

Universidade de Brasília – UnB  
Campus Gama – FGA  
Programa de Pós-Graduação em Engenharia Biomédica

**ABORDAGENS UNIVARIÁVEIS E MULTIVARIÁVEIS DE  
ÍNDICES AUTONÔMICOS CARDÍACOS EM MODELOS  
PREDITIVOS DE DIABETES TIPO 2:  
ANÁLISES COM TÉCNICAS ESTATÍSTICAS E DE  
APRENDIZADO DE MÁQUINA**

**ALEX ALVES BERNARDES**

Orientador: PROF<sup>a</sup> DR<sup>a</sup> FLÁVIA MARIA G. S. A. OLIVEIRA



UNB – UNIVERSIDADE DE BRASÍLIA

FGA – FACULDADE GAMA



Programa de  
Pós-Graduação em  
**ENGENHARIA BIOMÉDICA**

**ABORDAGENS UNIVARIÁVEIS E MULTIVARIÁVEIS DE  
ÍNDICES AUTONÔMICOS CARDÍACOS EM MODELOS  
PREDITIVOS DE DIABETES TIPO 2: ANÁLISES COM  
TÉCNICAS ESTATÍSTICAS E DE APRENDIZADO DE  
MÁQUINA**

**ALEX ALVES BERNARDES**

ORIENTADOR: PROF<sup>a</sup> DR<sup>a</sup> FLÁVIA MARIA G. S. A. OLIVEIRA

DISSERTAÇÃO DE MESTRADO EM  
ENGENHARIA BIOMÉDICA

PUBLICAÇÃO: 187A/2024

BRASÍLIA/DF, AGOSTO DE 2024

**UNB – UNIVERSIDADE DE BRASÍLIA**  
**FGA – FACULDADE GAMA**  
**PROGRAMA DE PÓS-GRADUAÇÃO**

**ABORDAGENS UNIVARIÁVEIS E MULTIVARIÁVEIS DE  
ÍNDICES AUTONÔMICOS CARDÍACOS EM MODELOS  
PREDITIVOS DE DIABETES TIPO 2: ANÁLISES COM  
TÉCNICAS ESTATÍSTICAS E DE APRENDIZADO DE  
MÁQUINA**

ALEX ALVES BERNARDES

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA BIOMÉDICA

APROVADA POR:

---

Prof<sup>a</sup> Dr<sup>a</sup> Flávia Maria G. S. A. Oliveira

(Orientador)

---

Prof Dr Cristiano Jacques Miosso

(Examinador interno)

---

Prof<sup>a</sup> Dr<sup>a</sup> Talía Simões dos Santos Ximenes

(Examinador externo)

**FICHA CATALOGRÁFICA**

BERNARDES, ALEX A.

ABORDAGENS UNIVARIÁVEIS E MULTIVARIÁVEIS DE ÍNDICES AUTONÔMICOS CARDÍACOS EM MODELOS PREDITIVOS DE DIABETES TIPO 2:

ANÁLISES COM TÉCNICAS ESTATÍSTICAS E DE APRENDIZADO DE MÁQUINA [Distrito Federal], 2024.

91p., 210 × 297 mm (FGA/UnB Gama, Mestrado em Engenharia Biomédica, 2024).

Dissertação de Mestrado em Engenharia Biomédica, Faculdade UnB Gama, Programa de Pós-Graduação em Engenharia Biomédica.

- |                                       |                             |
|---------------------------------------|-----------------------------|
| 1. Processamento de Sinais Biomédicos | 2. Diabetes Mellitus Tipo 2 |
| 3. Seleção de atributos               | 4. Aprendizado de máquina   |
| I. FGA UnB/UnB.                       | II. Título (série)          |

**REFERÊNCIA**

BERNARDES, ALEX A. (2024). ABORDAGENS UNIVARIÁVEIS E MULTIVARIÁVEIS DE ÍNDICES AUTONÔMICOS CARDÍACOS EM MODELOS PREDITIVOS DE DIABETES TIPO 2: ANÁLISES COM TÉCNICAS ESTATÍSTICAS E DE APRENDIZADO DE MÁQUINA. Dissertação de mestrado em engenharia biomédica, Publicação 187A/2024, Programa de Pós-Graduação, Faculdade UnB Gama, Universidade de Brasília, Brasília, DF, 91p.

**CESSÃO DE DIREITOS**

AUTOR: Alex Alves Bernardes

TÍTULO: ABORDAGENS UNIVARIÁVEIS E MULTIVARIÁVEIS DE ÍNDICES AUTONÔMICOS CARDÍACOS EM MODELOS PREDITIVOS DE DIABETES TIPO 2: ANÁLISES COM TÉCNICAS ESTATÍSTICAS E DE APRENDIZADO DE MÁQUINA

GRAU: Mestre

ANO: 2024

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

## **AGRADECIMENTOS**

Agradeço a todos aqueles que direta ou indiretamente me ajudaram a ampliar meus conhecimentos para a formalização deste trabalho. À minha esposa, Deanívea, por ter sido amorosa e compreensiva nos momentos em que estive ocupado, e por ter me confortado nos momentos de estresse. Aos meus pais, Cláudio e Doralice, por terem me apoiado em todas as decisões que tomei na minha vida pessoal, profissional e acadêmica. E, por fim, agradeço à professora Flávia pelas orientações no decorrer deste trabalho, por toda paciência e gentileza, mesmo nos momentos mais difíceis.

*Alex Alves Bernardes.*

## RESUMO

A Diabetes Mellitus tipo 2 é uma doença metabólica cuja incidência ocorre majoritariamente em adultos e está em crescimento contínuo em todo o mundo. Seus sintomas afetam o sistema nervoso autônomo (SNA), causando disfunções autonômicas. Diversas evidências indicam que é possível prever a presença da doença a partir de avaliações quantitativas da atividade autonômica sobre a regulação cardíaca.

Este trabalho tem como objetivo determinar índices quantitativos do SNA através de técnicas univariáveis e multivariáveis que consigam diferenciar indivíduos portadores de diabete mellitus (IPDM) e indivíduos controle (IC) utilizando sinais fisiológicos obtidos de maneira não invasiva como eletrocardiograma (ECG), pressão arterial (PA) contínua e fluxo respiratório (FR).

Adicionalmente, utilizando a abordagem *insightful*, deseja-se avaliar a eficácia de diferentes combinações de índices autonômicos. Estes índices serão utilizados como atributos de algoritmos de aprendizado de máquina na predição de IPDM e IC. Inicialmente, o estudo investiga o uso de modelos preditivos utilizando índices autonômicos obtidos a partir de um único sinal isoladamente, como o ECG ou a PA. Em seguida, são desenvolvidos modelos preditivos utilizando índices multivariáveis, que requerem associação de sinais fisiológicos complementares ao ECG, relacionando-os em modelos paramétricos cuja restrição de causalidade entre os sinais é imputada.

Os dados para este estudo foram obtidos da base de dados pública *“Cerebrovascular Disease in Elderly with Diabetes”*. A base é composta por dados de ECG, PA contínua e FR de indivíduos com idades entre 55 e 75 anos, com e sem diabetes tipo 2. Os agrupamentos de índices se basearam nos sinais fisiológicos de origem e em metodologias matemáticas de extração. Agrupamentos especiais foram propostos utilizando técnicas estatísticas como análise de correlação e análise de variância, além de técnicas de aprendizado de máquina. Quatro algoritmos de classificação foram utilizados: 1) *support vector machines* (SVM), 2) árvore de decisão (AD), 3) *k-nearest neighbor* (KNN), e 4) regressão logística (RL).

Oito conjuntos foram obtidos dos agrupamentos de índices autonômicos: **A**) VFC: índices relacionados exclusivamente à variabilidade da frequência cardíaca (VFC); **B**) VPA: índices espectrais da variabilidade da pressão arterial (VPA); **C**) FRF: combinação de índices

de função resposta em frequência (FRF) obtidos da análise espectral da VFC, VPA e índices de identificação de sistemas por resposta ao impulso (RI); **D**) RI: índices de identificação de sistemas por RI; **E**) VFC + VPA: a combinação dos índices nos conjuntos **A** e **B**; **F**) VFC + RI: a combinação dos índices nos conjuntos **A** e **D**; **G**) Especial 1: índices selecionados com base em resultados da análise de variância dos índices; **H**) Especial 2: índices que obtiveram alto desempenho na predição com índices individuais.

As métricas dos modelos de aprendizado de máquina utilizando índices autonômicos indicaram que o agrupamento de índices associados à VFC em conjunto com outros agrupamentos de índices tende a aumentar a acurácia, precisão e sensibilidade dos modelos. Entretanto, o agrupamento de índices associados exclusivamente à VFC apresentou um desempenho inferior aos demais agrupamentos na predição de indivíduos diabéticos.

Os agrupamentos de índices denominados especiais, **G** e **H**, se sobressaíram obtendo os melhores resultados de acurácia e medida-F1, dentre os demais agrupamentos em 3 dos AAM utilizados. O algoritmo **SVM** apresentou para o agrupamento **G** as respectivas métricas de acurácia (A), precisão (P), sensibilidade (S) e medida-F1 (F1): A=80,72%; P=79,47%; S=97,14%; F1=87,26%. Enquanto o agrupamento **H** apresentou os seguintes resultados: A=78,78%; P=77,53%; S=94,28%; F1=84,78%. Para o algoritmo **KNN**, o agrupamento **G** apresentou: A=80,72%; P=80,55%; S=93,80%; F1=86,5%. O agrupamento **H**: A=80,60%; P=79,47%; S=94,28%; F1=85,94%. Para o algoritmo **RL**, o agrupamento **G**: A=72,90%; S=88,57%; P=76,36%; F1=81,28%. O agrupamento **H**: A= 80,60%; P=80,47%; S=91,42%; F1=85,47%. A divergência de desempenho ocorreu com o algoritmo **AD**, onde as métricas de desempenho foram lideradas pelo agrupamento **E** contendo índices univariáveis da VFC e da VPA. As métricas do agrupamento **E** são: A=78,21%; P=86,66%; S=84,0%; F1=83,32%.

Dentre os índices individuais avaliados, os índices multivariáveis associados ao ganho dinâmico do acoplamento cardiorrespiratório (ACR) apresentaram os resultados mais robustos, como por exemplo o ganho dinâmico de alta-frequência 0,15 - 0,4Hz (AF) do ACR combinado com o acoplamento do barorreflexo arterial, com A=80,75% e F1=84,89%. A mesma tendência aparece nos agrupamentos de índices, onde diversos índices dos agrupamentos **G** e **H** são índices relacionados ao ACR, em combinação com índices univariáveis exclusivamente associados à VFC.

Os agrupamentos especiais, utilizando seleção de atributos, apresentaram melhora significativa na predição de IPDM e IC em relação a agrupamentos de índices puramente agrupados por origem de sinal e metodologia de extração. Os agrupamentos especiais também apresentaram melhora em algumas métricas para predição dos indivíduos em relação à predição utilizando índices individualmente, porém menos significativa.

**Palavras-chave:** Processamento de sinais biomédicos, Diabetes Mellitus tipo 2, Seleção de atributos, Aprendizado de máquina, Análise estatística.

## ABSTRACT

Type 2 Diabetes Mellitus is a metabolic disease whose incidence occurs mainly in adults and is continuously growing worldwide. Its symptoms affect the autonomic nervous system (ANS), causing autonomic dysfunctions. Several pieces of evidence indicate that it is possible to predict the presence of the disease through quantitative assessments of autonomic activity on cardiac regulation.

This work aims to determine quantitative indices of the ANS, using univariate and multivariate techniques, that can differentiate diabetes mellitus individuals (DMI) and control group (CG) using physiological signals obtained non-invasively such as electrocardiogram (ECG), continuous blood pressure (BP), and respiratory flow (RF).

Additionally, using an insightful approach, the goal is to evaluate the effectiveness of different combinations of autonomic indices for use as attributes in machine learning algorithms for predicting DMI and CG. Initially, the study investigates the use of predictive models employing autonomic indices obtained from a single signal in isolation, such as ECG or BP. Subsequently, predictive models are developed using multivariable indices, which require the combination of physiological signals complementary to the ECG, relating them in parametric models where the causality constraint between the signals is imposed.

The data for this study were obtained from the public database “*Cerebrovascular Disease in Elderly with Diabetes*”. The dataset consists of ECG, continuous BP, and RF data from individuals aged 55 to 75 years, with and without type 2 diabetes. The groupings of indices were based on the physiological signals of origin and mathematical extraction methodologies. Special groupings were proposed using statistical techniques such as correlation analysis and variance analysis, in addition to machine learning techniques. Four classification algorithms were used: 1) support vector machines (SVM), 2) decision tree (DT), 3) k-nearest neighbor (KNN), and 4) logistic regression (LR).

Eight sets were obtained from the groupings of autonomic indices: **A**) VFC/HRV - indices exclusively related to heart rate variability (HRV); **B**) VPA/BPV - spectral indices



of blood pressure variability (BPV); **C**) FRF - combination of FRF indices obtained from the spectral analysis of HRV, BPV, and system identification indices by impulse response (IR); **D**) RI/IR - system identification indices by IR; **E**) VFC/HRV + VPA/BPV - the combination of indices in sets **A** and **B**; **F**) VFC/HRV + RI/IR - the combination of indices in sets **A** and **D**; **G**) Special 1 – indices selected based on variance analysis results of the indices; **H**) Special 2 - indices that showed high performance in prediction with individual indices.

The metrics from the machine learning models using autonomic indices indicated that the grouping of indices associated with HRV in combination with other groupings of indices tends to increase the accuracy, precision, and sensitivity of the models. However, the grouping of indices associated exclusively with HRV performed worse than the other groupings in predicting diabetic individuals.

The groupings of indices called special, **G** and **H**, stood out by obtaining the best accuracy and F1-score results among the other groupings in 3 of the AAM used. The **SVM** algorithm presented the following metrics for grouping **G**: accuracy (A), precision (P), sensitivity (S), and F1-score (F1): A=80.72%; P=79.47%; S=97.14%; F1=87.26%. While grouping **H** presented the following results: A=78.78%; P=77.53%; S=94.28%; F1=84.78%. For the **KNN** algorithm, grouping **G** presented: A=80.72%; P=80.55%; S=93.80%; F1=86.5%. Grouping **H**: A=80.60%; P=79.47%; S=94.28%; F1=85.94%. For the **LR** algorithm, grouping **G**: A=72.90%; S=88.57%; P=76.36%; F1=81.28%. Grouping **H**: A=80.60%; P=80.47%; S=91.42%; F1=85.47%. The performance divergence occurred with the **DT** algorithm, where the performance metrics were led by grouping **E**, containing univariate indices of HRV and BPV. The metrics for grouping **E** are: A=78,21%; P=86,66%; S=84,0%; F1=83,32%.

Among the individual indices evaluated, multivariate indices associated with the dynamic gain of cardiorespiratory coupling (CRC) presented more robust results, such as the dynamic gain of high-frequency 0.15 - 0.4Hz (HF) from CRC combined with baroreflex mechanism coupling, with A=80.75% and F1=84.89%. The same trend appears in the groupings of indices, where the most indices of groupings **G** and **H** are related to the CRC, in combination with univariable indices exclusively associated with HRV.

The special groupings, using feature selection, showed a significant improvement in the prediction of DMI and CG compared to groupings of indices purely grouped by signal origin and extraction methodology. The special groupings also showed an improvement in some metrics for predicting individuals compared to using individual indices, but the improvement was less significant.

**Keywords:** Biomedical signal processing, Type 2 Diabetes Mellitus, Features selection, Machine Learning, Statistical analysis.

## SUMÁRIO

<b>Agradecimentos</b>	<b>v</b>
<b>Resumo</b>	<b>vi</b>
<b>Abstract</b>	<b>viii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Organização do trabalho . . . . .	4
<b>2 Fundamentação teórica</b>	<b>6</b>
2.1 Diabetes mellitus . . . . .	6
2.1.1 O contexto da diabetes mellitus no Brasil . . . . .	7
2.1.2 O contexto da diabetes mellitus nos Estados Unidos . . . . .	8
2.2 Sistema cardiovascular . . . . .	8
2.2.1 A atividade elétrica cardíaca . . . . .	9
2.2.2 Regulação do sistema cardiovascular . . . . .	10
2.3 Pressão arterial . . . . .	10
2.4 Fluxo respiratório . . . . .	13
2.5 Sistema nervoso autônomo . . . . .	15
2.5.1 Neuropatia autonômica diabética . . . . .	16
2.6 Índices fisiológicos . . . . .	17
2.6.1 Domínio do tempo . . . . .	18
2.6.2 Domínio da frequência . . . . .	18
2.6.3 Identificação de sistemas . . . . .	19
2.6.3.1 Expansão por bases de Laguerre . . . . .	21

2.6.3.2	Resposta ao impulso . . . . .	22
2.7	Análise de correlação . . . . .	22
2.8	Aprendizado de máquina . . . . .	23
2.8.1	Classificadores . . . . .	23
2.8.2	Avaliação dos modelos . . . . .	24
<b>3</b>	<b>Materiais e métodos</b>	<b>26</b>
3.1	Bases de dados . . . . .	28
3.2	Pré-processamento de sinais . . . . .	29
3.2.1	Conversão dos arquivos da base de dados . . . . .	30
3.2.2	Análise de sinais e separação de janelas de dados . . . . .	31
3.2.3	Criação de objetos do tipo <i>Paciente</i> . . . . .	35
3.2.4	Filtragem de sinais . . . . .	36
3.2.5	Alinhamento, interpolação, reamostragem de sinais e criação de sistemas parametrizados . . . . .	38
3.2.6	Estimação por recursão de parâmetros de sistemas paramétricos por expansão por bases de Laguerre . . . . .	40
3.2.7	Montagem de tabela de dados extraídos . . . . .	41
3.3	Análise de correlação . . . . .	42
3.3.1	Coefficiente de correlação de Pearson . . . . .	42
3.3.2	Coefficiente de correlação de Spearman . . . . .	43
3.3.3	Análise de dependência . . . . .	44
3.4	Análise estatística de atributos via One-way Anova . . . . .	44
3.4.1	One-way Anova on Ranks . . . . .	46
3.5	Classificação por aprendizado de máquina . . . . .	46
3.5.1	Índices individuais . . . . .	47
3.5.2	Agrupamentos de índices . . . . .	48
3.5.3	Avaliação dos modelos . . . . .	48
<b>4</b>	<b>Resultados e discussão</b>	<b>50</b>
4.1	Análise de correlação . . . . .	50

4.2	Análise estatística de atributos via One-way Anova . . . . .	56
4.3	Classificação por aprendizado de máquina . . . . .	59
4.3.1	Índices individuais . . . . .	59
4.3.2	Agrupamentos de índices . . . . .	62
<b>5</b>	<b>Conclusão</b>	<b>67</b>
5.1	Perspectivas futuras . . . . .	70
	<b>Lista de referências</b>	<b>71</b>
	<b>Apêndice A</b>	<b>79</b>
	<b>Apêndice B</b>	<b>84</b>
	<b>Apêndice C</b>	<b>87</b>
	<b>Apêndice D</b>	<b>89</b>

## LISTA DE TABELAS

3.1	Dados de interessa da base de dados CDED 1.0.1 . . . . .	28
3.2	Distribuição de dados desbalanceados após criação de janelas. . . . .	35
3.3	Distribuição de dados por indivíduos. . . . .	35
3.4	Mapa de recursão utilizado para identificação de parâmetros dos sistemas por expansão por bases de Laguerre para decaimento até zero. . . . .	41
4.1	Índices Considerados virtualmente iguais. . . . .	51
4.2	Apresentação das características dos índices utilizando método estatístico. . . . .	57
4.3	Apresentação de índices com características conflitantes entre janelas . . . . .	58
4.4	Ranking de desempenho individual dos índices para a classificação de grupos. . . . .	59
4.5	Desempenho individual dos melhores índices na classificação de grupos. . . . .	60
A.1	Marcadores obtidos durante a etapa de pré-processamento. . . . .	83
B.1	Combinação de índices por grupo para teste de desempenho na classificação de indivíduos portadores de diabete mellitus (IPDM) e indivíduos controle (IC). . . . .	86

## LISTA DE FIGURAS

2.1	Ciclo completo de eletrocardiograma. Adaptado de Silverthorn, 2017 [55].	10
2.2	Regulação do sistema cardiovascular. Adaptado de Silverthorn, 2017 [55].	10
2.3	Comportamento da Pressão Sistêmica. Decaimento da pressão com a distância do coração. Adaptado de Silverthorn, 2017 [55]. . . . .	12
2.4	Ciclo respiratório. Adaptado de Silverthorn, 2017 [55]. . . . .	14
2.5	Inervação dos ramos simpáticos e parassimpáticos do sistema nervoso autônomo e suas atuações sobre diversos órgãos do corpo. Adaptado de Silverthorn, 2017 [55]. . . . .	16
2.6	Diagrama de blocos do modelo de sistema acoplado MISO. Adaptado de Belozeroff, 2003 [7]. . . . .	20
2.7	Funções Bases de Laguerre. Extraído da Toolbox para MATLAB CRSIDLab. [17] . . . . .	21
2.8	Resposta ao impulso do sistema acoplado acoplamento cardiorrespiratório (ACR) e acoplamento do barorreflexo arterial (ABR). Obtido com Toolbox para MATLAB CRSIDLab. [17] . . . . .	22
3.1	Amostra de sinal de eletrocardiograma presente na base de dados CDED.	29
3.2	Amostra de sinal de pressão arterial presente na base de dados CDED. .	29
3.3	Amostra de sinal de fluxo respiratório presente na base de dados CDED.	29
3.4	Fluxo de Pré-processamento de dados. . . . .	30
3.5	Exemplo de batimento ectópico presente na base de dados CDED. . . . .	33
3.6	Exemplo de inconsistência de sinal presente no sinal de eletrocardiograma (ECG) da base de dados CDED. . . . .	33
3.7	Exemplo de inconsistência de sinal presente no sinal de fluxo respiratório (FR) da base de dados CDED. . . . .	33
3.8	Exemplo de recalibração de sinal presente no sinal de pressão arterial (PA) da base de dados CDED. . . . .	34

3.9	Exemplo de recalibração de sinal presente no sinal de FR da base de dados CDED. . . . .	34
3.10	Exemplo de sinais de registro descartado da base de dados CDED. . . . .	36
3.11	Alinhamento dos sinais após filtragem. a) Sinal de intervalos RR (IRR). b) Sinal de pressão arterial sistólica (PAS). c) volume pulmonar instantâneo (VPI). Obtido com Toolbox para MATLAB CRSIDLab. [17] . . . . .	38
3.12	Validação cruzada - <i>K-fold Cross-Validation</i> . . . . .	49
4.1	Grafos formados pelo agrupamento de índices utilizando correlação de Pearson. Janela 1 - Correlação entre índices do grupo diabetes mellitus tipo 2 (T2DM) (acima) e índices do grupo controle (abaixo). . . . .	53
4.2	Grafos formados pelo agrupamento de índices utilizando correlação de Spearman. Janela 1 - Correlação entre índices do grupo T2DM (acima) e índices do grupo controle (abaixo). . . . .	54
4.3	Resultado dos modelos de classificação por grupo de índices. Métricas integrais do desempenho dos agrupamentos de índices - apêndice C. . . .	63
4.4	Modelo com melhor ajuste (53,17%) em dados de validação na estimação do sinal de IRR no sistema de acoplamento mútuo. Polo em 0,9 e tamanho de memória de sistema em 300 amostras. . . . .	64
4.5	Modelo com pior ajuste (6,3%) em dados de validação na estimação do sinal de IRR no sistema de acoplamento mútuo. Polo em 0,78 e tamanho de memória de sistema em 150 amostras. . . . .	65
C.1	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para <i>support vector machines</i> (SVM). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias. . . . .	87
C.2	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para árvore de decisão (AD). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias. . . . .	87

C.3	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para <i>k-nearest neighbor</i> (KNN). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias. . . . .	87
C.4	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para regressão logística (RL). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias. . . . .	88
D.1	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para <i>support vector machines</i> (SVM). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias. . . . .	89
D.2	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para árvore de decisão (AD). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias. . . . .	90
D.3	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para <i>k-nearest neighbor</i> (KNN). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias. . . . .	90
D.4	Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para regressão logística (RL). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias. . . . .	91



## LISTA DE NOMENCLATURAS E ABREVIACÕES

- AAM** algoritmos de aprendizado de máquina
- ABR** acoplamento do barorreflexo arterial
- ACR** acoplamento cardiorrespiratório
- AD** árvore de decisão
- AF** alta-frequência 0,15 - 0,4Hz
- BF** baixa-frequência 0,04 - 0,15Hz
- DM** diabetes mellitus
- DEP** densidade espectral de potência
- ECG** eletrocardiograma
- FR** fluxo respiratório
- FRF** função resposta em frequência
- IC** indivíduos controle
- IPDM** indivíduos portadores de diabete mellitus
- IRR** intervalos RR
- KNN** *k-nearest neighbor*
- MISO** *multiple input single output*
- NA** neuropatia autonômica
- NAC** neuropatia autonômica cardiovascular
- NAD** neuropatia autonômica diabética
- OMS** Organização Mundial de Saúde
- PA** pressão arterial
- PAD** pressão arterial diastólica
- PAS** pressão arterial sistólica
- RI** resposta ao impulso
- RL** regressão logística
- SCV** sistema cardiovascular
- SNA** sistema nervoso autônomo
- SNC** sistema nervoso central
- SNP** sistema nervoso periférico

**SVM** *support vector machines*

**T2DM** diabetes mellitus tipo 2

**VFC** variabilidade da frequência cardíaca

**VPA** variabilidade da pressão arterial

**VPI** volume pulmonar instantâneo

# 1 INTRODUÇÃO

A diabetes mellitus (DM) pode ser definida como um grupo de doenças metabólicas caracterizadas por hiperglicemia - condição de altos níveis de concentração de açúcares no sangue. O estado de hiperglicemia pode ocorrer devido a diversos fatores, como a existência de defeitos na produção de insulina pelo corpo ou resistência a insulina [5].

Segundo a Organização Mundial de Saúde (OMS) em seu relatório técnico sobre diagnóstico e gerenciamento da Diabetes Mellitus Tipo 2, 2020, a diabetes é encontrada em todas as populações do mundo e em todas as regiões conhecidas. A OMS estima que o número de adultos portadores de diabetes no mundo em 2014 fosse de aproximadamente 422 milhões de pessoas [42]. A prevalência de adultos portadores de diabetes subiu significativamente em 2014 em comparação com dados anteriores, passando de aproximadamente 4,7% da população mundial em 1980 para 8,5% em 2014 [42].

O crescente aumento na quantidade de indivíduos diabéticos no mundo torna necessária a detecção precoce da doença, que possibilita a aplicação de manobras terapêuticas mais eficazes na redução da mortalidade e morbidez associada à DM [65]. A detecção tardia da DM leva a atrasos no início do tratamento e, por consequência, pode levar a danos permanentes aos órgãos de seus portadores [28].

Portadores de diabetes têm o dobro do risco de desenvolverem doenças vasculares, como doença cardíaca coronariana e acidente vascular cerebral [16]. A neuropatia autonômica cardiovascular (NAC), uma complicação grave, mas muitas vezes subdiagnosticada da diabetes, aumenta significativamente o risco de mortalidade cardiovascular em indivíduos diabéticos [3, 35]. A redução da variabilidade da frequência cardíaca (VFC) é um sinal precoce da NAC. Neste sentido, este trabalho propõe usar marcadores relacionados à variabilidade da frequência cardíaca como atributos de algoritmos de aprendizagem de máquina para a classificação de sujeitos diabéticos.

Com a popularização de dispositivos vestíveis inteligentes como relógios, anéis, e monitores de saúde portáteis, tornou-se comum a disponibilidade imediata de medições fisiológicas precisas. Combinados aos algoritmos de aprendizado de máquina (AAM), sinais como o eletrocardiograma (ECG) e a pressão arterial (PA) contínua obtidos por esses dispositivos podem ser utilizados para a predição precoce de patologias no dia-a-dia.

Os AAM, apesar de existentes desde a década de 1950, ressurgiram com bastante imponência nas últimas duas décadas. Estes algoritmos se tornaram ferramentas fundamentais na identificação de padrões e tendências em diversos campos de estudos. A medicina é uma das áreas da ciência com grande potencial para se beneficiar de ferramentas de inteligência artificial na caracterização de distúrbios e patologias.

A utilização de aprendizado de máquina nas áreas da saúde concentra-se em diagnósticos e identificação de padrões de patologias médicas. E apesar de poderosas, suas aplicações fundamentam-se na capacidade e na eficiência dos algoritmos em extraírem padrões e tendências de sinais fisiológicos e aspectos biométricos. Essa abordagem tem produzido resultados promissores, alcançando acurácias superiores a 99% para diversos algoritmos [24, 4]. Esta abordagem é conhecida com abordagem cega (*blind approach*) [21].

Neste trabalho será explorada a abordagem *insightful* [21], onde os sinais fisiológicos e aspectos biométricos de indivíduos portadores de diabete mellitus (IPDM) e indivíduos controle (IC) são pré-processados e refinados de forma a produzir índices quantitativos da função autonômica cardíaca. Estes índices são específicos e direcionados à diferenciação de pacientes diabéticos através de disfunções autonômicas presentes nesses indivíduos. Desta forma, índices autonômicos são utilizados como atributos em AAM também de maneira avaliativa, permitindo estimar a capacidade destes quantificadores em predizerem as classes.

A abordagem *insightful* contrapõe-se à abordagem cega, em que não há a preocupação em se utilizar atributos mais específicos. Especialmente na área médica, o uso de atributos relacionados a variáveis mais específicas, correlacionadas à doença investigada, é importante. Desta forma, a aprendizagem de máquina utilizada, ao invés de apenas classificar um indivíduo como diabético ou controle, pode também auxiliar o profissional da área da saúde a entender melhor a progressão da doença, a partir das variáveis avaliadas. Isto pode adicionalmente ajudar a abrir novas avenidas de investigação e a entender melhor a relação entre a diabetes e a evolução de disfunções autonômicas. O que pode futuramente contribuir para o desenvolvimento de métodos de triagem mais precisos para doenças cardiovasculares.

A DM é reconhecida por afetar significativamente a central de regulação cardiovascular através de alterações no controle autonômico. A neuropatia autonômica diabética (NAD) é uma condição associada à diabetes onde ocorre a degeneração progressiva nas fibras dos nervos do sistema nervoso autônomo (SNA) [49], afetando os ramos simpático e parassimpático, ocasionando disfunção na regulação cardiovascular, também chamada de NAC [62]. Alguns estudos indicam alterações nos índices autonômicos mesmo sem detecção clínica de neuropatia autonômica (NA) [65, 47].

A escolha deste trabalho é por averiguar índices quantitativos fisiológicos ligados à atuação do SNA na regulação cardiovascular através da busca por conjuntos de índices mais efetivos na predição do quadro de diabetes mellitus tipo 2 mediante algoritmos de aprendizado de máquinas. Ao investigar meios para distinguir IPDM e IC, encontra-se nas mudanças autonômicas possíveis características que possam ser combinadas para o aperfeiçoamento da classificação de indivíduos dentre os grupos estudados.

Os índices selecionados para descrever a atuação do SNA na regulação cardiovascular focam na caracterização da VFC. Porém, como complemento, a fim de englobar influências da respiração e da pressão arterial, e reduzir perturbações desconhecidas, quantificadores associados a Variabilidade da Pressão Arterial (BPV) e ao ciclo respiratório também são analisados na forma de 3 (três) sistemas acoplados. Destes sistemas foram produzindo índices que abrangem a relação de causa e efeito entre os sinais. Os sistemas acoplados são: a) acoplamento cardiorrespiratório (ACR), b) acoplamento do barorreflexo arterial (ABR) sobre a VFC, e c) Acoplamento da influência mútua do ACR e ABR sobre a VFC.

