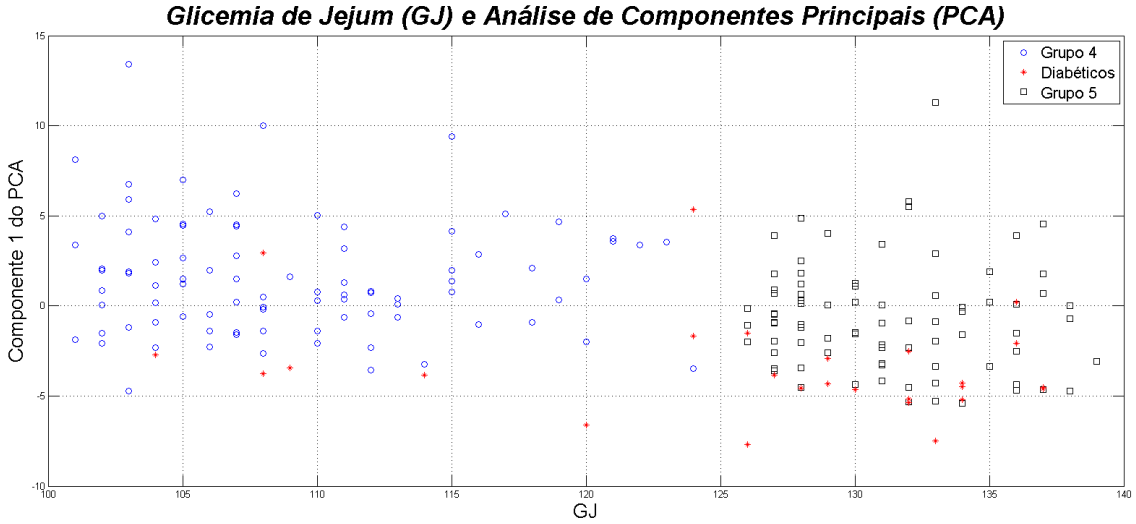


Figura 4.2: GJ e primeira componente da PCA

classificador possuem, no geral, valores menores de componente principal.

A Figura 4.3 ilustra a relação entre GJ com as duas primeiras componentes principais da PCA. Assim, as Figuras 4.2 e 4.3 exemplificam a dificuldade de se classificar indivíduos com valores de GJ caracterizados como pré-diabéticos, ou ainda, indivíduos com valores de glicemia intermediária na faixa de GJ entre 100 e 140 mg/dL.

Para a avaliação do modelo final do classificador SVM sobre os indivíduos dos grupos 3, 4 e 5, comparou-se os resultados de classificação com o exame de glicemia pós-prandial (TOTG), bem como com a indicação dada pela história médica pregressa (HMPA04).

Analisando os rótulos do classificador sobre o grupo 3, vemos que ele é composto por 21 diabéticos e 44 indivíduos saudáveis (totalizando 65 indivíduos). Da porção de 21 diabéticos, verifica-se que 15 indivíduos são indicados como sadios pelo exame TOTG (9 têm $110 < \text{GJ} \leq 125\,\text{mg/dL}$ e 4 têm $100 < \text{GJ} \leq 110\,\text{mg/dL}$). Porém 11 desses são identificados

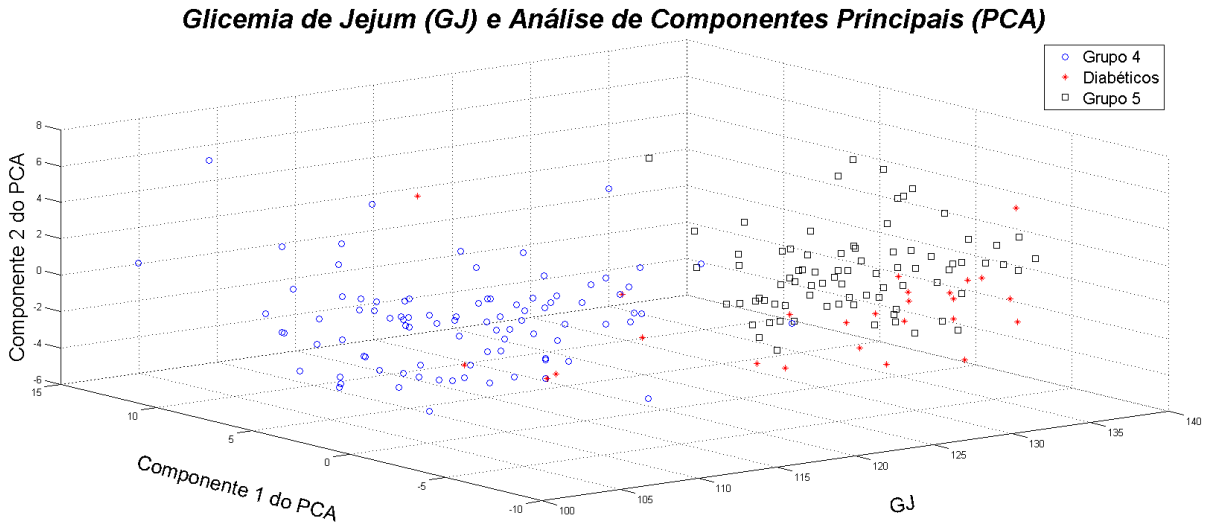


Figura 4.3: GJ e primeiras componentes da PCA

como diabéticos pela história médica pregressa (HMPA04 = 1). A ideia é que todo indivíduo com HMPA04 positivo (HMPA04 = 1) tem diabetes, e HMPA04 negativo (HMPA04 = 0) não garante que o indivíduo não seja diabético. Em relação à porção de 44 indivíduos sadios, verifica-se que 5 indivíduos são identificados como diabéticos pelo exame TOTG e confirmados pela história médica pregressa (4 desses possuem GJ < 100 mg/dL, a saber, $57, 87, 94$ e 94). Se compararmos os resultados de classificação do grupo 3 com o exame TOTG juntamente com a HMPA04, verifica-se que o classificador erra na identificação de 4 indivíduos sadios e 5 diabéticos, tal que a acurácia apresentada para o grupo 3 será de $86,15\%$.

De forma semelhante, o grupo 4 é composto por 8 diabéticos e 91 indivíduos saudáveis (totalizando 99 indivíduos). Da porção de 8 diabéticos, verifica-se que todos são indicados como sadios pelo exame TOTG e confirmados pelo história médica pregressa (HMPA04 = 0). Em relação à porção de 91 indivíduos sadios, verifica-se que 3 indivíduos são identificados como diabéticos pelo exame TOTG (3 possuem GJ próxima a 100 mg/dL, a saber, $105, 112$ e 115), porém não são confirmados pelo história médica pregressa (HMPA04 = 0). Se compararmos os resultados de classificação do grupo 4 como o exame TOTG e a HMPA04, verifica-se que o classificador erra apenas na identificação de 8 indivíduos sadios, tal que a acurácia apresentada para o grupo 4 será de $91,92\%$.

Relacionando GJ e HMPA04 nos grupos 3 e 4, vemos que:

- no grupo 3, 39 indivíduos apresentam história médica pregressa indicando a presença de diabetes (HMPA04 = 1). A priori, todo o grupo 3 seria classificado como sadio por

meio da análise individual de GJ, pois os valores de glicemia restringem-se à faixa de GJ < 125 mg/dL. Porém, ele possui 60% de diabéticos indicados por HMPA04;

- no grupo 4, apenas 2 indivíduos apresentam história médica pregressa indicando a presença de diabetes (HMPA04 $= 1$). De forma análoga ao grupo 3, a priori, todo o grupo 4 seria classificado como sadio por meio da análise individual de GJ (GJ $<$ 125 mg/dL). Assim, o grupo 4 possui $\approx 2,02\%$ de diabéticos indicados por HMPA04.

Ainda analisando os grupos 3 e 4, vemos que o classificador aprendeu em relação a GJ e HMPA4, em função da determinação dos grupos 1 e 2. Assim, valores de GJ próximos a 100 tendem a ser classificados como normais e, de alguma forma, a identificação de HMPA4 tem relação implícita com a VFC na determinação de indivíduos diabéticos. Além disso, a utilização de medicamentos (a exemplo do grupo 3) implica em alterações na VFC, que podem alterar a classificação.

Analisando os rótulos do classificador sobre o grupo 5, vemos que ele é composto por 18 diabéticos e 81 indivíduos saudáveis (totalizando 99 indivíduos). Da porção de 18 diabéticos, verifica-se que 11 são indicados como sadios pelo TOTG, dentro os quais apenas 1 é identificado como diabético pela história médica pregressa (HMPA04 $= 1$). Em relação à porção de 81 indivíduos sadios, verifica-se que 18 indivíduos são identificados como diabéticos pelo exame TOTG, ao passo que pela história médica pregressa 17 desses não são confirmados como diabéticos (HMPA04 $= 0$). Se compararmos os resultados de classificação do grupo 5 como o exame TOTG e a HMPA04, verifica-se que o classificador erra na identificação de 10 indivíduos sadios e 1 diabético, tal que a acurácia apresentada para o grupo 5 será de $88,89\%$.

4.5.4 Discussão

Antes dos cálculos dos índices de VFC, foram aplicados dois testes para avaliar a estacionariedade das séries RR utilizadas. Ambos os testes KPSS-ADF e RWS mostraram que as séries RR de 10 minutos são não estacionárias, ou seja, os momentos (média e desvio padrão) são dependentes do tempo. Teoricamente as séries RR da base de dados do projeto ELSA Brasil são de 10 minutos, porém ao se analisar os registros, muitos possuíam 9, 8 ou 7 minutos. Assim, as séries foram divididas em fragmentos de 5 minutos, e investigou-se novamente a estacionariedade nos trecho de 5 minutos iniciais, intermediários e finais das séries de dados. Observou-se que os trechos intermediários das séries RR eram os que apresentavam o maior número de registros estacionários (mesmo esse número sendo extremamente pequeno, em torno de 2%).

De acordo com o Task Force (1996), normalmente a estacionariedade não é avaliada para séries RR de 5 minutos de duração (séries de curta duração). Assim, os cinco minutos intermediários (normalmente de $2,5$ min a $7,5$ min) dos registros, mesmo sendo não estacionários em sua maioria, foram selecionados e utilizados na geração dos índices de VFC dos grupos de indivíduos da base de dados.

Após a geração dos índices de VFC, analisou-se a discriminância estatística entre o grupo de indivíduos saudáveis e diabéticos, por meio do teste t (Boneau, 1960). De fato, houve a aplicação do teste t apenas para os índices que apresentaram distribuição normal. A comparação entre grupos para índices com distribuição não normal foi feita pelo *Mann-Whitney U-test* (Hollander e Wolfe, 1999; Gibbons e Chakraborti, 2010). A ideia era obter a confirmação da hipótese não nula (análise do p-valor), ou seja, que os dados dos dois grupos eram amostras independentes e aleatórias, de distribuições com médias/medianas diferentes ao nível de significância de 5% (Tabelas 4.1, 4.2 e 4.3).

Embora o grupo de indivíduos diabéticos e sadios tenham a mesma distribuição de homens e mulheres, mesma faixa etária e aproximadamente a mesma frequência cardíaca, os diabéticos apresentam variabilidade da frequência cardíaca reduzida em relação aos sadios. Isto é indicado nas Tabelas 4.1 e 4.2 pelos menores valores de média dos índices de VFC nos indivíduos diabéticos.

A Tabela 4.1 mostra que os índices de VFC tradicionais no domínio do tempo, RR_{mean}, SDNN e RMSSD, são reduzidos para pacientes diabéticos, de acordo com (Javorka et al., 2008a). Além disso, a Tabela 4.2 mostra que a energia das bandas VLF, LF e HF são significativamente reduzidas para o grupo de indivíduos diabéticos (Javorka et al., 2008a). Em contrapartida, a razão LF/HF (mesmo não discriminante entre os grupos) é menor para o grupo de indivíduos saudáveis, ao contrário do obtido em (Faust et al., 2011) para registros de 60 minutos.

Ainda analisando os índices de VFC no domínio da frequência, de acordo com Pagani et al. (1988) a energia das bandas LF e HF, em valores absolutos, são menores para o grupo de diabéticos. Além disso, vê-se que LF_{norm}, HF_{norm} e LF/HF não apresentam diferença estatística significativa entre os grupos de sadios e diabéticos (Pagani et al., 1988).

Sabe-se que SD1, medida de variabilidade instantânea da série (*short-term variability*), correlaciona-se melhor com HF do que com LF (Contreras et al., 2007). De acordo com Contreras et al. (2007) os índices no domínio da frequência, bem como SD1, são menores para os indivíduos diabéticos. SD2, medida da variabilidade em registros de longa duração (*long-term variability*), também é reduzida nos indivíduos diabéticos, de acordo com Faust et al. (2011). Assim, a diminuição nos índices no domínio do tempo juntamente com a redução em SD1 e HF significam queda na modulação vagal em indivíduos diabéticos.

Os índices DET e L_{max} são mais elevados para o grupo de indivíduos diabéticos, de acordo com (Javorka et al., 2008b), (Faust et al., 2011). Além disso, observa-se que os demais índices obtidos dos gráficos de recorrência são dependentes do valor de REC (Javorka et al., 2008b), tal que, quanto maior o valor de REC maior será o valor de DET, L_{mean}, L_{max} e ShanEn, conforme observado na Tabela 4.3.

Observa-se que, a maioria dos parâmetros não lineares apresentam valores mais elevados para o grupo de indivíduos diabéticos, a exemplo dos valores médios de ApEn, α_1, FD, REC, DET, L_{mean}, L_{max} e ShanEn. Em contraste a (Faust et al., 2011), o índice SampEn é maior para o grupo de indivíduos normais, indicando maior entropia. A diferença no valor médio de ApEn e SampEn pode ser justificada, pois esses índices dependem da escolha dos parâmetros *m* e *r*. Além disso, o cálculo do valor de entropia pelos índices é feito de forma diferenciada (Acharya et al., 2006).

O grupo controle (grupo 2) foi definido pelo relato de ausência de diagnostico prévio de diabetes e presença de GJ menor que 100 mg/dL enquanto que o grupo "diabético" (grupo 1) foi definido pelo relato da presença da doença na história médica pregressa e por GJ superior a 140 mg/dL. Esse trabalho consistiu no desenvolvimento de um modelo capaz de separar indivíduos diabéticos e não diabéticos com base na glicemia de jejum (GJ). Para tal foi utilizado um classificador SVM identificando diferenças nos índices de VFC dos índivíduos de cada grupo, com acurácia de $75,69\%$. É importante ressaltar que, de acordo com a glicemia pós-prandial > 200 mg/dL (TOTG):

- o grupo 1 (diabéticos) possui 276 indivíduos diabéticos e 84 indivíduos saudáveis, tal que HMPA04 e GJ possuem acurácia de $76,67\%$, comparada ao exame TOTG;
- o grupo 2 (controle) possui 2 indivíduos diabéticos e 358 indivíduos saudáveis, tal que HMPA04 e GJ possuem acurácia de $94,44\%$, comparada ao exame TOTG.

Isso mostra que a associação de GJ e HMPA04 não corresponde fielmente à identificação dada pelo exame TOTG, principalmente para o grupo 1. Pode-se esperar que o classificador identifique melhor indivíduos sadios (maior especificidade), visto que GJ e HMPA04 têm maior correspondência com o exame TOTG no grupo 2. Ou ainda, os índices obtidos a partir da variabilidade da frequência cardíaca de indivíduos do grupo 2 são, de fato, característicos quase exclusivamente de indivíduos saudáveis, o que facilita o aprendizado do classificador para este grupo.

Analisando particularmente os grupos 3 e 4, vemos que o classificador aprendeu bem a relação a GJ < 100 mg/dL com HMPA4 $= 0$ e GJ > 140 mg/dL com HMPA4 $= 1$, em função da determinação dos grupos 1 e 2. Vale ressaltar que o grupo de diabéticos (grupo 1)

possui diversos indivíduos que utilizam medicamentos para controle de diabetes e hipertensão. Tais medicamentos alteram a variabilidade da frequência cardíaca e, consequentemente, dificultam a etapa de classificação. Os indivíduos que utilizam tais medicamentos não foram eliminados das análises, de tal forma que o grupo 1 fosse o maior possível (amostra de diabéticos mais representativa).