A escolha dos algoritmos de aprendizado de máquina utilizados neste trabalho concentrou-se em algoritmos fundamentais da literatura de inteligência artificial e que possibilitassem a exploração de características dos índices sem a necessidade de ajustes elaborados dos dados para o algoritmo, dado que a base de dados utilizada é finita e os índices fisiológicos utilizados são específicos para a abordagem do problema estudado. Utilizaram-se 4 (quatro) algoritmos que pudessem efetuar a classificação entre os IPDM e IC a partir de flutuações linear e não-linear, se detectáveis. Os algoritmos utilizados são: a) *support vector machines* (SVM) utilizando kernel *RBF* (radial basis function), b) árvore de decisão (AD) utilizando critério de seleção de atributos *GINI INDEX*, c) *k-nearest neighbor* (KNN) Classificador utilizando ponderação *DISTANCE*, e d) regressão logística (RL) utilizando solver *LIBLINEAR*.

As predições dos AAM de IPDM e IC foram avaliadas utilizando os índices individualmente e em conjuntos, agrupados utilizando critérios como sinais de origem, a partir de índices derivados da VFC ou variabilidade da pressão arterial (VPA), mas também com base nas metodologias de extração utilizada, como índices extraídos por métodos espectrais e de função resposta em frequência (FRF), e identificação de sistemas através de resposta ao impulso (RI) de sistemas acoplados. Variações destes conjuntos de índices foram criados para estimar a interação entre os grupos de índices, todos visando a caracterização das correlação dos sinais complementares com a VFC, produzindo conjuntos que combinam grupos de índices classificados como a VFC e VPA, e VFC e RI.

A análises estatísticas aplicadas, como análise de correlação auxiliam na seleção de índices que indiquem características similares e evite a utilização de índices virtualmente

iguais no mesmo conjunto, produzindo tendência incorretas para a predição desejada. O método estatístico *One-way Anova*, foi utilizado para que fosse possível identificar quais índices, dentre todos os índices extraídos da base de dados, possam indicar diferença estatística para uma melhor predição dos grupos de indivíduos estudados. Dois agrupamentos especiais foram derivados da análise de predição de IPDM e IC com índices individuais, e dos resultados obtidos com o método estatístico *One-way Anova* e análise de correlação, produzindo conjuntos de índices selecionados que complementem características do sinal a fim melhorar o desempenho na predição dos grupos de indivíduos.

Os resultados deste trabalho utilizam dados de indivíduos idosos entre 55 e 75 anos. Esses indivíduos são separados em portadores de diabetes mellitus tipo 2 e indivíduos controle. Não há na documentação da base de dados utilizada indicadores de detecção clínica de neuropatia autonômica diabética ou variações. Assim, os resultados obtidos nestes trabalho são restritas e comparáveis apenas a esta faixa etária e sem relação direta com outras patologias. Foi determinado para esse trabalho que os dados utilizados devem ser apenas índices autonômicos, assim, apesar de disponíveis, dados biométricos como altura, peso e raça (branco ou afro-americanos).

## 1.1 ORGANIZAÇÃO DO TRABALHO

Este trabalho está dividido em 5 (cinco) capítulos incluindo a introdução. A *Introdução* discute sobre o problema de pesquisa, descrevendo-o e propondo a forma como será estudado. Esta seção inclui uma descrição breve sobre as ferramentas e descrição dos objetos de estudo, como quantificadores e sistemas associados a regulação cardiovascular.

A *Fundamentação teórica* introduz o conteúdo teórico para compreensão dos resultados discutidos por esse trabalho. Nele é introduzido informações sobre o funcionamento de sistemas fisiológicos como do sistema cardiovascular, da pressão arterial, e o ciclo respiratório. Elucida sobre os assuntos principais deste estudo como a Diabetes Mellitus e a ação de controle do Sistema Nervoso Autônomo, e descreve melhor as ferramentas utilizadas para a exploração dos dados, como análise de correlação, análise estatística com *One-way Anova/on Ranks*, identificação de sistemas e aprendizagem de máquina,

O capítulo de *Materiais e Métodos* descreve em detalhes os materiais e metodologias utilizadas nesta pesquisa, desde a descrição da base de dados pública utilizada e fluxo de pré-processamento de dados para extração dos índices fisiológicos, até o detalhamento dos métodos de análise dos dados.

Em *Resultados e Discussão* são apresentados os resultados deste trabalho, assim como a discussão técnica e explanatória sobre especificidades da pesquisa e comparações técnicas com informações relacionadas presentes na literatura.

Por fim, no capítulo de *Conclusão* é apresentado um resumo concreto dos resultados e do desfecho da investigação sobre o impacto dos índices na predição dos grupos. Também é discutido sobre perspectivas futuras deste trabalho, limitações e possíveis melhorias.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 DIABETES MELLITUS

A diabetes mellitus (DM) pertence à classificação das síndromes metabólicas. Diversas evidências indicam que as doenças pertencentes a esse grupo são decorrentes de desequilíbrios nos mecanismos do corpo humano resultantes de má alimentação, pouca prática de atividade física, alterações no controle autonômico cardiovascular, a relação glicose-insulina, e a interação entre estes mecanismos [30]. As características da síndrome metabólica incluem regularmente obesidade, hipertensão, dislipidemia, e em casos de diabetes, deficiência ou resistência à insulina.

A DM, apesar de ser descrita em documentos médicos desde o século XVI a.C., é uma doença crônica e não possui cura, porém, desde a descoberta da insulina e da caracterização da doença pela medicina moderna, o distúrbio se tornou tratável e não-letal. A diabetes, apesar de não-letal quando controlada e sob tratamento, ainda impõe uma constante preocupação a seus portadores por proporcionar agravos em quadros clínicos de outras patologias, aumentando o risco geral de morte prematura.

Segundo a Organização Mundial de Saúde (OMS) em seu relatório técnico sobre diagnóstico e gerenciamento da diabetes Tipo 2, 2020, a diabetes é encontrada em todas as populações do mundo e em todas as regiões conhecidas. A OMS estima que o número de adultos portadores de diabetes no mundo em 2014 era de aproximadamente 422 milhões [42]. A prevalência de adultos portadores de diabetes subiu significativamente em 2014 em comparação com dados anteriores, passando de 4,7% da população mundial em 1980 para 8,5% em 2014 [42].

Dois tipos de diabetes prevalecem sobre todas as condições conhecidas da doença, são elas a diabetes tipo 1 e tipo 2. A diabetes mellitus tipo 1, também referida como insulina-dependente ou diabetes-juvenil, não tem causa conhecida, e, portanto, não há formas conhecidas de prevenção. Ela caracteriza-se a partir da degradação das células  $\beta$  do pâncreas, responsável pela produção de insulina. A diabetes mellitus tipo 2 (T2DM) é o tipo mais prevalente da doença em relação a todos os tipos conhecidos, também é referenciada como insulina-não-dependente ou diabetes-adulta. Ela se desenvolve pelo processo de resistência do corpo à insulina presente no sangue. Este processo de resistência



pode ocorrer por diferentes fenômenos, e está altamente relacionada à obesidade.

O aumento contínuo de indivíduos com sobrepeso ou obesidade globalmente é preocupante. Atualmente, acredita-se que entre indivíduos maiores de 18 anos, pelo menos 1 a cada 3 esteja com sobrepeso. Enquanto 1 a cada 10 indivíduos possa ser considerado obeso [41]. Dada a alta correlação entre estes fatores, é de interesse saber também a proporção de indivíduos portadores de Diabetes no mundo. A Organização Mundial de Saúde afirma que não há estimativas globais sobre a prevalência individual para os tipos de diabetes tipo 1 e tipo 2 [41], porém alguns autores indicam que a distribuição seja entre 5-15% de portadores de diabetes mellitus tipo 1 [5, 29], enquanto outros apontam mais de 90% de portadores de diabetes mellitus para o tipo 2 [19, 37].

As complicações da DM tipo 2 são bastante relacionadas a intensidade da doença e a sua duração [19]; além de coincidirem com o aumento de eventos como infarto do miocárdio, doenças vasculares periféricas e acidente vascular cerebral. Mudanças no sistema circulatório provenientes da diabetes, como distúrbios na pressão sanguínea no ciclo circadiano e desaparecimento da regulação do fluxo sanguíneo impactam diretamente estes indivíduos para desenvolvimento de hipertensão. A ocorrência de hipertensão é de 2 (duas) a 3 (três) vezes mais comuns em portadores de DM tipo 2 [19], além intensificar os riscos de outras doenças macrovasculares

### **2.1.1 O contexto da diabetes mellitus no Brasil**

O Brasil ocupa a quarta posição no mundo em quantidade de pessoas portadoras de diabetes, e apresenta uma prevalência aproximada de 8% da população com algum quadro da doença [37]. Diversos fatores populacionais apresentam indícios de alta associação com o risco de complicação por diabetes, como pertencer ao sexo masculino, ter idade avançada e possuir outras comorbidades. Entretanto, o contexto da Diabetes no Brasil também está altamente relacionado com os níveis de desigualdade da população brasileira.

Um estudo transversal utilizando dados da Pesquisa Nacional de Saúde encontrou a partir de uma amostra de 6317 indivíduos, obtida dentre a população brasileira, uma distribuição das principais complicações da diabetes. Esta distribuição engloba problemas individuais e incidências de múltiplas complicações – duas ou mais – pelo mesmo indivíduo [37]. Entre as complicações mais indicadas foram problemas de visão, com 30,6%, e problemas nos rins, com 9,7%. Complicações como problemas na visão e nos rins frequentemente apareceram juntas. Entre os problemas menos frequentes estão derrame e acidente vascular cerebral, com 6,4%, úlceras ou amputações, com 6,0%, e coma diabético, com 2,2% [37]. Em relação à distribuição de incidência de complicação única ou múltiplas, 25,4% indicaram ter apenas um único tipo de complicação, enquanto 12,4% indicaram possuir duas ou mais complicações.

Neves et al., (2023), aponta também que há na população brasileira um aumento da ocorrência de complicações de diabetes conforme há a diminuição da escolaridade e de renda, implicando em uma probabilidade de até 30% mais complicações nos níveis socioeconômicos mais baixos. Uma possível explicação para a causa é o menor acesso aos serviços de saúde pelos indivíduos de menor renda, acarretando uma pior gestão e cuidado no controle da diabetes.

### **2.1.2 O contexto da diabetes mellitus nos Estados Unidos**

Dada a contextualização geográfica da base de dados utilizada para esse trabalho, é útil apresentar a situação atual dos Estados Unidos em relação à Diabetes. Atualmente, estima-se que mais de 15 milhões de estadunidenses adultos sejam portadores de Diabetes Mellitus, isto representa um percentual de aproximadamente 8% da população total do país [26].

Um relatório de 2014 dos centros de prevenção e controle de doenças americano apresentou um resultado promissor relatando que a quantidade de novos casos de Diabetes Mellitus entre adultos estava diminuindo nos Estados Unidos pela primeira vez desde 1990, porém um estudo de Mayer-Davis, 2017, utilizando dados de 2002 a 2012, indicam que as taxas de diabetes tipo 1 e diabetes tipo 2 cresceram entre os mais jovens, principalmente entre as etnias minoritárias [36]. O aumento de casos de detectado para T2DM foi de 4,8% ao ano.

## **2.2 SISTEMA CARDIOVASCULAR**

O funcionamento do coração, é intrinsecamente relacionado com a sua atividade eletrofisiológica. O coração é o principal componente e faz parte do sistema cardiovascular (SCV), um sistema dinâmico, cujas funções incluem bombear sangue, transportar nutrientes, gases, hormônio, calor, eliminar produtos metabólicos entre outros.

O coração fica posicionado na parte central do sistema circulatório, especificamente no mediastino, região de separação entre os pulmões. O sistema circulatório é majoritariamente um sistema em paralelo, onde o lado esquerdo recebe o sangue da circulação pulmonar, oxigenado, e bombeia para a parte sistêmica do corpo. A parte direita do sistema, recebe o sangue da circulação sistêmica e bombeia para a circulação pulmonar. Pareando estes lados da circulação estão os órgãos. Alguns órgãos podem se apresentar na configuração em série, com o fígado e o intestino.

A atividade do coração ocorre através da organização das fibras do coração, que são dispostas em espirais, provendo um movimento de contração. A junção comunicante

destas fibras é um sistema especializado em condução. Deste modo, as células cardíacas podem ser classificadas em duas categorias: elétricas e musculares. As células elétricas produzem potencial de ação e conduzem os sinais elétricos; estas são incapazes de contrair. As células musculares são especializadas em contração e são capazes de conduzir impulsos elétricos.

### **2.2.1 A atividade elétrica cardíaca**

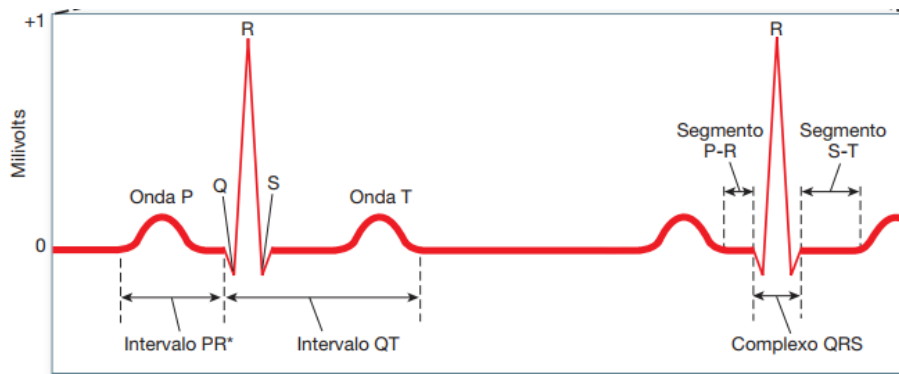
A atividade elétrica do coração é responsável por iniciar a contração do músculo cardíaco. O coração é um músculo cuja função se compara a uma bomba que é ativada de maneira ordenada e sincronizada a fim de circular o sangue pelo sistema circulatório. Essa atividade ordenada e sincronizada ocorre pela contração e relaxamento do músculo cardíaco, e são chamados de sístole e diástole, respectivamente.

A excitação elétrica do coração, normalmente originada no sistema marca-passo e no nó sinoatrial, é de origem não nervosa (não proveniente do sistema nervoso), sendo assim uma atividade elétrica cardíaca espontânea. Apesar de ser uma atividade espontânea e autônoma, ela é modulada em intensidade e frequência pelo sistema nervoso autônomo.

A modulação do sistema nervoso autônomo sobre a atividade cardíaca atua sobre duas condições. A primeira condição é a indução de uma resposta de luta ou fuga, em que o ramo simpático do sistema nervoso aumenta a força de contração das células musculares. A segunda atua diretamente sobre a frequência, onde o ramo parassimpático atua sobre as células marca-passo do coração, tanto no nó sinoatrial quanto no nó atrioventricular, induzindo a regulação hemodinâmica.

O método mais conhecido para se observar e descrever a eletrofisiologia do coração é através das ondas do ECG. Os dois componentes principais do eletrocardiograma são as ondas e os segmentos. Um modo de descrever a diferença entre eles é apresentando as ondas do electrocardiograma como as partes do sinal que se oscilam positivo ou negativamente em relação à linha de referência. Já os segmentos, são trechos da linha de referência se apresentam entre ondas do sinal.

Diferentes tipos de onda presentes no eletrocardiograma representam diferentes atividades ou grupos de atividades do músculo cardíaco. Os tipos de onda que compõem o eletrocardiograma são a onda P, o complexo QRS e por fim a onda T. A onda P representa a despolarização dos átrios, ou a contração do átrio. Em seguida, há um segmento P-R, equivalente à condução da corrente elétrica através do nó atrioventricular. O complexo QRS, equivale a despolarização dos ventrículos. Por último, há a onda T, estágio em que os ventrículos se repolarizam, enchendo-se de sangue e preparando-se para a contração do próximo ciclo. O ciclo completo é apresentado na Figura 2.1.

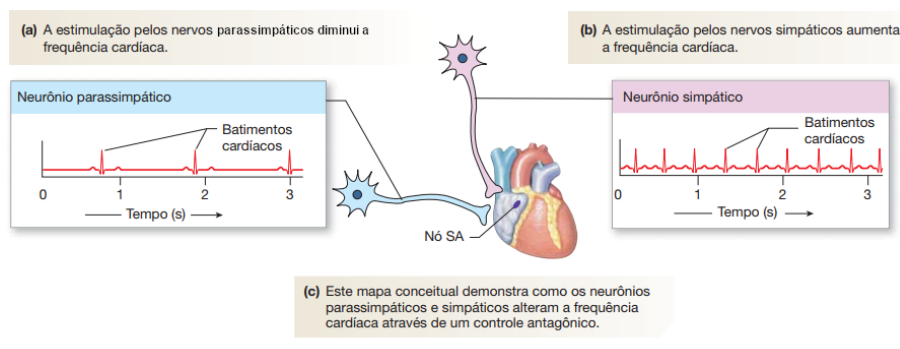


**Figura 2.1.** Ciclo completo de eletrocardiograma. Adaptado de Silverthorn, 2017 [55].

### 2.2.2 Regulação do sistema cardiovascular

A frequência cardíaca é iniciada pela atividade elétrica proveniente do nó sinoatrial, porém sabe-se que esta atividade, apesar de espontânea e autônoma, é modulada por estímulos neurais e hormonais. O sistema nervoso autônomo interfere na atividade cardíaca através dos ramos simpáticos e parassimpáticos - vagal.

As atividades destes dois ramos são normalmente apresentado como uma balança, uma vez que há em suas atuações oposição de um ao outro. A atividade do ramo parassimpático atua para a diminuição da frequência cardíaca, atrasando o potencial de ação das células marca-passo do coração. A atividade do controle simpático atua no aumento da frequência cardíaca, fazendo as células marca-passo dispararem mais potenciais de ação em um menor tempo.



**Figura 2.2.** Regulação do sistema cardiovascular. Adaptado de Silverthorn, 2017 [55].

## 2.3 PRESSÃO ARTERIAL

A pressão sanguínea tem a função de, através de um fluido carregado de nutrientes, gases e hormônios, levar suprimentos aos tecidos do corpo. A pressão sanguínea resulta

da regra do fluxo de fluidos da física, em que diferentes regiões conectadas entre si, com diferentes níveis de pressão, criam um gradiente que conduz matéria de um ponto de maior pressão para o ponto de mais baixa pressão. A regra do fluxo de fluidos é dada como diretamente proporcional ao gradiente de pressão entre dois pontos e inversamente proporcional à resistência do percurso e do próprio fluido [55].

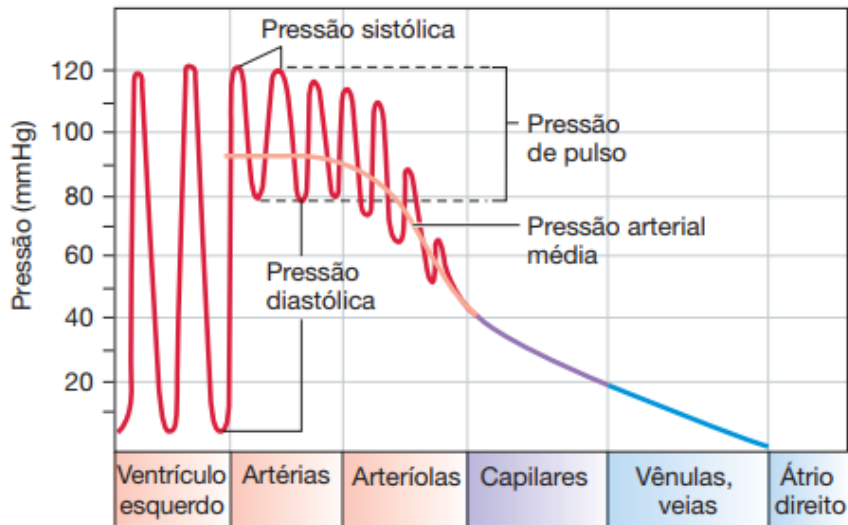
No corpo humano, o coração tem como atividades principais a contração e relaxamento do músculo cardíaco. Ao se contrair se torna uma área de alta pressão, forçando o fluxo de sangue para fora de si. Dentro de um caminho fechado, como é o sistema circulatório humano, o sangue percorre as artérias e os vasos sanguíneos, chegando aos pontos de menor pressão, as veias. As regiões de baixa pressão são próximas a reentrada do coração. Essa redução de pressão entre a saída e a reentrada do coração ocorre devido ao atrito do fluido com o próprio percurso nas artérias, veias e outros tecidos, como as próprias células sanguíneas.

O caminho da circulação fechada não é rígido, e assim como o fluido sanguíneo está sobre pressão, ele exerce pressão sobre o caminho, efeito conhecido como pressão hidrostática. As artérias e veias possuem uma elasticidade própria do tecido das quais são feitas, esta elasticidade é conhecida como resistência periférica. Esta característica que faz com que a aorta, as artérias e as veias se expandam para acomodar o sangue ejetado pelo coração.

Por causa da resistência periférica, as artérias funcionam como um reservatório de pressão, aplicando uma força elástica resultante da pressão hidrostática causada pelo sangue sobre o próprio sangue durante a fase de não contração do coração. Este efeito resulta na existência da pressão arterial média com fluxo sanguíneo constante. O fluxo sanguíneo é contraposto pela resistência do sistema, e é mais comumente expresso na forma de taxa de variação no tempo, como litros ou mililitros por minutos (L/min ou mL/min).

A PA é um trecho da circulação sistêmica, como apresentado na Figura 2.3. A pressão possui um comportamento ondulatório característico da atividade cardíaca, de contração e relaxamento. A PA é medida em milímetros de mercúrio (mmHg) e é útil por poder ser aferida de maneira não invasiva, além de possuir características qualitativas que permitem verificar, de maneira geral, a saúde do indivíduo. As duas características principais da PA, dada sua propriedade oscilatória, são seus pontos de máximo e de mínimo, denominados de pressão arterial sistólica (PAS) e pressão arterial diastólica (PAD), respectivamente.

No corpo humano, os valores referenciados como níveis saudáveis para a pressão arterial de um indivíduo são de 120 mmHg para a média de PAS, ponto de região de alta pressão ao longo da artéria; e a pressão de 80 mmHg para a média de PAD, ponto de menor pressão ao longo da artéria, valor mantido com grande contribuição da resistência



**Figura 2.3.** Comportamento da Pressão Sistêmica. Decaimento da pressão com a distância do coração. Adaptado de Silverthorn, 2017 [55].

periférica.

É importante ressaltar algumas características do comportamento da PA. Uma pressão muito alta ou muito baixa podem indicar problemas no sistema circulatório. Uma pressão muito baixa caracteriza uma hipotensão, caso em que a pressão sanguínea não é capaz de superar a ação de fatores externos, como a gravidade, para seu retorno ao coração; causando assim tonturas ou desmaios. Para o caso de uma pressão muito alta, se caracteriza uma hipertensão, caso em que uma pressão sanguínea alta por longos períodos, como a hipertensão crônica, fragiliza as paredes das artérias e vasos sanguíneos pelo excesso de pressão nos tecidos, possibilitando o surgimento de sangramentos e rupturas.

A pressão arterial, assim como a frequência cardíaca, é regulada pelo SNA. O SNA atua sobre o controle da pressão sanguínea como em um circuito de malha fechada, recebendo *feedback* sensorial dos barorreceptores, localizados nas artérias, que informam o cérebro sobre a necessidade de ajustes na atividade do SNA para manter a homeostase do organismo. Ambos os ramos simpático e parassimpático do SNA atuam nessa regulação. O ramo parassimpático atua sobre o aumento da pressão arterial, liberando neurotransmissores que causam a vasoconstrição dos vasos sanguíneos, que por consequência refletem no aumento da frequência cardíaca. O ramo simpático funciona de maneira similar, liberando neurotransmissores que causam vasodilatação, reduzindo assim, a pressão arterial.

Em indivíduos diabéticos, dois efeitos importantes devem ser considerados. O aumento da atividade simpática do SNA causada pela diabetes, que ocorre pela alta concentração de glicose no sangue ou da degradação do ramo parassimpático em casos clínicos

de NAD, que resulta na vasoconstrição dos vasos e artérias. Há também a contribuição da resistência a insulina, principal causa de T2DM, onde o efeito vasodilatador da insulina é diminuído, contribuindo para o aumento da pressão arterial. A resistência a insulina também é altamente associada à reabsorção de sódio pelos rim em uma conexão direta à hipertensão [56].

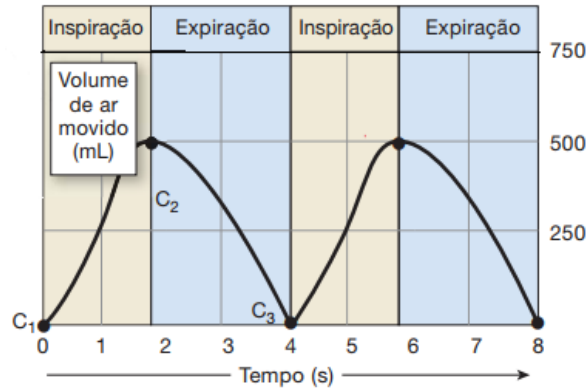
## 2.4 FLUXO RESPIRATÓRIO

A respiração é resultado da ação de músculos internos da cavidade torácica, que funcionam como uma bomba muscular que força um fluido a fluir em um gradiente de pressão, escoando de uma área de maior pressão para uma área de menor pressão; processo similar ao funcionamento da pressão sanguínea. O sistema respiratório humano possui 4 (quatro) funções principais: 1) Troca de gases entre a atmosfera e o sangue, 2) Regulação da acidez (pH) corporal, 3) Proteção contra materiais nocivos ao corpo aspirados por vias aéreas, e 4) Vocalização [55].

O fluxo de ar na respiração obedece à Lei dos Gases da física, porém não é um gás ideal, assim aproximações devem ser feitas para submeter a explicação da respiração às leis dos gases clássica. A descrição do gás ideal para a respiração, assume que dentro do corpo humano a temperatura e o número de moles são constantes, assim chega-se a relação em que o volume do fluido é proporcional ao inverso da pressão. A lei de Boyle reforça a relação de antagônica do volume e da pressão propondo que a compressibilidade do fluido é compensado com aumento da pressão e redução do volume.

Essas leis explicam como o ar é inspirado e exalado dos pulmões. As mudanças na musculatura do tórax causam um gradiente de pressão. Quando o volume do tórax aumenta, a pressão diminui e o ar é inspirado para os pulmões. Quando o volume do tórax diminui, a pressão aumenta e o ar é expirado. Esse mecanismo é importante para explicar a hipótese de interferência da respiração na regulação cardiovascular, não apenas pelo controle autônomo sobre a respiração, que tem fortes interferências do controle voluntário, mas pelo coração estar no mesmo meio que os pulmões, sofrendo variações com a pressão interna do tórax. É conhecido que a frequência cardíaca tende a reduzir quando há expiração, e a aumentar quando há inspiração do ar [12], evento conhecido como arritmia sinusal respiratória. O fluxo respiratório pode ser visualizado na Figura 2.4.

Os volumes pulmonares do ciclo respiratório podem ser descritos em 4 (quatro partes): 1) Volume corrente, 2) volume reserva inspiratório, 3) volume reserva expiratório, e 4) volume residual. O volume corrente é aquele que entra ou sai dos pulmões durante a respiração, com aproximadamente 500 mL. O volume de reserva inspiratório é o volume



**Figura 2.4.** Ciclo respiratório. Adaptado de Silverthorn, 2017 [55].

extra no pulmão, além do volume corrente quando se inspira profundamente. Contém aproximadamente 3000 mL. O volume de reserva expiratório é aquele que pode ser expulso dos pulmões ao fim de um ciclo respiratório regular (volume corrente). Possui aproximadamente 1100 mL. Por último, o volume residual é aquele que preenche os pulmões mas não pode ser expelido, impedindo seu colapso. Este possui aproximadamente 1200 ml [55]. Tais valores são referência para um homem adulto de aproximadamente 70 kg.

O controle da atividade respiratória é complexa, pois envolve uma grande porção de controle voluntário do indivíduos, porém o SNA possui contribuição significativa sobre a respiração. O SNA age tanto sobre a frequência respiratória quanto sobre a intensidade da respiração. Assim, como na pressão arterial, o SNA regula sua atividade simpática e parassimpática a partir de *feedback* recebidos do corpo. No caso da respiração o *feedback* pode ser feito tanto pelo próprio cérebro onde sua base informa ao SNA sobre a concentração de CO<sub>2</sub> e pH no líquido espinhal, quanto por detectores presentes na aorta, que informam sobre a concentração de O<sub>2</sub>, CO<sub>2</sub> e pH no sangue. Assim, o ramo simpático do SNA age sobre o aumento da frequência respiratória, enquanto o ramo parassimpático age sobre a diminuição da frequência respiratória.

Flutuações da respiração sobre a VFC vistas do ponto de vista da frequência são relacionadas a resposta dos barorreceptores. Em indivíduos diabéticos, a modulação da frequência respiratória sobre a VFC tende a se deslocar para frequências mais baixas que frequência original da respiração, indicando uma menor regulação do coração dada a menor variância da VFC. Tal efeito é relacionado ao efeito da diabetes sobre a atividade do ramo simpático, ligado aos barorreceptores [46].

A relação entre o acoplamento cardiorrespiratório é amplamente estudados em casos de apneia do sono. Estudos que relacionam apneia do sono com diabetes sugerem que há relação entre maiores quantidades de quadros de apneia com indivíduos diabéticos com NAD [63]. Outros estudos também sugerem a correlação entre mudanças no padrão



respiratório com NAD [45].

## 2.5 SISTEMA NERVOSO AUTÔNOMO

O sistema nervoso central (SNC) é um dos componentes principais da rede de controle do corpo humano formada pelo cérebro e pela medula espinhal. Dois reflexos de extremos que ocorrem no corpo humano podem ser descritos por uma situação conhecida como "luta ou fuga". Nesta situação o indivíduo está inicialmente em uma situação de equilíbrio e relaxamento, e por uma situação de estresse ou emergência, o indivíduo entra em um estado involuntário de reflexo, como uma luta ou fuga. Esta situação exemplifica o funcionamento do SNC, que coordena este reflexo, e do sistema nervoso periférico (SNP), que realiza a transmissão de informações do SNC para o corpo.

O SNP é constituída por duas partes: 1) sistema nervoso autônomo e 2) sistema nervoso somático. O SNA é atribuído a divisão de controle involuntária do corpo. Estas atribuições entretanto são generalizadas e o controle imputado as partes não são absolutos, podendo haver interferência ou ação de outros agentes, como o reflexo muscular.

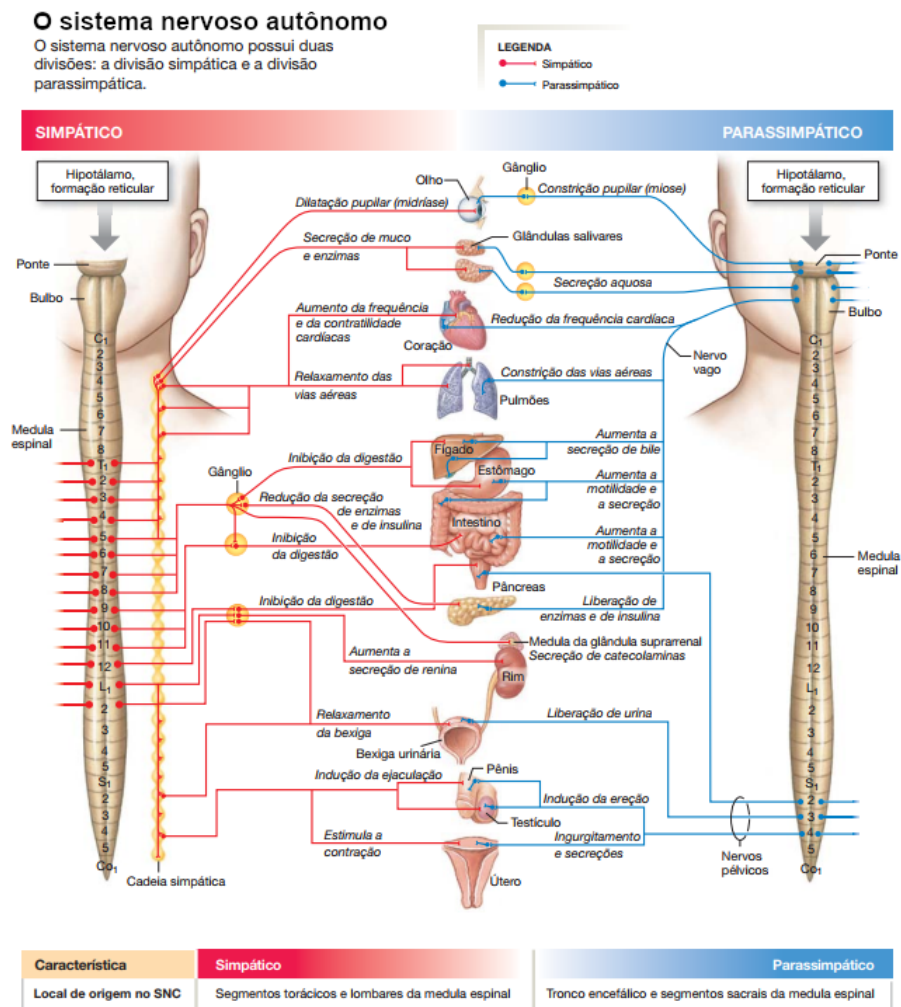
O SNA, também é conhecido como sistema nervoso vegetativo, dada a sua condição de atuação sobre reflexos e ações involuntárias, possui duas subdivisões de interesse para este estudo. São elas: a) o ramo simpático, e b) o ramo parassimpático. Estes ramos são fisicamente diferenciáveis na anatomia, porém suas ações geralmente atuam em conjunto, dificultando a sua identificação individual.