Os grupos 3, 4 e 5 são complicados de se analisar. Os indivíduos do grupo 3 utilizam medicamentos para controle de glicemia (que implicam em alterações na VFC) na faixa caracterizada como normal até hiperglicemia. Já os grupos 4 e 5 representam indivíduos com valores intermediários de glicemia na faixa de 100 a 140 $\mathrm{mg/dL}$.

Aparentemente o classificador identifica melhor indivíduos sadios (devido à maior especificidade) tal que indivíduos com valores de GJ próximos a 100 tendem a ser classificados como normais. O desempenho do classificador é relativamente baixo, especialmente nos grupos 3 e 5. Isso sugere que o classificador, da forma como foi ajustado (baseado apenas nos grupos 1 e 2 utilizando GJ, de acordo com a metodologia), é inadequado para a perfeita análise dos indivíduos com valores de glicemia intermediária e da influência da utilização de medicamentos na identificação de indivíduos com diabetes e neuropatia diabética.

Capítulo 5

Conclusões e Trabalhos Futuros

Este trabalho fornece subsídios para a identificação automatizada de diabetes, por meio das análises da variabilidade da frequência cardíaca (VFC). Sua originalidade está na proposição de um sistema diagnóstico completo para identificação de indivíduos diabéticos utilizando a VFC. Nestes termos, o presente estudo não se limita apenas à análise de discriminância estatística entre grupos. Além disso, em função dos mecanismos não lineares na geração da VFC e complexidade do sistema cardiovascular, são analisados os índices não lineares na análise das séries RR obtidas da base de dados brasileira do projeto ELSA.

Diversos índices de VFC, obtidos de séries de intervalos RR são utilizados no processo de identificação de indivíduos diabéticos, a saber: cinco medidas no domínio do tempo (média dos intervalos RR, SDNN, RMSSD, pNN50 e o Δ *index*), seis índices no domínio da frequência (VLF, LF, HF, LF_{norm}, HF_{norm} e LF/HF) e quinze índices não lineares (ApEn, SampEn, SD1, SD2, SD1/SD2, s, SDRR, α_1, α_2, FD, REC, DET, L_{mean}, L_{max} e ShanEn).

Nos testes realizados com o sistema, a base com as séries de intervalos RR serve de argumento à entrada do sistema, passando, inicialmente, pela etapa de extração de características (geração de índices de VFC) e, em seguida, pelo classificador. Após o resultado da classificação das séries RR, a priori, desconhecidas pelo sistema (conjunto de teste), medidas de desempenho do sistema foram tomadas. Além disso, o modelo do classificador SVM foi avaliado sobre outros conjuntos de dados (grupos 3, 4 e 5). Antes dessa última etapa, observou-se a relação entre índices de VFC e GJ com base nas Figuras 4.2 e 4.3, concluindo que não é possível inferir uma relação explícita do valor de GJ com as componentes da PCA, na faixa de glicemia intermediária. No caso, observou-se uma diminuição do valor das componentes principais, obtidas da PCA, em função do valor de glicemia.

Os resultados obtidos na Tabela 4.4 mostram que os índices utilizados são clinicamente aplicáveis no auxílio do diagnóstico do diabetes, uma vez que foram capazes de detectar as

diferenças entre os grupos (1 e 2). Assim, o sistema desenvolvido mostra que a identificação de pacientes com diabetes, por meio da variabilidade de frequência cardíaca, é viável. Tal sistema representa uma ferramenta complementar não invasiva, de fácil utilização e de relevância em análises científicas (pesquisa), aos exames convencionais de diagnóstico de DM, baseados em testes laboratoriais como glicemia de jejum (GJ), teste oral de tolerância a glicose (TOTG) e hemoglobina glicada (A1C). Além disso, vemos que o modelo de classificador desenvolvido pode ser utilizado, com determinadas limitações, à analise de outros grupos de indivíduos, como os grupos 3, 4 e 5, visto que apresentou acurácia sobre os grupos de $86,15\%$, $91,92\%$ e $88,89\%$, respectivamente.

Em trabalhos futuros pretende-se:

- agregar outros índices não lineares na análise das séries RR, como a dimensão de correlação (do inglês, *correlation dimension* ou CD);
- preparar o sistema para aquisição e processamento *online* das séries RR. Isto possibilitará que o sistema desenvolvido possa ser utilizado efetivamente no diagnóstico automatizado de diabetes no meio clínico;
- relacionar o tempo de presença de diabetes em indivíduos com a variabilidade da frequência cardíaca, avaliando alterações sobre os índices de VFC;
- analisar se em indivíduos com HMPA04 positivo o exame A1C ajuda a identificar os pacientes com diabetes descompensada.

Bibliography

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C., e Suri, J. (2006). Heart rate variability: a review. *Medical and Biological Engineering and Computing*, 44:1031–1051.

Acharya, U. R., Kannathal, N., e Krishnan, S. M. (2004). Comprehensive analysis of cardiac health using heart rate signals. *Physiological Measurement*, 25:1139.

Akaike, H. (1969). Power spectrum estimation through autoregression model fitting. *Annals of the Institute of Statistical Mathematics*, 21(1):407–419.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.

Akaike, H. (1977). On entropy maximization principle. Em *Applications of Statistics*, páginas 27–41, P. R. Krishnaiah, Ed. Amsterdam: North-Holland.

Akaike, H. (1978). A bayesian analysis of the minimum aic procedure. *Annals of the Institute of Statistical Mathematics*, 30(1):9–14.

Akselrod, S., Pinhas, I., Davrath, L. R., Shinar, Z., e Toledo, E. (2006). *Heart Rate Variability (HRV)*. John Wiley & Sons, Inc.

Almeida, D. L. F., Lima, R. R., e Carvalho, J. L. A. (2012). Análise das flutuações destendenciadas da variabilidade da frequencia cardíaca: Melhorias e aplicações. *CBEB 2012*, páginas 2086–2090.

Anderson, S. (1992). *Biophysical Measurement Series: Advanced Electrocardiography*. Spacelabs, Redmond, WA.

Aquino, E. M. L., Barreto, S. M., Bensenor, I. M., Carvalho, M. S., Chor, D., Duncan, B. B., Lotufo, P. A., Mill, J. G., del C. Molina, M., Mota, E. L. A., Passos, V. M. A., Schmidt, M. I., e Szklo, M. (2012). Brazilian longitudinal study of adult health (elsa-brasil): Objectives and design. *Am J Epidemiol.*, 175:315–324.

Association, A. D. (1997). Guide to diagnosis and classification of diabetes meliitus and other categories of glucose intolerance.

Association, A. D. (2003). Report of the expert committee on the diagnosis and classification of diabetes mellitus.

Association, A. D. (2011). Position statement of the american diabetes association: Diagnosis and classification of diabetes mellitus.

Baekkeskov, S., Aanstoot, H. J., Christgau, S., Reetz, A., Solimena, M., Cascalho, M., Folli, F., Richter-Olesen, H., e Camilli, P. D. (1990). Identification of the 64k autoantigen in insulin-dependent diabetes as the gaba-synthesizing enzyme glutamic acid decarboxylase. *Nature*, 347(6289):151–156.

Barceló, A., Aedo, C., Rajpathak, S., e Robles, S. (2003). The cost of diabetes in latin america and the caribean. *Bull World Health Organ*, 81(1):19–27.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society*, 160(901):268–282. Serie A.

Bennet, P. H. (1994). *Definition, diagnosis and classification of diabetes mellitus and impaired glucose tolerance*, páginas 193–215. EEUU: Editora Lea e Febiger Philadelphia, 13 edição. C. R. Kahn and G. C. Weir.

Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., e van der Molen, M. W. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34(6):623–648.

Bertsekas, D. P. (1995). *Dynamic Programming and Optimal Control*, volume 1 e 2. Athenas Scientific, Belmont, MA.

Boardman, A., Schlindwein, F. S., Rocha, A. P., e Leite, A. (2002). A study on the optimum order of autoregressive models for heart rate variability. *Physiol Meas*, 23:325–336.

Boneau, C. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57:49–64.

Bracewell, R. (1999). *The Fourier Transform and Its Applications*. McGraw-Hill, 3 edição.

Brennan, M., Palaniswami, M., e Kamen, P. (2001). Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. Biomed. Eng.*, 48:1342–1347.

Burton, A. C. (1977). *Fisiologia e Biofísica da Circulação*. Guanabara Koogan, Rio de Janeiro, RJ, segunda edição.

BVS (2002). *Ministério da Saúde. Plano de Reorganização da Atenção à Hipertensão arterial e ao Diabetes mellitus. Manual de Hipertensão arterial e Diabetes mellitus*. Brasília.

Carvalho, J. L. A., Rocha, A. F., Santos, I., Itiki, C., Jr, L. F. J., e Nascimento, F. A. O. (2003). Study on the optimal order for the auto-regressive time-frequency analysis of heart rate variability. *IEEE EMBS*, 3:2621–2624.

Chang, C.-C. e Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27.

Charles, M. A., Fontbonne, A., Thibult, N., Warnet, J. M., Rosselin, G. E., e Eschwege, E. (1991). Risk factors for niddm in white population. paris prospective study. *Diabetes*, 40(7):796–799.

Cohen, M. A. e Taylor, J. A. (2002). Short-term cardiovascular oscillations in man: measuring and modelling the physiologies. *J. Physiol*, 542(3):669–683.

Contreras, P., Canetti, R., e Migliaro, E. R. (2007). Correlations between frequency-domain hrv indices and lagged poincaré plot width in healthy and diabetic subjects. *Physiol. Meas.*, 28(1):85–94.

Coustan, D. R. (1995). *Gestational Diabetes*, chapter 35, páginas 703–717. National Institutes of Diabetes and Digestive and Kidney Diseases, NIH Publication No. 95-1468. Bethesda: NIDDK, 2nd edição.

Dantas, E. M., Sant'Anna, M. L., Andreão, R. V., Gonçalves, C. P., Morra, E. A., Baldo, M. P., Rodrigues, S. L., e Mill, J. G. (2012). Spectral analysis of heart rate variability with the autoregressive method: what model order to choose? *Comput Biol Med*, 42(2):164–170.

Dickey, D. A. e Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.

Duda, R. O. e Hart, P. (1973). *Pattern Classification and Scene Analysis*. Wiley, New York.

Durbin, J. (1960). The fitting of time-series models. *Review of the International Statistical Institute*, 28(3):233–244. Revue de l'Institut International de Statistique.

Eckberg, D. L. (2003). The human respiratory gate. *J. Physiol.*, 548(2):339–352.

Eckmann, J. P., Kamphorst, S., e D.Ruelle (1987). Recurrence plots of dynamical systems. *Europhys. Lett.*, 56(5):973–977.

Eckmann, J. P. e Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57:617–654.

Engelgau, M. M., Thompson, T. J., Herman, W. H., Boyle, J. P., Aubert, R. E., Kenny, S. J., Badran, A., Sous, E. S., e Ali, M. A. (1997). Comparison of fasting and 2-hour glucose and HbA1c levels for diagnosing diabetes. diagnostic criteria and performance revisited. *Diabetes Care*, 20(5):785–791.

Ewing, D. J. e Clarke, B. F. (1982). Diagnosis and management of diabetic autonomic neuropathy. *British Medical Journal*, 285:916–918.

Ewing, D. J., Martyn, C. N., Young, R. J., e Clarke, B. F. (1985). The value of cardiovascular autonomic function tests: 10 years experience in diabetes. *Diabetes Care*, 8(5):491–498.

Faust, O., Acharya, U., Ng, E., Ng, K.-H., e Suri, J. (2010). Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. *Journal of Medical Systems*, páginas 1–13.

Faust, O., Acharya, U. R., Molinari, F., Chattopadhyay, S., e Tamura, T. (2011). Linear and non-linear analysis of cardiac health in diabetic subjects. *Biomedical Signal Processing and Control*.

Fletcher, R. (1987). *Pratical Methods of Optimization*. Wiley, New York, 2 edição.

Fraser, A. M. e Swinney, H. L. (1986). Independent coordinates for strange attractor from mutual information. *Physics Review A*, 33(2):1134–1140.

Freeman, J. V., Dewey, F. E., Hadley, D. M., Myers, J., e Froelicher, V. F. (2006). Autonomic nervous system interaction with the cardiovascular system during exercise. *Prog. Cardiovasc. Dis.*, 48(5):342–362.

Freeman, R., Saul, J. P., Roberts, M. S., Berger, R. D., Broadbridge, C., e Cohen, R. J. (1991). Spectral analysis of heart rate in diabetic autonomic neuropathy. a comparison with standard tests of autonomic function. *Arch. Neurol.*, 48(2):185–190.

Friedman, N., Geiger, D., e Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29:131–163.

Frigo, M. e Johnson, S. G. (1997). The fastest Fourier transform in the west. Relatório Técnico MIT-LCS-TR-728, Massachusetts Institute of Technology.

Frigo, M. e Johnson, S. G. (1998). FFTW: An adaptive software architecture for the FFT. Em *Proc. 1998 IEEE Intl. Conf. Acoustics Speech and Signal Processing*, volume 3, páginas 1381–1384. IEEE.

Frigo, M. e Johnson, S. G. (2005). The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231. Special issue on "Program Generation, Optimization, and Platform Adaptation".

Frigo, M., Leiserson, C. E., Prokop, H., e Ramachandran, S. (1999). Cache-oblivious algorithms. Em *Proc. 40th Ann. Symp. on Foundations of Comp. Sci. (FOCS)*, páginas 285–297. IEEE Comput. Soc.

Fuller, J. H., Rose, M. J. S. G., Jarrett, R. J., e Keen, H. (1980). Coronary-heart-disease risk and impaired glucose tolerance. the whitehall study. *Lancet*, 1(8183):1373–1376.

Gibbons, J. D. e Chakraborti, S. (2010). *Nonparametric Statistical Inference*. CRC Press, 5 edição.

Grassberger, P. e Procaccia, I. (1983). Estimation of the kolmogorov entropy from a chaotic signal. *Phys Rev A*, 28:2591–2593.

Group, D. S. (1999). Glucose tolerance and mortality: comparison of who and american diabetes association diagnostic criteria. european diabetes epidemiology group. diabetes epidemiology: Collaborative analysis of diagnostic criteria in europe.

Grundy, S. M., Benjamin, I. J., Burke, G. L., Chait, A., Eckel, R. H., Howard, B. V., Mitch, W., Smith, J., Sidney, C., e Sowers, J. R. (1999). Diabetes and cardiovascular disease: a statement for healthcare professionals from the american heart association. *Circulation*, 100:1134–1146.

Guthrie, R. A. (1998). *The Diabetes Sourcebook*. Lowell House, Los Angeles.

Guyton, A. C. e Hall, J. E. (2002). *Tratado de Fisiologia Médica*. Editora Guanabara Koogan SA, 10 edição.

Hall, J. E. (2010). *Guyton and Hall Textbook of Medical Physiology*. Saunders Elsevier, 12 edição.

Hanna, F. W. e Peters, J. R. (2002). Screening for gestational diabetes; past, present and future. *Diabetic Medicine*, 19(5):351–358.

Hansen, J. (2011). *Atlas de Fisiologia Humana de Netter*. Elsevier Editora Ltda.

Harman-Boehm, I., Sosna, T., Lund-Andersen, H., e Porta, M. (2007). The eyes in diabetes and diabetes through the eyes. *Diabetes Research and Clinical Practice*, 78:S51–S58.

Haykin, S. (1994). *Neural Networks: A comprehensive foundation*. IEEE Press.

Hegger, R. e Kantz, H. (1998). Practical implementation of nonlinear time series methods: The tisean package. *Chaos*, 33:235–243.

Higuchi, T. (1988). Aproach to an irregular time series on the basis of the fractal theory. *Physica D*, 31:277–283.

Hogg, R. V. e Ledolter, J. (1987). *Engineering Statistics*. MacMillan.

Hollander, M. e Wolfe, D. A. (1999). *Nonparametric Statistical Methods*. John Wiley & Sons, Inc., 2 edição.

Holyst, J. A., Zebrowska, M., e Urbanowicz, K. (2000). Observations of deterministic chaos in financial time series by recurrence plots, can one control chaotic economy? *The European Physical Journal B*, 20:51–535.

Howorka, K., Pumprla, J., Haber, P., Koller-Strametz, J., Mondrzyk, J., e Schabmann, A. (1997). Effects of physical training on heart rate variability in diabetic patients with various degrees of cardiovascular autonomic neuropathy. *Cardiovascular research*, 34:206–214.