O ramo simpático do SNA é associado à situações intensas e de estresse, como ocorreria na situação de luta ou fuga. Porém, o ramo simpático também atua em situações cotidianas como no controle do fluxo sanguíneo. O ramo parassimpático do SNA é associado a situações de relaxamento e situações cotidianas como digestão. Vale lembrar que o controle intenso ou "individual" de qualquer um dos ramos se aplica apenas a situações extremas, para a maioria das atividades comuns e rotineiras os ramos atuam em conjunto, equilibrando e compensando a ação um do outro. Esta ação conjunta dos ramos simpático e parassimpático atuam com outros sistemas fisiológicos para manter a homeostase do corpo.

Esta ação de compensação entre os ramos do SNA possuem 4 (quatro) propriedades: 1) preservação das condições do meio interno, 2) regulação para cima ou para baixo do controle tônico, 3) controle antagonista, e 4) sinais químicos diversos [55].

A divisão simpática atua sobre a regulação cardíaca aumentando a frequência dos batimentos, enquanto a divisão parassimpática tende a reduzi-la. Esta relação antagônica se estende para diversos órgãos além do coração. Nos pulmões, a ação do ramo simpático dilata os brônquios, aumentando o fluxo respiratório, enquanto a ação do ramo paras-

simpático constrixe os brônquios, reduzindo o fluxo respiratório. As inervações conhecidas do SNA são apresentadas na Figura 2.5.



**Figura 2.5.** Inervação dos ramos simpáticos e parassimpáticos do sistema nervoso autônomo e suas atuações sobre diversos órgãos do corpo. Adaptado de Silverthorn, 2017 [55].

### 2.5.1 Neuropatia autonômica diabética

A neuropatia autonômica (NA) é uma das complicações microvasculares mais comuns em portadores de T2DM afetando aproximadamente 50% dos indivíduos após um período médio de 10 anos com a doença; pode ser estimado que aproximadamente 20% dos indivíduos já possuem sintomas de NA durante o diagnóstico de T2DM [22].

A NA afeta o sistema cardiovascular, gastrointestinal, órgão genitais e causa sudorese. Ele atua sobre os ramos simpático e parassimpático do SNA. Na regulação do SCV, o funcionamento irregular dos batimentos devido à NA é denominado como neuropatia autonômica cardiovascular (NAC), e é definida como o desparelhamento entre o SNA e o SCV. A NAD faz com que fatores de regulação não tenham o efeito correto sobre

a homeostase do organismo. Alguns dos sintomas da NAC são taquicardia em repouso, regulação anormal da pressão sanguínea, intolerância a exercícios e hipotensão ortostática, com fraqueza, tontura e síncope.

A diminuição da VFC é um dos principais indicativos da presença da NA. O diagnóstico para a doença pode ser feito através do padrão ouro CARTs (Cardiovascular Autonomic Reflex Tests). O CARTs inclui a avaliação da média da Variabilidade da Frequência Cardíaca, a resposta da frequência cardíaca à mudança postural para a posição em pé, resposta da frequência cardíaca à manobra de Valsalva, resposta da pressão arterial sistólica à mudança postural para a posição em pé e a resposta da pressão arterial diastólica para exercícios físicos isométricos. Análises da VFC no tempo e na frequência também podem ser aplicados.

A NAC pode ser detectada a partir do primeiro ano de diagnóstico de T2DM. Os principais sinais são a diminuição da VFC durante respiração profunda, diminuição da sensibilidade do barorreflexo, e hipotensão postural (queda súbita de pressão ao se levantar) [62]. É detectado, também, um aumento da atuação do ramo simpático, pois é comum os nervos do ramo parassimpático sofrerem os primeiros danos da doença.

A presença significativa de disfunção autonômica em indivíduos com diabetes, principalmente portadores de longa data da doença, fornece um meio para diferenciação dos grupos de IPDM e IC. Os índices utilizados para este estudo são quantificadores indiretos da atividade do SNA. A quantificação da atividade autonômica serve como indicação do funcionamento dos mecanismos fisiológicos envolvidos da regulação cardíaca, fornecendo embasamento para o flutuações da frequência cardíaca e no intervalos entre batimentos. Estas flutuações podem ser informações sobre o aumento ou diminuição da pressão arterial e/ou como os efeitos do ciclo respiratório se propaga pelo SNA e reflete na atividade do coração.

## 2.6 ÍNDICES FISIOLÓGICOS

Índices fisiológicos são medidas quantitativas que descrevem de alguma forma uma determinada característica de funcionamento do corpo humano. Os índices utilizados neste trabalho refletem o comportamento do sistema nervoso autônomo em IPDM e IC. Os índices da variabilidade da frequência cardíaca são apontados pela literatura como meio de acesso e quantificação da atuação dos ramos simpáticos e parassimpáticos na regulação cardíaca [12]. Além da VFC, obtida através dos dados de intervalos RR (IRR), também é sugerida a possibilidade de acesso à ação de controle autonômico cardiovascular utilizando modelos matemáticos que relacionem mudanças no PAS e na respiração às mudanças na VFC, criando assim um sistema virtualmente acoplado [7].

Para extração de índices fisiológicos que reflitam a atividade autonômica associada à atividade cardiovascular, é de prática comum a remoção de batimentos anormais, como batimentos ectópicos [14]. O aparecimento de batimentos ectópicos frequentemente aparecem antes ou depois da liberação do potencial de ação guiado pelo nó sinoatrial. Esse evento faz com que variações agudas apareçam no sinal de IRR, inserindo, de forma indesejada energia, e componentes de frequência ao sinal [14].

### 2.6.1 Domínio do tempo

Índices obtidos no domínio do tempo são, costumeiramente, valores simples quantificando o comportamento normal de variáveis como IRR, sem que haja transformações dos dados no domínio do tempo [onde foram originalmente extraídos] [44]. Neste trabalho, todos os índices obtidos no domínio do tempo estão relacionados a VFC. Estes valores estão associados a métricas estatísticas como a média do IRR, desvios padrão como o SDNN e o RMSSD, contagens de intervalos de valores, como o PNN50, máximo e mínimo do sinal. A unidade dos índices no tempo é dada em milissegundos. Embora sejam simples os cálculos dos índices no tempo, eles são menos sensíveis na distinção da ação dos ramos simpáticos e parassimpáticos. O apêndice A.1 apresenta os índices coletados para este estudo com uma breve descrição.

### 2.6.2 Domínio da frequência

Os índices no domínio da frequência são calculados através de métodos em frequência ou espectros potência. Índices relacionados a VFC são compreendidos como meios de acesso aos mecanismos de regulação [44]. A análise do espectro de potência da VFC, quando limitados à bandas de frequência como baixa-frequência 0,04 - 0,15Hz (BF) e alta-frequência 0,15 - 0,4Hz (AF) [33, 7], são interpretados como medidas da atuação do SNA sobre a regulação cardiovascular. A banda de alta-frequência 0,15 - 0,4Hz, é associado a contribuição da atividade do ramo parassimpático do SNA, enquanto a banda de BF possui contribuições de ambos os ramos parassimpático e simpático [31].

A obtenção destes índices pode ser feita pelo método de densidade espectral de potência (DEP) de Welch (periodograma janelado) cuja equação é apresentada em 2.1, e seus índices são dados em  $ms^2$ .

$$\widehat{DEP}_{uu}^{Welch}(\nu) = \frac{1}{K \cdot A} \cdot \sum_{i=1}^k \frac{1}{L} \cdot \left| \sum_{\lambda=0}^{L-1} u_{i,w}(\lambda) \cdot e^{-j2\pi \cdot \lambda \cdot \nu} \right|^2 = \frac{1}{K \cdot A} \cdot \sum_{i=1}^k |U_i(\nu)|^2 \quad (2.1)$$

O parâmetro  $\nu$  corresponde aos valores de frequência onde o DEP é aplicado. O parâmetro  $A$  corresponde ao fator para correção de escala causada pela aplicação de janela espectral, neste trabalho foi utilizada janela de Hanning. O parâmetro  $K$  representa o número de periodogramas resultantes da razão entre  $M$  (o número de dados amostrais totais) e  $L$  (o número de amostrais do segmento). O parâmetro  $\lambda$  representa os índices dos dados amostrais dos segmentos, enquanto o parâmetro  $k$  representa os índices dos dados amostrais totais. Por fim,  $u_{i,W}$  representa o segmento de dados após a aplicação da janela espectral. O passo-a-passo do desenvolvimento desta equação pode ser visto em [64].

A partir da análise com DEP, utilizando método de Welch, pode-se obter, além do espectro de potência de BF e AF do IRR, a razão entre as bandas,  $BF \div AF$ . Este índice de razão entre as bandas é compreendido como o balanço entre a modulação do ramo simpático e do ramo parassimpático. Outros índices espectrais, como aqueles obtidos através da VPA, também mensuram a atividade de regulação, em particular a sensibilidade do barorreflexo. Bandas como a BF do PAS e PAD, são compreendidas como representação da modulação simpática da resistência vascular periférica [7].

### 2.6.3 Identificação de sistemas

A identificação de sistemas é uma metodologia de análise que busca explorar [e caracterizar o comportamento de] um sistema desconhecido [66]. Nesta metodologia utiliza-se dados de entrada e saída, ou modelos aproximados para prever o comportamento do sistema. A identificação de sistemas pode assumir 3 (três) abordagens baseadas em diferentes fatores ou modelos de sistemas: 1) modelos *white-box*, 2) modelos *black-box*, e 3) modelos *grey-box*.

Em abordagens utilizando modelos *white-box* os sistemas são construídos a partir de princípios conhecidos do sistema, como leis da física, reações químicas e biológicas. Modelos *black-box* - caixa preta - são construídos exclusivamente a partir de medições de sinais de entrada e saída, estimando-se deles um modelo do sistema. Modelos *grey-box* utilizam de ambas as estratégias dos modelos *white-box* e *black-box*, utilizando todas as informações disponíveis para estimar o modelo real de um sistema. Neste trabalho os sistemas utilizados são estimados utilizando a estratégia *black-box* dada a alta complexidade de modelagem de reações químicas e biológicas.

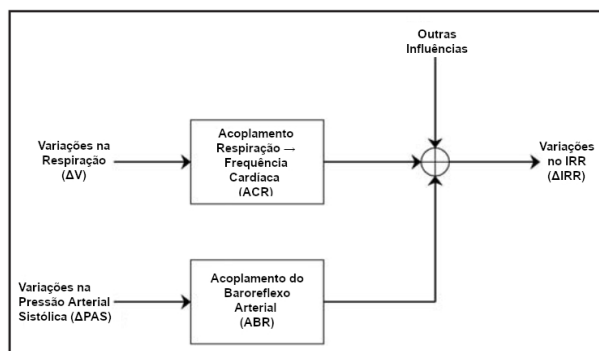
Os sistemas acoplados, termo utilizado neste trabalho para descrever sistemas causais entre dois sistemas fisiológicos, são baseados em modelos matemáticos como uma tentativa de estimar o comportamento de um sistema composto por pares de variáveis fisiológicas. Funcionando como o conceito de caixa preta, tem-se apenas os sinais de entrada e de saída, onde assume-se que haja um efeito de causalidade entre estes sinais e que as flutuações do sinal de saída seja causado ou afetado pelas flutuações do sinal de entrada.

A utilização de identificação de sistemas por RI, busca compensar a limitação presente nas análises utilizando índices apenas no domínio do tempo ou na frequência, que são análises univariáveis. Ao utilizar análises multivariáveis, espera-se avaliar influências dos sinais complementares, como PA e respiração sobre a regulação cardíaca. O modelo matemático para um sistema acoplado possuindo duas entradas, como a PAS e a respiração, e ejetando na saída um sinal de IRR é conhecido na identificação de sistemas como *multiple input single output* (MISO). A Equação 2.2 apresenta um modelo matemático utilizado para este tipo de análise:

$$\Delta IRR(t) = \sum_{\tau=0}^{M-1} h_{acr} \Delta VPI(t - \tau - T_{acr}) + \sum_{\tau=0}^{M-1} h_{abr} \Delta PAS(t - \tau - T_{abr}) + w(t) \quad (2.2)$$

Os parâmetros  $\Delta VPI$  e  $\Delta PAS$  representam as flutuações mediadas respectivamente pela respiração – volume pulmonar instantâneo (VPI) – e pela PAS. O parâmetro  $T_{acr}$  e  $T_{abr}$  representam as latências associadas aos acoplamentos do acoplamento cardiorrespiratório e acoplamento do barorreflexo arterial. O parâmetro  $M$  é o número de amostras de dados inseridos no modelo. Já a função  $w(t)$  representa flutuações ou componentes estocásticas do IRR não contempladas pelo componentes específicos do modelo.

A Equação 2.2 possui dois parâmetros específicos,  $h_{acr}$  e  $h_{abr}$ . Estes parâmetros são as resposta ao impulso características dos respectivos modelos de acoplamento cardiorrespiratório e acoplamento do barorreflexo arterial estimados a partir dos dados mensurados. O modelo utilizado, Figura 2.6, é considerado linear, logo para um sistema MISO encontra-se uma RI para cada entrada, conforme a definição de linearidade de independência dos sinais de entrada.



**Figura 2.6.** Diagrama de blocos do modelo de sistema acoplado MISO. Adaptado de Belozeroff, 2003 [7].

### 2.6.3.1 Expansão por bases de Laguerre

A expansão do modelo matemático da Equação 2.2 utiliza o método de estimação de modelo de resposta ao impulso por expansão por bases de Laguerre. A utilização das bases de Laguerre visa reduzir no número de parâmetros a serem estimados, as bases são usadas ponderadas, e aproveita-se de seu decaimento suave até zero para replicar o comportamento de uma resposta ao impulso. Os modelos de resposta ao impulso estimadas por bases de Laguerre estão descritas nas Equações 2.3 e 2.4, e 2.5 e 2.6:

$$h_{acr}(t) = \sum_{j=0}^{q_{acr}-1} c_j^{acr} L_j(t) \quad (2.3)$$

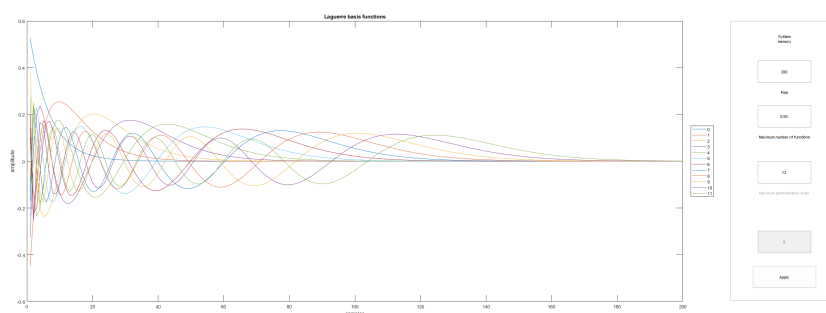
$$h_{abr}(t) = \sum_{j=0}^{q_{abr}-1} c_j^{abr} L_j(t). \quad (2.4)$$

Os parâmetros a serem estimados são as ponderação das bases ortogonais de Laguerre, representadas como  $c_j^{sistema}$ , enquanto as bases de Laguerre são apresentadas como  $L_j(t)$ . As Equações 2.5 e 2.6, apresentam como calcular as bases utilizadas.

$$L_0(t) = \sqrt{\alpha^t(1 - \alpha)} \quad (2.5)$$

$$L_i(t) = \sqrt{\alpha}L_i(t - 1) + \sqrt{\alpha}L_{i-1}(t - 1) - l_i, \quad 0 \leq j \leq q_{acr}, q_{abr} \quad (2.6)$$

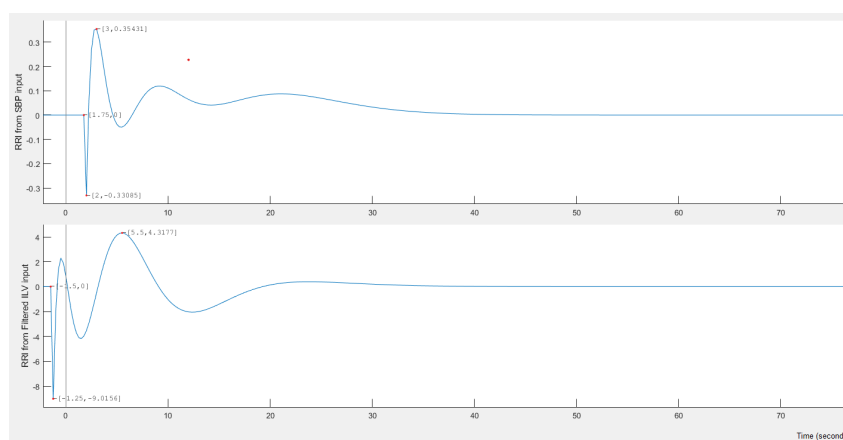
Nas bases de Laguerre,  $\alpha$  representa a taxa exponencial de decaimento, como apresentado na Figura 2.7.



**Figura 2.7.** Funções Bases de Laguerre. Extraído da Toolbox para MATLAB CRSIDLab. [17]

### 2.6.3.2 Resposta ao impulso

O conceito de sistemas acoplados é utilizado aqui exclusivamente para extrair índices fisiológicos dos modelos estimados através da identificação de sistemas utilizando resposta ao impulso (RI). O modelo, aqui chamado de sistema acoplado, é um modelo matemático que assume que as flutuações presentes no sinal de IRR que representam a VFC possuem dois componentes em sua composição: 1) flutuações decorrentes da respiração, resultado do acoplamento cardiorrespiratório, e 2) flutuações decorrentes da pressão, resultando do acoplamento do barorreflexo arterial [7]. A RI sofre uma transformação ao passar pelo sistema acoplado, apresentando uma resposta (Figura 2.8) que pode ser analisada no tempo e na frequência.



**Figura 2.8.** Resposta ao impulso do sistema acoplado ACR e ABR. Obtido com Toolbox para MATLAB CRSIDLab. [17]

Os índices obtidos através do método de identificação são a amplitude da resposta ao impulso, medindo a diferença entre o seu máximo e o seu mínimo. Os ganhos dinâmicos da resposta também são obtidos, o ganho dinâmico da banda total entre 0,04Hz - 0,4Hz, os ganhos nos valores de baixa-frequência 0,04 - 0,15Hz e alta-frequência 0,15 - 0,4Hz. Para os índices de ganho dinâmico, é feita a transformada de Fourier da resposta ao impulso estimada, após a obtenção da função de transferência da resposta, é feita a média dos ganhos desta função nas bandas de frequência [7].

## 2.7 ANÁLISE DE CORRELAÇÃO

A análise de correlação é utilizada em estudos para se compreender a relação entre duas variáveis aleatórias e a sua interdependência. A correlação não é capaz de indicar qual das variáveis analisadas possui dependência ou associação e qual é a autônoma, ou se ambas são codependentes. O coeficiente de correlação resultante da análise indica, em um intervalo de -1 a 1, a força do relacionamento das variáveis. Coeficientes negativos



indicam uma expectativa estatística de que o valor de uma das variáveis diminua quando o valor da outra aumenta, enquanto valores positivos indicam a expectativa estatística de que o valor de uma das variáveis aumente quando o valor da outra também aumenta.

O grau de correlação entre os índices estudados fornece uma base de identificação de similaridade entre os índices, indicando a possibilidade de representarem informações equivalentes. As equações para cálculo dos coeficientes de Pearson e Spearman pode ser vistos nas Equações 3.1 e 3.4.

## 2.8 APRENDIZADO DE MÁQUINA

### 2.8.1 Classificadores

Classificadores são modelos matemáticos implementados em computador que visam obter previsões de um determinado problema com base em ocorrências passadas [34]. Diversas tentativas de aplicação de algoritmos de aprendizado de máquina em estudos sobre diabetes surgiram nos últimos anos almejando auxiliar profissionais da saúde a tomarem medidas rápidas e seguras em relação a detecção da doença.

Diferentes tipos de algoritmos são utilizados almejando melhorias ou novas abordagens na detecção de padrões fisiológicos dos indivíduos. Desde os algoritmos mais estabelecidos e normalmente mais utilizados, até algoritmos complexos com diversas etapas e níveis de cálculos. Também é de interesse avaliar a necessidade do algoritmo em receber dados pré-processados ou dados “crus”, restringindo os padrões a serem detectados ou generalizando-os. Neste estudo os índices foram submetidos a pré-processamentos e os atributos foram gerados a partir de dados filtrados, reamostrados e transformados por diferentes métodos matemáticos.

Segundo [40], em uma tabela disponibilizada com uma breve revisão bibliográfica, diferentes níveis de pré-processamento foram aplicados em diversos estudos para seleção de atributos. Os métodos de pré-processamento e seleção de atributos mais comuns foram a remoção/substituição de valores faltantes, análises de correlação ou extração de índices. Porém, métodos mais rebuscados também foram aplicados, como algoritmo de *Random Forest* e análise de componentes. Dentre os algoritmos mais utilizados estão regressão logística, *support vector machines*, *k-nearest neighbor*, árvore de decisão e *Multi Layer Perceptrons* (Redes Neurais).

## 2.8.2 Avaliação dos modelos

Os critérios de avaliação dos modelos neste trabalho são calculados a partir da combinação de quatro métricas possíveis. Destas métricas é possível montar uma tabela de desempenho conhecida como matriz de confusão, contendo um resumo crítico do comportamento apresentado pelo classificador [34].

1. Verdadeiro Positivo (VP) - Quando o resultado do algoritmo indica que a classe é positiva, e ao validar-se com a resposta correta, pré-identificada, a classe é realmente positiva.
2. Verdadeiro Negativo (VN) - Quando o resultado do algoritmo indica que a classe é negativa, e ao validar-se com a resposta correta, pré-identificada, a classe é realmente negativa.
3. Falso Positivo (FP) - Quando o resultado do algoritmo indica que a classe é positiva, e ao validar-se com a resposta correta, pré-identificada, a classe é negativa. Erroneamente classificada como positiva.
4. Falso Negativo (FN) - Quando o resultado do algoritmo indica que a classe é negativa, e ao validar-se com a resposta correta, pré-identificada, a classe é positiva. Erroneamente classificada como negativa.

Ao se treinar algoritmos de aprendizado de máquina, estamos modificando parâmetros internos de modelo matemático ou lógico a fim de ajustar-se aos dados para identificar características presente nestes dados. Isto pode ser feito de diversas maneiras, porém a problemática onde se aplica estes algoritmos devem ser estudadas para que os resultados emitidos pelo algoritmo não causem um problema na vida real.

Por exemplo, a predição de um indivíduo no qual pretende-se prever ou diagnosticar a disfunção Diabete Mellitus. Não basta que dentre 10 (dez) indivíduos acerte-se 9 (nove). Um problema que pode apresentar-se na vida real é a tendência de erro do algoritmo se para prever que indivíduos que realmente possuam Diabetes Mellitus, sejam classificadas como não portadoras. Este erro pode levar a um quadro inesperado de saúde do indivíduo por acreditar-se estar saudável. Para que este tipo de situação seja controlada, utiliza-se dos possíveis resultados apresentados na lista de possíveis resultados para observações de aprendizado de máquina, Lista 1 da Seção 2.8.2, para cria-se métricas do algoritmo. Para esse trabalho, 4 (quatro) métricas de desempenho serão utilizadas, como descrito a seguir:

1. Acurácia (A):  $\left(\frac{VP+VN}{VP+VN+FP+FN}\right)$

2. Precisão (P):  $(\frac{VP}{VP+FP})$
3. Sensibilidade (S):  $(\frac{VP}{VP+FN})$
4. Medida F1 (F1):  $(2 \times \frac{P \times S}{P+S})$

A acurácia é uma métrica que considera o número de acertos dentre o total de predições feitas pelo algoritmo. Esta métrica é a mais utilizada para avaliar o desempenho de modelos treinados, almejando-se sempre obter o melhor desempenho possível. A unidade da acurácia é a porcentagem, variando entre 0% (nenhuma predição do algoritmo coincide com a classe real da observação) e 100% (a predição de todas as observações feitas pelo algoritmo coincide com a classe real das observações).

A precisão avalia a capacidade do algoritmo em acertar a predição de uma classe como positiva. A precisão é importante para evidenciar o desempenho do modelo treinado em bases desbalanceadas. Como citado no exemplo da diabetes, é indesejado que um indivíduo que possua diabetes, uma classe minoritária dentre a população, seja classificado com não possuindo a doença. Para este problema, uma alta precisão é a métrica desejável de desempenho, além da acurácia, evitando que entenda-se que o modelo treinado possua alto desempenho ao acertar apenas observações da classe prevalente (IC) em bases de dados desbalanceadas. Ao utilizar a precisão, é possível reduzir as chances de aceitar-se um modelo que identifique como saudável um indivíduo da classe positiva, neste caso um indivíduo diabético, baseado apenas na acurácia.

A sensibilidade ou *recall*, é uma métrica complementar à precisão, ela avalia a sensibilidade do modelo treinado em distinguir a classe positiva, IPDM, dentre todas as observações. Nela, verifica-se se o modelo a proporção de acertos da classe positiva em relação ao total real de observações da classe positiva. Diferentemente da precisão que avalia o total de acertos da classe positiva em relação ao total de predições do modelos para a classe positiva (independente da classe correta da observação), a sensibilidade é utilizada para evidenciar o grau de confusão do modelo treinado em relação as classes presentes nos dados. Ao utilizar a sensibilidade, é possível identificar a capacidade do modelo treinado em identificar observações da classe positiva dentre as observações existentes da classe positiva. Isto é, sensibilidade para identificar uma classe.

A medida F1 pode ser considerada como um resumo da matriz de confusão e de desempenho do modelo treinado, com exceção da acurácia. A medida F1 resulta de uma média harmônica entre os valores de precisão e de sensibilidade do modelo. Por ser uma média harmônica, a medida prioriza o menor valor entre precisão e sensibilidade. Portanto, uma medida F1 alta representa um modelo equilibrado e com alta precisão e alta sensibilidade.

### 3 MATERIAIS E MÉTODOS

Os elementos desta pesquisa visam analisar como índices fisiológicos relacionados à regulação cardíaca podem se complementar na composição de conjuntos de atributos a fim de aperfeiçoar a distinção entre indivíduos portadores de diabetes mellitus e indivíduos controle idosos.

É de interesse deste trabalho apresentar propostas de agrupamento de índices fisiológicos para a classificação de indivíduos diabéticos e de controle utilizando algoritmos de inteligência artificial. Os índices utilizados são derivados de sinais fisiológicos obtidos por métodos não invasivos. São eles: eletrocardiograma, pressão arterial contínua e fluxo respiratório. Os agrupamentos destes índices ocorrem com base em sua origem de sinal e em sua metodologia de extração.

As origens dos índices podem ser os sinais de VFC ou VPA. Enquanto grupos formados a partir do metodologia de extração visam agrupar índices obtidos a partir de técnicas como DEP, possibilitando avaliar índices de modelos univariáveis, e identificação de sistemas utilizando RI, produzindo índices multivariáveis, e permitindo caracterizar sistemas acoplados, que representam a atividade autonômica que relaciona pares de sinais fisiológicos que estejam sob controle do SNA. Os índices agrupados são considerados indicadores da ação do controle autonômico na regulação cardiovascular, e seus agrupamentos buscam formar conjuntos de informações complementares da atividade autonômica.

Os índices utilizados são resultados de modelos univariáveis e multivariáveis da fisiologia humana. Os índices obtidos de modelos univariáveis, como índices da VFC e da VPA, no domínio do tempo e da frequência, fornecem informações da atividade autonômica associada ao sistema fisiológico de origem do sinal. Índices relacionados à VFC, caracterizam a atividade autonômica sobre a atividade cardíaca. Em índices relacionados à VPA, caracterizam a atividade autonômica sobre as características da PA, como a resistência periférica vascular.

A atividade cardíaca possui uma atividade basal regulada pelas próprias células do coração, sem interferência do SNC. Assim a atividade do SNA associada à VFC é interpretada como uma ação modulatória da atividade cardíaca, onde a atividade basal do coração é alterada de acordo com as necessidades do corpo, a fim de atingir a ho-

meostase do organismo. Para a VPA, a atividade do SNA atua sobre a liberação de neurotransmissores que atuam sobre a resistência vascular.

Os índices que quantificam a atividade do SNA extraídas por modelos univariáveis, não são capazes de determinar como a atividade autonômica obtém as informações que determinam como ela deve modular a regulação cardíaca, quando aumentar ou diminuir a frequência do coração ou alterar a força dos batimentos. A mesma explicação se aplica à interpretação dos índices que quantificam a atividade autonômica em relação a VPA. Esses índices explicam como a atividade autonômica modifica o comportamento elástico do sistema vascular, através de liberação de hormônios, mas não determinam como essa informação é obtida pelo SNA.

Ao utilizar índices produzidos a partir de modelos multivariáveis, informações complementares de outros sistemas fisiológicos reduzem a limitação de causalidade existente nos índices de modelos univariáveis sobre a natureza da atividade autonômica. Assim, a utilização de sinais complementares para a criação de sistemas onde os dados de entrada e saídas são acoplados, permitem que, através do SNA, avalie-se a influência da pressão arterial e da respiração sobre a regulação cardíaca. Em indivíduos diabéticos, a presença de disfunções autonômicas permite que a diferença entre os índices dos grupos de IPDM e IC seja usada para a diferenciação dos grupos de indivíduos.

A utilização de pré-processamento de sinais previamente ao treinamento de AAM apresenta duas vantagens. Ao extrair índices diretamente associados ao SNA, aumenta-se a confiabilidade das informações utilizadas como atributos para o aprendizado de máquina, em maior ou menor grau. Essa abordagem direcionada de seleção de atributos possibilita a utilização de algoritmos de inteligência artificial sem modificações complexas, em contraponto a algoritmos em que se utilizam dados “crus” na tentativa de extrair padrões de informações imprecisas. A segunda vantagem aborda a utilização de bases de dados pequenas, frequentemente encontradas na área de engenharia biomédica e dados de saúde, já que se utiliza de índices direcionados, em uma abordagem conhecida como “*insightful approach*” [21].

Os índices fisiológicos extraídos do pré-processamento foram utilizados como atributos individuais para o treinamento dos AAM a fim de identificar quais índices apresentavam melhor acurácia, sensibilidade e precisão ao distinguir entre os grupos de indivíduos estudados. Os índices também foram agrupados com base em sua origem de sinal e na metodologia de extração utilizada. Os grupos criados estão descritos no Apêndice B.1.

Índices da VFC, no domínio do tempo e no domínio da frequência, foram reunidos em um único conjunto. Assim como índices da VPA que também foram reunidos em um conjunto. Diferentes conjuntos também foram criados reunindo índices relacionados a FRF dos dados de VFC, VPA e de modelos multivariáveis. Índices extraídos através

de identificação de sistemas por RI também foram agrupados. Além de conjuntos combinando agrupamentos existentes, como conjuntos que reúnem índices dos agrupamentos da VFC e da VPA, e VFC e RI. Conjuntos chamados especiais, de índices selecionados a partir de análises estatísticas e da classificação utilizando índices individualmente, foram propostos. Os dados necessários para a análise de classificação e que devem ser obtidos na etapa de pré-processamento de dados estão descritos no apêndice A.1.