Ince, N. F., Arica, S., e Tewfik, A. (2006). Classification of single trial motor imagery eeg recordings with subject adapted non-dyadic arbitrary time-frequency tailings. *Journal of Neural Engineering*, 3:235–244.

Javorka, M., Trunkvalterova, Z., Tonhajzerova, I., e Baumert, J. J. K. J. M. (2008a). Short-term heart rate complexity is reduced in patients with type 1 diabetes mellitus. *Clinical Neurophysiology*, 119:1071–1081.

Javorka, M., Trunkvalterova, Z., Tonhajzerova, I., Lazarova, Z., Javorkova, J., e Javorka, K. (2008b). Recurrences in heart rate dynamics are changed in patients with diabetes mellitus. *Clin. Physiol. Funct. Imaging*, 28(5):326–331.

Johnsen, S. J. e Andersen, N. (1978). On power estimation in maximum entropy spectral. *Geophysics*, 43(4):681–690.

Johnson, S. G. e Frigo, M. (2007). A modified split-radix FFT with fewer arithmetic operations. *IEEE Trans. Signal Processing*, 55(1):111–119.

Johnson, S. G. e Frigo, M. (2008). Implementing FFTs in practice. Em Burrus, C. S., editor, *Fast Fourier Transforms*, chapter 11. Connexions, Rice University, Houston TX.

Katz, M. J. (1988). Fractals and analysis of waveforms. *Comput Biol Med*, 18(3):145–156.

Kay, S. M. (1988). *Modern spectral estimation: Theory and application*. Prentice Hall, Englewood Cliffs, NJ. Chapters 7 and 13.

Kennel, M. B., Brown, R., e Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, 45.

Khan, M. G. (2007). *Rapid ECG Interpretation (Contemporary Cardiology)*. Humana Press, 3rd edição.

Khandoker, A. H., Jelinek, H. F., e Palaniswami, M. (2009). Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis. *BioMedical Engineering OnLine*, 8(3).

Kuhl, C. (1991). Insulin secretion and insulin resistance in pregnancy and gdm. implications for diagnosis and management. *Diabetes*, 40(2):18–24.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., e Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

Laederach-Hofmann, K., Mussgay, L., e Rúddel, H. (2000). Autonomic cardiovascular regulation in obesity. *The Journal of endocrinology*, 164(1):59–66.

Leite, F. S., Carvalho, J. L. A., e da Rocha, A. F. (2010). Matlab software for detrended fluctuation analysis of heart rate variability. Em *Biosignals*, páginas 225–229. INSTICC Press.

Levene, H. (1960). *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press, 1 edição. Ingram Olkin.

Levinson, N. (1947). The wiener rms (root-means-square) error criterion in filter design and prediction. *J Math Phys*, 25(4):261–278.

Li, J. K. J. (2004). *Dynamics of the Cardiovascular System*. World Scientific Publishing Co Inc, Singapore. Series on Bioengineering & Biomedical Engineering - Vol. 1.

Liao, D., Sloan, R. P., Cascio, W. E., Folsom, A., Liese, A. D., Evans, G. W., Cai, J., e Sharrett, A. R. (1998). Multiple metabolic syndrome is associated with lower heart rate variability. *Diabetes Care*, 21(12):2116–2122.

Lombardi, F. (2000). Chaos theory, heart rate variability, and arrhythmic mortality. *Circulation*, 101:8–10.

Magagnin, V., Bassani, T., Bari, V., Turiel, M., Maestri, R., Pinna, G. D., e Porta, A. (2011). Non-stationarities significantly distort short-term spectral, symbolic and entropy heart rate variability indices. *Physiol Meas*, 32(11):1775–1786.

Malerbi, D. A. e Franco, L. J. (1992). Multicenter study of the prevalence of diabetes mellitus and impaired glucose tolerance in the urban brazilian population aged 30-69 yr. the brazilian cooperative. *Diabetes Care*, 15:1509–1516.

Malliani, A., Pagani, M., e Lombardi, F. (1994). Physiology and clinical implications of variability of cardiovascular parameters with focus on heart rate and blood pressure. *Am. J. Cardiol.*, 73(10):3–9.

Mandelbrot, B. B. e Freeman, W. H. (1983). *The Fractal Geometry of Nature*. W. H. Freeman, San Francisco.

March, T. K., Chapman, S. C., e Dendy, R. O. (2004). Recurrence plot statistics and the effects of embedding. *Physica D*, 200:171–184.

Marple, S. L. (1987). *Digital Spectral Analysis: With Applications*. Prentice Hall, Englewood Cliffs, NJ.

Marsaglia, G., Tsang, W., e Wang, J. (2003). Evaluating kolmogorov's distribution. *Journal of Statistical Software*, 8(18).

Marwan, N. (2003). *Encounters with Neighbours*. Tese de Doutorado, Universitat Potsdam.

Marwan, N. e Kurths, J. (2004). Line structure in recurrence plots. *Physics Letters A*, 336:349–357.

Marwan, N., Romano, M. C., Thiel, M., e Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5-6):237–329.

Massey, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78.

Meddins, B. e Meddins, R. (2000). *Introduction to digital signal processing*. Essential Electronics Series. Newnes, Oxford, Auckland, Boston, Johannesburg, Melbourne and New Delhi, 1 edição.

Mendez, M. O., Bianchi, A. M., Villiantieri, O. P., Cerutti, S., e Penzel, T. (2006). Time- variant spectral analysis of the heart rate variability during sleep in healthy and obstructive sleep apnoea subjects. *Computers in Cardiology*, 33:741–744.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, 1 edição.

Norman, A. E. . (1991). *Rapid ECG Interpretation*. McGraw-Hill Professional, 1 edição.

Novak, V., Novak, P., Champlain, J., e and, R. N. (1994). Altered cardiorespiratory transfer in hypertension. *Hipertension*, 23:104–113.

on the Diagnosis, E. C. e of Diabetes Mellitus, C. (2003). Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*, 26(1):5–20.

Orfanidis, S. J. (1985). *Optimum signal processing - An introduction*. Macmillan Publishing Company, New York, NY.

Ott, E. (1993). *Chaos in dynamical systems*. Cambridge University Press, 2 edição.

Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., Sandrone, G., Malfatto, G., Dell'Orto, S., Piccaluga, E., Turiel, M., Baselli, G., Cerutti, S., e Malliani, A. (1986). Power spectral analysis of heart rate and arterial pressure as a marker of sympathovagal interaction in man and concious dog. *Circulation Research*, 59:178–193.

Pagani, M., Malfatto, G., Pierini, S., Casati, R., Masu, A. M., Poli, M., Guzzetti, S., Lombardi, F., Cerutti, S., e Malliani, A. (1988). Spectral analysis of heart rate variability in the assessment of autonomic diabetic neuropathy. *J. Auton. Nerv. Syst.*, 23(2):143–153.

Palmer, J. P., Asplin, C. M., Clemons, P., Lyen, K., Tatpati, O., Raghu, P. K., e Paquette, T. L. (1983). Insulin antibodies in insulin-dependent diabetics before insulin treatment. *Science*, 222(4630):1337–1339.

Parzen, E. (1974). Some recent advances in time series modeling. *IEEE Transactions on Automatic Control*, 19:723–730.

Patwardhan, A. (2006). *Respiratory Sinus Arrhythmia*. John Wiley & Sons, Inc.

Peng, C. K., Havlin, S., Stanley, E. H., e Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):82–87.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA*, 88:2297–2301.

Polychronaki, G. E., Ktonas, P. Y., Gatzonis, S., Siatouni, A., Asvestas, P. A., Tsekou, H., Sakas, D., e Nikita, K. S. (2010). Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. *J Neural Eng*, 7(4).

Porta, A., D'Addio, G., Guzzetti, S., Lucini, D., e Pagani, M. (2004). Testing the presence of non stationarities in short heart rate variability series. Em *Computers in Cardiology*, páginas 645–648.

Rabin, D. U., Pleasic, S. M., Shapiro, J. A., Yoo-Warren, H., Oles, J., Hicks, J. M., Goldstein, D. E., e Rae, P. M. (1994). Islet cell antigen 512 is a diabetes-specific islet autoantigen related to protein tyrosine phosphatases. *J Immunol*, 152(6):3183–3188.

Ribeiro, J. P. e Filho, R. S. M. (2005). Heart rate variability as a tool for the investigation of the autonomic nervous system. *Brazilian Journal of Hypertension*, 12(1):14–20.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2):416–431.

Rissanen, J. (1984). Universal coding, information prediction and estimation. *IEEE Transactions on Information Theory*, 30:629–636.

Rissanen, P., Franssila-Kallunki, A., e Rissanen, A. (2001). Cardiac parasympathetic activity is increased by weight loss in healthy obese women. *Obesity research*, 9(10):637–643.

Robertson, D., Biaggioni, I., Burnstock, G., Low, P. A., e Paton, J. F. R. (2011). *Primer on the Autonomic Nervous System*. Elsevier Academic Press, 3 edição.

Schlögl, A., Lee, F., Bischof, H., e Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition. *Journal of Neural Engineering*, 2(4):14–22.

Schroeder, E. B., Chambles, L. E., Prineas, R. J., Evans, G. W., Rosamond, G. W., e Heiss, G. (2005). Diabetes, glucose, insulin, and heart rate variability: the atherosclerosis risk in communities (aric) study. *Diabetes Care*, 28(3):668–674.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shannon, C. E. (1948). A mathematical theory of communication. *The bell system technical ournal*, 27:379–423.

Singh, J. P., Larson, M. G., Donnell, C. J. O., Wilson, P. F., Tsuji, H., Lloyd-Jones, D. M., e Levy, D. (2000). Association of hyperglycemia with reduced heart rate variability (the framingham heart study). *The American Journal of Cardiology*, 86:309–312.

Singh, J. P., Larson, M. G., Tsuji, H., Evans, J., O'donnell, C. J., e Levy, D. (1998). Reduced heart rate variability and new-onset hypertension: insights into pathogenesis of hypertension: the framingham heart study. *Hypertension*, 32(2):293–297.

Souza, E. G. (2008). *Caracterização de Sistemas Dinâmicos através de Gráficos de Recorrência*. Tese de Doutorado, Universidade Federal do Paraná.

Tarvainen, M., Georgiadis, S., e Karjalainen, P. (2005). Time-varying analysis of heart rate variability with kalman smoother algorithm. *Conf Proc IEEE Eng Med Biol Soc*, 3:2718–2721.

Task Force (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation*, 93(5):1043–1065.

Thiel, M., Romano, M. C., Read, P. L., e Kurths, J. (2004). Estimation of dynamical invariants without embedding by recurrence plots. *Chaos*, 14(2):234–243.

Todd, J. A., Bell, J. I., e McDevitt, H. O. (1987). Hla-dq beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature*, 329(6140):599–604.

Tulppo, M. P., Mäkikallio, T. H., Seppänen, T., e Laukkanen, R. T. (1988). Vagal modulation of heart rate during exercise: effects of age and physical fitness. *Am J Physiol*, 274:H424–9.

Ursino, M. e Magosso, E. (2003). Short-term autonomic control of cardiovascular function: a mini-review with the help of mathematical models. *J. Integr. Neurosci.*, 2(2):219–247.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.

Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley, New York.

Walker, G. (1931). On periodicity in series of related terms. *Proceedings of the Royal Society of London*, 131(818):518–532. Series A, Containing Papers of a Mathematical and Physical Character.

Webber, C. e Zbilut, J. (1994). Dynamical assesment of physiological systems and states using recurence plots strategies. *J. Appl. Physiol.*, 76:965–973.

WHO (2011). World health organization. fact sheet n°312. Disponível em: `http://www.who.int/mediacentre/factsheets/fs312/en/index.html`. Acessado em 10/06/2011.

Wild, S., Roglic, G., Green, A., Sicree, R., e King, H. (2004). Global prevalence of diabetes. estimates for the year 2000 and projections for 2030. *Diabetes Care*, 27(5):1047–1053.

Xiao, X., Mukkamala, R., Sheynberg, N., Grenon, S. M., Ehrman, M. D., Mullen, T. J., Ramsdell, C. D., Williams, G. H., e Cohen, R. J. (2004). Effects of simulated microgravity on closed-loop cardiovascular regulation and orthostatic intolerance: analysis by means of system identification. *J. Appl. Physiol.*, 96(2):489–497.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London*, 226:267–298. Series A, Containing Papers of a Mathematical or Physical Character.

Apêndice A

Modelo AR e Processos Estocásticos

A.1 Modelo AR

O pressuposto básico é que o processo a ser estudado é estocástico e estacionário. O modelo AR prevê os valores atuais de uma série temporal a partir de valores anteriores da mesma. A dependência futura dos valores passados pode ser demonstrada através da função de autocorrelação. A autocorrelação é a média de somas do produto do conjunto de dados $x[n]$ com ele mesmo deslocado (*lag*). A função de autocorrelação é descrita pela equação:

$$r_{xx}[k] = \frac{1}{N} \times \sum_{n=1}^{N-k} x[n] \cdot x[n+k]. \tag{A.1}$$

Onde, $r_{xx}[k]$ é o valor da função de autocorrelação de x para um *delay* de k amostras, e N é o número de elementos do conjunto de dados.

Algumas considerações sobre a função de autocorrelação:

- para pequenos deslocamentos o sinal será similar o que implica em alto valo da função de autocorrelação;
- à medida que o *lag* aumenta tem-se uma diminuição do valor da função de autocorrelação;
- a autocorrelação é útil na separação de sinais periódicos de ruído aleatório.

Basicamente, o modelo AR pode ser considerado como um conjunto de funções de autocorrelação. A modelagem AR de uma série temporal é baseada na suposição de que os

elementos mais recentes dos dados contêm mais informações que os demais elementos, e que cada valor da série pode ser previsto como uma soma ponderada dos valores anteriores da mesma série, mais o erro. O modelo AR é definido por:

$$x[n] = \sum_{i=1}^{M} a_i x[n-i] + \varepsilon[n], \tag{A.2}$$

onde, $x[n]$ é o valor atual da série temporal, $a_1, a_2, a_3, \ldots, a_M$ são os coeficientes de predição (vistos como pesos da Equação A.2), M é a ordem do modelo indicando o número de valores passados utilizados para prever o valor atual, e $\varepsilon[n]$ representa o erro de predição de um passo, por exemplo, a diferença entre o valor predito e o valor atual em um dado ponto do conjunto de dados.

O modelo AR determina um filtro de análise, através do qual a série temporal é filtrada. Isso produz a sequência de erro de predição. Na identificação do modelo, o filtro AR usa os valores de entrada atuais e passados para obter o valor atual de saída (Vide Figura A.1).

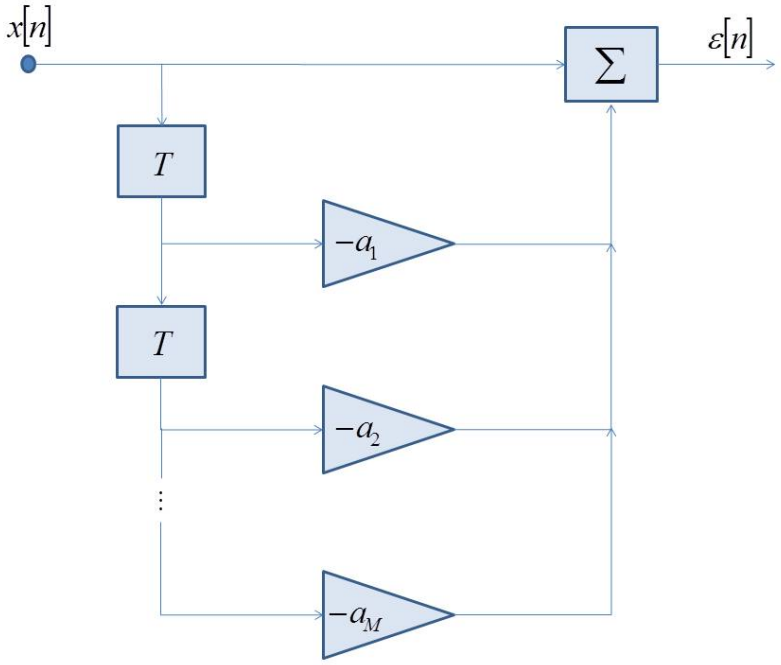


Figura A.1: Filtro de análise AR. $x[n]$ e $\varepsilon[n]$ representam as sequencias de entrada e saída, respectivamente. T é um *delay* de um período de amostragem (delay unitário), $a_1, a_2, \ldots, a_M$ são os coeficientes de predição, e o bloco $\sum$ realiza a soma dos valores $a_i x[n-iT]$, $i = 1, \ldots, M$.

Reescrevendo a equação anterior, como:

$$\varepsilon[n] = x[n] - \sum_{i=1}^{M} a_i x[n-i], \tag{A.3}$$

obtemos um filtro com resposta ao impulso $[1, -a_1, -a_2, \ldots, -a_M]$, o qual produz a sequencia de erro de predição $\varepsilon[n]$. Os coeficientes de predição normalmente são estimados usando a técnica de minimização pelos mínimos quadrados, que produz o menor erro $\varepsilon[n]$. Expandindo a Equação A.2, temos:

$$x[n] = a_1 x[n-1] + a_2 x[n-2] + \ldots + a_n x[n-M] + \varepsilon[n]. \tag{A.4}$$

Se utilizarmos a Equação A.4 para escrever expressões para várias estimativas de $x[n]$, obtemos um conjunto de equações lineares:

$$\mathbf{x} = \begin{bmatrix} x[M] & X[M-1] & \cdots & x[1] \\ x[M+1] & X[M] & \cdots & x[2] \\ \vdots & \vdots & \cdots & \vdots \\ x[N-1] & X[N-2] & \cdots & x[N-M] \end{bmatrix} \mathbf{a} + \varepsilon \Rightarrow \mathbf{x} = \mathbf{X}\mathbf{a} + \varepsilon \tag{A.5}$$

onde, $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix}$ e $\varepsilon = \begin{bmatrix} \varepsilon[M+1] \\ \vdots \\ \varepsilon[N] \end{bmatrix}$.

Portanto, precisamos de M equações para descobrir os M coeficientes desconhecidos $a_i,\ i = 1, 2, \ldots, M$. A solução pelos mínimos quadrados é facilmente obtida pelo cálculo matricial. Em outras palavras, $\mathbf{X}$ é uma matriz quadrada $M \times M$, e $\mathbf{a}$ e ε são matrizes coluna $M \times 1$.

Os coeficientes de predição ótimos (a_{opt}) podem ser obtidos pela aplicação do princípio da ortogonalidade na técnica de minimização pelos mínimos quadrados. Isso significa que os coeficientes de predição são selecionados de tal forma que o vetor coluna ε seja ortogonal a cada vetor coluna na matriz $\mathbf{X}(\mathbf{x_i},\ i = 1, 2, \ldots, M)$. Como na análise convencional de regressão, isto minimiza o erro médio quadrático. Portanto, o vetor ε é independente dos dados $\mathbf{X}$, ou seja, o vetor contém a parte da série temporal que não pode ser gerada pelos M valores anteriores da série de dados.

Em vez de resolver estas M equações $\mathbf{x_i^T}\varepsilon$ separadamente, utiliza-se das vantagens do cálculo matricial para resolução simultânea das equações. Isto é feito pela transposição da matriz $\mathbf{X}(\mathbf{X^T})$, assim:

$$\mathbf{X^T}\varepsilon = \mathbf{X^T}(\mathbf{x} - \mathbf{X}\mathbf{a}_{\text{opt}}) = \mathbf{0} \Leftrightarrow \mathbf{X^T}\mathbf{X}\mathbf{a}_{\text{opt}} = \mathbf{X^T}\mathbf{x} \Rightarrow \mathbf{a}\mathbf{a}_{\text{opt}} = (\mathbf{X^T}\mathbf{X})^{-1}\mathbf{X^T}\mathbf{x}. \tag{A.6}$$

Essa solução direta é denominada de método da covariância (Orfanidis, 1985).

Quando estudamos as novas matrizes formadas $\mathbf{X}\mathbf{a}^\mathbf{T}\mathbf{X}$ e $\mathbf{X^T}\mathbf{x}$, verificamos que elas consistem de somas semelhantes a função de autocorrelação com diferentes *lags*. $\mathbf{X^T}\mathbf{X}$ pode ser

vista no formato matricial como:

$$\frac{\mathbf{X^T X}}{N} \approx \mathbf{R} = \begin{bmatrix} r(0) & r(-1) & \cdots & r(1-M) \\ r(1) & r(0) & \cdots & \vdots \\ \vdots & \vdots & \cdots & r(-1) \\ r(M-1) & \cdots & r(1) & r(0) \end{bmatrix} \tag{A.7}$$

. Ou seja, $\mathbf{X^T X}$ é próxima a matriz de autocorrelação de $\mathbf{x}$ ($\mathbf{R}$). No caso de um sinal real de única variável, os valores da matriz são simétricos em relação a diagonal principal. Na prática, uma sequência de dados medidos de tamanho N está disponível, e a autocorrelação do processo original é substituída pela autocorrelação amostral correspondente, calculada a partir da dada sequência de dados.

A outra matriz gerada $\mathbf{X^T x}$ é semelhante ao vetor de autocorrelação ($\mathbf{r}$) e deve ser calculado a partir da sequência de dados como:

$$\frac{\mathbf{X^T x}}{N} \approx \mathbf{r} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(M) \end{bmatrix} \tag{A.8}$$

. Combinando as equações anteriores, temos:

$$\mathbf{a}_{\text{opt}} = (\mathbf{X^T X})^{-1}\mathbf{X^T x} \therefore \mathbf{a}_{\text{opt}} = \mathbf{R^{-1}r}. \tag{A.9}$$

As equações de autocorrelação são de *Yule-Walker* (Yule, 1927), (Walker, 1931), tal que solução é de Wiener. Por causa do custo da R^{-1}, explorando as propriedades da matriz de autocorrelação, o método recursivo, geralmente usado para resolver a equação de *Yule-Walker*, é chamado de algoritmo de *Levinson-Durbin* (Levinson, 1947), (Durbin, 1960). A vantagem do método recursivo é que produz soluções para todas as ordens de modelos inferiores ao M escolhido, tornando assim a seleção de M mais fácil.

As Equações A.9 podem ser usadas para estimar os parâmetros do modelo. Existem também outros métodos para extrair estimativas razoáveis dos parâmetros do modelo usando uma sequência de dados, mas todos os outros métodos são derivados a partir destes dois métodos básicos.

O modelo AR é identificado pela determinação de M, ajustando o modelo aos dados, e verificando a variância do erro de predição. Existem vários métodos para cada uma dessas etapas. Na prática, vários modelos são ajustados à série temporal, enquanto varia-se o M e o melhor modelo é escolhido. O critério mais comum para a seleção da ordem M do modelo

AR é o critério de *Informação Akaike* (AIC) (Akaike, 1974), que pode ser apresentado como: $\mathrm{AIC}(M) = N \times \ln(\sigma_p^2) + 2M$, onde σ_p^2 é a variância do erro de predição associada a M. O valor selecionado de M minimiza o valor do critério. Existem outros critérios como: FPE, MDL, CAT e BIC.

Depois que o modelo foi identificado, sua validade deve ser verificada. Os testes são geralmente baseados nas propriedades estatística dos erros de predição. As propriedades verificadas são: a presença de todas as componentes de frequência e se a distribuição é normal. Se a ordem M for correta, a parcela de erro deve ser um ruído branco (gaussiano) com média zero.

A.2 Ordem do Modelo AR

A.2.1 Introdução

Estudos recentes são divergentes quanto o critério de determinação da ordem do modelo AR (do inglês, *AR model order* ou ARMO) preferindo utilizar uma ordem fixa. Assim, (Schlögl et al., 2005) utilizou uma ordem fixa de 3, (Ince et al., 2006) utilizou ordem fixa de 6, e (Tarvainen et al., 2005) utilizou ordem fixa de 16.

Em contrapartida, (Boardman et al., 2002) comparou quatro critérios (FPE, AIC, CAT e RIS - descritos nas seções seguintes), comumente utilizados, para estimar a ordem "ótima" de modelo a partir de séries RR de curta duração, mostrando que todos os critérios subestimam a ordem "ótima" do modelo AR, quando analisou uma série sintética obtida de um modelo AR com ordem conhecida. Em (Boardman et al., 2002) concluiu-se que a escolha de uma ordem (ARMO) fixa é recomendada, sugerindo, do ponto de vista qualitativo, que a ordem esteja na faixa de 16 a 22. O (Task Force, 1996) recomenda a utilização de ordens de modelo entre 8 a 20. Já (Mendez et al., 2006) sugere que a ordem deve estar entre 8 e 16.

Finalmente, com o intuito de analisar de forma eficaz a influência da escolha da ordem do modelo AR, (Dantas et al., 2012) mostrou que a estimação da ordem ótima de modelo AR não é um fator determinante em análises espectrais em registros curtos (*short term*) de VFC para ordens entre 13 e 25. No caso, a seleção da ordem do modelo AR deve alterar os valores das bandas de frequência, mas não afeta significativamente as componentes normalizadas LF e HF, ou ainda, não altera os valores normalizados na análise espectral. Assim, ordens entre 9 a 25 podem ser utilizadas na análise de índices espectrais normalizados, e ordens entre 13 a 25 não afetam significativamente os valores absolutos dos índices de VFC no domínio da frequência.

A.2.2 Critérios para a Seleção da Ordem do Modelo AR

Existem alguns critérios para a seleção da ordem do modelo AR entre eles temos: FPE, AIC, MDL, CAT, RIS e BIC. Estes critérios levam em conta a variância do erro de previsão do modelo ao se determinar sua ordem. Assim, o modelo AR "ótimo" é o modelo de menor ordem que minimiza o erro de previsão, sem levar a carga computacional desnecessária. O número de amostras do sinal de entrada também é levado em conta.

- *FPE*. Critério Akaike de erro de previsão final (do inglês, *Akaike's final prediction error*) (Akaike, 1969). O FPE, seleciona a ordem do modelo AR de tal forma que a variância do erro de previsão é minimizada sem implicar em elevada ordem desnecessária. No caso, σ_p^2 é a variância do erro de previsão para a ordem p, e N é o número de amostras.

$$\mathrm{FPE}[p] = \sigma_p^2 \left(\frac{N+(p+1)}{N-(p+1)} \right) \tag{A.10}$$

- *AIC*. Critério de informação Akaike (do inglês, *Akaike's information criterion*) (Akaike, 1974). No AIC, a ordem ótima para o modelo autorregressivo é a que minimiza a Equação A.11.

$$\mathrm{AIC}[p] = N \times \ln(\sigma_p^2) + 2p \tag{A.11}$$

  Se o número de amostras (N) é grande, os critérios FPE e AIC apresentam resultados similares.

- *MDL*. Menor comprimento de descrição (do inglês, *minimum description length*) (Rissanen, 1983). O MDL foi desenvolvido para corrigir os critérios FPE e AIC, que superestimam a ordem ótima quando sinais com elevados números de amostras são utilizados.

$$\mathrm{MDL}[p] = N \times \ln(\sigma_p^2) + p\ln(N) \tag{A.12}$$

- *CAT*. Critério de Parzen da função de transferência autorregressiva (do inglês, *Parzen's criterion of autoregressive transfer function*) (Parzen, 1974). O CAT apresenta resultados semelhantes ao FPE e AIC.

$$\mathrm{CAT}[p] = \left(\frac{1}{N} \times \sum_{j=1}^{p} \frac{1-j/N}{\sigma_j^2} \right) - \frac{1-p/N}{\sigma_p^2} \tag{A.13}$$

- *RIS*. Método de descrição de comprimento mínimo de Rissanen (do inglês, *Rissanen's minimum description length method*) (Rissanen, 1984).

$$RIS[p] = \sigma_p^2 \left(1 + \left(\frac{p+1}{N} \right) \ln(N) \right) \tag{A.14}$$

- *BIC*. Critério bayesiano de informação (do inglês, *bayesian information criterion*) (Akaike, 1977), (Akaike, 1978), (Schwarz, 1978). O BIC é descrito na Equação A.15, onde σ_x^2 é a variância do sinal de saída (por exemplo, o sinal de VFC).

$$BIC[p] = N\ln(\sigma_p^2) - (N-p)\ln\left(1-\frac{p}{N}\right) + p\ln(N) + p\ln\left[\frac{1}{p}\left(\frac{\sigma_x^2}{\sigma_p^2}-1\right)\right] \quad \text{(A.15)}$$

A.3 Análise Espectral AR

O espectro do sinal mostra como a energia é distribuída como uma função da frequência. Teoricamente, a energia total, medida no domínio da frequência, é matematicamente idêntica à variância, medida no domínio do tempo. A análise espectral AR pode fornecer o número, a frequência central e a energia de componentes oscilatórias em uma série temporal. Quando o modelo AR é ajustado a uma série temporal, o modelo é invertido, a sequência de erro de previsão é considerada como uma entrada e a série temporal como uma saída para um filtro de síntese AR. Vide Figura A.2.

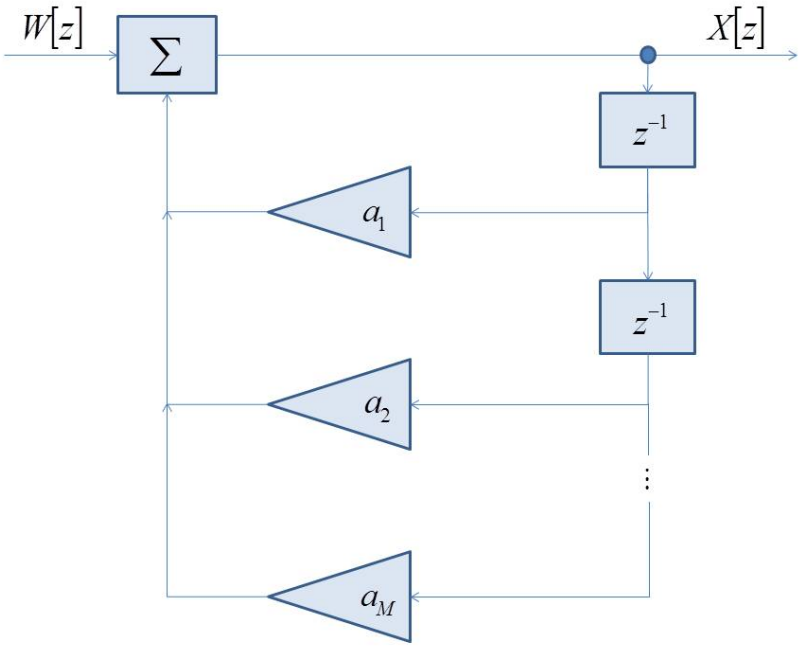


Figura A.2: Filtro de síntese AR. $W[z]$ e $X[z]$ representam as sequencias de entrada e saída, respectivamente. z^{-1} é um *delay* de um período de amostragem no domínio z, $a_1, a_2, \ldots, a_M$ são os coeficientes de predição, e o bloco $\sum$ realiza a soma dos valores $a_i X[z^{-1}]$, $i = 1, \ldots, M$.

Assim, as propriedades do filtro de síntese AR são então usadas para determinar o espectro de potência da série temporal.

Quando um sinal passa por um filtro, os valores de amplitude de determinadas frequências são modificadas. Diferentes frequências que compõem o sinal também são atrasadas por diferentes atrasos, ou seja, sinais de frequências diferentes são deslocados e depois colocados de volta juntos de uma forma diferente (Meddins e Meddins, 2000). A sequência de saída de qualquer filtro pode ser encontrada por meio da convolução, no domínio do tempo, ou pela multiplicação em frequência (realizada pela multiplicação das DFT's das séries de dados). O filtro de síntese pode ser visto através da Transformada Z, assim a Equação A.2 dada por $x[n] = \sum_{i=1}^{M} a_i x[n-i] + \varepsilon[n]$ transforma-se em:

$$X(z) = \sum_{i=1}^{M} a_i X(z) z^{-i} + W(z) \Leftrightarrow X(z) = \left(a_1 z^{-1} + a_2 z^{-2} + \cdots + a_M z^{-M}\right) X(z) + W(z).$$

Logo,

$$X(z)\left(1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M}\right) = W(z) \Leftrightarrow \frac{X(z)}{W(z)} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M}}.$$

Portanto,

$$H(z) = \frac{1}{1 - \sum_{i=1}^{M} a_i z^{-i}}. \tag{A.16}$$

Onde, $H(z)$ é a função de transferência do filtro de síntese AR. $X(z)$ deve ser considerada como a saída do filtro cuja entrada é a sequencia de erro de predição $W(z)$.