### 3.1 BASES DE DADOS

Os dados utilizados para a geração dos índices quantitativos do sistema nervoso autônomo, necessários para treinamento de AAM, foram obtidos a partir da base de dados pública *Cerebrovascular Disease in Elders with Diabetes*, CDED, disponível em <https://physionet.org/content/cded/1.0.1/> [39, 38]. A base de dados CDED contém sinais contínuos de eletrocardiograma, pressão arterial e fluxo respiratório sob diferentes atividades, como testes de hipocapnia, hipercapnia, manobra de Valsalva, exame de inclinação postural passiva e testes de regulação cardiovascular ao sentar-se e levantar-se, além de trechos de sinais dos indivíduos em repouso. A CDED foi originalmente criada para avaliar o impacto dos efeitos da diabetes na vasorregulação cerebral e fez o acompanhamento de 69 indivíduos com idades entre 55 e 75 anos. Dos 69 indivíduos, apenas 41 completaram o acompanhamento de 2 anos do estudo.

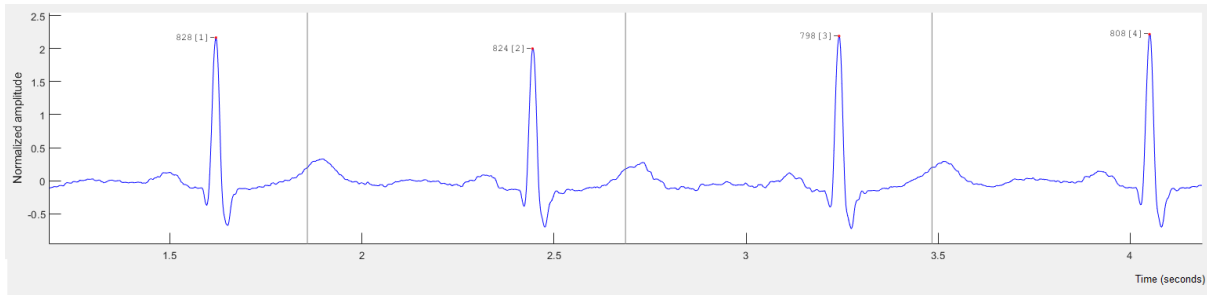
Sinal	Descrição
Marker	Indicador de trocas de etapas para as atividades
ECG	Dados de eletrocardiograma
ABP	Dados contínuos de pressão arterial
FLOW_RATE	Dados contínuos da taxa de fluxo respiratório

**Tabela 3.1.** Dados de interessa da base de dados CDED 1.0.1

Dentre os sinais presentes na base de dados CDED, capítulo 3.1, o interesse deste estudo está centrado em 3 (três) sinais: **a) ECG**, **b) ABP**, e **c) FLOW\_RATE**, Figura 3.3, conforme descrito na tabela 3.1.

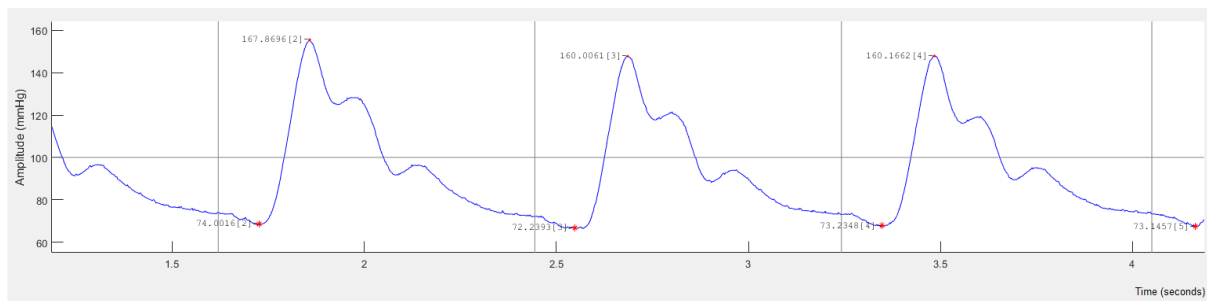
A Figura 3.1 apresenta uma amostra de dados de ECG utilizada para extração de índices quantitativos do SNA. No trecho apresentado é possível ver com clareza todas as ondas e segmentos que compõem o eletrocardiograma, assim como sua amplitude em unidades normalizadas. Também é indicado, os valores de IRR, indicados pelos valores numéricos no pico das ondas R.

A Figura 3.2 apresenta uma amostra da forma de onda da PA contínua. Ambas as Figuras 3.1 e 3.2 estão no mesmo intervalo de tempo, onde é possível verificar a correlação entre o pico de contração da atividade cardíaca e a resposta da pressão alguns instantes



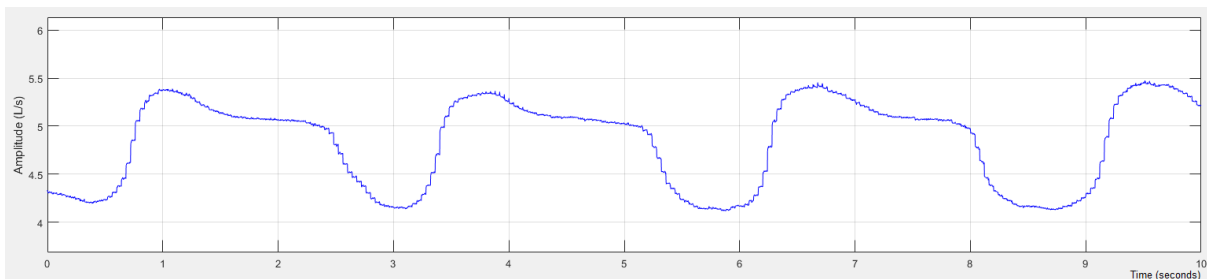
**Figura 3.1.** Amostra de sinal de eletrocardiograma presente na base de dados CDED.

depois. A imagem também aponta para os dados característicos de PAS, com um valor médio 160 mmHg, e a PAD, com um valor médio de 73 mmHg.



**Figura 3.2.** Amostra de sinal de pressão arterial presente na base de dados CDED.

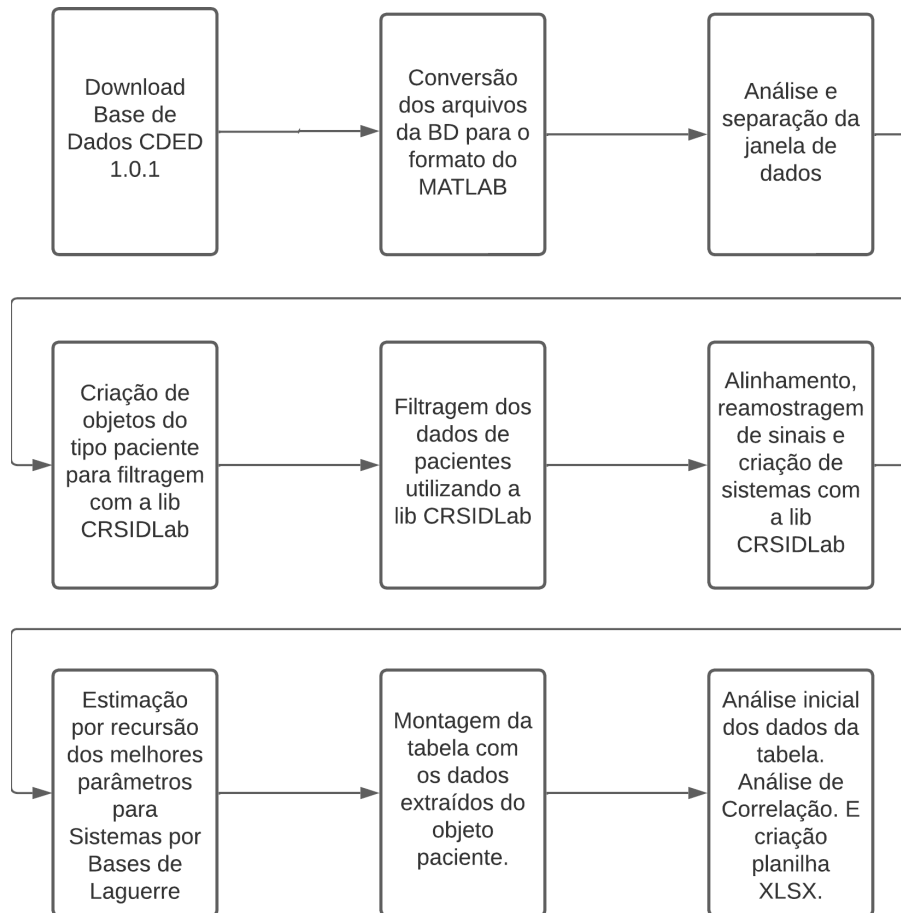
A Figura 3.3 apresenta uma amostra do sinal de fluxo respiratório (FR) e apresenta ciclos maiores de sinal em relação ao ECG e à PA, contando com aproximadamente 3 segundos.



**Figura 3.3.** Amostra de sinal de fluxo respiratório presente na base de dados CDED.

## 3.2 PRÉ-PROCESSAMENTO DE SINAIS

O panorama geral do fluxo de pré-processamento de dados e extração de índices fisiológicos quantitativos do SNA pode ser visto na Figura 3.4, com a descrição detalhada de cada etapa nos subcapítulos subsequentes.



**Figura 3.4.** Fluxo de Pré-processamento de dados.

### 3.2.1 Conversão dos arquivos da base de dados

A base de dados CDED é disponibilizada no formato de diretórios, separando dados, descrição dos dados e artigos relacionados. Também está presente o arquivo com a descrição do protocolo seguido pela pesquisa original. Os dados dos pacientes estão separados conforme o protocolo entre ECG (visita de avaliação), Labview-v2 (visita de acompanhamento ano 1) e Labview-v8 (visita de acompanhamento ano 2). Para este trabalho os dados escolhidos para análise estão presentes em Labview-v2 durante a avaliação de hiperventilação e reabsorção de  $CO_2$ , onde se utilizará apenas o trecho de sinais onde o indivíduo é mantido por 10 minutos na posição supina em repouso, sendo monitorado sem qualquer outra intervenção.

Os dados dos pacientes são apresentados como um par de arquivos com formatos “.dat” e “.hea”. Onde os arquivos “.dat” possuem os dados dos pacientes codificados e os arquivos “.hea” possuem os cabeçalho de estrutura e codificação dos sinais dos pacientes. Scripts para MATLAB foram criados para todas as etapas de pré-processamento, e os arquivos de pacientes foram submetidos à função *wfdb2mat* da toolbox WFDB Application [51, 52] para MATLAB, responsável pela conversão dos arquivos.



Após a etapa de conversão, as informações dos indivíduos então convertidas para unidades físicas características de cada sinal, como mV, mmHg e L/s. Estes valores estão disponíveis junto ao cabeçalho de indivíduo. Por fim, os dados são salvos no formato para MATLAB “.mat” contendo o vetor de tempo dos sinais do indivíduo, uma matriz com os sinais do indivíduo, um vetor de células contendo os rótulos dos sinais e por fim a frequência em que os dados foram amostrados.

### 3.2.2 Análise de sinais e separação de janelas de dados

Para essa etapa da pesquisa é introduzido o conceito de janela de sinais. Uma janela de sinais é definida como um intervalo de dados contendo múltiplos sinais fisiológicos capturados simultaneamente. Os sinais utilizados simultaneamente são o ECG, a PA e o FR.

Segundo as recomendações da Força Tarefa da Sociedade Europeia de Cardiologia e a Sociedade Norte-americana de Eletrofisiologia, que estabelecem padrões de procedimentos para análise e interpretação de dados de fisiologia cardíaca, recomenda-se para a análise da variabilidade da frequência cardíaca que o sinal tenha pelo menos 10 vezes a duração do menor componente de frequência presente no sinal [1].

A definição das bandas de frequência associadas às atividades simpáticas (ambas as bandas de AF e BF) e parassimpáticas (banda de AF) do SNA resultam de estudos que empregam metodologias de reação do SNA à perturbações agudas, além de estudos farmacológicos com bloqueadores de receptores autonômicos.

Strauss, H. (2007) elaborou uma revisão dos mecanismos fisiológicos que influenciam a VFC em indivíduos saudáveis e identificou que essas alterações ocorrem da inervação eferente do SNA no nó sinoatrial e de atrasos entre a ação autonômica na modulação da atividade cardíaca.

Os atrasos nos mecanismos de propagação da ação autonômica ocorrem devido a mecanismos celulares, tornando a ação da modulação parassimpática sobre o coração mais rápida que a ação da modulação simpática. Flutuações rápidas da VFC, na mesma banda de frequência que a respiração, são mediadas pelo ramo parassimpático. Enquanto flutuações mais lentas da VFC são moduladas por ambos os ramos. Assim, a banda de AF se faz relacionada a atividade do ramo parassimpático, enquanto BF, mais especificamente a razão entre BF e AF, são compreendidos como índices do equilíbrio entre os ramos simpático e parassimpático [57].

Deste modo, para análise de ganhos na banda de alta-frequência 0,15 - 0,4Hz (AF), seria necessário 1 minuto de sinal, enquanto para ganhos na banda de baixa-frequência 0,04 - 0,15Hz (BF), 2 minutos seriam suficientes. Porém, o padrão estabelecido pela

Força Tarefa é um máximo de 5 minutos, de forma a respeitar a característica de estacionariedade do sinal.

Foi escolhido um intervalo de 4 minutos para a definição de janelas de dados desse estudo. Esse tamanho de intervalo de dados também é apontado por Cavalcanti (2022) como acima do tamanho mínimo, de 3 minutos de segmento de sinais, para distinção entre grupos de IPDM e IC [9]; e está dentro dos parâmetros sugeridos pela Força Tarefa.

Um segmento de dados de 10 minutos dos pacientes em repouso na posição supina foi escolhida como fonte para criação das janelas de sinais. O segmento apresenta eventos naturais de regulação como batimentos ectópicos dos indivíduos, mas também eventos artificiais como falhas nas capturas dos sinais ou recalibrações automáticas do equipamento, causando perda de informações em trechos de sinais, tornando necessária de uma seleção minuciosa das janelas dentro deste segmento.

A escolha da posição das janelas de sinais dentro do segmento de 10 minutos utilizado é importante para se evitar a maior quantidade de imperfeições possíveis, pois a quantificação da atividade autonômica através destes dados pode ser afetados por suas características. A exemplo de batimentos ectópicos, apesar de serem eventos naturais, esses batimentos podem levar a falsas interferências de sinais de frequência no sinal de IRR derivado do ECG. Assim, tanto eventos naturais e artificiais precisam ser removidos e tratados antes da extração dos índices desejados.

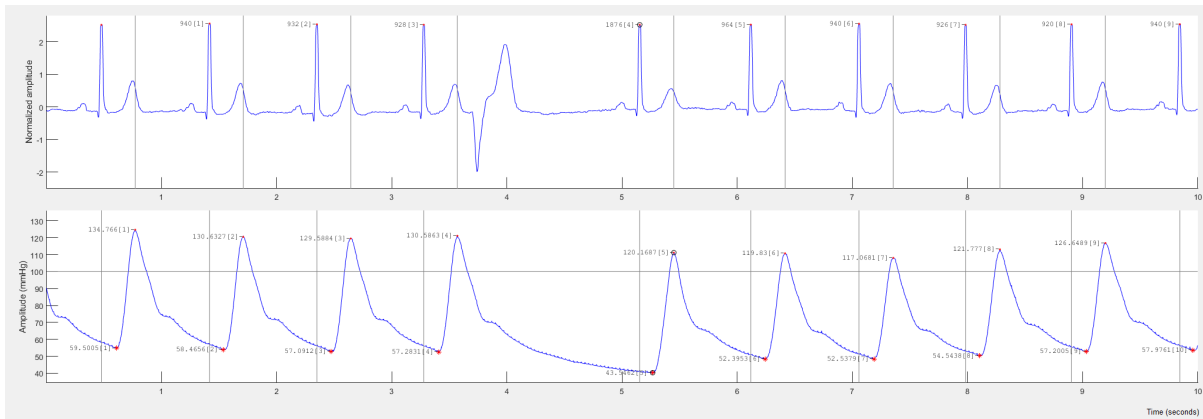
A posição dos segmentos para as janelas de sinais de cada indivíduos foi escolhida após verificação manual de todos os sinais, a marcação de batimentos ectópicos no sinal de ECG (Figura 3.5), a marcação das inconsistências de sinais (Figuras 3.6 e 3.7) e recalibrações (Figuras 3.8 e 3.9).

Batimentos ectópicos precisam ser removidos e interpolados para a obtenção do sinal de IRR que reflita a atividade do SNA sobre o coração. Inconsistências também afetam os sinais estudado, porém em casos como a inconsistência do sinal de ECG podem ser aceitos quando possível a identificação do pico da onda R, como apresentado no exemplo da Figura 3.6. Em outros casos, como a inconsistência da Figura 3.7, toma-se cuidado para que não afete múltiplos ciclos.

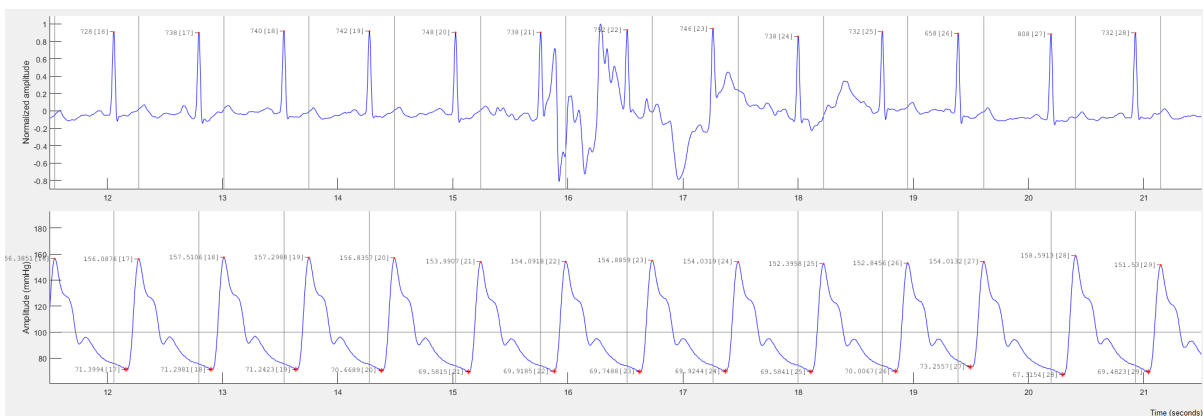
Recalibrações de sinais não devem estar presentes no sinal após o pré-processamento. Nos pares de sinais ECG e PA, é possível remover uma recalibração pequena da PA com menos de 3 batimentos associados, marcando os batimentos como ectópicos e interpolando os sinais de IRR, PAS e PAD. No caso da respiração, foram aceitas recalibrações cujo intervalo fosse significativamente menores que o período do FR.

A interpolação de sinais é utilizada para estimar valores intermediários. Comumente feita uma média entre os valores laterais ou utilizado algoritmos que estime uma curva

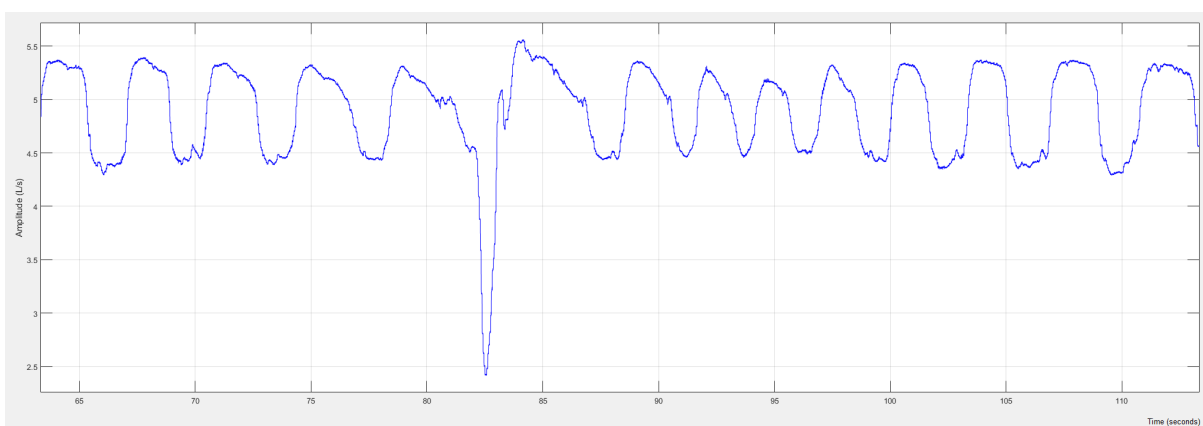
que passe pelos pontos do sinal e selecione um ponto da curva para repor o dado que foi removido.



**Figura 3.5.** Exemplo de batimento ectópico presente na base de dados CDED.



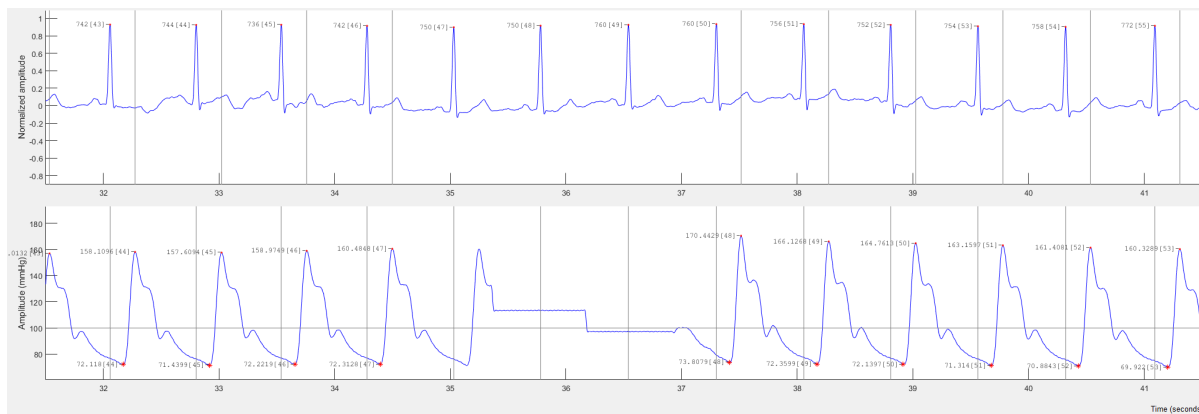
**Figura 3.6.** Exemplo de inconsistência de sinal presente no sinal de ECG da base de dados CDED.



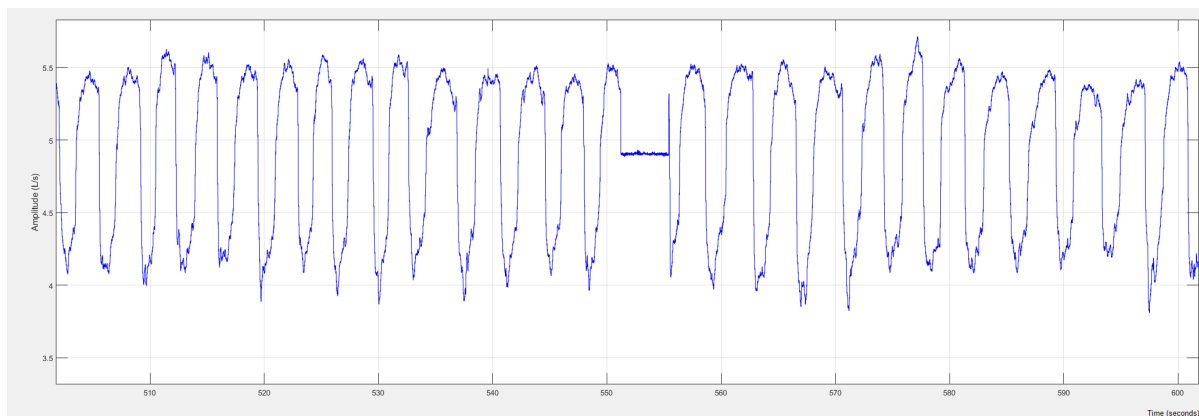
**Figura 3.7.** Exemplo de inconsistência de sinal presente no sinal de FR da base de dados CDED.

Para dados do eletrocardiograma foram aceitos intervalos contendo um máximo de até 3 batimentos ectópicos para toda a janela. Para estes batimentos, foi feita a marcação e interpolação dos dados.

Para os dados de pressão arterial, qualquer recalibração ou inconsistência que interferisse com mais de 3 batimentos foram descartados. Recalibrações e inconsistências curtas foram interpoladas, respeitando a regra definida para o número de batimentos. Para o FR, apenas trechos longos de recalibração foram desprezados, trechos curtos em relação a um período completo de inspiração-expiração foram aceitos desde que não ocorressem repetidamente.



**Figura 3.8.** Exemplo de recalibração de sinal presente no sinal de PA da base de dados CDED.



**Figura 3.9.** Exemplo de recalibração de sinal presente no sinal de FR da base de dados CDED.

Para a etapa de análise de desempenho dos índices no treinamento de algoritmos de aprendizado de máquina, foi determinada a necessidade de uma maior quantidade de dados. Por este motivo, para aumentar a quantidade de dados disponíveis para o treinamento com a base de dados selecionada, foram extraídas até duas janelas de sinais de 4 (quatro) minutos por indivíduos, evitando-se sobreposição entre as janelas.

As janelas de sinais devem conter, se possível, sinais de ECG, PA e FR, para que seja possível a produção completa de todos os índices estudados. Caso não fosse possível, decidiu-se a criação de janelas com as seguintes combinações de sinais: a) ECG e PA, b)

ECG e FR, ou c) ECG isoladamente; a fim de reduzir descarte de dados e possibilitar a produção parcial de índices que dependessem apenas dos sinais contidos na janela.

Essa produção de dados utilizando janelas que não contivessem todos os 3 (três) sinais produz uma característica no conjunto de dados conhecido como desbalanceamento de classes e ausência de valores. Ao utilizar apenas 1 (um) ou 2 (dois) sinais para composição da janela de sinais. Restringe-se os índices que podem ser produzidos. A Tabela 3.2 apresenta o desbalanceamento das classes, onde há diferenças significativa na quantidade de dados de cada classe, após o pré-processamento e antes da remoção de *outliers*. A remoção de *outliers* aumenta ainda mais o desbalanceamento.

Sinais	$N^\circ$ DM	$N^\circ$ Controle	Total de janelas
Contém ECG	62	52	114
Contém ABP	37	23	60
Contém apenas ECG + ABP	2	7	9
Contém apenas ECG + RF	23	12	35
Contém ECG + ABP + RF	35	25	60

**Tabela 3.2.** Distribuição de dados desbalanceados após criação de janelas.

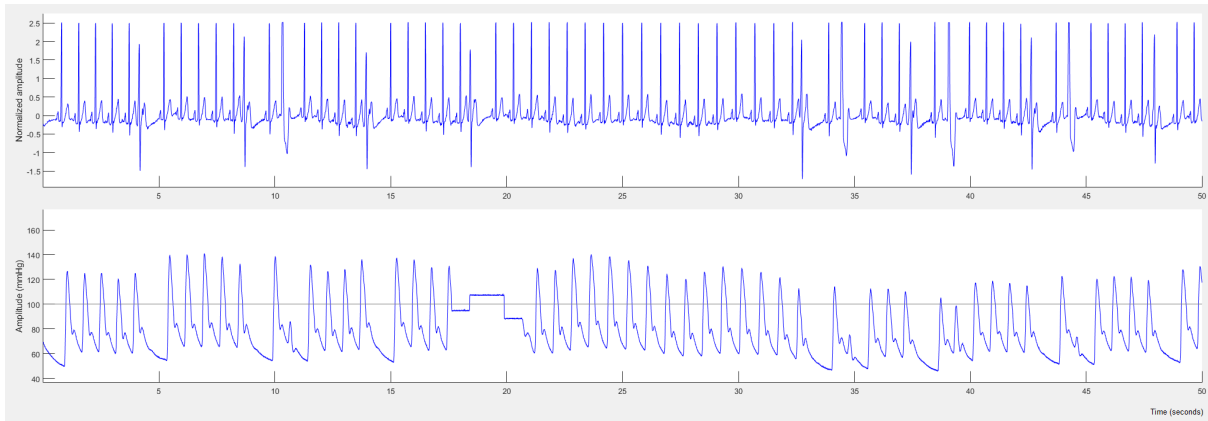
Dos 69 indivíduos disponíveis para análise, 7 foram descartados por apresentarem excessos de ectópicos ou recalibrações, como o sinal apresentado na Figura 3.10, tornando-os inaptos para o estudo, ou pela ausência de registros adequados na base de dados. Para o restante dos indivíduos, ao menos uma janela foi extraída. A Tabela 3.3 indica as quantidades de janelas por classe de indivíduos, IPDM e IC.

Distribuição de Dados	DM	Controle
Total de Pacientes	34	28
Janela 1	32	26
Janela 2	30	26
Janelas Totais	62	52

**Tabela 3.3.** Distribuição de dados por indivíduos.

### 3.2.3 Criação de objetos do tipo Paciente

O CRSIDLab é uma toolbox de processamento de sinais fisiológicos para MATLAB, [17]. A toolbox foi criada na Universidade de Brasília, por Silva, Luisa S. C. B. em colaboração com Oliveira, Flavia M. G. S. A., [53, 54] e foi atualizada com novos módulos em [43, 32]. Suas funcionalidades abrangem a identificação do sistema cardiorrespiratório através de modelos paramétricos como o autorregressivo (AR), o autorregressivo com entradas exógenas (ARX), modelos baseados em funções de bases de Laguerre e modelos baseados em funções de bases de Meixner. A ferramenta também permite análises espectrais e estimação de funções de transferência.



**Figura 3.10.** Exemplo de sinais de registro descartado da base de dados CDED.

A toolbox CRSIDLab necessita de uma estrutura própria de dados, contendo um objeto do tipo “dataPkg.patientData”. A criação desse objeto é explicada passo-a-passo no tutorial do CRSIDLab [17].

Para lidar com o tratamento de grandes volumes de dados, foi necessária a automação de partes do processo executado pela *toolbox* CRSIDLab. Optou-se assim pela criação de scripts de MATLAB para a execução automatizada da ferramenta através de acessos às interfaces.

Para este trabalho, dado o volume considerável de indivíduos analisados, trechos do fluxo de pré-processamento foram automatizados, sendo intercalados com etapas de validação manual. A criação de objetos do tipo pacientes faz parte da etapa de processamento de sinais da *toolbox*. *Scripts* para MATLAB foram desenvolvidos para acessar elementos da interface do software, possibilitando a automação da criação de objetos do tipo paciente para todos os indivíduos da base de dados.

### 3.2.4 Filtragem de sinais

A filtragem dos sinais dos indivíduos tem como objetivo reduzir a proporção de energia do sinal em regiões do espectro consideradas menos relevantes, que possam interferir negativamente na obtenção dos índices do SNA. Os filtros de frequência são funções matemáticas aplicadas a um sinal, analógico ou digital, que modificam o sinal no domínio da frequência, determinado por bandas passantes de frequência ou de rejeição de frequência, eliminando ou amenizando a interferência de determinadas componentes de frequência no sinal.

Um exemplo comum de interferências de frequência é o da rede de distribuição elétrica, possuindo um padrão estabelecido em 50Hz ou 60Hz, por serem consideradas de maior eficiência na transmissão de energia em longas distâncias. Componentes alimentados ou

conectados a redes elétricas tendem a possuir amplificação dos sinais nestas frequências. Para efeitos de atenuação deste componente, um filtro notch em 60Hz com tolerância de largura de banda de rejeição de 20% foi aplicado ao sinal de eletrocardiograma e tolerância de largura de banda de rejeição de 1% foi aplicada ao sinal de PA para menores distorções. Tais porcentagem são determinadas a partir da largura de banda rejeição de frequência a partir de frequências normalizadas [54].