A resposta em frequência do filtro é dada por:

$$H\left(z|_{z=e^{j\omega T}}\right) = \frac{1}{1 - a_1 e^{-j\omega T} - \cdots - a_M e^{-j\omega MT}} \therefore H(e^{j\omega T}) = \frac{1}{1 - \sum_{i=1}^{M} a_i e^{-j\omega iT}}. \tag{A.17}$$

Ou ainda,

$$H(z) = \frac{1}{(1 - p_1 z^{-1}) \times (1 - p_2 z^{-1}) \times \cdots \times (1 - p_M z^{-1})}. \tag{A.18}$$

O espectro da série temporal modelada, $R\left(e^{j\omega}\right)$, é obtido pela multiplicação da variância do erro de predição com o quadrado da função de transferência. Em outras palavras, temos:

$$R\left(e^{j\omega}\right) = \left|H\left(e^{j\omega}\right)\right|^2 \times \sigma_p^2 \Rightarrow R\left(e^{j\omega}\right) = \frac{\sigma_p^2}{\left|1 - a_1 e^{-j\omega} - \cdots - a_M e^{-j\omega M}\right|^2}. \tag{A.19}$$

Resolução em frequência é uma das principais diferenças entre a DFT e a análise espectral baseada no modelo AR. Na DFT, a resolução em frequência é determinada pelo número de pontos utilizados para o cálculo, enquanto que nos métodos baseados em modelos o espectro pode ser avaliado em um número arbitrário de pontos e a resolução em frequência não é afetada pelo tamanho da sequência de dados. A estimação do espectro do sinal com

o modelo AR também permite que as componentes de frequência sejam determinadas mais precisamente que na DFT, desde que o modelo seja bom o suficiente.

Ao analisar os sinais de VFC, por exemplo, um valor de M maior que o obtido pelo critério AIC é frequentemente utilizado. Quanto maior o M escolhido, melhor será o ajuste do modelo aos dados. Uma ordem (M) elevada oferece melhor resolução frequencial, mas implica em menor robustez na estimativa de energia. Ordens muito altas também podem levar a divisão de linhas e a picos espúrios (Kay, 1988). Divisão de linha significa que um único pico no espectro se divide em dois picos. Picos espúrios são falsos picos espectrais em uma frequência onde não deveria existir um pico. Assim, apenas um pequeno número de parâmetros podem ser determinados, com segurança a partir de uma quantidade limitada de dados.

A.4 Modelagem Paramétrica de uma Série Temporal

Os métodos paramétricos de modelagem de séries temporais tem algumas vantagens sobre os métodos não paramétricos (baseados na DFT).

Existem muitos algoritmos para a obtenção de estimativas dos parâmetros AR. Por exemplo, métodos baseados em estimativa da sequência de autocorrelação (*Yule-Walker*), o algoritmo de *Burg*, e os algoritmos de predição linear baseados nos mínimos quadrados (incluindo o método de covariância modificado) (Kay, 1988), (Marple, 1987). Existem também algoritmos adaptativos, tais como, o mínimo quadrado médio (do inglês, *least mean square* ou LMS) e o mínimo quadrado recursivo (do inglês, *recursive least square* ou RLS), que atualizam as estimativas dos parâmetros a medida que novos dados tornam-se disponíveis (Marple, 1987).

A.4.1 Estimativa do Espectro AR

O espectro de potência do processo é dado por $P_x(z) = H(z)H\left(\frac{1}{z}\right)$, o qual pode ser escrito como:

$$P_x(z) = \frac{\sigma_\varepsilon^2}{A(z)A\left(\frac{1}{z}\right)}, \tag{A.20}$$

onde $A(z)$ é obtido da função de transferência do processo AR no domínio Z, dada pela equação A.18, e $P_\varepsilon(z) = \sigma_\varepsilon^2$ é o espectro da série temporal residual.

Considere a função de transferência de um processo AR de ordem p,

$$H(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k \times z^{-k}}. \tag{A.21}$$

Assumindo o processo AR de ordem p, o espectro de potência de $x(n)$ pode ser escrito como:

$$\hat{P}_{\text{AR}}(z) = \frac{\hat{\sigma}_e^2}{\prod_{k=1}^{p} (z - p_k)(z^{-1} - p_k^*)}, \tag{A.22}$$

onde $\hat{\sigma}_\varepsilon^2$ é a variância estimada de $\varepsilon(n)$, e p_k são os pólos do modelo. O espectro de potência também pode ser obtido utilizando as estimativas $\hat{a}_k$ e escrevendo $z = \mathrm{e}^{j\omega}$, tal que:

$$\hat{P}_{\text{AR}}(\omega) = \frac{\hat{\sigma}_\varepsilon^2}{\left| 1 + \sum_{k=1}^{p} \hat{a}_k \times \mathrm{e}^{-j\omega k} \right|^2}. \tag{A.23}$$

O espectro de uma única componente pode ser estimado pela seguinte expressão:

$$\hat{P}_k(z) = \frac{c_k}{(z - p_k)(z^{-1} - p_k^*)}. \tag{A.24}$$

Assumindo que $\omega = \omega_k$ e $z = \mathrm{e}^{j\omega}$. A constante c_k pode ser escrita como:

$$c_k \approx \frac{\hat{\sigma}_\varepsilon^2}{\prod_{j \neq k} (z - p_k)(z^{-1} - p_k^*)},\ z = \mathrm{e}^{j\omega_k}. \tag{A.25}$$

Agora a frequência ω_k está relaciona ao pólo p_k. Considera-se que a parte da estimativa de $\hat{P}_{\text{AR}}(z)$, c_k, será constante quando $\omega \approx \omega_k$. A estimativa do espectro AR deve seguir a relação $\hat{P}_{\text{AR}}(z) \approx \sum_k \hat{P}_k(z)$, ou seja, o espectro é a soma das estimativas espectrais das componentes.

A.4.2 Estimativa da Energia Associada às Componentes

Acima, definiu-se a estimativa do espectro de uma única componente relacionada com o pólo p_k. A energia de uma componente observada na frequência ω_k pode ser estimada utilizando o resíduo da função analisada (Johnsen e Andersen, 1978), (Marple, 1987).

$$\hat{P}(\omega_k) = q \cdot \Re \left\{ \text{RES}_{p_k} \left\{ \frac{\hat{P}_{\text{AR}}(z)}{z} \right\} \right\}, \tag{A.26}$$

onde o resíduo de $\hat{P}_{\text{AR}}(z)/z$ é determinado para o pólo p_k, e o coeficiente $q = 1$ para pólos reais e $q = 2$ para pólos complexos. A operação $\Re\{\cdot\}$ obtém a parte real da função. A potência total do espectro e os resíduos de uma função no domínio z são relacionados por:

$$\frac{1}{2\pi j} \oint f(z)dz = \sum_{k=1}^{n} \text{RES}_{p_k} f(z). \tag{A.27}$$

Além disso, a potência total do sinal deve ser igual a soma das estimativas de potência das componentes, $P_{\text{TOT}} \approx \sum_k \hat{P}(\omega_k)$. Assim, a estimativa de potência de uma única componente pode ser calculada por:

$$\hat{P}(\omega_k) = q \cdot \Re \left\{ \frac{(z - p_k)\hat{\sigma}_e^2}{z\hat{A}(z)\hat{A}\left(\frac{1}{z}\right)} \right\}. \tag{A.28}$$

Onde, $z = p_k$ e $\hat{A}(z) = \prod_{k=1}^{p} \left(1 - p_k z^{-1}\right)$, q é igual a 1 para pólos reais e 2 para pólos complexos, e $\hat{\sigma}_e^2$ é a variância do erro de predição da estimativa para uma dada ordem de modelo.

Na análise da VFC, as estimativas de potência são algumas vezes apresentadas em unidades normalizadas [n.u.] em vez de unidades absolutas [ms^2]. A estimativa de potência em unidades normalizadas será (Pagani et al., 1986):

$$\hat{P}(\omega_k)_{\text{n.u.}} = 100 \cdot \frac{\hat{P}(\omega_k)}{P_{\text{TOT}} - \hat{P}_0}. \tag{A.29}$$

Onde, $\hat{P}_0$ representa a estimativa de potência das flutuações muito lentas no sinal ($f < 0,03$ Hz) e P_{TOT} representa a potência total do sinal. Logo, $0 < \hat{P}(\omega_k)_{\text{n.u.}} \leq 100$.

A.5 Coeficiente de Correlação, Processos Estocásticos e Estacionariedade

A.5.1 Coeficiente de Correlação entre duas Variáveis Aleatórias

O momento conjunto de duas variáveis aleatórias X e Y fornece informação sobre o comportamento conjunto das variáveis. O jk^{th} momento conjunto de X e Y é definido por

$$\mathrm{E}\left[X^j Y^k\right] = \begin{cases} \int_0^\infty \int_{-\infty}^\infty x^j y^k f_{X,Y}(x,y)\,dx\,dy & X,Y \text{ Conjuntamente continuas} \\ \sum_i \sum_n x_i^j y_n^k p_{X,Y}(x_i, y_n) & X,Y \text{ Discretas} \end{cases} \tag{A.30}$$

Na engenharia elétrica, bem como em outras áreas, denomina-se o momento com $j = 1$ e $k = 1$, $\mathrm{E}[XY]$, como correlação entre X e Y.

O jk^{th} momento central de X e Y é definido como o momento conjunto das variáveis aleatórias centralizadas, $X - \mathrm{E}[X]$ e $Y - \mathrm{E}[Y]$:

$$\mathrm{E}\left[(X - \mathrm{E}[X])^j (Y - \mathrm{E}[Y])^k\right].$$

Onde a covariância de X e Y é definida como o momento central com $j = 1$ e $k = 1$:

$$\mathrm{cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]. \tag{A.31}$$

O coeficiente de correlação entre duas variáveis aleatórias X e Y é definido por

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\rho_X \rho_Y} = \frac{\text{E}[XY] - \text{E}[X]\text{E}[Y]}{\rho_X \rho_Y}. \tag{A.32}$$

A.5.2 Especificando um Processo Estocástico

Processo estocástico é uma família de variáveis aleatórias que descrevem a evolução no tempo e no espaço de um processo físico. Ou ainda, uma função $Z_t(\xi) : S \to \mathbb{R}$ que associa um número real a cada resultado ξ do experimento realizado ao longo do tempo. Geralmente representamos um processo estocástico por $\{X(t,\xi),\ t \in I\}$, onde o conjunto S é o espaço amostral dos possíveis resultados (ξ) de um experimento, e o conjunto I é o conjunto de índices, ou espaço paramétrico do processo estocástico.

Um processo estocástico $\{X(t,\xi),\ t \in I\}$ é completamente descrito pela coleção de funções de distribuição cumulativa conjunta de k^{th} ordem.

Amostrando $X(t,\xi)$ nos tempos $t_1, t_2, \ldots, t_k$, obtemos k variáveis aleatórias $X_1, X_2, \ldots, X_k$, tal que:

$$X_1 = X(t_1,\xi), X_2 = X(t_2,\xi), \ldots, X_k = X(t_k,\xi).$$

Logo, para processos estocásticos contínuos, temos:

$$F_{X_1,X_2,\ldots,X_k}(x_1,x_2,\ldots,x_k) = P[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k], \tag{A.33}$$

e para processos estocásticos discretos, temos:

$$p_{X_1,X_2,\ldots,X_k}(x_1,x_2,\ldots,x_k) = P[X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k]. \tag{A.34}$$

De forma correspondente, o processo estocástico $\{X(t,\xi),\ t \in I\}$ também é completamente descrito pelos momentos do conjunto de variáveis aleatórias $X_1, X_2, \ldots, X_k$. Como:

- Média: $m_X(t) = \text{E}[X(t)] = \int_{\infty}^{\infty} x \cdot f_{X(t)}(x)dx$, para t fixo;
- Autocorrelação: $R_X(t_1,t_2) = \text{E}[X(t_1) \cdot X(t_2)]$;
- Autocovariância: $C_X(t_1,t_2) = \text{E}[\{X(t_1) - m_X(t_1)\} \cdot \{X(t_2) - m_X(t_2)\}]$, ou $C_X(t_1,t_2) = R_X(t_1,t_2) - m_X(t_1) \cdot m_X(t_2)$;
- Variância: $\text{var}[X(t)] = \text{E}\left[(X(t) - m_X(t))^2\right] \Rightarrow \text{var}[X(t)] = C_X(t,t)$.

A.5.3 Processos Estocásticos Estacionários no Sentido Estrito

Um processo estocástico $\{X(t,\xi),\ t \in I\}$ é estacionário no sentido estrito se a distribuição conjunta das variáveis aleatórias $\{X(t_1), X(t_2), \ldots, X(t_k)\}$ for a mesma de:

$$\{X(t_1+\tau), X(t_2+\tau)\ldots X(t_k+\tau)\},$$

ou ainda, para todos τ, k e todos $t_1, t_2, \ldots, t_k$, tem-se:

$$F_{X(t_1),X(t_2),\ldots,X(t_k)}(x_1,x_2,\ldots,x_k) = F_{X(t_1+\tau),X(t_2+\tau),\ldots,X(t_k+\tau)}(x_1,x_2,\ldots,x_k). \tag{A.35}$$

Ou seja, um processo estocástico, contínuo ou discreto no tempo, $X(t)$ é estacionário se a distribuição conjunta de qualquer conjunto de amostras não depender do instante inicial em que a amostra foi obtida.

Isso implica que a média e a variância de um processo estacionário é constante e independente do tempo: $m_X(t) = \mathrm{E}[X(t)] = m,\ \forall t$ e $\mathrm{var}[X(t)] = \mathrm{E}\left[(X(t)-m)^2\right] = \sigma^2,\ \forall t$. E, a função de autocorrelação e a autocovariância de $X(t)$ dependem somente do intervalo de tempo $t_2 - t_1$: $R_X(t_1,t_2) = R_X(t_2-t_1),\ \forall t_1,t_2$ e $C_X(t_1,t_2) = C_X(t_2-t_1),\ \forall t_1,t_2$; ou ainda, depende somente do atraso (do inglês, *lag*) dado por: $\tau = t_2 - t_1$.

A.5.4 Processos Estocásticos Estacionários no Sentido Amplo

Um processo estocástico $\{X(t,\xi),\ t \in I\} \equiv \{X(t),\ t \in I\}$ é estacionário no sentido amplo ou fracamente estacionário (do inglês, *Wide Sense Stationarity* ou WSS) se:

- $m_X(t) = \mathrm{E}[X(t)] = m,\ \forall t$;
- $C_X(t_1,t_2) = C_X(t_2-t_1),\ \forall t_1,t_2$.

Ou ainda, o processo estocástico é dito fracamente estacionário, ou estacionário de segunda ordem, se a sua média é constante e se a sua função de autocovariância (e, autocorrelação) depende apenas do *lag* ($\tau = t_2 - t_1$).

A função de autocovariância, representada por: $C_X(\tau)$; $\tau = t_2 - t_1$, de um processo estacionário discreto $\{X_1, X_2, \ldots, X_n\}$ com média zero ($R_X(\tau) = C_X(\tau)$), satisfaz as seguintes propriedades:

1. $R_X(0) = \mathrm{E}[X(t)\cdot X(t)] \equiv R_X(0) = \mathrm{E}\left[X^2(t)\right] > 0$, para todo t. A função de autocorrelação, em $\tau = 0$, fornece a potência média do processo (momento de 2ª ordem);

2. $R_X(\tau) = \mathrm{E}\left[X(t+\tau)\cdot X(t)\right] = \mathrm{E}\left[X(t)\cdot X(t+\tau)\right] = R_X(-\tau) \equiv R_X(\tau) = R_X(-\tau)$. A função de autocorrelação é simétrica (par);

3. $\mathrm{P}\left[|X(t+\tau)\cdot X(t)| > \varepsilon\right] = \mathrm{P}\left[(X(t+\tau)\cdot X(t))^2 > \varepsilon^2\right] \leq \frac{\mathrm{E}\left[(X(t+\tau)-X(t))^2\right]}{\varepsilon^2} = \frac{2\{R_X(0)-R_X(\tau)\}}{\varepsilon^2}$
A função de autocorrelação é uma medida da taxa de mudança de um processo estocástico, considerando a mudança no processo de t a $t+\tau$. Ou ainda, se $R_X(0) - R_X(\tau)$ é pequeno, isto é, $R_X(\tau)$ varia lentamente, portanto a probabilidade de uma grande mudança em $X(t)$, em τ segundos, é pequena;

4. $R_X(0) \geq |R_X(\tau)|$, ou seja, a função de autocorrelação é máxima em $\tau = 0$;

5. Se $R_X(0) = R_X(d)$, então $R_X(\tau)$ é periódica com período d e $X(t)$ é periódico com média quadrática (do inglês, *mean square periodic*), ou seja,

$$\mathrm{E}\left[(X(t+d) - X(t))^2\right] = 0.$$

6. Seja $X(t) = m + N(t)$, onde $N(t)$ é um processo estocástico com média nula para o qual $R_N(\tau) \to 0$ quando $\tau \to \infty$. Então, $R_X(\tau) \to m^2$ quando $\tau \to \infty$. Em outras palavras, $R_X(\tau)$ tende ao quadrado da média de $X(t)$ à medida que $\tau \to \infty$.