Dada a natureza dos sinais e o seu método de aquisição, é comum que eletrodos ou equipamentos utilizados sobre a pele, e conseqüentemente sobre músculos, possuam interferência de sinais mioelétricos – sinais elétricos presentes nos músculos – e apresentem interferência nas frequências em torno de 35Hz. Assim, a frequência de corte de um filtro passa-baixa de 2° ordem do tipo Butterworth foi definida com frequência de corte em 35Hz [54]. Os sinais submetidos à filtragem foram o sinal de eletrocardiograma e de pressão sanguínea arterial.

Para a atenuação da flutuação da linha de base, foi utilizado um filtro passa-alta frequência de 2° ordem do tipo Butterworth com frequência de corte em 0,001Hz. Flutuações de linha de base ocasionadas no eletrocardiograma podem ter como origem a própria respiração do indivíduo ou pequenas movimentações do corpo [54]. O valor de 0,001Hz é escolhido por ser significativamente menor que a menor componente de frequência considerada do controle do sistema nervoso autônomo sobre a modulação cardíaca.

Na etapa de filtragem dos dados de eletrocardiograma e de pressão arterial, é executada a extração da VFC, representada pelo sinal de IRR, e das pressões sistólicas e diastólicas, representando o sinal de VPA. A extração do IRR é feita a utilizando a marcação das ondas R do eletrocardiograma. A toolbox CRSIDLab possui 2 (dois) algoritmos para a extração de IRR denominados algoritmo rápido e algoritmo lento. Para este trabalho a detecção foi efetuada utilizando o algoritmo lento.

Para a detecção das pressões sistólicas e diastólicas nos ciclos da pressão arterial, é possível que a detecção seja feita utilizando exclusivamente a forma de onda do sinal ou que seja combinada com a detecção anterior das ondas R no eletrocardiograma a fim de alinhar os ciclos, pois como entende-se, são estritamente relacionados. Assim, para a detecção da PAS e da PAD, o sinal foi derivado da forma de onda, combinado com validação manual.

Na etapa de pré-processamento do sinal de respiração, representado pelo fluxo-respiratório, é necessária não somente a filtragem do sinal, mas também uma integração numérica para conversão do fluxo-respiratório em volume pulmonar instantâneo. Após a integração numérica, 2 (dois) tratamentos são aplicados ao sinal. 1) Um *detrend*, aqui escolhido como polinomial de ordem 5 (cinco); e 2) Uma filtragem utilizando um filtro passa-baixa

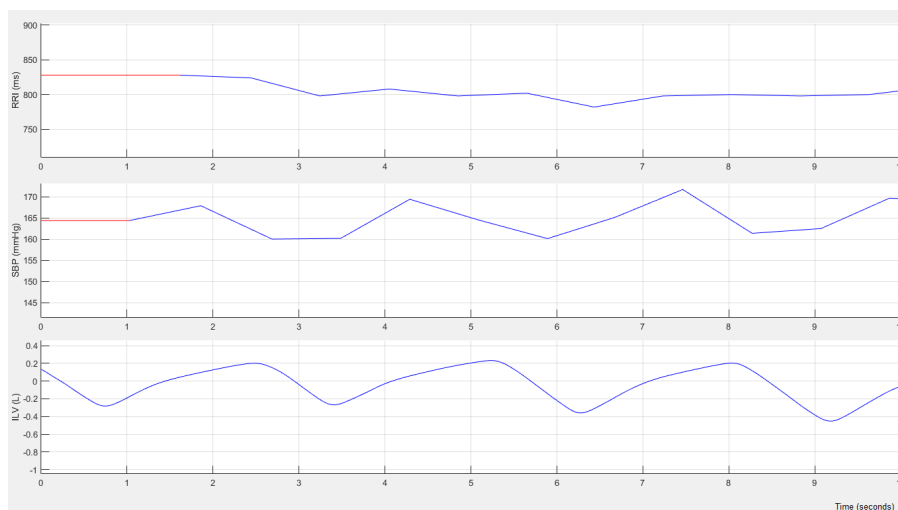
frequência em 4Hz e um filtro para remoção de ruído de alta-frequência.

### 3.2.5 Alinhamento, interpolação, reamostragem de sinais e criação de sistemas parametrizados

Devido à natureza dos sinais, é importante notar que os sinais desejados obtidos, como IRR, PAS, PAD e VPI possuem durações diferentes. A correção nas diferenças de tamanho entre os sinais de uma mesma janela de sinais é chamada neste trabalho de alinhamento. Esta correção completa as bordas de um sinal com repetição de valores ou inserção de zeros.

O instante no tempo em que uma onda RR ocorre não é o mesmo instante onde se ocorre a medição da pressão arterial sistólica ou diastólica. Além disto, por se tratarem de dados extraídos de trechos dos ciclos fisiológicos contidos nos sinais, os sinais derivados de IRR, PAS, PAD e VPI não possuem o mesmo tamanho, pois a primeira onda R ocorre antes ou após a detecção do primeiro pico de pressão sistólica.

O sinal de volume respiratório instantâneo, por ser uma integração numéricas e não depender de um instante específico de seu ciclo, costuma ser o sinal mais longo, começando primeiro que os demais e terminando por último. A Figura 3.11 ilustra o exemplo descrito.



**Figura 3.11.** Alinhamento dos sinais após filtragem. a) Sinal de IRR. b) Sinal de PAS. c) VPI. Obtido com Toolbox para MATLAB CRSIDLab. [17]

Para o alinhamento dos sinais e para manter a conformidade do tamanho de janela de sinais entre os diversos pacientes, optou-se por manter os sinais alinhados ao começo e ao final do sinal de volume instantâneo da respiração, padronizando-se assim, os 4 minutos de janela de sinais também para os sinais derivados. A ausência de sinais no começo e no final dos dados de IRR, PAS e PAD, foram completados utilizando um padding constante reproduzindo o valor do próximo dado disponível. A aplicação do padding não afetará



significativamente o sinal, pois para a existência desejada de simetria entre o início e o final do sinal um janelamento de Hanning é aplicado, funcionando como uma espécie de máscara e suavizando as extremidades do sinal.

Análises no domínio da frequência como a transformada de Fourier, exigem que os dados utilizados estejam amostrados de forma regular e equidistantes. Por este motivo, torna-se necessário que o sinal seja reamostrado, e conseqüentemente, dada a característica discreta no tempo, uma etapa de interpolação dos dados deve ser executada. A interpolação de dados é possibilitada por três algoritmos disponíveis na toolbox: 1) interpolação linear; 2) interpolação cúbica; e 3) algoritmo de Berger;

A escolha deste trabalho foi pela interpolação cúbica. A interpolação cúbica foi escolhida por três motivos: 1) pela suavidade da curva de interpolação, tornando-a derivável em qualquer ponto; 2) para padronização do método de interpolação dos sinais analisados, já que caso utilizássemos o algoritmo de Berger – mais aconselhável e com melhores resultados em relação a distorção espectral – para o IRR e para o PAS/PAD, o algoritmo cúbico seria utilizado automaticamente para a respiração, pois o algoritmo de Berger é projetado para sinais amostrados não igualmente espaçados, o que não é o caso da respiração [17]; e 3) Por utilizar o método de *spline interpolation*, que utilizar interpolações de trechos ao invés do sinal completo, obtendo assim, polinômios de baixa ordem e com menos oscilações erráticas para o sinal [27].

A frequência de reamostragem escolhida para os sinais fisiológicos utilizados nesse estudo foi de 4Hz. Baseada na frequência considerada normal para a operação do músculo cardíaco. A literatura aponta um intervalo de frequência de reamostragem para a variabilidade da frequência cardíaca de 4 Hz [20] a 7 Hz [15], sendo compatível com a frequência máxima característica de 100 batimentos por minuto do coração em repouso e a frequência máxima de aproximadamente 220 batimentos por minuto para o coração em atividade [23, 50], cumprindo-se o critério de Nyquist de frequência mínima de um sinal reamostrado utilizando interpolação, sendo esta frequência no mínimo duas vezes a frequência máxima presente no sinal a fim de se evitar o efeito de *aliasing* no sinal.

Após o alinhamento e reamostragem dos sinais dos indivíduos do estudo, introduz-se o metodologia de sistemas do CRSIDLab [17]. Os sistemas apresentados na toolbox tem o propósito de criar um modelo de entrada e saída de dados com o intuito de identificar e estimar o modelo paramétrico, de preferência apresentando o comportamento de causalidade.

Além da análise espectral de potência, a identificação dos modelos também ocorre através da estimação da resposta ao impulso unitário introduzido na entrada do sistema, analisando-se o comportamento da saída. Os modelos paramétricos utilizados para isto são o Autorregressivo (AR), o Autorregressivo Exógeno (ARX), linear utilizando pon-

derações por bases de Laguerre e linear utilizando ponderações por bases de Meixner.

### 3.2.6 Estimação por recursão de parâmetros de sistemas paramétricos por expansão por bases de Laguerre

Os modelos para este trabalho são feitas utilizando sistemas estimados por expansões de funções de bases de Laguerre, cujos parâmetros são selecionados utilizando o **critério de melhor ajuste** presente na toolbox, alternativa que seleciona entre os métodos de seleção de parâmetros MDL (*Minimum Description Length*) e AIC (*Akaike Information Criterion*). As ordens e os atrasos dos modelos são estabelecidos de forma que não seja gerado um modelo não-causal e que os atrasos das amostras dos dados sejam condizentes com a realidade. É necessário, também, que haja restrição entre o polo do modelo e o tamanho da memória, dada em número de amostras. Ambos devem ser compatíveis para que as bases de Laguerre utilizadas para a ponderação dos sinais de entrada não sejam truncadas e garantam que haja um decaimento até zero, como recomendado pelo manual "CRSIDLab 2.0 – Cardiorespiratory System Identification Lab User Guide" [48].

Para que os modelos sejam condizentes com a realidade, foi necessário que uma faixa adequada de valores para as ordens e os atrasos fosse estabelecida, assim o critério de seleção de parâmetros deve escolher dentre os valores disponíveis, aqueles que apresentem o melhor ajuste.

As escolhas dos intervalos permitidos para atrasos das amostras foram diferentes para os tipos de entradas possíveis: 1) para o modelo com entrada PAS, por possuir uma característica direta e de curto atraso entre a saída IRR, não foi permitida atrasos negativos. De forma a explicitar que o efeito que queremos visualizar sobre as flutuações do IRR, são de fato causados pelo PAS. O intervalo de atrasos possíveis para o PAS como entrada foi definido entre 1 a 12 amostras; 2) para o modelo com entrada a respiração, foi admitido que houvessem atrasos negativos, permitidos pela ferramenta, como descrito em como descrito em [48]. Estes atrasos aparecem como uma forma de atribuir parte do efeito da respiração do IRR, por meio do longo ciclo da respiração e a necessidade da preparação de outros mecanismos fisiológicos para que o evento ocorra, como a diminuição da pressão na caixa torácica causada pelo diafragma, que ocorre antes momentos antes da mudança no volume de ar nos pulmões. Assim, o intervalo permitido para atrasos para a respiração como sinal de entrada foi de -8 a 12 [54].

A escolha do polo de cada sistema e do tamanho de memória adequada para as funções de base de Laguerre são escolhas individuais para cada sistema, de forma empírica. Para este trabalho, optou-se pela automatização da estimação do modelo, variando dentre uma tabela pré-definida de polos e tamanhos de memória, Tabela 3.4, para que cada um dos modelos estimados fosse estimado dentre essas múltiplas combinações e selecionasse os

valores de polos e tamanho de memória que causassem o melhor ajuste.

O melhor ajuste do modelo pelo método recursivo é definido pelo critério de maior valor, definida como soma do ajuste do trecho de sinal de estimação onde é utilizado 80% dos dados e do ajuste do trecho de sinal de validação onde é utilizado 20% dos dados. Assim, que obtiver o maior valor nominal a partir desta soma, será escolhido como parâmetros finais do modelo.

<b>Polo</b>	<b>Tamanho de Memória</b>
0.9	300
0.88	260
0.86	240
0.84	220
0.82	190
0.80	170
0.78	150
0.76	140
0.74	130
0.72	120
0.70	110
0.68	100
0.66	90
0.64	85
0.62	80
0.60	75
0.58	70
0.56	70
0.54	65
0.52	60

**Tabela 3.4.** Mapa de recursão utilizado para identificação de parâmetros dos sistemas por expansão por bases de Laguerre para decaimento até zero.

### 3.2.7 Montagem de tabela de dados extraídos

Após o fim da preparação dos dados e do pré-processamento, o objeto do tipo paciente já possui todos os índices desejados para este trabalho. A criação de uma tabela única agregando todos os índices necessário é uma etapa necessária para a simplificação da análise dos dados, que poderão ser utilizados em outros softwares.

Para a criação da tabela é necessário que o objeto do tipo paciente de cada indivíduo seja acessado e cada um dos campos de dados desejados sejam extraídos para a inserção na tabela. O objeto de tipo paciente pode facilmente ser navegado, já que se organiza numa estrutura de dados do tipo struct. Assim, o caminho de cada um dos dados é acessado e armazenado em uma variável.

Alguns índices calculados no domínio do tempo, como a média, mediana, SDNN entre outras medidas do IRR, são calculados utilizando o IRR extraído do objeto tipo paciente. Dois índices geométricos, plot de Poincaré, conhecidos da literatura também foram obtidos: SD1 e SD2. Estes foram calculados fora da *toolbox* do CRSIDLab.

A etapa de pré-processamento acaba com a produção de uma tabela de extensão “.csv” contendo todos os índices alvo de análise desta pesquisa. É de interesse ressaltar que a tabela de índices está organizada por código de identificação dos indivíduos, janela de dados do indivíduo, tipos de sinais originais presentes nas janelas, índices e o indicador de classe do indivíduo – 0 para indivíduos controle; 1 para indivíduos portadores de DM;

### 3.3 ANÁLISE DE CORRELAÇÃO

A análise de correlação aplicada aos índices estudados foram obtidos através de *script* desenvolvido para MATLAB. Partindo da posse de uma tabela organizada contendo os índices listados no Apêndice A.1, foram aplicado 2 (dois) algoritmos para a criação da matriz com os coeficientes de correlação obtidos par-a-par (Pairwise) dos valores de cada índice fisiológico. Os dois algoritmos escolhidos foram: a) Pearson, e b) Spearman. A análise de correlação visa estudar o relacionamento entre variáveis, medindo de forma quantitativa a dependência ou associação entre duas variáveis aleatórias [60].

#### 3.3.1 Coeficiente de correlação de Pearson

A análise de correlação de Pearson é baseada na suposição de que os dados utilizados são normalmente distribuídos, e o seu coeficiente, normalmente denotado como  $r$ , é compreendido como uma medida da associação linear entre estas variáveis. Apesar de ser modelado como uma medida linear, estudos mostram que o coeficiente de Pearson pode apresentar maior robustez na investigação de variáveis com distribuição não-normal [60]. O coeficiente de correlação de Pearson é calculado na equação 3.1 na sua forma amostral. A matriz dos coeficientes de correlação foi obtida pelo uso da função do MATLAB *corrcoef*, fornecendo, além da matriz com os coeficientes, uma matriz com o valor-P associado ao cálculo de cada par de variáveis X e Y.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3.1)$$

O coeficiente de Pearson é tido como sensível à *outliers*, por este motivo, para o cálculo da matriz de correlação utilizando Pearson, a matriz contendo os índices do estudo foi submetida a uma filtragem de *outliers* utilizando como regra o método de amplitude

interquartil, apresentada nas Equações 3.2 e 3.3:

$$Lim_{sup} = \bar{x} - 1,5 \times (Q_3 - Q_1) \quad (3.2)$$

$$Lim_{sup} = \bar{x} + 1,5 \times (Q_3 - Q_1). \quad (3.3)$$

Nestas equações,  $\bar{x}$  representa a média das amostras, enquanto  $Q_1$  e  $Q_3$  representam respectivamente os quartis 1 e 3 das amostras de dados.

Após a filtragem por *outliers*, uma nova filtragem foi executada com o intuito de selecionar apenas coeficientes considerados confiáveis, dado o tamanho diminuto do conjunto de dados. Esta filtragem foi feita utilizando o valor-P do cálculo par-a-par das variáveis analisadas. Assim, somente foram mantidos coeficientes de correlação cujo valor-P associado ao coeficiente tenha sido  $\leq 0,05$ .

### 3.3.2 Coeficiente de correlação de Spearman

A análise de correlação de Spearman é apresentada como um contraponto ao coeficiente de Pearson, contornando a necessidade de distribuição normal das variáveis. Porém, esta vantagem também está associadas a uma menor robustez do índice quando comparado com o coeficiente de Pearson, muito embora seja mais apropriados do que o coeficiente de Pearson quando aplicado sobre dados contendo *outliers* e assimetrias. O coeficiente de correlação de Spearman, normalmente denotado como  $\rho$ , é baseada no posto (rank) dos dados, e compreendido como uma medida de concordância entre as variáveis. Esta medida tende a quantificar como a variação de uma variável coincide com variações de outra variável [60].

O coeficiente de Spearman é calculado em termos de pares concordantes e discordantes [60] e é calculado aplicando-se coeficiente de Pearson ao posto das variáveis estudadas. O coeficiente de Spearman é apresentado na equação 3.4 na sua forma amostral.

$$\rho = 1 - \frac{6}{n \times (n^2 - 1)} \sum_{i=1}^n [R_i^X - R_i^Y]^2. \quad (3.4)$$

Os parâmetros  $R_i^X$  e  $R_i^Y$  representam os postos (*rank*) de  $X_i$  e  $Y_i$  entre todos os X e Y, separadamente.

Diferentemente do coeficiente de Pearson, o coeficiente de Spearman é resiliente a *outliers* e assimetria nos dados, por este motivo, para obtenção do coeficiente de Spearman, não é aplicada uma filtragem para remoção de *outliers*, somente a filtragem para

seleção de coeficientes considerados confiáveis é efetuada, seguindo-se as mesmas regras apresentadas nas equações 3.2 e 3.3.

### 3.3.3 Análise de dependência

A análise de dependência entre os índices fisiológicos estudados baseia-se na avaliação da amplitude do coeficiente de correlação obtido entre os índices. Para isto, 3 (três) regras empíricas foram consideradas, utilizando-se como fundamentação a amplitude interquartil do módulo do coeficiente.

1. Coeficientes virtualmente equivalentes:  $|r|$  ou  $|\rho| \geq 0.9$
2. Coeficientes com provável forte dependência:  $|r|$  ou  $|\rho| \geq 0.75$
3. Coeficientes com provável fraca dependência:  $|r|$  ou  $|\rho| \leq 0.25$

Utilizando as regras estabelecidas acima, os valores de correlação foram retratados como grafos ponderados, onde os vértices são representados pelos índices analisados e as arestas (ponderações) são representadas pelo valor de correlação entre os índices. Desta forma foi possível observar como, para os diferentes grupos de IPDM e IC, haviam diferentes conexões e como estes índices se agrupavam dada sua similaridade ou possível dependência, mesmo que não fosse possível determinar qual índice é o "influenciador" e qual índice o "influenciado".

A análise de dependência permite identificar quais índices podem representar a mesma informação. Em uma etapa de aprendizado de máquinas, é possível através do grau de similaridade entre os índices, selecionar um único índice, dentre um grupo que represente a mesma informação. Assim, evita-se o reforço de tendências por inserção de atributos de AAM que representem a mesma informação.

## 3.4 ANÁLISE ESTATÍSTICA DE ATRIBUTOS VIA ONE-WAY ANOVA

A análise estatística de variância de um par de variáveis (*One-way Anova*) é um método para verificar se existe diferença estatística entre as médias dos grupos de duas variáveis, assumindo-se a presunção de normalidade em sua distribuição e independência entre as observações da população [8]. Esta metodologia foca na variabilidade individual dos grupos de dados estudados, mas também na variabilidade dos grupos entre si.

O *One-way Anova* necessita de duas variáveis, uma variável chamada de variável independente ou categórica, e a outra chamada de variável dependente, que deve ser

contínua. Todas as observações utilizadas para a metodologia ANOVA devem ser independentes. Esta metodologia avalia a diferença entre os grupos partindo de duas hipóteses a serem testadas. A hipótese nula ( $H_0$ ), onde não há significância estatística entre as médias dos grupos de variáveis apresentando um valor-P superior a 0,05. E a hipótese alternativa,  $H_1$ , onde é identificada significância estatística entre as médias dos grupos de variáveis, e resulta em um valor-P igual ou inferior a 0,05. O valor-P de 0,05 é utilizado como limiar de decisão para os testes.

O modelo matemático utilizado para descrever o método do *One-way Anova* [18] é apresentado na Equação 3.5. A equação à esquerda é apresentada como a decomposição das observações, enquanto a direita corresponde ao modelo.

$$x_{ij} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i) = \mu + \alpha_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \quad (3.5)$$

A variável  $\bar{x}$  equivale a média de todas as observações, enquanto  $\bar{X}_i$  equivale a média do grupo  $i$ . O termo  $X_{ji}$ , apresenta a  $j$ -ésima observação no grupo  $i$ .

Para esta seção da análise, foi utilizado o software SigmaPlot 15.0. Carregados os valores, 5 (cinco) testes foram executados.

1. Teste de normalidade de Shapiro-Wilk (Necessário para o requisito de normalidade do One-way Anova).
2. Teste de igualdade de variâncias de Brown-Forsythe (Necessário para o requisito de normalidade do One-way Anova).
3. All Pairwise Multiple Comparison Procedure de Holm-Sidak (Teste para validação sobre a significância de diferença entre as médias. [18]).
4. Kruskal-Wallis One-way Anova on Ranks (Utilizado quando não é possível utilizar o método original por falta de normalidade ou distinção entre as variâncias).
5. All Pairwise Multiple Comparison Procedure por Método de Dunn (Teste para validação quando há possibilidade de rejeição da hipótese nula por Kruskal-Wallis. Também efetivo por trabalhar com dados faltantes).

O fluxo do One-way Anova depende da normalidade e da variação da variância, para isso, faz parte do fluxo de cada variável analisada, os testes dos itens 1 e 2. Caso aprovado nos requisitos, então o teste do item 3 é utilizado para estimar onde ocorre a diferença entre as médias.

Caso algum dos testes necessário para o One-way Anova falhe, é possível executar o teste de Kruskal-Wallis, item 4, como substituto para o One-way Anova. Por fim, caso

aprovado neste teste, o método de Dunn no teste do item 5 é utilizado para estimar a posição de diferença entre as médias.

### 3.4.1 One-way Anova on Ranks

O método do *One-way Anova on Ranks* tem o mesmo propósito do método original, porém é modificado para verificar grupos cuja hipótese de distribuição dos dados não seja garantida. Neste método alternativo, os dados são substituídos por seus postos (*rank*) sem se preocupar com seu grupo pertencente. E por fim, a soma quadrada dos postos entre os grupos é calculada para a probabilidade de rejeição da hipótese nula.

## 3.5 CLASSIFICAÇÃO POR APRENDIZADO DE MÁQUINA

Os algoritmos de classificação por aprendizado de máquina são ferramentas poderosas e apresentam diversas características na intenção de encontrar funções matemáticas ou métricas capazes de distinguir conjuntos de dados. Porém, apesar de poderosas, é essencial que exista uma seleção de atributos adequada ao problema estudado, de forma a não criar tendências ou divergências que conflitem ou anulem as informações que se busca.

No campo de estudo da engenharia biomédica é possível encontrar diversas medidas quantitativas ou categóricas para indicar padrões em diferentes quadros de enfermidades. Neste trabalho, busca-se avaliar o comportamento de 38 (trinta e oito) índices fisiológicos obtidos através de uma variedade de métodos matemáticos. Porém, a utilização de todos os índices possíveis não é desejável, dado que muitos podem refletir a mesma informação ou informações contraditórias, reforçando tendências incorretas ou incoerentes na classificação dos grupos.

Os algoritmos selecionados para a investigação dos índices são algoritmos robustos, porém de baixa complexidade de utilização e necessidade de ajustes. São eles a) *Support Vector Machines*, b) *Árvore de Decisão*, c) classificador *K-Nearest Neighbour*, e d) *Regressão Logística*.

Os algoritmos de *support vector machines (SVM)* são ferramentas úteis na classificação de dados. A classificação de dados utilizando SVM busca por planos capazes de separar as classes contidas nos dados. É um modelo supervisionado, então, para o treinamento e para a validação do modelo, é necessário saber de antemão qual é a real classe dos dados. A SVM utiliza modelos matemáticos chamados *kernels* para mapear os dados em espaços dimensionais mais altos, para assim, encontrar um hiperplano que maximize a diferença entre os dados das classes contida nos dados [13]. Para este trabalho, o *kernel*



RBF (*Radial Basis Function*, também chamado de *kernel* Gaussiano) foi utilizado por conseguir mapear amostras de dados não-linearmente, adequada para análise de dados fisiológicos. Os parâmetros utilizados para o modelos foram aqueles padronizados pela biblioteca utilizada *sklearn* para *Python 3.9*, com  $C = 1$  e  $\gamma = auto$ .

A árvore de decisão (AD) se baseia na divisão de de conjuntos onde produz uma estrutura de decisão de dados em formato de árvore (invertida) onde os nós possuem regras de decisão binárias sobre cada um dos atributos. A AD utilizando seleção de atributos por *GINI INDEX* é utilizada para medir a “impureza” dos dados [6]. O *GINI INDEX* calcula a probabilidade de um atributo ser classificado incorretamente quando selecionado aleatoriamente. A escolha por esse critério leva em consideração seu melhor desempenho sobre variáveis consideradas contínuas, quando comparada com o critério *Information Gain*.

O *k-nearest neighbor (KNN)* utiliza do princípio de de agrupamento de amostras utilizando como métrica a distância entre os dados, assumindo que há similaridade entre os dados quanto mais próximo forem. Na classificador KNN, todos os dados de treinamento estão contidos no modelo e possuem voto sobre os dados de validação que entrar passam pelo classificador. Diferente do critério de ponderação *uniforme*, onde todos os dados de treinamento possuem mesmo peso ao votarem pela classe da amostra a ser classificada. No critério **distância**, os dados que possuem mais proximidade física, possuem peso maior ao votarem. Assim, o classificador consegue distinguir amostras do “núcleo” do conjunto associada a classe, daquelas que estão nas bordas do conjunto [11].

A regressão logística (RL) é um modelo de AAM baseado em um método estatístico de predição de classes binárias. Utilizando como curva base para separação a função *sigmoide*. O uso do solver *liblinear* como algoritmo de otimização ocorre por sua indicação para pequenos conjuntos de dados como ocorre para dados fisiológicos.

### 3.5.1 Índices individuais

A análise individual foi executada como meio para entender as capacidades individuais de cada um dos índices na distinção entre os grupos de IPDM e IC. Todos os índices estão de alguma forma atrelados à regulação cardiovascular através da VFC. Assim, espera-se que todos reflitam com maior ou menor intensidade a atuação do Sistema Nervoso Autônomo na regulação cardíaca.

A classificação dos grupos através de índices individuais buscou reunir em um conjunto determinado como especial, vide Apêndice A.1, aqueles índices que demonstraram individualmente alto desempenho na classificação das classes. Espera-se identificar quais métodos ou influências dos sinal de origem se complementam melhor a fim de formar

um agrupamento de índices com alta acurácia, precisão e sensibilidade na predição dos indivíduos.

### 3.5.2 Agrupamentos de índices

A inspeção do desempenho dos índices quando utilizados em conjuntos para o treinamento e validação do algoritmos de aprendizado de máquina tem por objetivo avaliar a possível melhoria da predição das classes. A hipótese é de que ao reunir índices de diferentes categorias como índices relacionados a VFC e índices relacionados VPA ou a respiração, as informações sejam complementares e apontem tendências correlatas entre si, intensificando características que melhorem a classificação dos grupos de IPDM e IC.

### 3.5.3 Avaliação dos modelos

Para avaliar o desempenho dos modelos treinados por cada tipo de algoritmo de classificação são utilizadas as seguintes métricas: a) acurácia, b) precisão, c) sensibilidade, e d) medida F1. Estas métricas são necessárias para dizer o quão bom um modelo pode ser dado um determinado tipo de problema. É de interesse, compreender se o modelo treinado atende a necessidade do problema, não bastando ter uma alta taxa de acertos, mas também entender quando os erros ocorrem.

Ao treinar modelos de ML, é essencial que uma parte dos dados seja guardada para a validação do modelo. A proporção mais utilizada é de 80% dos dados para treinamento e 20% dos dados para a validação. Para evitar que o algoritmo “decore” os dados do treinamento e resulte em um *overfitting* é importante que se verifique a curva de desempenho do modelo. Um bom modelo possui desempenho entre as métricas de treinamento e de validação.

Um dos problemas para validação de desempenho em bases de dados pequenas está no desbalanceamento dos dados – alta divergência entre o número de dados para cada classe – e a maior possibilidade de *overfitting*. Ao lidar com dados médicos, este é um problema recorrente. Para lidar com este obstáculo, e reduzir a variância das métricas de avaliação de desempenho dos modelos, foi aplicado ao AAM a técnica de validação cruzada – *K-fold Cross-validation* – onde os dados são divididos em  $n$  segmentos, e o algoritmo é treinado em  $n - 1$  segmentos de dados e validado no segmento restante. O treinamento é executado isoladamente em  $n$  vezes, alternando-se o segmento de validação. Para cada treinamento, as métricas de avaliação do modelo são geradas, e ao final, a média das métricas para os  $n$  treinamentos resulta nas métricas finais do modelo.

Para este trabalho utilizou-se 5 segmentos (valor do parâmetro  $k$ ) refletindo a pro-

porção comum de 80% e 20%. Esta proporção, apesar de não ser necessária para a metodologia *K-fold Cross-validation*, permitiu para a base de dados usada que sempre houvesse ao menos 1 elemento de cada classe sendo representada. Para esta metodologia *K-fold Cross-validation* o valor do parâmetro  $k$  é tipicamente igual à 10 [25]. A Figura 3.12 apresenta graficamente a metodologia da validação cruzada.



**Figura 3.12.** Validação cruzada - *K-fold Cross-Validation*.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 ANÁLISE DE CORRELAÇÃO

Para análise inicial dos 38 índices univariáveis e multivariáveis do sistema nervoso autônomo, obtidos a partir do processamento e análise dos sinais de eletrocardiograma, pressão arterial contínua e fluxo respiratório, decidiu-se utilizar a análise de correlação, com o objetivo de se explorar as características de índices extraídos de modelos univariáveis e multivariáveis na utilização como atributos para AAM.

Através da análise de correlação, pretende-se verificar o relacionamento entre os índices, identificando quais podem ser considerados como representação da mesma informação, sendo virtualmente iguais, ou apenas determinar o grau de relacionamento entre os índices, interpretando-os como pares de índices com forte influência ou fraca influência entre si. Essa informação será importante para a proposta de conjuntos de índices que serão utilizados no treinamento de AAM, servindo como fator de seleção entre índices de mesma origem ou metodologia de extração, a fim de evitar inserção de tendências nos modelos.

A análise de correlação foi a metodologia escolhida para a análise inicial dos 38 índices obtidos a partir do pré-processamento de sinais fisiológicos e a partir da estimação de modelos de acoplamento multivariáveis. A correlação é uma forma estatística de mensurar a dependência ou associação entre duas variáveis aleatórias e contínuas. Há três coeficientes popularmente utilizados na literatura [60]. São eles: 1) Pearson's  $r$ , 2) Spearman's  $\rho$ , e 3) Kendall's  $\tau$ . Existem diferentes interpretações sobre a abrangência e especificidades entre os coeficientes de análise de correlação apresentados.

Atualmente, entende-se que a análise de correlação na forma de Pearson, modelo que se baseia na distribuição normal das variáveis, contemple melhor a associação linear entre as medidas, enquanto as análises de correlação nas formas de Spearman e Kendall meçam melhor a associação monotônica das variáveis por basear-se no posto (*rank*) das variáveis analisadas.

Como diversos índices falharam no teste de normalidade, optou-se por utilizar duas formas de análise de correlação foram aplicadas: 1) Pearson, e 2) Spearman. Para am-

bas as análises, por causa das características dos registros, após remoção de *outliers* e formação com apenas 1 (um) ou 2 (dois) sinais, em possuírem dados faltantes para diversos índices, as matrizes dos coeficientes de correlação foram calculadas utilizando *Pairwise*, onde considera-se apenas pares de dados existentes em ambas as variáveis, tornando possível a análise com ausências de dados ou classes desbalanceadas.

Os índices dos grupos de indivíduos diabéticos e indivíduos controle foram submetidos separadamente à análise de correlação e separados por janelas. Como citado na metodologia, por causa do tamanho da base de dados, foram extraídos de cada indivíduos até dois conjuntos de dados – denominados janelas 1 e 2 – que foram analisados separadamente e serão reunidas para treinamentos de algoritmos de aprendizado de máquinas.

Como cada grupo de indivíduos, IPDM e IC potencialmente apresenta uma diferença na correlação entre os índices obtidos, optou-se por realizar a análise de correlação dos índices de modo separado para cada grupo. Após a obtenção das matrizes com os coeficientes de correlação para cada uma das situações de interesses, as matrizes foram filtradas para que apenas os coeficientes de correlação cujo valor-P relacionado ao cálculo das hipóteses de correlação dos índices fosse igual ou inferior a 0,05. Garantindo que apenas índices cujo volume de dados pudessem fornecer informações confiáveis sobre suas relações fossem mantidos.

Um dos fatores considerados na análise de correlação dos índices é a presença de índices considerados virtualmente equivalentes. Utilizando uma das regras empíricas apresentadas em 3.3.3, índices cujas correlações em módulo fossem  $\geq 0,99$ , valor próximo ao utilizado por Pham et al., (2021), para considerar índices como representações da mesma informação [44]. A Tabela 4.1 apresenta os índices contemplados por essa regra.

Índice 1	Índice 2
MEDIA	MEDIANA
MEDIA	QUARTIL_1
MEDIANA	QUARTIL_1
MEDIA	QUARTIL_3
MEDIANA	QUARTIL_3
RMSSD	SD_1
SDNN	SD_2
HF_DG_ABR	ABR_DYNAMIC_GAIN
HF_DG_RCC	DYNAMIC_GAIN_RCC

**Tabela 4.1.** Índices Considerados virtualmente iguais.

Os valores de corte para as regras empíricas 3.3.3 foram selecionados baseando-se nos valores de segmentação de quartis para visualização dos níveis de interação entre os índices. Utilizando estas regras como guias, as relações entre os índices são apresentadas como um grafo cujos nós são os índices e as arestas ponderadas são os valores dos coefici-

entes de correlação. As Figuras 4.1 e 4.2 apresentam uma visão geral do comportamento de associação e interação dos índices.

A utilização da análise de correlação serve como ferramenta de seleção de atributos para o aprendizado de máquina. Ao selecionar atributos, busca-se informações não redundantes que se somem. Algumas abordagens sugerem a utilização de atributos que possuam baixa correlação entre si, assumindo que possuam uma relação "não óbvia" [10]. Porém, busca-se também a utilização de índices similares em conjuntos, porém sem que haja índices virtualmente iguais, na tentativa de refocar a identificação por uma característica "bem definida".

As Figuras 4.1 e 4.2 são grafos formados apenas pelas janelas 1 de sinais dos indivíduos. Nelas é apresentada conexões entre índices, onde todas são consideradas significativas, porém aquelas, em vermelho, são muito fortes com  $|r| \geq 0,9$ . Esse valor foi escolhido em referência ao valor de 0,999 utilizado por Pham (2021) [44], porém atenuado para representar a significância das relações entre os índices dado o volume de dados disponíveis na base CDED.

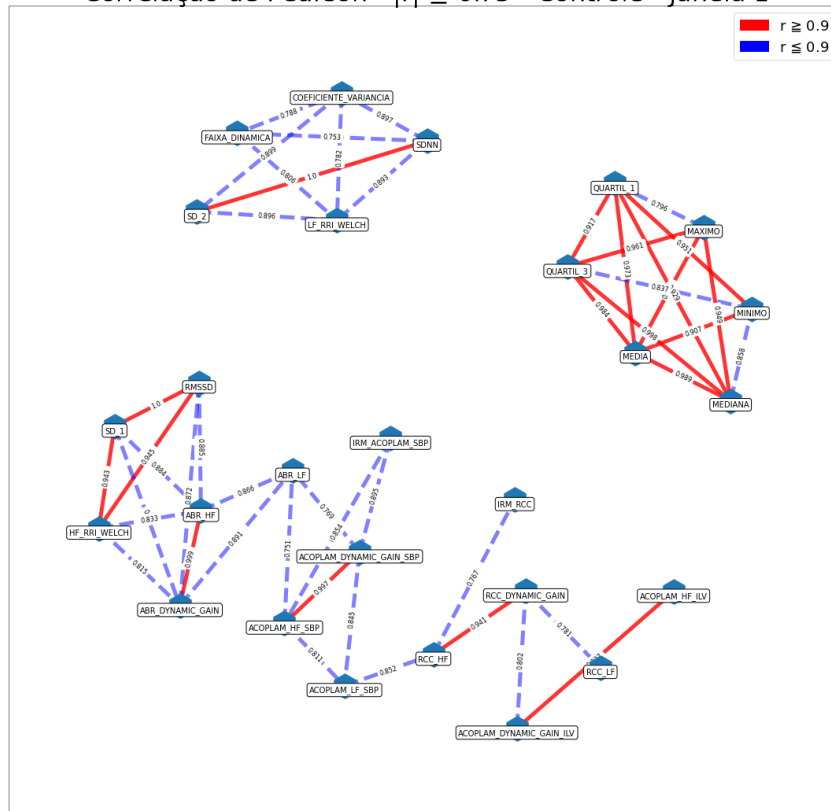
Ao visualizar a relação entre os índices como um grafo, algumas informações se apresentam e podem ser interpretadas. A relação mais evidente ocorre quando se compara o grafo produzido ao utilizar apenas dados dos indivíduos de controle *versus* apenas dados dos indivíduos portadores de diabetes mellitus tipo 2. Utilizando a regra de coeficientes de correlação maior ou igual a  $|0.75|$  diferentes aglomerações de índices se apresentam.

Para os indivíduos de controle, as aglomerações de índices são maiores, conectando muitos índices com centros concisos, com múltiplas conexões entre os índices no centro do aglomerado. As bordas dos aglomerados de IC, apresentam bordas menos concisas, com poucas conexões entre os índices do aglomerado. Entende-se das figuras com grafos formados de índices de IC, que os sistemas fisiológicos interagem amplamente no grupo controle, e que suas interação possuem mais força, ou um maior acoplamento.

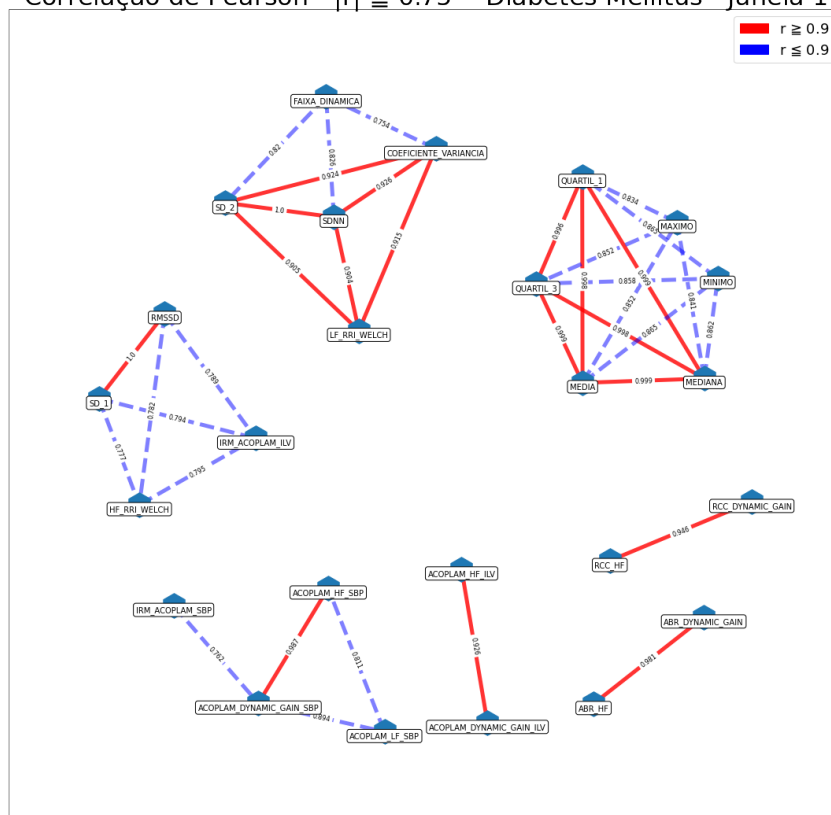
Quando se compara as figuras de grafos de IC com as figuras de grafos dos IPDM, verifica-se que aglomerados menores se formam, porém em maior número. Os aglomerados são bastante concisos, com múltiplas conexões entre os índices no centro do aglomerado, porém menos interligados com outros agrupamentos e sem a presença de índices de borda, com poucas conexões. Entende-se que índices cujo sinal de origem sejam os mesmos, mantenham fortes interações em indivíduos diabéticos, porém a interação entre diferentes sistemas fisiológicos seja mais fraca.

Uma possível interpretação para esta característica pode ser fundamentada pelo efeito de desacoplamento dos sistemas regulatórios através de disfunções autonômicas ou pela presença de NAD. Em indivíduos portadores de Diabetes Mellitus, especialmente a tipo 2, pode ocorrer no intervalo de até 1 ano após o diagnóstico da doença, porém é associada

Correlação de Pearson -  $|r| \geq 0.75$  - Controle - Janela 1

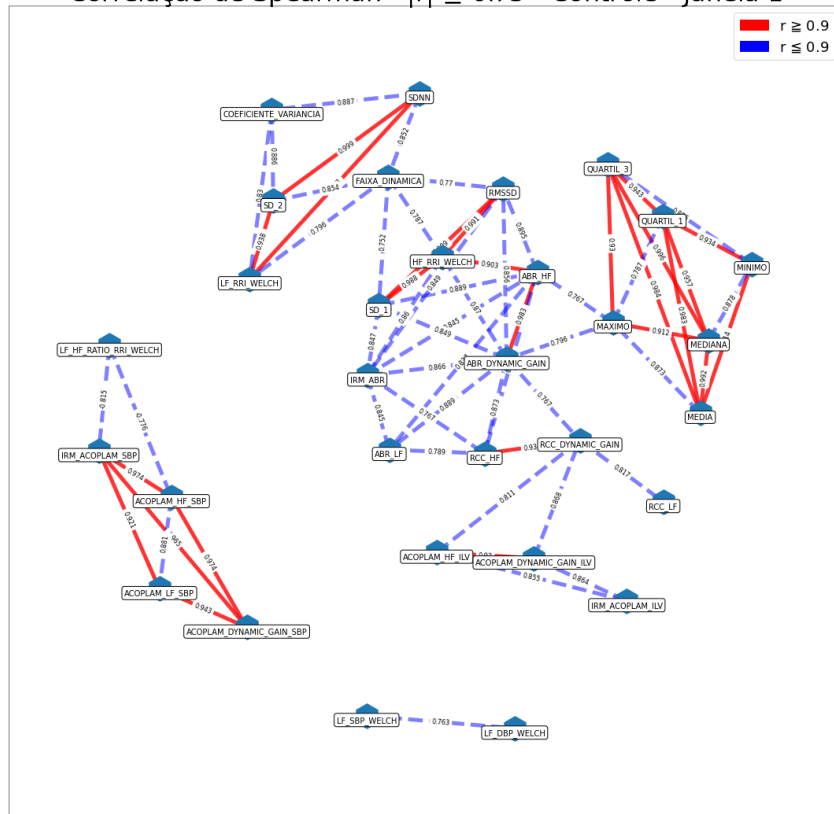


Correlação de Pearson -  $|r| \geq 0.75$  - Diabetes Mellitus - Janela 1

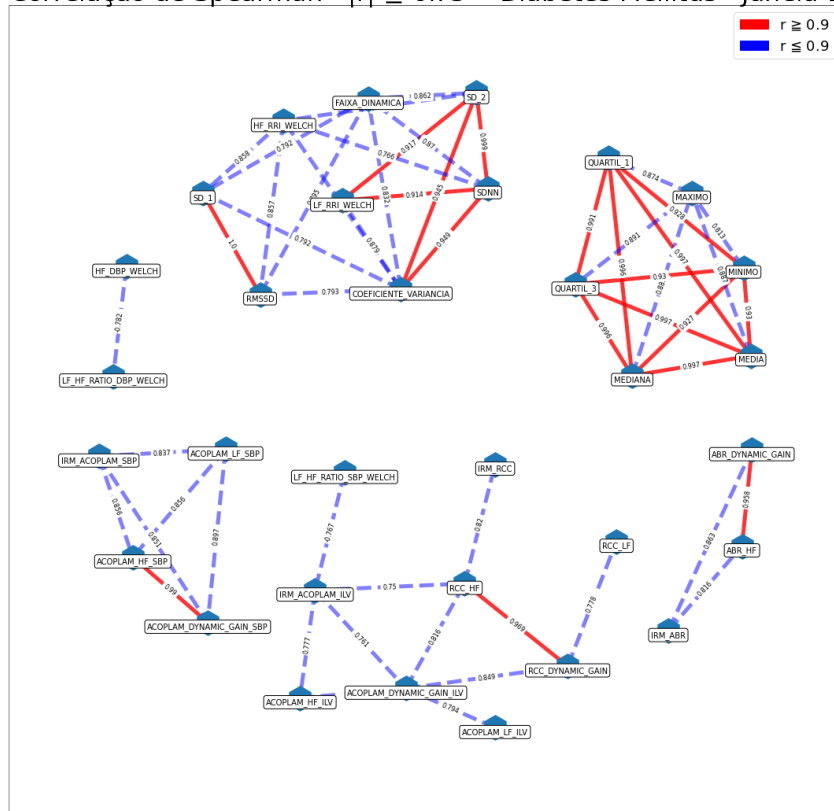


**Figura 4.1.** Grafos formados pelo agrupamento de índices utilizando correlação de Pearson. Janela 1 - Correlação entre índices do grupo T2DM (acima) e índices do grupo controle (abaixo).

Correlação de Spearman -  $|r| \geq 0.75$  - Controle - Janela 1



Correlação de Spearman -  $|r| \geq 0.75$  - Diabetes Mellitus - Janela 1



**Figura 4.2.** Grafos formados pelo agrupamento de índices utilizando correlação de Spearman. Janela 1 - Correlação entre índices do grupo T2DM (acima) e índices do grupo controle (abaixo).



com a longa duração da diabetes [3]. A NAD clínica danifica os nervos do músculo cardíaco e os vasos sanguíneos do corpo, causando anormalidades na dinâmica regulatória. Em indivíduos diabéticos sem NAD, a presença de disfunções autonômicas produzem efeitos similares na ação autonômica.

Disfunções autonômicas causam um desacoplamento do sistema nervoso autônomo tido como o ator central de regulação e controle da homeostase corporal. Alguns dos sintomas principais são a diminuição da variabilidade da frequência cardíaca e a diminuição da sensibilidade do mecanismo de barorreflexo, causando conseqüentemente uma regulação defeituosa da pressão sanguínea [62].

Outro efeito ocorrido ao visualizar os agrupamentos que surgem após a verificação de índices com alta correlação é que, naturalmente, índices cujo sinais de origem sejam os mesmos tendem a se manter unidos. Um exemplo disso são os índices cuja origem são os sinais de ECG/IRR, PA/PAS/PAD e FR/VPI que tendem a se agrupar.

Os intervalos RR são dados obtidos a partir do eletrocardiograma e atribuídos à atuação do SNA. As pressões pressão arterial sistólica e pressão arterial diastólica representam as alterações de máximo e mínimo no sinal de PA, característica que é influenciada pelo SNA. O VPI é o produto final da integração numérica do sinal de FR, que reflete indiretamente a relação de pressão dentro do tórax, que pode influenciar a regulação cardiovascular.

É interessante notar que os índices não necessariamente se aglomeram quando avaliarmos a metodologia matemática utilizada para produção do índices, tal aglomeração de índices nos grafos correm apenas para índices no domínio do tempo. Verifica-se, também, que índices obtidos no domínio da frequência, como índices da VFC e da VPA ou índices obtidos através de modelos multivariáveis não necessariamente se aglomeram, ocorrendo com frequência conexões de alta-correlação entre índices obtidos por diferentes tipos de metodologias de extração.

Um dos objetivos da análise de correlação é a verificação de índices com baixa interação. O principal objetivo da análise de correlação era a verificação de índices virtualmente independentes, que pudessem complementar informações para formar um conjunto de índices de alto desempenho para classificação dos grupos de IPDM e IC a partir de relações não triviais. Porém, não houveram resultados significativos em relação a índices com baixa correlação. A regra empírica definida em 3.3.3, onde índices com fraca correlação deveriam possuir  $|r|$  ou  $|\rho| \leq 0,25$ , não apresentou coeficientes de correlação em limiares tão baixos. Aumentando o limiar da regra para 0,4 e 0,45, conexões apareceram em grandes volumes, acima de 150 conexões, impedindo que qualquer dado de importância pudesse ser obtido.

## 4.2 ANÁLISE ESTATÍSTICA DE ATRIBUTOS VIA ONE-WAY ANOVA

A aplicação do método estatístico *One-way Anova* foi utilizada para validar a capacidade dos índices estudados em distinguir entre as classes IPDM e IC. Como descrito na Seção 3.4, o método avalia a variância entre os grupos, com o intuito de identificar diferenças significativas nas médias dos grupos. A técnica estatística *One-way Anova* não comporta múltiplas amostras de um mesmo indivíduo, pois exige que as observações utilizadas sejam independentes, sendo inadequado para aplicação em um conjunto de dados reunindo janela 1 e 2, onde os dados são extraídos do mesmo indivíduo.

A técnica *One-way Anova* foi então aplicada nos índices da janela 1 e da janela 2 separadamente. Dessa forma, foram produzidos dois resultados estatísticos para cada índice estudado. As variáveis utilizadas para o teste estatístico utiliza duas variáveis, uma variável contínua e uma variável categórica. A variável contínua são os índices extraídos dos modelos univariáveis e multivariáveis, enquanto a variável categórica foi um valor numérico representando as classes dos indivíduos estudados. Valor 0 para indivíduos controle e valor 1 para IPDM.

Algumas suposições são assumidas para a utilização correta da técnica *One-way Anova*. Os dados utilizados como variável contínua devem ser observações independentes. Os valores devem seguir uma distribuição normal, e a variância dos grupos devem ser aproximadamente iguais, possuindo uma dispersão em torno de suas médias similar entre as classes.

A técnica estatística utiliza um conjunto de testes em sua metodologia, buscando garantir as suposições necessárias, para que sejam avaliada duas hipóteses. A hipótese nula -  $H_0$ , onde as médias dos grupos são iguais e não se pode identificar diferenças entre as classes. E a hipótese alternativa -  $H_1$ , onde a média dos grupos são diferentes, indicando a possibilidade estatística de distinguir entre os grupos a partir dos dados utilizadas como variável contínua. As hipóteses levam em consideração o valor-P, aqui utilizado 0,05, onde a hipótese alternativa somente é considerada caso possua uma significância estatística menos que 0,05.

A aplicação da metodologia Anova possui uma alternativa para casos onde os dados utilizados não atendem os requisitos de normalidade e variância. O *One-way Anova on Ranks* é utilizado para estes casos. Nele utiliza-se a transformação do dados pelos seus postos (*ranks*), avaliando-se assim, não a média, mas a mediana dos grupos de diabéticos e de controle. Normalmente a abordagem "*on Ranks*" utiliza o teste de Kruskal-Wallis, onde as hipóteses são as mesmas da abordagem *One-way Anova* comum, e também baseando-se na significância estatística abaixo de 0,05.

Cinco tipo de resultados podem ser obtidos para este trabalho. A variedade de resul-

tados é causado pela combinação dos resultados das metodologias Anova e da existência de duas observações de cada indivíduo, nas janelas 1 e 2. Assim, a utilização das metodologia do *One-way Anova* comum e *"on Ranks"* fornecem quatro resultados possíveis, e a separação dos dados por janelas pode resultar no estado de indeterminação dos resultados, que ocorre quando um índices apresenta resultados discrepantes entre janelas.

Os resultados da metodologia são: 1) Índices que atendem às suposições de normalidade e variância e tiveram a  $H_1$  aceita (há diferença estatística entre os grupos). 2) Índices que atendem às suposições de normalidade e variância e tiveram a  $H_0$  aceita (não há diferença estatística entre os grupos).

3) Os índices que **NÃO** atendem às suposições de normalidade e variância e tiveram a  $H_1$  aceita (há diferença estatística entre os grupos). 4) Índices que **NÃO** atendem às suposições de normalidade e variância e tiveram a  $H_0$  aceita (não há diferença estatística entre os grupos). A tabela 4.2 apresenta um resumo das características dos índices sob a ótica destes 4 (quatro) resultados.

	Atende suposição (Anova comum)	Não atende suposição (On Ranks)
$H_1$ aceita (Aprovado)	MEDIA	IRM_RCC
	MEDIANA	LF_DG_RCC
	QUARTIL_3	HF_DG_RCC
	LF_DG_ILV_ACOPLAM	
	HF_DG_ILV_ACOPLAM	
	DYNAMIC_GAIN_ILV_ACOPLAM	
	LF_DG_ABR	
	DYNAMIC_GAIN_RCC	
$H_0$ aceita (Reprovado)	MINIMO	pNN50
	MAXIMO	RMSSD
	QUARTIL_1	FAIXA_DINAMICA
	LF_SBP_WELCH	SD_1
	LF_HF_RATIO_SBP_WELCH	SD_2
	HF_DBP_WELCH	HF_SBP_WELCH
	ACOPLAM_LF_SBP	

**Tabela 4.2.** Apresentação das características dos índices utilizando método estatístico.

Dentre os índices verificados, 11 (onze) obtiveram êxito estatístico em distinguir entre as classes de indivíduos. Os índices considerados aprovados, tiveram a  $H_1$  aceita e apresentaram relevância estatística em ambas as janelas. Os índices classificados como reprovados, tiveram a  $H_0$  aceita em ambas as janelas. Índices classificados como indefi-

nidos, Tabela 4.3, onde houve aceitação da  $H_1$  e uma das janelas e aceitação da  $H_0$  em outra, não podem ser completamente descartados, pois tal efeito pode ser decorrente da quantidade de dados de cada janela ou da qualidade dos índices de uma das janela.

Diversos índices foram classificados como indefinidos, cuja hipótese  $H_1$  foi aceita em apenas uma das janelas, sendo a hipótese  $H_0$  aceita na outra. Porém, a divergência entre os valor-P foi levada em consideração. Casos onde uma janela apresentasse significância estatística muito forte e outra uma significância muito fraca, os índices eram considerados como tendo hipótese  $H_0$  aceita. Casos onde as janelas apresentassem  $H_1$  aceita em uma e a  $H_0$  aceita em outra, porém passassem nos testes de normalidade, variância, com significância estatística próxima a da janela cuja hipótese  $H_1$  fosse aceita, foram considerados.

Índices indefinidos
SDNN
COEFICIENTE_VARIANCIA
LF_RRI_WELCH
HF_RRI_WELCH
LF_HF_RATIO_RRI_WELCH
LF_DBP_WELCH
LF_HF_RATIO_DBP_WELCH
IRM_SBP_ACOPLAM
HF_DG_SBP_ACOPLAM
DYNAMIC_GAIN_SBP_ACOPLAM
IRM_ILV_ACOPLAM
IRM_ABR
HF_DG_ABR
DYNAMIC_GAIN_ABR

**Tabela 4.3.** Apresentação de índices com características conflitantes entre janelas

Os índices apresentados em 4.3 foram assim classificados, não apenas pela divergência no resultado entre suas janelas, mas também por apresentar divergências em seus testes de normalidade, onde uma janela de dados apresentava distribuição normal dos dados e a outra janela não apresentava. Estratégias como a metodologia *One-way Anova* podem ser consideradas para análise quantitativa ou qualitativa das variáveis que devem ser consideradas para um agrupamento de índices na busca pelo aperfeiçoamento da classificação de grupos como IPDM e IC.

## 4.3 CLASSIFICAÇÃO POR APRENDIZADO DE MÁQUINA

### 4.3.1 Índices individuais

O treinamento dos 4 (quatro) algoritmos de classificação selecionados para o teste de desempenho individual dos índices almeja validar o impacto na predição como atributos das classes de IPDM e IC. Utilizando o método *K-fold Cross-validation*, Figura 3.12, o desempenho final de cada modelo é o resultado da média de 5 treinamentos isolados do classificador. A tabela 4.4 apresenta os 10 índices com melhores desempenhos individuais para cada algoritmo na etapa de validação com *k-fold Cross-validation*. O quesito utilizado para o ranking foi o desempenho em acurácia, embora precisão, sensibilidade e medida F1 também tenham sido calculados.

Posição	SVM	AD	KNN	RL
1°	DYNAMIC.GAIN_ILV_ACOPLAM	HF_DG_RCC	HF_DG_ILV_ACOPLAM	DYNAMIC.GAIN_ILV_ACOPLAM
2°	HF_DG_ILV_ACOPLAM	HF_DG_ILV_ACOPLAM	ACOPLAM.DYNAMIC.GAIN_ILV	HF_DG_ILV_ACOPLAM
3°	DYNAMIC.GAIN_RCC	QUARTIL_3	QUARTIL_3	DYNAMIC.GAIN_RCC
4°	LF_DG_RCC	MEDIANA	DYNAMIC.GAIN_RCC	LF_DG_RCC
5°	HF_DG_RCC	MEDIA	HF_DG_RCC	HF_DG_RCC
6°	LF_DG_ILV_ACOPLAM	ACOPLAM.DYNAMIC.GAIN_ILV	LF_DG_ILV_ACOPLAM	LF_DG_ILV_ACOPLAM
7°	QUARTIL_3	DYNAMIC.GAIN_RCC	MEDIA	IRM_RCC
8°	IRM_RCC	MAXIMO	MEDIANA	MAXIMO
9°	MEDIANA	SD_1	IRM_SBP_ACOPLAM	QUARTIL_3
10°	IRM_ILV_ACOPLAM	IRM_RCC	ABR_LF	MEDIANA

**Tabela 4.4.** Ranking de desempenho individual dos índices para a classificação de grupos.

A acurácia e medida F1 (utilizada por abranger e simplificar a representação das medidas de precisão e sensibilidade) dos 15 (quinze) índices presentes na tabela de desempenho 4.4 são descritas na tabela 4.5 a fim de comparação.

Um padrão se estabeleceu entre os diferentes AAM, em que os índices que aparecem entre as 10 melhores classificações para acurácia, também aparecem entre os melhores índices da classificação para precisão e medida F1. Porém, o mesmo efeito não ocorre para a sensibilidade, onde diversos índices com as melhores sensibilidades possuem piores resultados para acurácia, precisão e medida F1. Curiosamente, a medida F1 não é prejudicada, ao se averiguar, entende-se que os índices com melhores resultados de acurácia e precisão também possuem alta sensibilidade, embora essa sensibilidade não esteja no topo do ranking.

Os 6 (seis) primeiros índices apresentados na tabela 4.5 aparecem entre os 10 (dez) melhores índices em todos os 4 (quatro) algoritmos de aprendizado de máquina na tabela 4.4. A recorrência dos índices no ranking entre os algoritmos utilizados é importante para explicitar a utilização de índices direcionados ao problema estudado como atributos de AAM, em uma metodologia *"insightful approach"*. Assim, verifica-se quais índices possuem características significativas na distinção dos grupos de IPDM e IC independentemente do algoritmo de aprendizado utilizado.

É interessante ressaltar também que dentre os 6 (seis) primeiros índices da tabela 4.5, 4 (quatro) estão relacionados ao ACR, enquanto outros 2 (dois) são exclusivos de medidas da VFC.

Índice	Algoritmo	Maior acurácia	Maior medida F1	N.º Algorit. (Máx: 4)
HF_DG_ILV_ACOPLAM	KNN	80,75 %	84,95 %	4
DYNAMIC_GAIN_ILV_ACOPLAM	SVM/KNN/RL	79,09 %	84,14 %	4
DYNAMIC_GAIN_RCC	SVM	76,40 %	83,28 %	4
HF_DG_RCC	AD	75,16 %	80,9 %	4
QUARTIL3	KNN	73,63 %	77,17 %	4
MEDIANA	AD	71,89 %	75,48 %	4
LF_DG_ILV_ACOPLAM	SVM	69,09 %	78,05 %	3
IRM_RCC	RL	67,19 %	75,34 %	3
LF_DG_RCC	SVM	71,89 %	80,93 %	2
MEDIA	KNN	69,20 %	73,53 %	2
MAXIMO	RL	66,08 %	70,61 %	2
LF_DG_ABR	KNN	66,26 %	68,52 %	1
IRM_ILV_ACOPLAM	SVM	67,27 %	77,58 %	1
SD_1	AD	64,28 %	67,22 %	1
IRM_SBP_ACOPLAM	KNN	66,54 %	73,52 %	1

**Tabela 4.5.** Desempenho individual dos melhores índices na classificação de grupos.

O desempenho de índices relacionados à respiração pode indicar variações significativas na VFC devido ao aumento da pressão intratorácica causada pela respiração. Para modelos acoplados, principalmente aqueles relacionados ao ACR, os ganhos alterados de BF e AF do componente de respiração são consistentes com o alterações diretamente proporcionais da atividade simpática do SNA. Tal indicador pode evidenciar os efeitos de disfunções autonômicas em diabéticos, pela degradação do ramo parassimpático, reduzindo a presença da sua atuação.

A utilização de AAM para exploração do desempenho individual dos índices na predição de IPDM e IC, ressalta a eficiência dos índices extraídos a partir de modelos multivariáveis. Ao estimar um modelo no domínio do tempo que relacione dois ou mais sinais fisiológicos, portanto disponibilizando mais informações sobre os indivíduos do que índices baseados em um único sinal fisiológico, impõe-se restrições de causalidade temporal entre o sinal utilizado como entrada, IRR ou PAS, e o sinal utilizado como saída, IRR. Isso impede com que dados futuros da entrada afetem dados presentes da saída.

Em sistemas no domínio da frequência, como índices univariáveis obtidos por análise espectral ou multivariáveis obtidos pela estimação de funções de transferência entre dois sinais por análise espectral cruzada, é possível identificar padrões de influência de um sinal em outro, porém não é possível identificar o efeito de causalidade com restrição temporal. Esta restrição temporal é apresentada como vantagem de índices do SNA extraídos a partir de identificação de sistemas por resposta ao impulso.

A predominância de índices derivados do ACR dos sistemas acoplados no *ranking* da

Tabela 4.5 evidencia como abordagens com modelos multivariáveis são adequados para investigação da atividade autonômica, considerando a influência mútua de diferentes sistemas da fisiologia humana. Por esta abordagem, exclui-se a influência de fatores desconhecidos em sistemas complexos com influências de diferentes mecanismos, a exemplo da regulação cardíaca, como citado em [2, 31].

Curiosamente, a mesma banda de frequência, alta-frequência 0,15 - 0,4Hz, associada à atividade do ramo parassimpático, e atribuída à influência da respiração sobre a regulação cardíaca em índices de sistemas univariáveis, aparecem aqui, para índices de sistemas multivariáveis, representados índices de ganho dinâmicos do ACR. Todos os 4 (quatro) índices com melhores desempenhos individuais em todos os AAM refletem diretamente a atividade do ramo parassimpático, seja com diferença significativa no ganho dinâmico da banda de AF (HF\_DG\_ILV\_ACOPLAM - ganho dinâmico da banda de AF da entrada ILV do sistema acoplado de influência mútua [ACR + ABR], e HF\_DG\_RCC - ganho dinâmico da banda de AF do sistema do ACR) ou com diferença significativa no ganho dinâmico total, entre 0,04-0,4 Hz (DYNAMIC\_GAIN\_ILV\_ACOPLAM - ganho dinâmico total do sistema acoplado de influência mútua [ACR + ABR], e DYNAMIC\_GAIN\_RCC - ganho dinâmico total do sistema do ACR. Apenas um índice relacionado ao ABR apareceu no ranking dos índices com melhores desempenhos, porém apenas em um dos AAM. Entende-se por isso que a influência dos barorreceptores é importante quando combinada com outros fatores como a respiração, porém não representa grande influência quando utilizado isoladamente.