Anexo A

Métodos de Classificação

Nas seções a seguir são descritos brevemente os classificadores utilizados para a identificação de padrões em sinais de VFC. Especificamente são avaliados três classificadores, os k vizinhos mais próximos, o *naive bayes*, e as máquinas de vetor suporte.

A Seção A.3 tratará mais detalhadamente das máquinas de vetor suporte. Nesta seção, serão analisados as características das máquinas de vetor suporte, as funções os parâmetros de cada núcleo utilizado na estrutura do classificador.

A.1 k Vizinhos mais Próximos

Existem diversos métodos de classificação, amplamente utilizados na prática, que não utilizam os dados de treinamento para obter um modelo de classificação. O processo de classificação utilizado por estes classificadores, em especial o k vizinhos mais próximos, pode ser entendido da seguinte forma: a cada nova instância que se deseja classificar, utiliza-se os dados de treinamento para verificar quais são as instâncias nesta base de dados que mais se assemelham à nova instância. Portanto, a instância será classificada dentro da classe mais comum a que pertencem as instâncias mais similares a ele, ou ainda, a classificação é feita por analogia tal que nenhum modelo de classificação é criado.

Este classificador requer técnicas eficientes de armazenamento dos dados de tal forma que a determinação da classe das novas instancias, por meio da comparação, seja eficiente. São adequados à implementação em ambientes de computação paralela, suportando aprendizado incremental.

O classificador k vizinhos mais próximos (do inglês, *k-Nearest Neighbor* ou KNN) utiliza os elementos do conjunto de treinamento para verificar quais destes mais se assemelham

(utilizando alguma métrica) ao elemento a ser classificado. Assim, o elemento é classificado dentro da classe com mais elementos com menor métrica relativa a ele.

Suponha um conjunto D de vetores de treinamento. Cada elemento de D é um vetor $(x_1, x_2, \ldots, xn, c)$, onde c é a classe à qual pertence o vetor de características $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Seja $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ um dado ainda não classificado. Para classificá-lo, calcula-se a distância de $\mathbf{y}$ a todos os elementos de D e considera-se os k elementos mais próximos a $\mathbf{y}$ (Mitchell, 1997).

A.2 *Naive Bayes*

O classificador *Naive Bayes* (NB) é um dos classificadores probabilísticos mais simples. O modelo construído por este classificador é um conjunto de probabilidades. As probabilidades são estimadas pela contagem da frequência de cada valor de característica para as instancias dos dados de treino. Dada uma nova instância, o classificador estima a probabilidade de essa instância pertencer a uma classe específica, baseada no produto das probabilidades condicionais individuas para os valores característicos da instância.

O cálculo utiliza o Teorema de Bayes e é por essa razão que o algoritmo é denominado um classificador de Bayes. O termo "*naive*", origina-se da suposição, como hipótese de trabalho, que o efeito do valor de uma característica não influencia o valor das demais, ou ainda, considera que as variáveis de entrada são condicionalmente independentes (uma característica não é relacionada com a outra) (Friedman et al., 1997). Esta hipótese tem como objetivo facilitar os cálculos envolvidos na tarefa de classificação. A Figura A.1 mostra a estrutura da rede *naive bayes*.

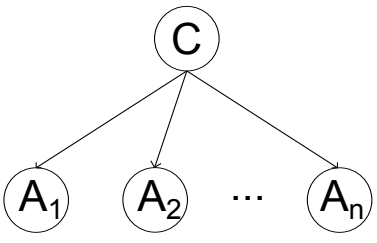


Figura A.1: Rede *naive bayes*

A classificação é então feita aplicando o Teorema de Bayes para calcular a probabilidade de C dado uma instância particular de $A_1, A_2, \ldots, A_n$ e então predizendo a classe com a maior probabilidade a posteriori.

Considere uma base de dados com m classes distintas: $C_1, C_2, \ldots, C_m$. Suponha que X é um dado de entrada desconhecido. O NB vai classificar X numa classe C para a qual a probabilidade condicional $P[C|X]$ é a mais alta possível. Repare que os valores dos índices de X podem ser encarados como um evento conjunto. Assim, X será classificado na classe C se a probabilidade condicional de C acontecer dado que X acontece é maior do que a probabilidade de qualquer outra classe C' acontecer dado que X acontece. Logo, X será classificada na classe C_i se $P[C_i|X] > P[C_j|X]$ para todas as classes C_j, $C_j \neq C_i$.

Pelo Teorema de Bayes, conhecida como probabilidade posterior, a probabilidade $P[C|X]$ é calculada por:

$$P[C|X] = \frac{P[X|C] \times P[C]}{P[X]}. \tag{A.1}$$

A probabilidade incondicional das classes, $P[C]$, pode ser obtida por meio do conhecimento de um especialista ou atribuindo probabilidades iguais para todas as classes $\left(P[C] = \frac{1}{m}\right)$.

A.3 Máquinas de Vetor Suporte

A.3.1 Introdução

As máquinas de vetor suporte (do inglês, *supported vector machines* ou SVM), propostas inicialmente por Vapnik (Vapnik, 1995), podem ser usadas para o processo de classificação de padrões e para regressão linear, da mesma forma que as redes perceptrons de múltiplas camadas e as redes de função de base radial. A Figura A.2 ilustra esta relação de semelhança, bem como mostra algumas características inerentes às SVM que serão descritas mais adiante.

No contexto de classificação de padrões, e analisando classes linearmente separáveis, a ideia principal de uma máquina de vetor suporte é construir um hiperplano como superfície de decisão de tal forma que a margem de separação entre exemplos positivos e negativos, ou dados de duas classes distintas, seja máxima, de acordo com a abordagem baseada na teoria da aprendizagem estatística (Haykin, 1994). Mais precisamente a máquina de vetor suporte é uma implementação do método de minimização estrutural de risco. Este princípio indutivo é baseado no fato de que a taxa de erro de uma máquina de aprendizagem sobre dados de teste (por exemplo, a taxa de erro de generalização) é limitada pela soma da taxa de erro de treinamento e por um termo que depende da dimensão de *Vapnik-Chervonenkis (V-C)*; no caso de padrões separáveis, uma máquina de vetor suporte produz um valor zero para o primeiro termo e minimiza o segundo termo.

Consequentemente, a máquina de vetor suporte pode fornecer um bom desempenho em

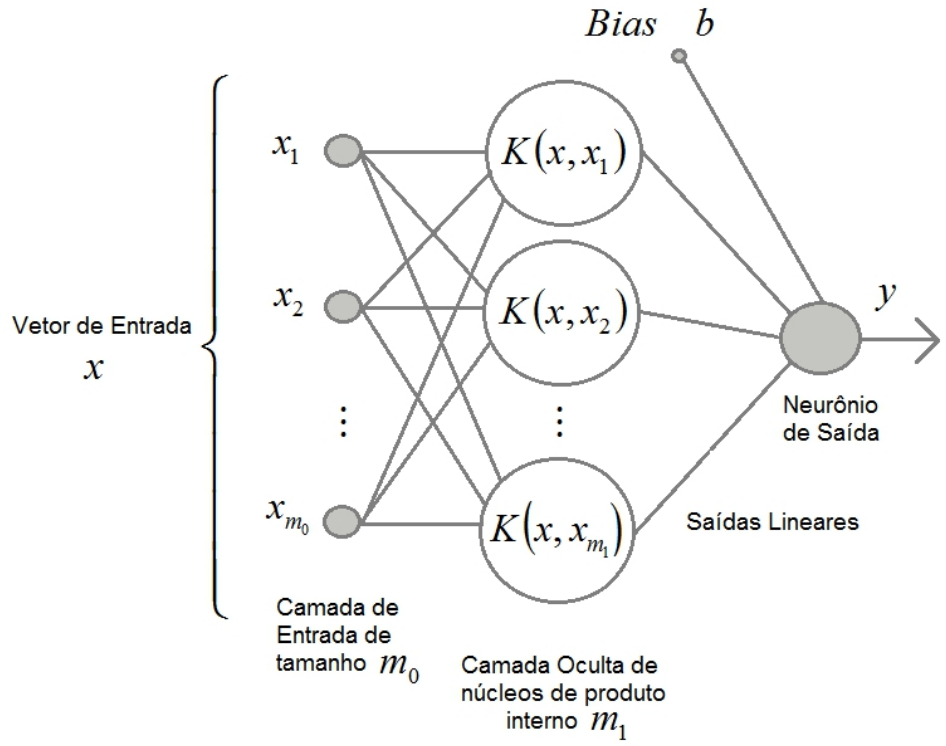


Figura A.2: Arquitetura da máquina de vetor de suporte.

problemas de classificação de padrões, apesar de não incorporar conhecimento do domínio do problema que será resolvido.

A.3.2 Hiperplano Ótimo para Padrões Linearmente Separáveis

Considere uma amostra de treinamento $(\mathbf{x}_i, d_i)_{i=1}^{N}$, onde $\mathbf{x}_i$ é o padrão de entrada para o i^{th} exemplo e d_i é a classe correspondente a este padrão. Assume-se que o padrão representado pelo subconjunto $d_i = +1$ e o padrão representado pelo conjunto $d_i = -1$ são "linearmente separáveis". A Equação A.2 representa a superfície de decisão na forma de um hiperplano que realiza a separação entre as classes:

$$\mathbf{w}^T\mathbf{x} + b = 0, \tag{A.2}$$

onde $\mathbf{x}$ é um vetor de entrada, $\mathbf{w}$ é um peso ajustável e b é um viés. Podemos assim escrever a Equação A.3

$$\begin{aligned} \mathbf{w}^T\mathbf{x}_i + b \geq 0 \text{ , para } d_i = +1 \\ \mathbf{w}^T\mathbf{x}_i + b < 0 \text{ , para } d_i = -1. \end{aligned} \tag{A.3}$$

Verifica-se que, se as classes de dados forem linearmente separáveis, ou ainda, se a Equação A.3 for válida, pode-se sempre ajustar os valores de $\mathbf{w}$ e b para obter uma nova configuração de modo que a Equação A.2 para os valores ótimos $\mathbf{w}_0^T x + b_0 = 0$ permaneça válida.

Para um dado vetor de peso $\mathbf{w}$ e viés b, a separação entre o hiperplano definido na Equação A.2 e o elemento do vetor de característica mais próximo é denominada margem de separação, representada por ρ. O objetivo de uma máquina de vetor suporte é encontrar o hiperplano particular para o qual a margem de separação é máxima. Sob esta condição, a superfície de decisão é referida como *hiperplano ótimo*. A Figura A.3 ilustra a construção geométrica de um hiperplano ótimo para um espaço de entrada bidimensional.

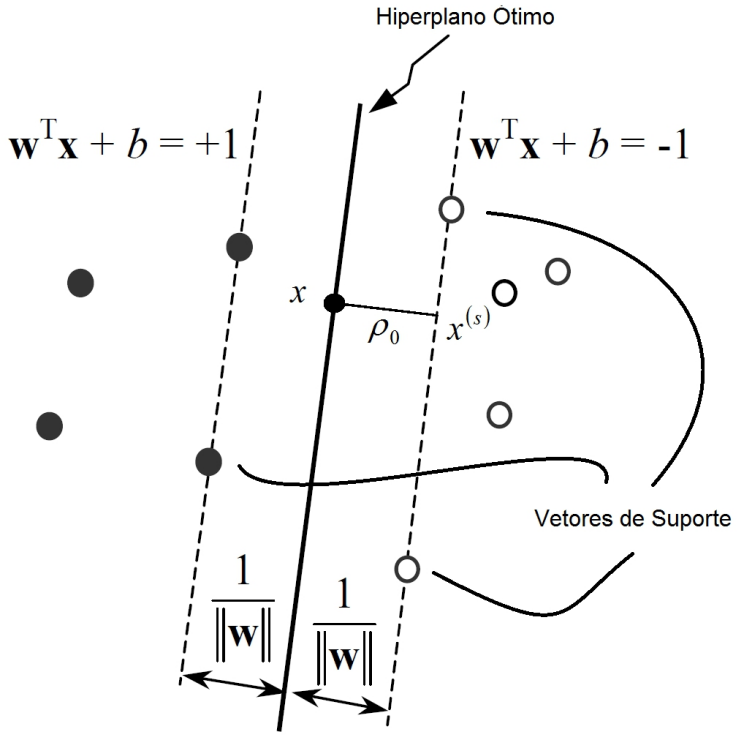


Figura A.3: Hiperplano ótimo para padrões linearmente separáveis.

Considere que $\mathbf{w}_0$ e b_0 representem os valores ótimos do vetor de pesos e do viés, respectivamente. Consequentemente, o hiperplano ótimo, representando uma superfície de decisão linear multidimensional no espaço de entrada, é definido pela Equação A.4, onde, basicamente, aplica-se estes valores ótimos na Equação A.2.

$$\mathbf{w}_0^T \mathbf{x} + b_0 = 0. \tag{A.4}$$

A função discriminante, descrita abaixo, fornece uma medida algébrica da distância de $\mathbf{x}$ até o hiperplano (Duda e Hart, 1973):

$$g(\mathbf{x}) = \mathbf{w}_0^T \mathbf{x} + b_0. \tag{A.5}$$

Seja $\mathbf{x}$ descrito como:

$$\mathbf{x} = \mathbf{x}_P + r\frac{\mathbf{w}_0}{\|\mathbf{w}_0\|}, \tag{A.6}$$

onde $\mathbf{x}_P$ é a projeção normal de $\mathbf{x}$ sobre o hiperplano ótimo, e r é a distância algébrica desejada, tal que r é positivo se $\mathbf{x}$ estiver no lado positivo do hiperplano ótimo e negativo se $\mathbf{x}$ estiver no lado negativo. Pela definição de hiperplano ótimo, temos que $g(\mathbf{x}_P) = 0$, ou ainda, $g(\mathbf{x}_P) = \mathbf{w}_0^T\mathbf{x}_P + b_0 = 0$ tal que a distância algébrica pode ser definida,

$$g(\mathbf{x}) = \mathbf{w}_0^T\left(\mathbf{x}_P + r\frac{\mathbf{w}_0}{\|\mathbf{w}_0\|}\right) + b_0 = \mathbf{w}_0^T\mathbf{x}_P + b_0 + r\|\mathbf{w}_0\| \therefore g(\mathbf{x}) = r\|\mathbf{w}_0\| \Rightarrow r = \frac{g(\mathbf{x})}{\|\mathbf{w}_0\|} \tag{A.7}$$

Logo,

$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}_0\|}. \tag{A.8}$$

Em particular, a distância da origem (considerando a origem em $\mathbf{x} = 0$) até o hiperplano ótimo é dada por $\frac{b_0}{\|\mathbf{w}_0\|}$. Se $b_0 > 0$, a origem está no lado positivo do hiperplano ótimo; se $b_0 < 0$, ela está no lado negativo; e se $b_0 = 0$, o hiperplano ótimo passa pela origem. Uma interpretação geométrica destes resultados algébricos é dada na Figura A.4.

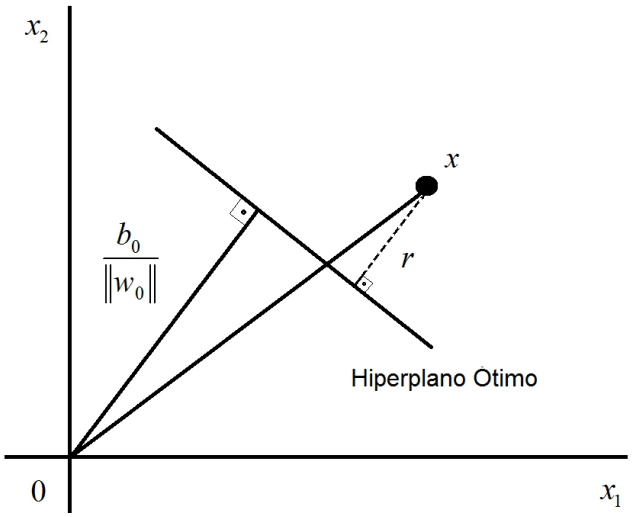


Figura A.4: Interpretação geométrica das distâncias algébricas de pontos até o hiperplano ótimo para um caso bidimensional.

Sendo assim, o problema se resume a encontrar os parâmetros $\mathbf{w}_0$ e b_0 para o hiperplano ótimo, dado o conjunto de treinamento $\psi = (\mathbf{x}_i, d_i)$. Com base na Figura A.4, vemos que o par $(\mathbf{w}_0, b_0)$ deve satisfazer à restrição:

$$\begin{aligned} \mathbf{w}_0^T\mathbf{x} + b_0 &\geq 1 \text{ , para } d_i = +1 \\ \mathbf{w}_0^T\mathbf{x} + b_0 &< 1 \text{ , para } d_i = -1. \end{aligned} \tag{A.9}$$

Os pontos particulares $(\mathbf{x}_i, d_i)$ para os quais a Equação A.9 é satisfeita com o sinal de igualdade são denominados de vetores suporte, por isso o nome "*máquina de vetor suporte*". Em termos conceituais, os vetores suportes são os vetores que se encontram mais próximos da superfície de decisão e de fato são os vetores mais difíceis de serem classificados. Sendo assim, verifica-se que eles tem grande influência na localização da superfície de decisão adequada.

Considere um vetor suporte $\mathbf{x}^{(s)}$ para o qual $d^{(s)} = +1$. Utilizando a Equação A.5 e analisando o posicionamento dos vetores suporte na Figura A.3 obtém-se, por definição, a equação abaixo:

$$g(\mathbf{x}^{(s)}) = \mathbf{w}_0^T \mathbf{x}^{(s)} \mp b_0 \mp 1 \text{ , } para\ d^{(s)} = \mp 1. \tag{A.10}$$

De acordo com a Equação A.7, verifica-se que a distância algébrica do vetor de suporte $\mathbf{x}^{(s)}$ até o hiperplano ótimo é dada por

$$r = \frac{g(\mathbf{x}^{(s)})}{\|\mathbf{w}_0\|} = \begin{cases} \frac{1}{\|\mathbf{w}_0\|}, & d^{(s)} = +1 \\ \frac{-1}{\|\mathbf{w}_0\|}, & d^{(s)} = -1 \end{cases}$$

onde o sinal positivo indica que $\mathbf{x}^{(s)}$ se encontra no lado positivo do hiperplano ótimo e o sinal negativo indica que $\mathbf{x}^{(s)}$ está no lado negativo do hiperplano ótimo. Considere que ρ represente o valor ótimo da margem de separação entre as duas classes que constituem o conjunto de treinamento ψ. Assim,

$$\rho = 2r = \frac{2}{\|\mathbf{w}_0\|}. \tag{A.11}$$

Em outras palavras, a Equação A.11 afirma que maximizar a margem de separação entre classes é equivalente a minimizar a norma euclidiana do vetor de pesos $\mathbf{w}$.

Em resumo, o hiperplano ótimo definido pela Equação A.4 é único no sentido de que o vetor de peso $\mathbf{w}_0$ fornece a máxima separação possível entre exemplos de classes distintas, tal que esta condição ótima é obtida com a minimização da norma do vetor de pesos $\mathbf{w}$.

A.3.3 Otimização Quadrática para Encontrar o Hiperplano Ótimo

A otimização quadrática é basicamente um procedimento eficiente computacionalmente para, utilizando a amostra de treinamento $\psi = \{(\mathbf{x}_i, d_i)\}_{i=1}^N$, encontrar o hiperplano ótimo, sujeito à seguinte restrição:

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \text{ , para } i = 1, 2, \ldots, N. \tag{A.12}$$

Esta restrição combina as duas linhas da Equação A.9 com $\mathbf{w}$ usado no lugar de $\mathbf{w}_0$. O problema de otimização restrito agora pode ser formulado como:

Teorema A.1 *Dada a mostra de treinamento* $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$*, encontre os valores ótimos do vetor de pesos* $\mathbf{w}$ *e viés b de modo que satisfaçam as seguintes restrições:*

- *obedeça à Equação A.12;*
- *o vetor de pesos* $\mathbf{w}$ *minimize a função de custo:* $\Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$.

O fator de escala $\frac{1}{2}$ é incluído por conveniência da formulação. Este problema de otimização restrito é denominado de problema primordial, sendo caracterizado como:

- a função de custo $\Phi(\mathbf{w})$ é uma função *convexa* de $\mathbf{w}$;
- as restrições são *lineares* em relação a $\mathbf{w}$.

Consequentemente, pode-se resolver o problema de otimização restrito usando o método dos multiplicadores de Lagrange (Haykin, 1994). Primeiro constrói-se a função *lagrangiana*:

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \left[d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right], \tag{A.13}$$

onde as variáveis auxiliares não-negativas α_i são chamadas de multiplicadores de Lagrange.

A solução para o problema de otimização restrito é determinada pelo ponto de sela da função lagrangiana $J(\mathbf{w}, b, \alpha)$, que deve ser minimizada, em relação a $\mathbf{w}$ e a b, igualando os resultados a zero. Realizando este procedimento obtemos as seguintes condições de otimização:

1. $\frac{\partial J(\mathbf{w},b,\alpha)}{\partial \mathbf{w}} = 0$;
2. $\frac{\partial J(\mathbf{w},b,\alpha)}{\partial b} = 0$.

A aplicação da condição de otimização 1 à função lagrangiana da Equação A.13 produz:

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i. \tag{A.14}$$

A aplicação da condição de otimização 2 à função lagrangiana da Equação A.13 produz:

$$\sum_{i=1}^{N} \alpha_i d_i = 0. \tag{A.15}$$

O vetor solução $\mathbf{w}$ é definido em termos de uma expansão que envolve os N exemplos de treinamento. Note, entretanto, que, embora esta solução seja única em virtude da convexidade lagrangiana, o mesmo não pode ser dito sobre os coeficientes de lagrange, α_i.

Também é importante notar que no ponto de sela, para cada multiplicador de Lagrange (α_i), o produto daquele multiplicador pela sua restrição correspondente desaparece, como mostrado por:

$$\alpha_i \left[d_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \right] = 0 \text{ , para } i = 1, 2, \ldots, N \tag{A.16}$$

Dessa forma, apenas aqueles multiplicadores que satisfazem exatamente a Equação A.16 podem assumir valores não-nulos. Esta propriedade resulta das condições de *Kuhn-Tucker* da teoria da otimização (Fletcher, 1987), (Bertsekas, 1995).

Como observado anteriormente, o problema primal lida com uma função de custo convexa e com restrições lineares. Dado um problema de otimização restrito como este, é possível construir um outro problema chamado de problema dual. Este segundo problema tem o mesmo valor ótimo do problema primordial, mas com os multiplicadores de Lagrange fornecendo a solução ótima. Em particular, pode-se formular o seguinte teorema da dualidade (Bertsekas, 1995):

- se o problema primordial tem uma solução ótima, então o problema dual também tem uma solução ótima, e os valores ótimos correspondentes são iguais;
- para que $\mathbf{w}_0$ seja uma solução primordial ótima e α_0 seja uma solução dual ótima, é necessário e suficiente que $\mathbf{w}_0$ seja realizável para o problema primordial, e

$$\Phi(\mathbf{w}_0) = J(\mathbf{w}_0, b_0, \alpha_0) = \min_{\mathbf{w}} J(\mathbf{w}_0, b_0, \alpha_0).$$

Para postular o problema dual para o referido problema primordial, primeiro deve-se expandir a Equação A.13, termo a termo, como segue:

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i d_i \mathbf{w}^T \mathbf{x}_i - b \sum_{i=1}^{N} \alpha_i d_i + \sum_{i=1}^{N} \alpha_i, \tag{A.17}$$

O terceiro termo no lado direito da Equação A.17 é zero devido à condição de otimização da Equação A.15. Além disso, da Equação A.14 temos:

$$\mathbf{w}^T\mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i \mathbf{w}^T \mathbf{x}_i = b \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j^T.$$

Consequentemente, fazendo a função objetivo $J(\mathbf{w}, b, \alpha) = Q(\alpha)$, pode-se reformular a Equação A.17 como

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j^T. \tag{A.18}$$

Onde os α_i são não-negativos.

Podemos agora formular o problema dual:

Teorema A.2 *Dada a amostra de treinamento* $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$, *encontre os multiplicadores de Lagrange* $\{\alpha_i\}_{i=1}^{N}$ *que maximizam a função objetivo*

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j^T$$

sujeita às restrições:

1. $\sum_{i=1}^{N} \alpha_i d_i = 0$;
2. $\alpha_i \geq 0$,*para* $i = 1, 2, \ldots, N$.

Note que o problema dual é formulado inteiramente em termos dos dados de treinamento. Além disso, a função $Q(\alpha)$ a ser maximizada depende apenas dos padrões de entrada na forma de um conjunto de produtos escalares, $\left\{(\mathbf{x}_i^T, \mathbf{x}_j)\right\}_{(i,j)=1}^{N}$.

Havendo determinado os multiplicadores de Lagrange ótimos, representados por $\alpha_{0,i}$, podemos calcular o vetor de pesos ótimo $\mathbf{w}_0$ usando a Equação A.14 e assim escrever

$$\mathbf{w}_0 = \sum_{i=1}^{N} \alpha_{0,i} d_i \mathbf{x}_i. \tag{A.19}$$

Para calcular o bias ótimo b_0, pode-se usar o $\mathbf{w}_0$ assim obtido e tirar vantagem da Equação A.10 relativa ao vetor de suporte positivo, e assim escrever

$$b_0 = 1 - \mathbf{w}_0^T x^{(s)} \text{ , para } d^{(s)} = 1. \tag{A.20}$$

A.3.4 Propriedades Estatísticas do Hiperplano Ótimo

Da teoria estatística da aprendizagem apresentada em (Haykin, 1994), verifica-se que a dimensão $V - C$ de uma máquina de aprendizagem determina o modo como uma estrutura de funções aproximativas deve ser usada, e observa-se que a dimensão $V - C$ de um conjunto de hiperplanos de separação em um espaço de dimensionalidade m é igual a $m + 1$. Entretanto, para se aplicar o método da minimização estrutural de risco, descrita em (Haykin, 1994), precisa-se construir um conjunto de hiperplanos de separação de dimensão $V - C$ variável tal que o risco empírico (por exemplo, o erro de classificação de treinamento) e a dimensão $V - C$ sejam minimizados ao mesmo tempo. Em uma SVM é imposta uma estrutura sobre o conjunto de hiperplanos de separação restringindo a norma euclidiana do vetor peso $\mathbf{w}$. Especificamente, pode-se formular o seguinte teorema (Vapnik, 1995), (Vapnik, 1998):

Teorema A.3 *Considere que D represente o diâmetro da menor esfera contendo todos os vetores de entrada* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$. *O conjunto de hiperplanos ótimos descrito pela Equação A.4 tem dimensão* $V - C$, *limitada acima por h:*

$$h \leq \min\left\{ \left\lceil \frac{D^2}{\rho^2}, m_0 \right\rceil \right\} + 1, \tag{A.21}$$

onde o sinal de máximo $\lceil \cdot \rceil$ *representa o menor inteiro maior ou igual ao número abrangido por ele,* ρ *é a margem de separação igual a* $\frac{2}{\|\mathbf{w}_0\|}$ *e* m_0 *é a dimensionalidade do espaço de entrada.*

Este teorema nos diz que podemos exercer controle sobre a dimensão $V - C$ do hiperplano ótimo, independentemente da dimensionalidade m_0 do espaço de entrada, escolhendo adequadamente a margem de separação ρ.

Suponha então que se tenha uma estrutura aninhada, descrita em termos dos hiperplanos de separação, como:

$$S_k = \left\{ \mathbf{w}^T \mathbf{x} + b : \|\mathbf{w}_0\| \leq c_k \right\}, \quad k = 1, 2, \ldots \tag{A.22}$$

Em virtude do limite superior h da dimensão $V - C$ definido na Equação A.21, a estrutura descrita na Equação A.22 pode ser reformulada em termos das margens de separação na forma equivalente

$$S_k = \left\{ \left\lceil \frac{r^2}{\rho^2} \right\rceil + 1 : \rho^2 \geq a_k \right\}, \quad k = 1, 2, \ldots \tag{A.23}$$

onde os α_k e c_k são constantes.

Observa-se que, para obter uma boa capacidade de generalização, devemos selecionar a estrutura particular com menor dimensão $V - C$ e erro de treinamento, de acordo com o princípio da minimização estrutural de risco (Haykin, 1994). Das Equações A.21 e A.23 vemos que esta exigência pode ser satisfeita usando-se o hiperplano ótimo. Equivalentemente, considerando a Equação A.11, devemos usar o vetor de pesos ótimo, tendo a norma euclidiana mínima. Assim, a escolha do hiperplano ótimo como a superfície de decisão para um conjunto de padrões linearmente separáveis concorda plenamente com o princípio de minimização estrutural de risco de uma máquina de vetor de suporte.

A.3.5 Hiperplano Ótimo para Padrões Não Separáveis

Considere agora o caso de padrões não separáveis. Dado um conjunto de dados de treinamento como a base de dados, na Seção 4.1, não é possível construir um hiperplano de

separação sem que haja erros de classificação. Apesar disso, deseja-se encontrar um hiperplano ótimo que minimize a probabilidade de erro de classificação, calculada como a média sobre o conjunto de treinamento.

Diz-se que a margem de separação entre classes é suave se um elemento do vetor de características $(\mathbf{x}_i, d_i)$ violar a condição dada pela Equação A.24.

$$d_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) \geq +1 \text{ , para } i = 1, 2, \ldots, N. \tag{A.24}$$

Esta violação pode surgir de duas formas:

- o ponto de dado $(\mathbf{x}_i, d_i)$ se encontra dentro da região de separação, mas do lado correto da superfície de decisão, como ilustrado na Figura A.5(a);
- o ponto de dado $(\mathbf{x}_i, d_i)$ se encontra dentro da região de separação, mas do lado incorreto da superfície de decisão, como ilustrado na Figura A.5(b).

Note que para o primeiro caso temos um processo de classificação correto, e no segundo um incorreto.

Para um tratamento matemático formal posterior para o caso de classes não separáveis, seja $\{\xi_i\}_{i=1}^N$ um novo conjunto de variáveis escalares não-negativas na definição do hiperplano de separação (superfície de decisão) como:

$$d_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) \geq 1 - \xi_i \text{ , para } i = 1, 2, \ldots, N. \tag{A.25}$$

As ξ_i, denominadas de variáveis soltas, medem o desvio de um elemento do vetor de características da condição ideal de separatibilidade de padrões, ou melhor, as variáveis soltas medem o quão distante os elementos do vetor de características estão da superfície de decisão. Para $0 \leq \xi_i \leq 1$, o elemento do vetor de características se encontra dentro da região de separação, mas do lado correto da superfície de decisão, como ilustrado na Figura A.5(a). Para $\xi_i > 1$, ele se encontra no lado errado do hiperplano de separação como ilustrado na Figura A.5(b). Os vetores de suporte são aqueles pontos de dados particulares que satisfazem a Equação A.24 precisamente, mesmo se $\xi_i > 0$. Observe que, se um exemplo com $\xi_i > 0$ for deixado fora do conjunto de treinamento, a superfície de decisão não muda. Assim, os vetores de suporte são definidos exatamente do mesmo modo, tanto para o caso linearmente separável como para o caso não separável.