Por fim, 2 (dois) índices de modelos univariáveis associados unicamente à VFC também aparecem no ranking de melhores desempenhos em todos os 4 (quatro) AAM, reforçando suas características de quantificação do SNA independentemente do algoritmo utilizado. Foram ele a MEDIADA e o QUARTIL\_3 dos IRR. A MEDIA também apareceu entre os índices com melhor desempenho, em posição inferior à MEDIADA e ao QUARTIL\_3, e somente em 2 (dois) dos AAM. Esse resultado era esperado quando levada em consideração a análise estatística *One-way Anova*. O índice de MEDIANA foi aprovado com significância em ambas as janelas onde a metodologia foi aplicada, enquanto o índice da MEDIA apresentou significância abaixo de 0,05 em apenas uma janela, A MEDIA foi reprovada por significância em sua outra janela, mas possuindo normalidade e variâncias adequadas.

É conhecida da estatística que a mediana de um conjunto de valores é mais resiliente a variações que a média do mesmo conjunto de valores. Este fato, combinado a significância do terceiro quartil, que representa o ponto na distribuição de valores onde 75% dos dados estão abaixo, pode ser interpretado a partir do efeito da diabetes em reduzir a VFC de seus indivíduos. A variabilidade dos intervalos dos batimentos são mais regulares em indivíduos diabéticos, representando uma menor capacidade de regulação

da atividade cardíaca, enquanto em indivíduos controle, essa variabilidade é maior e com uma maior dispersão de valores, o que resultaria em diferentes medianas e quartis entre as classes de IPDM e IC.

Em resumo, as métricas individuais dos índices avaliados individualmente apresentaram bons desempenhos, com o ganho de alta frequência do acoplamento do ACR no sistema mútuo apresentando acurácia superior às acurácias dos agrupamentos de índices quando treinado com *k-nearest neighbor*. Porém, em todos os outros índices e algoritmos, o desempenho médio das métricas foram ligeiramente inferiores aos agrupamentos, que se mostraram mais resilientes aos algoritmos.

As métricas integrais sobre o desempenho dos índices individuais em cada um dos algoritmos está disponível no Apêndice 5.1. Nele é possível visualizar em unidades unitárias o desempenho não apenas de acurácia e medida-F1, utilizada aqui para meio de equiparação de desempenho, mas também as métricas individualizadas como a precisão e a sensibilidade. O detalhamento do tamanho de amostras para cada classe também está descrita no apêndice.

### 4.3.2 Agrupamentos de índices

Nesta etapa, os índices estudados foram agrupados em 9 (nove) diferentes conjuntos. Os índices foram agrupados visando comparar seu desempenho como um conjunto de informações complementares. O agrupamento dos índices se deu, inicialmente, com base em seus métodos de extração e sinal de origem. Foram estes o agrupamento **A** - Índices da VFC, o agrupamento **B** - índices da VPA, o agrupamento **C** - referente aos índices produzidos a partir de análises espectrais, e o agrupamento **D** - referente aos índices produzidos a partir da identificação de sistemas por RI.

Em seguida, os grupos formados inicialmente, foram combinados, formando novos agrupamentos de índices, fornecendo indiretamente aos AAM, informações complementares a partir da reunião de diferentes índices de sistemas univariáveis. Os grupos criados foram: o conjunto **E** - contendo os índices presentes nos agrupamentos A (VFC) e B (VPA), o conjunto **F** - contendo os índices presentes nos agrupamento A (VFC) e D (RI).

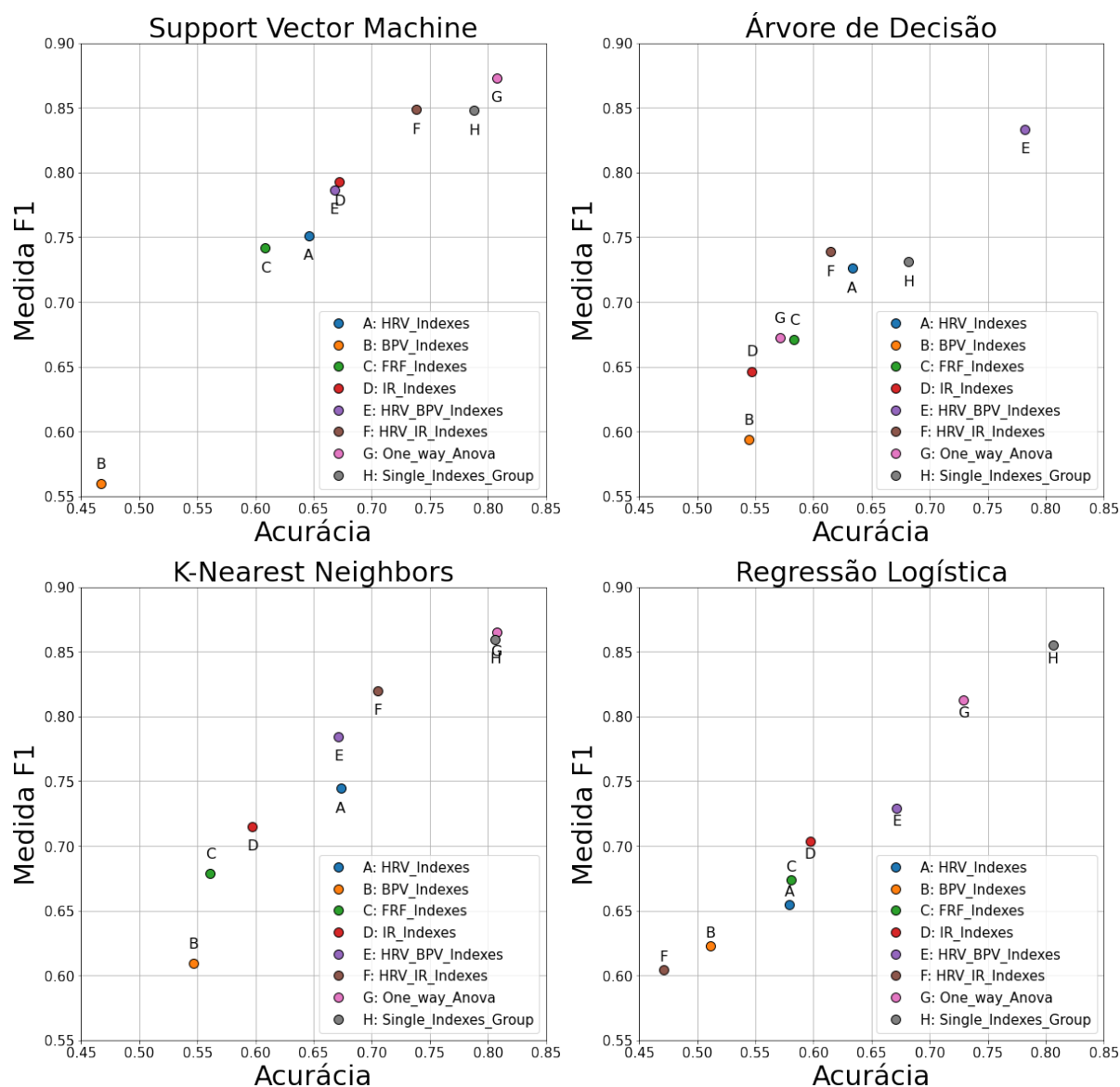
Dois conjuntos, denominados especiais, foram propostos baseados nos resultados obtidos da análise estatística *One-way Anova*, Seção 4.1, e do desempenho individual dos índices na classificação dos grupos de IPDM e IC, Seção 4.3.1. Os conjuntos especiais são: o agrupamento **G** - contém os índices identificados pela metodologia *One-way Anova* como capazes de distinguir entre os grupos de indivíduos, e o agrupamento **H** - contém os índices com os melhores desempenhos a partir do desempenho individual dos índices na predição de indivíduos com T2DM e controle, e que apareçam em todos os 4 (quatro)



AAM utilizados.

Os conjuntos de índices utilizados no treinamento dos 4 (quatro) algoritmos de classificação estão descritos no Apêndice B.1. Coincidentemente, o conjunto que designava a combinação dos grupos de índices da FRF combinados com a VFC, resultaram em arranje idêntico ao conjunto que combinava os grupos de índices da VPA com índices da VFC. Por este motivo, a combinação dos grupos de índices FRF e VFC foi descartada.

O resultado dos conjuntos treinados está disposto na Figura 4.3. O gráfico de desempenho das classificação dos índices em grupos foi feita utilizando como eixos a acurácia e a medida F1 dos modelos.



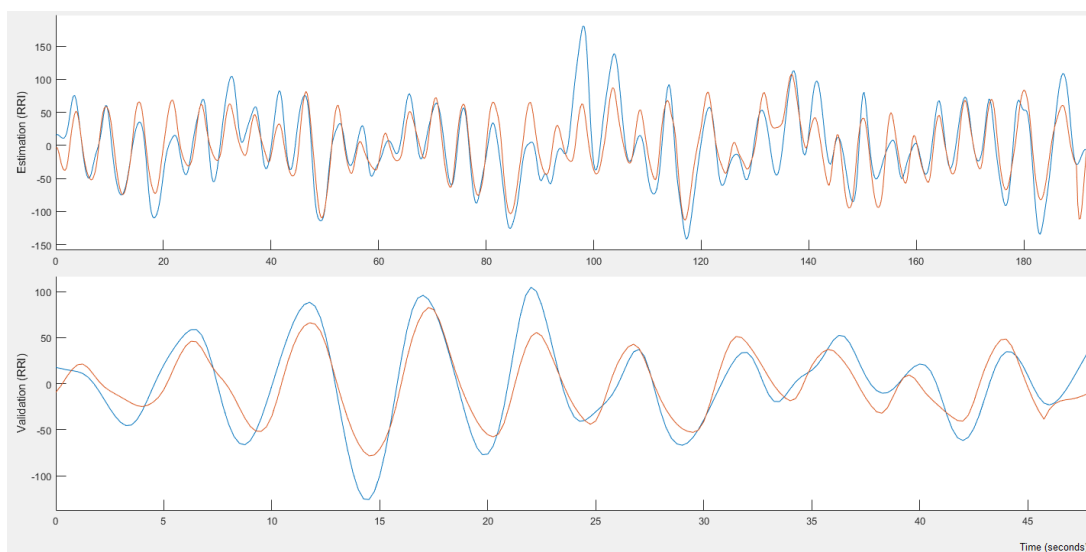
**Figura 4.3.** Resultado dos modelos de classificação por grupo de índices. Métricas integrais do desempenho dos agrupamentos de índices - apêndice C.

Dois conjuntos se destacam na predição de indivíduos e possuem algumas carac-

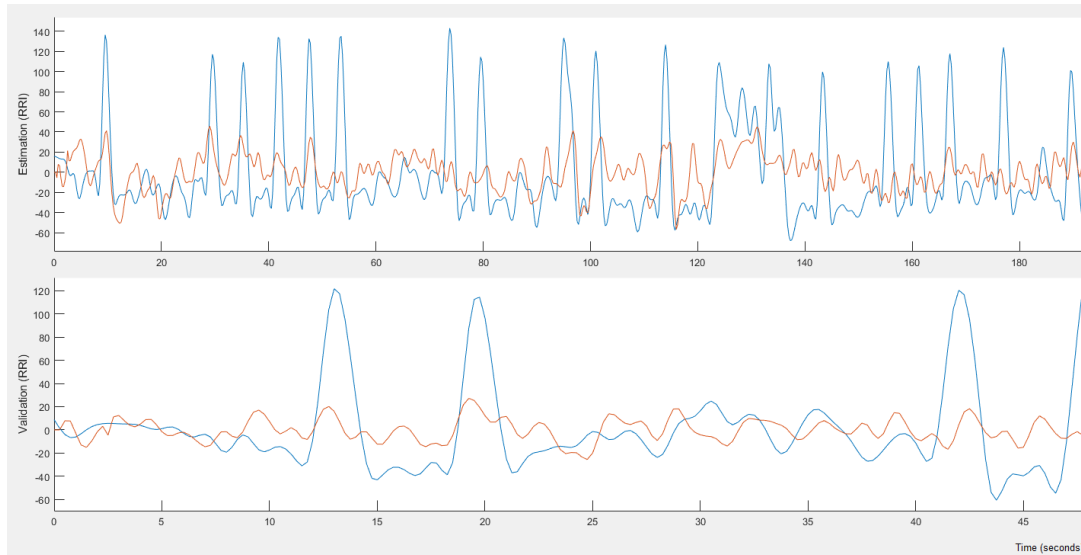
terísticas interessantes, são eles os grupos **G** obtido a partir da análise estatística Anova e **H** obtido a partir do desempenho individual dos índices na classificação dos grupos. Eles possuem diversos índices associados à respiração. O grupo **G** possui 63,63% dos índices com relação direta com a respiração, seja através do ganho dinâmico total, ganho dinâmico de BF ou ganho dinâmico de AF do sistema acoplado da influência mútua - MISO, e do sistema de ACR. O mesmo vale para o Grupo **H**, que é composto por 66,66% de índices relacionados a respiração.

Os algoritmos que tiveram melhor desempenho foram o SVM e o KNN. Porém, apesar de treinados utilizando a metodologia *K-fold Cross-validation*, para reduzir o risco de *overfitting* dos sistemas. O modelo baseado no algoritmo KNN aparenta ser excessivamente robusto para o pequeno volume de dados utilizados, indicando *overfitting* durante o treinamento. As métricas de treinamento para o modelo de aprendizado utilizando KNN indicam acurácia, precisão, sensibilidade, e medida F1 máxima (igual à 1) para todos os conjuntos de índices, muito embora para a validação as métricas sejam comparáveis entre os outros algoritmos como SVM, AD e RL.

Os conjuntos compostos por índices extraídos a partir da identificação de sistemas com RI de sistemas acoplados obtiveram baixo desempenho quando comparados com todos os outros grupos com abordagens multivariáveis (como o grupo *E*, *G* e *H*), ao contrário do que se esperava ao unir as influências de diferentes sistemas da regulação cardíaca. O baixo desempenho pode ser justificado pela caracterização dos sistemas estimados cujos ajustes dos modelos aos dados, tanto para estimação, quanto para validação, foram significativamente baixos, apesar de seguirem adequadamente os formatos dos sinais de resposta, como mostra a Figuras 4.4 e 4.5.



**Figura 4.4.** Modelo com melhor ajuste (53,17%) em dados de validação na estimação do sinal de IRR no sistema de acoplamento mútuo. Polo em 0,9 e tamanho de memória de sistema em 300 amostras.



**Figura 4.5.** Modelo com pior ajuste (6,3%) em dados de validação na estimação do sinal de IRR no sistema de acoplamento mútuo. Polo em 0,78 e tamanho de memória de sistema em 150 amostras.

A média de ajustes para estimação e validação dos modelos acoplados sem *outliers* são respectivamente: a) sistema acoplado de influência mútua - MISO [29,26% e 26,75], b) acoplamento cardiorrespiratório [18,46% e 23,02%], e c) acoplamento do barorreflexo arterial [21,75% e 18,37%]. A utilização de funções de transferência por convolução pode ser uma alternativa para um sistema acoplado mais ajustado aos dados, porém perde-se a vantagem do efeito de causalidade temporal dos modelos paramétricos.

Os índices do grupo **A**, relacionados diretamente à variabilidade da frequência cardíaca, possuem um desempenho adequados quando utilizado isoladamente, mas parecem estar limitados a uma acurácia média de  $\sim 62,27\%$ , necessitando de informações adicionais para uma possível melhora na classificação das classes. Uma possível alternativa é a verificação e remoção de índices conflitivos dos agrupamentos de índices como os índices classificados como virtualmente equivalentes, com correlações muito altas ou índices com cuja hipótese nula tenha sido aceita em pelo menos um das janelas de sinais (dados das janelas 1 ou dados da janela 2) a partir da análise com metodologia *One-way Anova*, independentemente de possuírem distribuição normal, variância adequada e possuírem significância estatística na análise da outra janela de sinais (dados da janela 1 ou dados da janela 2).

Como esperado, a adição de índices diretamente relacionados à VFC tende a aumentar o desempenho de outros conjuntos de índices utilizados na classificação. Alguns exemplos são os aumentos de desempenho dos grupos **B**  $\rightarrow$  **E** (média da acurácia entre os algoritmos de 51,72%  $\rightarrow$  69,87%); **D**  $\rightarrow$  **F** (média da acurácia entre os algoritmos de 60,35%  $\rightarrow$  63,21%).

Em relação aos 2 (dois) grupos especiais, o grupo **G** - com índices relacionados a

análise com *One-way Anova*, composto por 11 índices, teve o desempenho médio de acurácia de 73,86% nos algoritmos, alcançando o topo da classificação duas vezes, com os algoritmos SVM e KNN, e alcançando 80,72% de acurácia. Neste conjunto apenas um índice relacionado ao ABR foi utilizado, o LF\_DG\_ABR.

O grupo **H** - com índices relacionados ao desempenho individual dos índices na predição de indivíduos, composto por 6 índices, teve um desempenho similar com 5 índices a menos que o grupo **G**. O grupo **H** apresentando uma média de acurácia de 77,95% nos algoritmos, porém alcançando o topo da classificação apenas uma vez, como RL.

Comparando-se as métricas dos agrupamentos de índices por AAM, apresentadas em detalhes no Apêndice C, observa-se resultados semelhantes para os algoritmos SVM, KNN e RL. Porém, para o algoritmo de AD, agrupamentos de índices com presença de índices de modelos univariáveis forneceram métricas similares ou melhores (como é o caso do agrupamento **E** - VFC + VPA). Novamente, a análise do agrupamento **B** - associados à VPA, reforça a interpretação de que os índices da VPA isoladamente não apresentam robustez para a distinção entre os grupos de IPDM e IC quando avaliados por algoritmos clássicos de aprendizagem de máquina, permanecendo entre as piores métricas em todos os 4 (quatro) algoritmos.

em resumo, os agrupamentos de índices denominados especiais, **G** e **H**, se sobressaíram obtendo os melhores resultados de acurácia e medida-F1, dentre os demais agrupamentos em 3 dos AAM utilizados. O algoritmo **SVM** apresentou para o agrupamento **G** as respectivas métricas de acurácia (A), precisão (P), sensibilidade (S) e medida-F1 (F1): A=80,72%; P=79,47%; S=97,14%; F1=87,26%. Enquanto o agrupamento **H** apresentou os seguintes resultados: A=78,78%; P=77,53%; S=94,28%; F1=84,78%. Para o algoritmo **KNN**, o agrupamento **G** apresentou: A=80,72%; P=80,55%; S=93,80%; F1=86,5%. O agrupamento **H**: A=80,60%; P=79,47%; S=94,28%; F1=85,94%. Para o algoritmo **RL**, o agrupamento **G**: A=72,90%; S=88,57%; P=76,36%; F1=81,28%. O agrupamento **H**: A= 80,60%; P=80,47%; S=91,42%; F1=85,47%. A divergência de desempenho ocorreu com o algoritmo **AD**, onde as métricas de desempenho foram lideradas pelo agrupamento **E** contendo índices univariáveis da VFC e da VPA. As métricas do agrupamento **E** são: A=78,21%; P=86,66%; S=84,0%; F1=83,32%.

## 5 CONCLUSÃO

Na Seção 4.1, foi apresentada a análise dos índices sob a perspectiva da análise de correlação. Nela foi possível visualizar os relacionamentos entre as variáveis como um embasamento para a seleção de índices na composição de conjuntos de índices para a classificação de grupos. Ao apresentar os índices correlacionados utilizando regras empíricas baseadas em quartis 3.3.3, verificou-se que os índices provenientes de indivíduos portadores de diabete mellitus apresentam aglomerações de índices mais desconexos quando comparado ao grupo de controle. Isso reforça a estimativa de *Verrotti, 2014*, onde espera-se que grupos de indivíduos com T2DM possuam entre 20 e 70% portadores de neuropatia autonômica diabética [62]. O número de portadores da doença pode ser mais comum, como apontado por *Sukla, 2016*, referenciando estudos que apontam entre 50 e 77% dos indivíduos [58]. Esta informação não se encontra disponível na base de dados CDED, mas reforça que indivíduos diabéticos, mesmo quando não diagnosticados clinicamente com NA, apresentam disfunções autonômicas, onde se é possível utilizar índices quantitativos do sistema nervoso autônomo para predição entre IPDM e indivíduos controle.

Os resultados da Seção 4.2 se baseiam na análise estatística do método *One-way Anova* para explorar as características dos dados em relação a suas capacidades de distinguir com significância estatística entre os grupos de indivíduos. A partir disto foi possível identificar quais índices individualmente eram capazes de distinguir entre os grupos analisados. O *One-way Anova* depende da normalidade dos dados para sua análise, aplicando assim também um teste de normalidade e de variância sobre as variáveis a fim de verificar se há diferenças significativa entre as médias dos grupos analisados. Caso estes testes falhem, é possível tentar a normalização forçada dos dados aplicando  $\log_{10}X$  ou optando-se pela variação do método, o *One-way Anova on Ranks*, onde os requisitos de normalidade e variância dos dados não são obrigatórios. Esta análise resultou em aceitação ou não da hipótese nula, onde não há diferença entre as médias, porém por haver duplicação dos dados dos pacientes, e a metodologia exigir independência das observações, o *One-way Anova* foi aplicada duas vezes, uma para cada conjunto de dados presentes nos dois conjuntos de janelas de sinais dos indivíduos. Após comparação do resultado de suas hipóteses e sua significância estatística, aceitou-se ou não o índice como capaz de distinguir entre os grupos.

Daqui surgiram um conjunto de 11 (onze) índices que foram agrupados para integrar a predição dos grupos de indivíduos utilizando algoritmos de aprendizado de máquina. Os índices cuja hipótese alternativa foi aceita abrigam índices clássicos da literatura, como aqueles obtidos no domínio do tempo **MEDIA**, **MEDIANA** e **QUARTIL\_3**, conhecidos por sua robustez na identificação de indivíduos diabéticos, mas também índices resultantes de sistemas acoplados, como sistemas de influência mútua - MISO, acoplamento cardiorrespiratório e acoplamento do barorreflexo arterial. Nesta análise surgem os primeiros indícios da influência da respiração sobre a VFC. Infelizmente os resultados do *One-way Anova* não foram conclusivos para os índices mais conhecidos da literatura sobre acesso ao SNA como a BF e AF dos intervalos RR [62, 61, 12, 7]. Estes índices foram considerados indefinidos por apresentarem resultados amplamente incompatível entre as janelas de dados dos indivíduos. Aceitando a hipótese nula em uma janela e aceitando a hipótese alternativa na outra.

Na Seção 4.3.1 é introduzida a avaliação do desempenho individual dos índices na classificação dos grupos de IPDM e IC. 4 (quatro) AAM foram utilizados: 1) SVM, 2) AD, 3) KNN, e 4) RL. Os resultados obtidos aqui reforçam parcialmente os resultados da análise do *One-way Anova*, apresentando como índices com melhor desempenho os mesmos índices da VFC obtidos do domínio do tempo, como **MEDIANA** e **QUARTIL\_3**, além de diversos outros associados ao acoplamento da Respiração com a VFC. Os 6 (seis) índices de melhor desempenho individual foram combinados em um grupo especial para serem utilizados na seção de classificação utilizando agrupamentos de índices.

Na etapa final, apresentada na Seção 4.3.2, a classificação utilizando agrupamentos de índices mostrou como a combinação de índices extraídos de diferentes sinais de origens e metodologias matemáticas podem se complementar no aumento da acurácia e melhora nas métricas da matriz de difusão das predições dos indivíduos, fundamentando-se na métrica da medida F1. Como esperado, índices com origem unicamente em sistemas univariáveis, como o agrupamento de índices exclusivamente da VFC ou da VPA não obtiveram desempenhos significativos, porém quando combinados com outros índices de sistemas univariáveis ou multivariáveis, ampliaram as acurácias e medida F1 do classificador. Alguns exemplos de grupos com melhora significativa são os grupos **B** → **E**, e os grupos **D** → **F**. Porém o uso indiscriminado de índices por origem de sinal e metodologia de extração não são completamente eficazes, já que ao comparar os resultados obtidos pelos grupos especiais, **G** e **H**, contendo apenas selecionados dos índices obtidos no domínio do tempo e associados a respiração, foi possível melhorar o desempenho das métricas e alcançar o topo da classificação entre todos os algoritmos, exceto a AD, cuja combinação dos índices associados à VFC combinado com VPA alcançou o topo de acurácia e medida F1.

Por não conter informações sobre a presença de NAD nos indivíduos da base de da-

dos CDED, e apoiando nas evidências de disfunções autonômicas, há um menor grau de confiabilidade na busca por índices de diabéticos que divirjam dos índices de controle através do acesso exclusivo ao desacoplamento do controle autonômico. Mesmo para a base de dados CDED onde todos os indivíduos são de idosos, indicando a possibilidade de maior presença de disfunções autonômicas pela duração da diabetes, não foram evidentes as características esperadas como a diminuição da sensibilidade do acoplamento do mecanismo do barorreceptor.

Estima-se que a neuropatia autonômica cardiovascular esteja presente entre 20 a 70% dos indivíduos com T2DM [62], outros, como [58], apontam estudos com faixas mais estreitas, 50 a 77%. Ao avaliar valores associados com a NAC em indivíduos IPDM, Tang et al., (2014) [59] aponta que há uma tendência de presença de NAC em indivíduos com frequência cardíaca aumentada quando em repouso. Esta informação se refletiria uma menor média da VFC em indivíduos. Outros sinais da presença da NAC são a regulação anormal da pressão sanguínea e a diminuição da atuação do acoplamento do mecanismo barorreceptor na regulação, dada a sua baixa sensibilidade [62].

Tang et al., (2014) [59], também indica que nenhuma diferença significativa foi encontrada a partir do gênero dos indivíduos com NAC ou sem NAC, assim, ao não considerar diferenças entre os gêneros dos indivíduos nos índices de acesso ao controle autonômico, não haveria prejuízo significativo a este estudo. Entretanto, é afirmado que a VFC diminui com o aumento da idade. Assim, para uma base contendo dados de indivíduos idosos, espera-se uma medida da VFC menor que a população geral, como foi indicado pelas variáveis da VFC.

Van Dijk et al., (1991) [61], analisou os efeitos da pressão sanguínea e da frequência cardíaca sobre o balanço autonômico cardíaco, e mostrou a prevalência de atividade simpática em indivíduos portadores de disfunções autonômicas. O motivo deste desbalanço é a pode ser a manifestação da NAD, que ataca primeiro nervos longos, como o nervo vago que produz a maior parte da atividade parassimpática do coração [62], resultando numa elevação da atividade simpática na regulação. Infelizmente este resultado não pode ser confirmado, pois para esta base de dados, os índices relacionados ao domínio da frequência apresentaram baixo desempenho na classificação e nas análises estatísticas.

Na respiração, a central de controle responsável pela ventilação também pode ser afetada por disfunções autonômicas, afetando a efetividade ação autônoma sobre a respiração. Obviamente, a respiração possui uma atuação forte do controle voluntário do indivíduos, porém a deficiências de sua ação autonômica pode levar ao desenvolvimento de outras doenças como a síndrome da apneia do sono (SAS), momento onde a respiração está majoritariamente no "modo automático". A SAS é bastante prevalente em indivíduos com disfunções autonômicas, especialmente com NAD, e bastante prevalente

em indivíduos com T2DM.

Khoo et al., (2016) [31], aponta que para modelos relacionados às síndromes respiratórias, similares aos modelos acoplados utilizados neste trabalho, as componentes de BF e AF dos ganhos dos sistemas ACR e do sistema de influência mútua - MISO são compatíveis com o desacoplamento dos mecanismo cardiorrespiratório quando há aumento dos ganhos em ambas as bandas de frequência, juntamente com a presença de um aumento da atividade do ramo parassimpático.

A seleção de atributos é essencial para a utilização de AAM. Em problemas como classificação e diagnósticos de patologias em indivíduos, os dados utilizados para treinamento do modelo são cruciais e podem resultar em desempenhos significativos sem a utilização de modelos mais complexos. A compreensão e exploração de como marcadores de sistemas fisiológicos estão ligados a diabetes, e como suas alterações resultam em maior impacto na saúde geral do indivíduo, pode indicar quais dados devem ser considerados com maior atenção ao desenvolver novos modelos de aprendizado de máquina, e quais podem ser dispensados a depender do objetivo.

## 5.1 PERSPECTIVAS FUTURAS

Para uma investigação mais detalhada, sugere-se aplicar a metodologia utilizada neste trabalho em bases de dados diversas relacionadas à Diabetes Mellitus Tipo 2. Este trabalho utilizou a base CDED cujo enfoque é a Diabetes Mellitus Tipo 2 em indivíduos idosos, entre 55 e 75 anos. A comparação dos resultados deste trabalho pode ser validados também contra dados de indivíduos jovens e até mesmo em dados de portadores de Diabetes Mellitus Tipo 1.

Ao utilizar uma base de dados com dados de indivíduos idosos, fica evidente a existência de divergências entre IPDM e IC em flutuações fisiológicas relacionadas ao desacoplamento cardiorrespiratório. Essas divergências foram avaliadas por diferentes métodos como uma categoria de índices interessante na distinção entre os grupos estudados.

Por fim, sugere-se um aperfeiçoamento metodológico em relação a estimação de parâmetros utilizados para estimação dos sistemas acoplados utilizando expansões por bases de Laguerre. Entende-se que índices extraídos de sistemas multivariáveis, como aqueles estimados por resposta ao impulso contenha flutuações que possibilitem uma caracterização mais aprofundada da ação autonômica em indivíduos com diabetes, principalmente em relação a interferência da respiração sobre a regulação cardiovascular.

Os modelos multivariáveis utilizados para este trabalho apresentaram baixo ajuste ao sinal de IRR em sua saída, mesmo com a implementação recursiva de seleção de polos



e tamanho de memória para os modelos combinado com a seleção de parâmetros pelo melhor método (MDL e AIC). Para contornar o problema de ajuste dos modelos paramétricos, sugere-se a exploração de diferentes modelos ou diferentes critérios de seleção de parâmetros para o modelo.

## LISTA DE REFERÊNCIAS

- [1] 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, Mar 1996.
- [2] Raphael Martins de Abreu, Beatrice Cairo, e Alberto Porta. On the significance of estimating cardiorespiratory coupling strength in sports medicine. *Frontiers in Network Physiology*, 2, January 2023. <http://dx.doi.org/10.3389/fnetp.2022.1114733>.
- [3] Shruti Agashe e Steven Petak. Cardiac autonomic neuropathy in diabetes mellitus. *Methodist DeBakey Cardiovascular Journal*, 14(4):251, October 2018. <http://dx.doi.org/10.14797/mdcj-14-4-251>.
- [4] Kalaiarasi Sonai Muthu Anbananthen, Mikail Bin Muhammad Azman Busst, Rajkumar Kannan, e Subarmaniam Kannan. A comparative performance analysis of hybrid and classical machine learning method in predicting diabetes. *Emerging Science Journal*, 7(1):102–115, October 2022. <http://dx.doi.org/10.28991/ESJ-2023-07-01-08>.
- [5] American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 35(Supplement\_1):S64–S71, December 2011. <http://dx.doi.org/10.2337/dc12-s064>.
- [6] Saba Bashir, Usman Qamar, e Farhan Hassan Khan. A multicriteria weighted vote-based classifier ensemble for heart disease prediction. *Computational Intelligence*, 32(4):615–645, June 2015. <http://dx.doi.org/10.1111/coin.12070>.
- [7] Vasily Belozeroff, Richard B. Berry, e Michael C.K. Khoo. Model-based assessment of autonomic control in obstructive sleep apnea syndrome. *Sleep*, 26(1):65–73, January 2003. <http://dx.doi.org/10.1093/sleep/26.1.65>.
- [8] Bernard. *Fundamentals of Biostatistics*. CENGAGE Learning Custom Publishing, Mason, OH, 8 edição, August 2015.