O objetivo deste classificador é encontrar um hiperplano de separação para o qual o erro de classificação, como média sobre o conjunto de treinamento, é minimizado. Pode-se fazer isto minimizando o funcional

$$\Phi(\xi) = \sum_{i=1}^{N} I(\xi_i - 1).$$

Em relação ao vetor de pesos $\mathbf{w}$ sujeito a restrição descrita na Equação A.24 e a restrição sobre $\|\mathbf{w}\|^2$. A função $I(\xi)$ é uma função indicadora, definida por

$$I(\xi) = \begin{cases} 0, & \xi \leq 0 \\ 1, & \xi > 0 \end{cases}$$

Infelizmente, a minimização de $\Phi(\xi)$ em relação a $\mathbf{w}$ é um ponto de otimização não-convexo no qual, aparentemente, nenhum algoritmo computacional eficiente pode ser encontrado. Para tornar o problema matematicamente tratável, aproxima-se o funcional $\Phi(\xi)$ escrevendo

$$\Phi(\xi) = \sum_{i=1}^{N} \xi_i.$$

Além disso, simplifica-se os cálculos formulando o funcional a ser minimizado em relação ao vetor de pesos $\mathbf{w}$ como:

$$\Phi(w, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N} \xi_i. \tag{A.26}$$

Como anteriormente, a minimização do primeiro termo da Equação A.26 está relacionada com a minimização da dimensão $V-C$ da SVM. Assim, de forma análoga ao segundo termo $\sum_i \xi_i$, ele é um limite superior para o número de erros de teste. Assim, a formulação da função de custo $\Phi(\mathbf{w}, \xi)$ dada pela Equação A.26 está de acordo com o princípio da minimização estrutural de risco. O parâmetro C controla o compromisso entre a complexidade da máquina e o número de elementos não separáveis. Por isso, ele pode ser visto como uma forma de parâmetro de "regularização". Portanto, o parâmetro C deve ser selecionado pelo usuário. Isto pode ser feito de duas formas:

- o parâmetro C é determinado experimentalmente através do uso padrão de um conjunto de treinamento/teste (validação);
- ele é determinado analiticamente estimando a dimensão $V-C$ através da Equação A.21 e então usado-se limites de desempenho de generealização da máquina baseados na dimensão $V-C$.

De qualquer forma, o funcional $\Phi(\mathbf{w}, \xi)$ é otimizado em relação a $\mathbf{w}$ e $\{\xi_i\}_{i=1}^{N}$, sujeito à restrição descrita na Equação A.25 para $\xi_i \geq 0$. Fazendo isso, a norma quadrada de $\mathbf{w}$ é tratada como uma quantidade a ser minimizada simultaneamente em relação aos elementos não-separáveis, e não como uma restrição imposta sobre a minimização do número de elementos não separáveis.

O problema de otimização de padrões não separáveis assim formulado inclui o problema de otimização para padrões linearmente separáveis, como um caso especial. Especificamente, fazer $\xi_i = 0$ para todo i nas Equações A.25 e A.26 implica na redução destas às

formas correspondentes para o caso linearmente separável. Pode-se agora formalizar o problema primordial para o caso não separável como:

Teorema A.4 *Dada a amostra de treinamento* $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$*, encontre os valores ótimos do vetor de pesos* $\mathbf{w}$ *e do viés b de modo que satisfaçam a restrição*

$$d_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1 - \xi_i \text{ ,para } i = 1,2,\ldots,N,$$

onde $\xi_i \geq 0$, $\forall i$*. E de modo que o vetor de pesos* $\mathbf{w}$ *e as variáveis soltas* ξ_i *minimizem o funcional de custo*

$$\Phi(\mathbf{w},\xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i,$$

onde C é um parâmetro positivo especificado pelo usuário.

A.3.6 Máquina de Vetor de Suporte para Reconhecimento de Padrões

Tendo em mente a ideia simplificada de como encontrar o hiperplano ótimo para padrões não separáveis, pode-se agora descrever formalmente a construção de uma máquina de vetor de suporte para uma tarefa de reconhecimento de padrões.

Basicamente, a ideia da SVM depende de duas operações matemáticas:

1. o mapeamento não linear de um vetor de entrada para um espaço de características de alta dimensionalidade, que é oculto da entrada e da saída;
2. a construção de um hiperplano ótimo para separar as características descobertas no passo 1.

A Figura A.6 ilustra estas operações matemáticas.

A operação 1 é realizada de acordo com o teorema de Cover sobre a separabilidade de padrões (Haykin, 1994). Considerando um espaço de entrada constituído de padrões não linearmente separáveis, o teorema de Cover afirma que este espaço multidimensional pode ser transformado em um novo espaço de características onde os padrões são linearmente separáveis com alta probabilidade, desde que duas condições sejam satisfeitas. Primeiro, a transformação é não linear. Segundo, a dimensionalidade do espaço de características é suficientemente alta. Observe que estas duas condições são incorporadas na operação 1. Note, entretanto, que o teorema de Cover não discute se o hiperplano de separação é ótimo. No entanto, apenas através da utilização de um hiperplano de separação ótimo que a dimensão $V - C$ é minimizada e a generalização é alcançada.

É nesta última questão que entra a segunda operação. Especificamente, a operação 2 explora a ideia de construir um hiperplano de separação ótimo de acordo com a teoria descrita para padrões não separáveis, mas com uma diferença fundamental: o hiperplano de separação agora é definido como uma função linear de vetores retirados do espaço de características em vez do espaço de entrada original. O mais importante é que a construção deste hiperplano é realizada de acordo com o princípio da minimização estrutural de risco (de acordo com teoria de dimensão $V-C$) e depende do cálculo do núcleo de um produto interno.

Núcleo do Produto Interno

Considere que $\mathbf{x}$ represente um vetor retirado do espaço de entrada de dimensão m_0. Considere que $\{\varphi_j(\mathbf{x})\}_{j=1}^{m_1}$ represente um conjunto de transformações não lineares do espaço de entrada para o espaço de características, onde m_1 é a dimensão do espaço de características. Assume-se que $\varphi_j(\mathbf{x})$ é definido, a priori, para todo j. Dado este conjunto de transformações não lineares, podemos definir um hiperplano atuando como superfície de decisão de acordo com a Equação A.27.

$$\sum_{j=1}^{m_1} \mathrm{w}_j \varphi_j(\mathbf{x}) + b = 0. \tag{A.27}$$

Onde $\{\mathrm{w}_j\}_{j=1}^{m_1}$ representa um conjunto de pesos lineares conectando o espaço de características com o espaço de saída, e b é o viés. Podemos simplificar o desenvolvimento escrevendo

$$\sum_{j=0}^{m_1} \mathrm{w}_j \varphi_j(\mathbf{x}) = 0, \tag{A.28}$$

onde foi assumido que $\varphi_0(\mathbf{x}) = 1$ para todo $\mathbf{x}$, de modo que w_0 represente o viés b. A Equação A.28 define a superfície de decisão calculada no espaço de características em termos dos pesos lineares da máquina. A quantidade $\varphi_j(\mathbf{x})$ representa a entrada fornecida ao peso w_j através do espaço de características. Tal que,

$$\varphi(\mathbf{x}) = [\varphi_0(\mathbf{x}), \varphi_1(\mathbf{x}), \ldots, \varphi_{m_1}(\mathbf{x})]^T, \tag{A.29}$$

onde, por definição, temos:

$$\varphi_0(\mathbf{x}) = 1\,, \quad \forall \mathbf{x}. \tag{A.30}$$

Na verdade, o vetor $\varphi(\mathbf{x})$ representa a "imagem" induzida no espaço de características pelo vetor de entrada $\mathbf{x}$, como ilustrado na Figura A.6. Assim, em termos desta imagem, pode-se definir a superfície de decisão na forma simplificada:

$$\mathbf{w}^T \varphi(\mathbf{x}) = 0. \tag{A.31}$$

Adaptando a equação $\mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i \mathbf{x}_i$, relacionada à resolução do problema de otimização restrito para encontrar o hiperplano ótimo por meio da função lagrangiana, à situação envolvendo um espaço de características onde se procura a separabilidade "linear" de características, pode-se escrever

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i \varphi(\mathbf{x}_i), \tag{A.32}$$

onde o vetor de características $\varphi(\mathbf{x}_i)$ corresponde ao padrão de entrada $\mathbf{x}_i$ no i^{th} exemplo. Dessa forma, substituindo a Equação A.32 em A.31, pode-se definir a superfície de decisão calculada no espaço de características como:

$$\sum_{i=1}^{N} \alpha_i d_i \varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}) = 0. \tag{A.33}$$

O termo $\varphi^T(\mathbf{x}_i) \varphi(\mathbf{x})$ representa o produto interno de dois vetores induzidos no espaço de características pelo vetor de entrada $\mathbf{x}$ e o padrão de entrada $\mathbf{x}_i$ relativo ao i^{th} exemplo. Podemos então introduzir o *núcleo do produto interno* representado por

$$K(\mathbf{x}, \mathbf{x}_i) = \varphi^T(\mathbf{x}) \varphi(\mathbf{x}_i) = \sum_{j=0}^{m_1} \varphi_j^T(\mathbf{x}) \varphi_j(\mathbf{x}_i) \text{ , para } i = 1, 2, \ldots, N. \tag{A.34}$$

Desta definição verifica-se que o núcleo do produto interno é uma *função simétrica* de seus argumentos, ou seja, $K(\mathbf{x}, \mathbf{x}_i) = K(\mathbf{x}_i, \mathbf{x})$, $\forall i$.

O mais importante é que pode-se usar o núcleo do produto interno $K(\mathbf{x}, \mathbf{x}_i)$ para construir o hiperplano ótimo no espaço de características sem ter que considerar o próximo espaço de características de forma explícita. Isto é visto facilmente usando-se a Equação A.34 em A.33, de onde resulta que o hiperplano é agora definido por

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) = 0. \tag{A.35}$$

Projeto ótimo de uma SVM

A expansão do núcleo de produto interno $K(\mathbf{x}, \mathbf{x}_i)$ na Equação A.34 nos permite construir uma superfície de decisão que é não linear no espaço de entrada, mas cuja imagem no espaço de característica é linear. Com base nesta expansão, pode-se agora formular a forma dual para a otimização restrita de uma SVM como:

Teorema A.5 *Dada a amostra de treinamento* $\{(\mathbf{x}_i, d_i)\}_{i=1}^{N}$, *encontre os multiplicadores de Lagrange* $\{\alpha_i\}_{i=1}^{N}$ *que maximizam a função objetivo*

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(\mathbf{x}_i, \mathbf{x}_j),$$

sujeita às restrições

1. $\sum_{i=1}^{N} \alpha_i d_i = 0$;

2. $0 \leq \alpha_i \leq C$,*para* $i = 1,2,\ldots,N$. *Onde C é um parâmetro positivo especificado pelo usuário.*

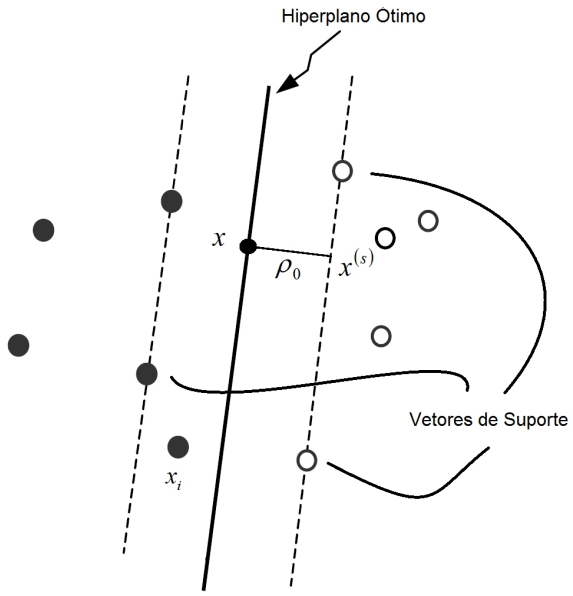
Note que a restrição 1 surge da otimização do lagrangiano $Q(\mathbf{x})$ em relação ao viés $b = \mathrm{w}_0$ para $\varphi_0(\mathbf{x}) = 1$. O problema dual formulado tem a mesma forma como no caso de padrões não separáveis, exceto pelo fato de que o produto interno usado anteriormente foi substituído pelo núcleo do produto interno $K(\mathbf{x}_i, \mathbf{x}_j)$. Podemos ver $K(\mathbf{x}_i, \mathbf{x}_j)$ como o elemento ij de uma matriz simétrica de dimensão $N \times N$, K, como mostrado por

$$K = \{K(\mathbf{x}_i, \mathbf{x}_j)\}_{(i,j)=1}^{N}. \tag{A.36}$$

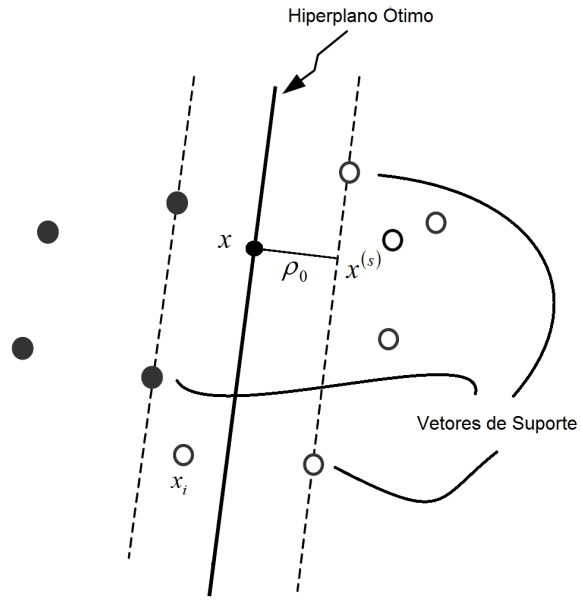
Tendo encontrado os valores ótimos dos multiplicadores de Lagrange, representados por $\alpha_{0,i}$, pode-se determinar o valor ótimo correspondente do vetor linear de pesos $\mathbf{w}_0$, que conecta o espaço de características ao espaço de saída adaptando a fórmula $\mathbf{w}_0 = \sum_{i=1}^{N} \alpha_{0,i} d_i \varphi(\mathbf{x}_i)$ à nova situação. Especificamente, reconhecendo que a imagem $\varphi(\mathbf{x}_i)$ desempenha o papel de entrada para o vetor de pesos $\mathbf{w}$, pode-se definir $\mathbf{w}_0$ como

$$\mathbf{w}_0 = \sum_{i=1}^{N} \alpha_{0,i} d_i \varphi(\mathbf{x}_i), \tag{A.37}$$

onde $\varphi(\mathbf{x}_i)$ é a imagem induzida no espaço de características devido a $\mathbf{x}_i$. Note que a primeira componente de $\mathbf{w}_0$ representa o viés ótimo b_0.



(a)



(b)

Figura A.5: (a) O elemento $\mathbf{x}_i$, pertencente à primeira classe, se encontra dentro da região de separação, mas no lado correto da superfície de decisão. (b) O ponto de dado $\mathbf{x}_i$, pertencente à segunda classe, se encontra no lado errado da superfície de decisão

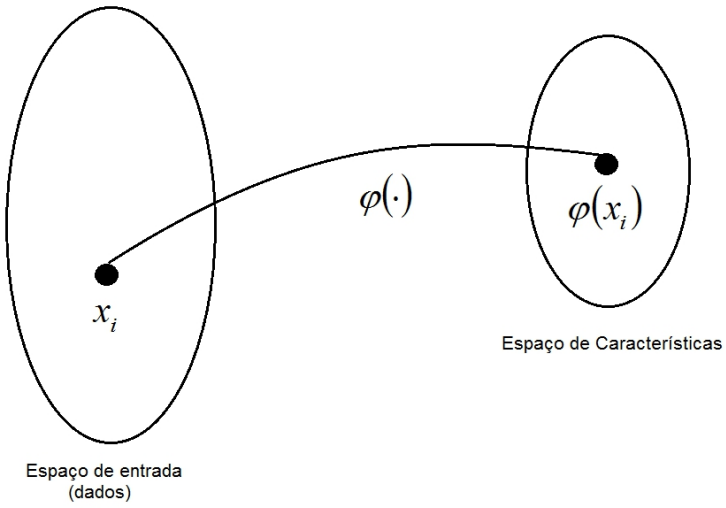


Figura A.6: Mapa não linear do espaço de entrada para o espaço de características.