- [9] Sandro Muniz Cavalcanti. Quantificação do sistema nervoso autônomo no diabetes mellitus tipo 2. <https://bdm.unb.br/handle/10483/34504>, September 2022.
- [10] Victor Chang, Jozeene Bailey, Qianwen Ariel Xu, e Zhili Sun. Pima indians diabetes mellitus classification based on machine learning (ML) algorithms. *Neural Computing and Applications*, 35(22):16157–16173, March 2022. <http://dx.doi.org/10.1007/s00521-022-07049-z>.
- [11] Xuewei Chao e Yang Li. Semisupervised few-shot remote sensing image classification based on knn distance entropy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:8798–8805, 2022. <http://dx.doi.org/10.1109/JSTARS.2022.3213749>.
- [12] Zhe Chen, E.N. Brown, e R. Barbieri. Assessment of autonomic control and respiratory sinus arrhythmia using point process models of human heart beat dynamics. *IEEE Transactions on Biomedical Engineering*, 56(7):1791–1802, July 2009. <http://dx.doi.org/10.1109/TBME.2009.2016349>.
- [13] Chih-Jen Lin Chih-Wei Hsu, Chih-Chung Chang. A practical guide to support vector classification. Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, apr 2010. Guide and Instructions.
- [14] G Clifford. *Signal processing methods for heart rate variability*. PhD thesis, Oxford University, UK, 2002.
- [15] G.D. Clifford e L. Tarassenko. Quantifying errors in spectral estimates of hrv due to beat replacement and resampling. *IEEE Transactions on Biomedical Engineering*, 52(4):630–638, April 2005. <http://dx.doi.org/10.1109/TBME.2005.844028>.
- [16] The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *The Lancet*, 375(9733):2215–2222, June 2010. [http://dx.doi.org/10.1016/S0140-6736\(10\)60484-9](http://dx.doi.org/10.1016/S0140-6736(10)60484-9).
- [17] Luisa Santiago C. B. da Silva e Flavia Maria G. S. Oliveira. CRSIDLab: A toolbox for multivariate autonomic nervous system analysis using cardiorespiratory identification. *IEEE Journal of Biomedical and Health Informatics*, 24(3):728–734, March 2020. <http://dx.doi.org/10.1109/JBHI.2019.2914211>.
- [18] Peter Dalgaard. *Analysis of variance and the Kruskal–Wallis test*, page 127–143. Springer New York, 2008. [http://dx.doi.org/10.1007/978-0-387-79054-1\\_7](http://dx.doi.org/10.1007/978-0-387-79054-1_7).

- [19] Ralph A. DeFronzo, Ele Ferrannini, Leif Groop, Robert R. Henry, William H. Herman, Jens Juul Holst, Frank B. Hu, C. Ronald Kahn, Itamar Raz, Gerald I. Shulman, Donald C. Simonson, Marcia A. Testa, e Ram Weiss. Type 2 diabetes mellitus. *Nature Reviews Disease Primers*, 1(1), July 2015. <http://dx.doi.org/10.1038/nrdp.2015.19>.
- [20] J. C. Echeverría, J. A. Crowe, M. S. Woolfson, e B. R. Hayes-Gill. Application of empirical mode decomposition to heart rate variability analysis. *Medical and Biological Engineering and Computing*, 39(4):471–479, July 2001. <http://dx.doi.org/10.1007/BF02345370>.
- [21] Mohamed Elgendi, Yongbo Liang, e Rabab Ward. Toward generating more diagnostic features from photoplethysmogram waveforms. *Diseases*, 6(1):20, March 2018. <http://dx.doi.org/10.3390/diseases6010020>.
- [22] Charles Faselis, Alexandra Katsimardou, Konstantinos Imprialos, Pavlos Deligkaris, Manolis Kallistratos, e Kiriakos Dimitriadis. Microvascular complications of type 2 diabetes mellitus. *Current Vascular Pharmacology*, 18(2):117–124, January 2020. <http://dx.doi.org/10.2174/1570161117666190502103733>.
- [23] Samuel M. Fox e John P. Naughton. Physical activity and the prevention of coronary heart disease. *Preventive Medicine*, 1(1–2):92–120, March 1972. [http://dx.doi.org/10.1016/0091-7435\(72\)90079-5](http://dx.doi.org/10.1016/0091-7435(72)90079-5).
- [24] Orlando Iparraguirre-Villanueva, Karina Espinola-Linares, Rosalynn Ornella Flores Castañeda, e Michael Cabanillas-Carbonell. Application of machine learning models for early detection and accurate classification of type 2 diabetes. *Diagnostics*, 13(14):2383, July 2023. <http://dx.doi.org/10.3390/diagnostics13142383>.
- [25] Aliferis Constantin J. Simon, Gyorgy. *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences Best Practices and Pitfalls*. Springer International Publishing, 2024. <http://dx.doi.org/10.1007/978-3-031-39355-6>.
- [26] Vardhmaan Jain, Abdul Mannan Khan Minhas, Robert W. Ariss, Salik Nazir, Safi U. Khan, Muhammad Shahzeb Khan, Mahmoud Al Rifai, Erin Michos, Anurag Mehta, Arman Qamar, Elizabeth M. Vaughan, Laurence Sperling, e Salim S. Virani. Demographic and regional trends of cardiovascular diseases and diabetes mellitus-related mortality in the united states from 1999 to 2019. *The American Journal of Medicine*, 136(7):659–668.e7, July 2023. <http://dx.doi.org/10.1016/j.amjmed.2023.03.002>.
- [27] Dae-Geun Jang, Minsoo Hahn, Jae-Keun Jang, U. Farooq, e Seung-Hun Park. A comparison of interpolation techniques for rr interval fitting in ar spectrum estima-

- tion. In *2012 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, November 2012. <http://dx.doi.org/10.1109/BioCAS.2012.6418424>.
- [28] Devi Kalyan Karumanchi, Elizabeth R. Gaillard, e James Dillon. Early diagnosis of diabetes through the eye. *Photochemistry and Photobiology*, 91(6):1497–1504, October 2015. <http://dx.doi.org/10.1111/php.12524>.
- [29] Anastasia Katsarou, Soffia Gudbjörnsdottir, Araz Rawshani, Dana Dabelea, Ezio Bonifacio, Barbara J. Anderson, Laura M. Jacobsen, Desmond A. Schatz, e Åke Lernmark. Type 1 diabetes mellitus. *Nature Reviews Disease Primers*, 3(1), March 2017. <http://dx.doi.org/10.1038/nrdp.2017.16>.
- [30] Michael C. K. Khoo, Flavia M. G. S. Oliveira, e Limei Cheng. Understanding the metabolic syndrome: A modeling perspective. *IEEE Reviews in Biomedical Engineering*, 6:143–155, 2013. <http://dx.doi.org/10.1109/RBME.2012.2232651>.
- [31] Michael C.K. Khoo e Patjanaporn Chalacheva. Model-derived markers of autonomic cardiovascular dysfunction in sleep-disordered breathing. *Sleep Medicine Clinics*, 11(4):489–501, December 2016. <http://dx.doi.org/10.1016/j.jsmc.2016.07.003>.
- [32] Ayllah A. Lopes. Estimaco da sensibilidade do barorreflexo variante no tempo. <https://bdm.unb.br/handle/10483/33650>, May 2022.
- [33] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, e P. J. Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, March 1996. <http://dx.doi.org/10.1093/oxfordjournals.eurheartj.a014868>.
- [34] Diego Mariano. *Mtricas de avaliao em machine learning: acurcia, sensibilidade, preciso, especificidade e F-score*. Alfahelix, July 2021. <http://dx.doi.org/10.51780/978-6-599-275326-15>.
- [35] Raelene E. Maser, Braxton D. Mitchell, Aaron I. Vinik, e Roy Freeman. The association between cardiovascular autonomic neuropathy and mortality in individuals with diabetes. *Diabetes Care*, 26(6):1895–1901, June 2003. <http://dx.doi.org/10.2337/diacare.26.6.1895>.
- [36] Elizabeth J. Mayer-Davis, Jean M. Lawrence, Dana Dabelea, Jasmin Divers, Scott Isom, Lawrence Dolan, Giuseppina Imperatore, Barbara Linder, Santica Marcovina, David J. Pettitt, Catherine Pihoker, Sharon Saydah, e Lynne Wagenknecht. Incidence trends of type 1 and type 2 diabetes among youths, 2002–2012. *New England Journal of Medicine*, 376(15):1419–1429, April 2017. <http://dx.doi.org/10.1056/nejmoa1610187>.

- [37] Rosália Garcia Neves, Elaine Tomasi, Suele Manjourany Silva Duro, Elizabet Saes-Silva, e Mirelle de Oliveira Saes. Complicações por diabetes mellitus no brasil: estudo de base nacional, 2019. *Ciência & Saúde Coletiva*, 28(11):3183–3190, November 2023. <http://dx.doi.org/10.1590/1413-812320232811.11882022>.
- [38] Vera Novak e Rodrigo Quispe. Cerebromicrovascular disease in elderly with diabetes, 2022. <https://physionet.org/content/cded/1.0.1/>.
- [39] Vera Novak, Peng Zhao, Brad Manor, Ervin Sejdić, David Alsop, Amir Abduljalil, Paula K. Roberson, Medha Munshi, e Peter Novak. Adhesion molecules, altered vasoreactivity, and brain atrophy in type 2 diabetes. *Diabetes Care*, 34(11):2438–2441, October 2011. <http://dx.doi.org/10.2337/dc11-0969>.
- [40] Chollette C. Olisah, Lyndon Smith, e Melvyn Smith. Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Computer Methods and Programs in Biomedicine*, 220:106773, June 2022. <http://dx.doi.org/10.1016/j.cmpb.2022.106773>.
- [41] World Health Organization. *Global report on diabetes*. World Health Organization, 2016.
- [42] World Health Organization. *Hearts d: diagnosis and management of type 2 diabetes*. Technical documents, World Health Organization, 2020.
- [43] Amanda dos Santos Pereira. Estimaco da sensibilidade do barorreflexo variante no tempo. <https://bdm.unb.br/handle/10483/34429>, October 2022.
- [44] Tam Pham, Zen Juen Lau, SH Annabel Chen, e Dominique Makowski. Unveiling the structure of heart rate variability (HRV) indices: A data-driven meta-clustering approach. October 2021. <http://dx.doi.org/10.31234/osf.io/mwa6x>.
- [45] P J Rees, J G Prior, G M Cochrane, e T J H Clark. Sleep apnoea in diabetic patients with autonomic neuropathy. *Journal of the Royal Society of Medicine*, 74(3):192–195, March 1981. <http://dx.doi.org/10.1177/014107688107400306>.
- [46] Ana Leonor Rivera, Bruno Estaol, Ruben Fossion, Juan C. Toledo-Roy, Jos A. Callejas-Rojas, Jos A. Gien-Lpez, Guillermo R. Delgado-García, e Alejandro Frank. Loss of breathing modulation of heart rate variability in patients with recent and long standing diabetes mellitus type II. *PLOS ONE*, 11(11):e0165904, November 2016. <http://dx.doi.org/10.1371/journal.pone.0165904>.
- [47] Charushila Rukadikar, Atul Rukadikar, e Surekha Kishore. A review on autonomic functional assessment in diabetic patients. *Cureus*, February 2023. <http://dx.doi.org/10.7759/cureus.34598>.

- [48] Ayllah A.; Silva Luisa S. C. B. Santos, Amanda; Lopes. *CRSIDLab 2.0 – Cardiorespiratory System Identification Lab User Guide*. University of Brasilia, Campus Universitário Darcy Ribeiro Asa Norte, Brasília-DF, CEP 70910-900, 2 edição, 2022.
- [49] Emily B. Schroeder, Lloyd E. Chambless, Duanping Liao, Ronald J. Prineas, Gregory W. Evans, Wayne D. Rosamond, e Gerardo Heiss. Diabetes, glucose, insulin, and heart rate variability. *Diabetes Care*, 28(3):668–674, March 2005. <http://dx.doi.org/10.2337/diacare.28.3.668>.
- [50] Daniel Shookster, Bryndan Lindsey, Nelson Cortes, e Joel R Martin. Accuracy of commonly used age-predicted maximal heart rate equations. *Int. J. Exerc. Sci.*, 13(7):1242–1250, September 2020.
- [51] Ikaro Silva, Benjamin Moody, e George Moody. Waveform database software package (WFDB) for MATLAB and Octave, 2021. <https://physionet.org/content/wfdb-matlab/0.10.0/>.
- [52] Ikaro Silva e George B. Moody. An open-source toolbox for analysing and processing physionet databases in MATLAB and Octave. *Journal of Open Research Software*, 2, September 2014. <http://dx.doi.org/10.5334/jors.bi>.
- [53] Luisa S. C. B. SILVA. Interface gráfica para análise de sinais do sistema cardiorrespiratório. <https://bdm.unb.br/handle/10483/15299/>, July 2015.
- [54] Luisa S. C. B. SILVA. Quantitative analysis of the autonomic nervous system : toolbox development and application. <http://repositorio2.unb.br/jspui/handle/10482/31424>, November 2017.
- [55] D.U. Silverthorn e V. Tradutores. *Fisiologia Humana: UMA ABORDAGEM INTEGRADA*. ARTMED, 2017. <https://books.google.com.br/books?id=mWpcswEACAAJ>.
- [56] Manoocher Soleimani. Insulin resistance and hypertension: new insights. *Kidney International*, 87(3):497–499, March 2015. <http://dx.doi.org/10.1038/ki.2014.392>.
- [57] Harald M Stauss. Physiologic mechanisms of heart rate variability. *Rev Bras Hipertens*, 14(1):8–15, 2007.
- [58] Pradeep Sukla, Saurabh RamBihariLal Shrivastava, Prateek Saurabh Shrivastava, e Nambaru Lakshmana Rao. Assessment of the cardiac autonomic neuropathy among the known diabetics and age-matched controls using noninvasive cardiovascular reflex tests in a south-indian population: A case-control study. *Avicenna Journal of Medicine*, 06(03):81–85, July 2016. <http://dx.doi.org/10.4103/2231-0770.184067>.

- [59] Zi-Hui Tang, Fangfang Zeng, Zhongtao Li, e Linuo Zhou. Association and predictive value analysis for resting heart rate and diabetes mellitus on cardiovascular autonomic neuropathy in general population. *Journal of Diabetes Research*, 2014:1–7, 2014. <http://dx.doi.org/10.1155/2014/215473>.
- [60] Edwin van den Heuvel e Zhuozhao Zhan. Myths about linear and monotonic associations: Pearson’s  $r$ , Spearman’s  $\rho$ , and Kendall’s  $\tau$ . *The American Statistician*, 76(1):44–52, January 2022. <http://dx.doi.org/10.1080/00031305.2021.2004922>.
- [61] J.Gert van Dijk, Monique Koenderink, Aeilko H. Zwinderman, Joost Haan, Cor G.S. Kramer, e Jan C. den Heijer. Autonomic nervous system tests depend on resting heart rate and blood pressure. *Journal of the Autonomic Nervous System*, 35(1):15–24, July 1991. [http://dx.doi.org/10.1016/0165-1838\(91\)90034-z](http://dx.doi.org/10.1016/0165-1838(91)90034-z).
- [62] Alberto Verrotti, Giovanni Prezioso, Raffaella Scattoni, e Francesco Chiarelli. Autonomic neuropathy in diabetes mellitus. *Frontiers in Endocrinology*, 5, December 2014. <http://dx.doi.org/10.3389/fendo.2014.00205>.
- [63] M. P. Villa, G. Multari, M. Montesano, J. Pagani, M. Cervoni, F. Midulla, E. Cerone, e R. Ronchetti. Sleep apnoea in children with diabetes mellitus: effect of glycaemic control. *Diabetologia*, 43(6):696–702, June 2000. <http://dx.doi.org/10.1007/s001250051365>.
- [64] S. Villwock, M. Pacas, e T. Eutebach. Application of the Welch-method for the automatic parameter identification of electrical drives. In *31st Annual Conference of IEEE Industrial Electronics Society, 2005. IECON 2005*. IEEE, 2005. <http://dx.doi.org/10.1109/IECON.2005.1569118>.
- [65] Philip J Weston, Martin A James, Ronnie B Panerai, Paul G McNally, John F Potter, e Herbert Thurston. Evidence of defective cardiovascular regulation in insulin-independent diabetic patients without clinical autonomic dysfunction. *Diabetes Research and Clinical Practice*, 42(3):141–148, December 1998. [http://dx.doi.org/10.1016/S0168-8227\(98\)00094-1](http://dx.doi.org/10.1016/S0168-8227(98)00094-1).
- [66] Yung-Jhe Yan, Chien-Nan Chen, e Mang Ou-Yang. Using system identification to construct an inherent model of pupillary light reflex to explore diabetic neuropathy. *Brain Sciences*, 11(7):852, June 2021. <http://dx.doi.org/10.3390/brainsci11070852>.



## APÊNDICE A

### ÍNDICES RESULTANTES DO PRÉ-PROCESSAMENTO

Índice	Descrição
MEDIA_RRI	Média do sinal filtrado e reamostrado dos dados do intervalos RR. (ms)
MEDIANA_RRI	Mediana do sinal filtrado e reamostrado dos dados do intervalos RR. (ms)
SDNN	Desvio padrão da média de todos os intervalos RR normalizados. (ms)
MINIMO_RRI	Valor mínimo do sinal filtrado e reamostrado contendo o intervalos RR. (ms)
MAXIMO_RRI	Valor máximo do sinal filtrado e reamostrado contendo o intervalos RR. (ms)
QUARTIL_1_RRI	Medida estatística de dispersão do sinal contendo o IRR. (ms)
QUARTIL_3_RRI	Medida estatística de dispersão do sinal contendo o IRR. (ms)
PNN50	Porcentagem de intervalos RR sucessivos com diferença na duração maior que 50 ms. (%)
RMSSD	É a raiz quadrada da média dos quadrados das diferenças sucessivas de um IRR. (ms)
COEFICIENTE_VARIANCIA_RRI	Variabilidade (dispersão) do IRR em relação a média. (%)
FAIXA_DINAMICA_RRI	Diferença entre o valor mais alto e o valor mais baixo (amplitude) do sinal de IRR.(ms)
SD_1	Parâmetro da plotagem de Poincaré. Avalia a variabilidade da variabilidade dos dados. (ms)
SD_2	Parâmetro da plotagem de Poincaré. Reflete dos ramos simpáticos e parassimpáticos do SNA. (ms)

<b>Índice</b>	<b>Descrição</b>
LF_RRI_WELCH	Área da banda de frequência de baixa frequência (0,04Hz-0.15Hz) do sinal de IRR utilizando o método de Welch para análise espectral de potência. ( $ms^2$ ou n.u.)
HF_RRI_WELCH	Área da banda de frequência de alta frequência (0,15Hz-0.4Hz) do sinal de IRR utilizando o método de Welch para análise espectral de potência. ( $ms^2$ ou n.u.)
LF_HF_RATIO_RRI_WELCH	Razão entre as área de baixa frequência e alta frequência do sinal de IRR utilizando o método de Welch para análise espectral de potência. (n.u.)
LF_SBP_WELCH	Área da banda de frequência de baixa frequência (0,04Hz-0.15Hz) do sinal de pressão arterial sistólica utilizando o método de Welch para análise espectral de potência. ( $mmHg^2$ ou n.u.)
HF_SBP_WELCH	Área da banda de frequência de alta frequência (0,15Hz-0.4Hz) do sinal de pressão arterial sistólica utilizando o método de Welch para análise espectral de potência. ( $mmHg^2$ ou n.u.)
LF_HF_RATIO_SBP_WELCH	Razão entre as área de baixa frequência e alta frequência do sinal de pressão arterial sistólica utilizando o método de Welch para análise espectral de potência. (n.u.)
LF_DBP_WELCH	Área da banda de frequência de baixa frequência (0,04Hz-0.15Hz) do sinal de pressão arterial diastólica utilizando o método de Welch para análise espectral de potência. ( $mmHg^2$ ou n.u.)

<b>Índice</b>	<b>Descrição</b>
HF_DBP_WELCH	Área da banda de frequência de alta frequência (0,15Hz-0,4Hz) do sinal de pressão arterial diastólica utilizando o método de Welch para análise espectral de potência. ( <i>mmHg</i> <sup>2</sup> ou n.u.)
LF_HF_RATIO_DBP_WELCH	Razão entre as área de baixa frequência e alta frequência do sinal de pressão arterial diastólica utilizando o método de Welch para análise espectral de potência. (n.u.)
IRM_SBP_ACOPLAM	Amplitude da resposta do sistema acoplado (acoplamento do barorreflexo arterial + acoplamento cardiorrespiratório) a um impulso unitário aplicado na entrada da PAS. (ms/mmHg)
LF_DG_SBP_ACOPLAM	Média dos ganhos da função de transferência entre 0,04Hz - 0,15Hz obtida da FFT da resposta ao impulso submetida a entrada PAS sistema acoplado. (ms/mmHg)
HF_DG_SBP_ACOPLAM	Média dos ganhos da função de transferência entre 0,15Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida a entrada PAS sistema acoplado. (ms/mmHg)
DYNAMIC_GAIN_SBP_ACOPLAM	Média dos ganhos da função de transferência entre 0,04Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida a entrada PAS sistema acoplado. (ms/mmHg)
IRM_ILV_ACOPLAM	Amplitude da resposta do sistema acoplado (acoplamento do barorreflexo arterial + acoplamento cardiorrespiratório) a um impulso unitário aplicado na entrada da volume pulmonar instantâneo. (ms/L)

<b>Índice</b>	<b>Descrição</b>
LF_DG_ILV_ACOPLAM	Média dos ganhos da função de transferência entre 0,04Hz - 0,15Hz obtida da FFT da resposta ao impulso submetida a entrada do VPI sistema acoplado. (ms/L)
HF_DG_ILV_ACOPLAM	Média dos ganhos da função de transferência entre 0,15Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida a entrada do VPI sistema acoplado. (ms/L)
DYNAMIC_GAIN_ILV_ACOPLAM	Média dos ganhos da função de transferência entre 0,04Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida a entrada do VPI sistema acoplado. (ms/L)
IRM_ABR	Amplitude da resposta do sistema acoplado (acoplamento do barorreflexo arterial) a um impulso unitário aplicado na entrada. (ms/mmHg)
LF_DG_ABR	Média dos ganhos da função de transferência entre 0,04Hz - 0,15Hz obtida da FFT da resposta ao impulso submetida ao ABR. (ms/mmHg)
HF_DG_ABR	Média dos ganhos da função de transferência entre 0,15Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida ao ABR. (ms/mmHg)
DYNAMIC_GAIN_ABR	Média dos ganhos da função de transferência entre 0,04Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida ao ABR. (ms/mmHg)
IRM_RCC	Amplitude da resposta do sistema acoplado (ACR) a um impulso unitário aplicado na entrada. (ms/L)

<b>Índice</b>	<b>Descrição</b>
LF_DG_RCC	Média dos ganhos da função de transferência entre 0,04Hz - 0,15Hz obtida da FFT da resposta ao impulso submetida ao ACR. (ms/L)
HF_DG_RCC	Média dos ganhos da função de transferência entre 0,15Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida ao ACR. (ms/L)
DYNAMIC_GAIN_RCC	Média dos ganhos da função de transferência entre 0,04Hz - 0,4Hz obtida da FFT da resposta ao impulso submetida ao ACR. (ms/L)

**Tabela A.1.** Marcadores obtidos durante a etapa de pré-processamento.

## APÊNDICE B

### AGRUPAMENTO DE ÍNDICES PARA TESTE DE CLASSIFICAÇÃO EM GRUPO

Grupo	Índices	Descrição
<b>VFC [HRV] (a)</b>	MEDIA	Índices relacionados exclusivamente à variabilidade da frequência cardíaca. No domínio do tempo, no domínio da frequência.
	MEDIANA	
	SDNN	
	MINIMO	
	MAXIMO	
	QUARTIL_1	
	QUARTIL_3	
	pNN50	
	RMSSD	
	COEFICIENTE_VARIANCIA	
	FAIXA_DINAMICA	
	SD_1	
	SD_2	
	LF_RRI_WELCH	
HF_RRI_WELCH		
LF_HF_RATIO_RRI_WELCH		
<b>VPA [BPV] (b)</b>	LF_SBP_WELCH	Índices relacionados à variabilidade da pressão arterial. Abrange índices no domínio da frequência das pressões sistólicas e diastólicas.
	HF_SBP_WELCH	
	LF_HF_RATIO_SBP_WELCH	
	LF_DBP_WELCH	
	HF_DBP_WELCH	
	LF_HF_RATIO_DBP_WELCH	
<b>FRF [FRF] (c)</b>	LF_RRI_WELCH	Índices obtidos através a análise espectral dos sinais de IRR, PAS, PAD, e das respostas ao impulso dos sistemas acoplados.
	HF_RRI_WELCH	
	LF_HF_RATIO_RRI_WELCH	
	LF_SBP_WELCH	
	HF_SBP_WELCH	
	LF_HF_RATIO_SBP_WELCH	
	LF_DBP_WELCH	
	HF_DBP_WELCH	
	LF_HF_RATIO_DBP_WELCH	
	LF_DG_SBP_ACOPLAM	

Grupo	Índices	Descrição
	HF_DG_SBP_ACOPLAM	
	LF_DG_ABR	
	HF_DG_ABR	
	LF_DG_RCC	
	HF_DG_RCC	
<b>RI [IR] (d)</b>	IRM_SBP_ACOPLAM	Índices relacionados exclusivamente às Respostas ao Impulso dos sistemas acoplados.
	LF_DG_SBP_ACOPLAM	
	HF_DG_SBP_ACOPLAM	
	DYNAMIC_GAIN_SBP_ACOPLAM	
	IRM_ILV_ACOPLAM	
	LF_DG_ILV_ACOPLAM	
	HF_DG_ILV_ACOPLAM	
	DYNAMIC_GAIN_ILV_ACOPLAM	
	IRM_ABR	
	LF_DG_ABR	
	HF_DG_ABR	
	DYNAMIC_GAIN_ABR	
	LF_DG_RCC	
	HF_DG_RCC	
DYNAMIC_GAIN_RCC		
<b>VFC + VPA [HRV + BPV] (e)</b>	Grupo VFC	Combinação de grupos de índices relacionados à variabilidade da frequência cardíaca e à variabilidade da pressão arterial. A junção destes grupos implicam no acoplamento do baroreflexo arterial.
	Grupo VPA	
<b>VFC + RI [HRV + IR] (f)</b>	Grupo VFC	Combinação de grupos de índices relacionados à variabilidade da frequência cardíaca e respostas ao impulso dos sistemas acoplados.
	Grupo RI	

<b>Grupo</b>	<b>Índices</b>	<b>Descrição</b>
<b>One-way Anova (Especial 1 - g)</b>	MEDIA	Índices identificados pelo One-way Anova como capazes de distinguir entre as classes IPDM e IC.
	MEDIANA	
	QUARTIL_3	
	LF_DG_ILV_ACOPLAM	
	HF_DG_ILV_ACOPLAM	
	DYNAMIC_GAIN_ILV_ACOPLAM	
	LF_DG_ABR	
	IRM_RCC	
	LF_DG_RCC	
	HF_DG_RCC	
	DYNAMIC_GAIN_RCC	
<b>Classificação Individual [Single] (Especial 2 - h)</b>	HF_DG_ILV_ACOPLAM	Índices que obtiveram alto desempenho individual na classificação das classes de IPDM e IC e apareceram em entre os melhores índices para todos os algoritmos de aprendizado de máquina.
	DYNAMIC_GAIN_ILV_ACOPLAM	
	DYNAMIC_GAIN_RCC	
	HF_DG_RCC	
	QUARTIL_3	
	MEDIANA	

**Tabela B.1.** Combinação de índices por grupo para teste de desempenho na classificação de IPDM e IC.



## APÊNDICE C

### MÉTRICAS INTEGRAIS POR ALGORITMOS DE APRENDIZADO DE MÁQUINA DOS AGRUPAMENTOS DE ÍNDICES

Support Vector Machines								
Marcador	Algoritmo	Acurácia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE	
G - One_way_Anova	SVM	0,807272727	0,794761905	0,971428571	0,8726546	34	17	
H - Single_Indexes_Group	SVM	0,787878788	0,77531746	0,942857143	0,84780112	35	21	
F - HRV_IR_Indexes	SVM	0,738095238	0,738095238	1	0,848484848	25	9	
D - IR_Indexes	SVM	0,672222222	0,694047619	0,933333333	0,792967033	29	13	
E - HRV_BPV_Indexes	SVM	0,667857143	0,672619048	0,96	0,786573427	23	13	
A - HRV_Indexes	SVM	0,645833333	0,648661616	0,908333333	0,751086835	44	32	
C - FRF_Indexes	SVM	0,608333333	0,628333333	0,92	0,741538462	26	15	
B - BPV_Indexes	SVM	0,466666667	0,513333333	0,626666667	0,559487179	27	22	

**Figura C.1.** Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para *support vector machines* (SVM). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias.

Árvore de Decisão							
Marcador	Marcador	Acurácia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE
E - HRV_BPV_Indexes	RL	0,782143	0,866667	0,84	0,833232	23	13
H - Single_Indexes_Group	RL	0,681818	0,75881	0,714286	0,730989	35	21
A - HRV_Indexes	RL	0,633333	0,677692	0,797222	0,725969	44	32
F - HRV_IR_Indexes	RL	0,614286	0,73619	0,76	0,738788	25	9
C - FRF_Indexes	RL	0,583333	0,66619	0,693333	0,670909	26	15
G - One_way_Anova	RL	0,570909	0,71873	0,647619	0,672619	34	17
D - IR_Indexes	RL	0,547222	0,686667	0,613333	0,646465	29	13
B - BPV_Indexes	RL	0,544444	0,610952	0,593333	0,593939	27	22

**Figura C.2.** Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para árvore de decisão (AD). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias.

K-Nearest Neighbor							
Marcador	Algoritmo	Acurácia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE
G - One_way_Anova	KNN	0,807272727	0,805555556	0,938095238	0,865	34	17
H - Single_Indexes_Group	KNN	0,806060606	0,794761905	0,942857143	0,859467787	35	21
F - HRV_IR_Indexes	KNN	0,704761905	0,748571429	0,92	0,82	25	9
A - HRV_Indexes	KNN	0,673333333	0,728156288	0,794444444	0,744343434	44	32
E - HRV_BPV_Indexes	KNN	0,671428571	0,721904762	0,88	0,784242424	23	13
D - IR_Indexes	KNN	0,597222222	0,694285714	0,753333333	0,715198135	29	13
C - FRF_Indexes	KNN	0,561111111	0,633333333	0,733333333	0,678787879	26	15
B - BPV_Indexes	KNN	0,546666667	0,572619048	0,66	0,60952381	27	22

**Figura C.3.** Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para *k-nearest neighbor* (KNN). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias.

Regressão Logística							
Marcador	Algoritmo	Acurácia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE
H - Single_Indexes_Group	RL	0,806060606	0,804761905	0,914285714	0,854761905	35	21
G - One_way_Anova	RL	0,729090909	0,763650794	0,885714286	0,812879767	34	17
E - HRV_BPV_Indexes	RL	0,671428571	0,773333333	0,71	0,729206349	23	13
D - IR_Indexes	RL	0,597222222	0,7	0,726666667	0,703549784	29	13
C - FRF_Indexes	RL	0,580555556	0,644285714	0,72	0,673650794	26	15
A - HRV_Indexes	RL	0,579166667	0,633690476	0,702777778	0,654629515	44	32
B - BPV_Indexes	RL	0,511111111	0,544444444	0,74	0,623076923	27	22
F - HRV_IR_Indexes	RL	0,471428571	0,66	0,56	0,604444444	25	9

**Figura C.4.** Métricas de acurácia, precisão, sensibilidade e Medida F1 dos agrupamentos de índices para regressão logística (RL). A escala de cores varia do verde ao vermelho. Quanto mais verde, maior o valor. Quanto mais vermelho, menor o valor. Valores em branco são valores centrais. Os dados estão em unidades unitárias.

## APÊNDICE D

### MÉTRICAS INTEGRAIS POR ALGORITMOS DE APRENDIZADO DE MÁQUINA DOS ÍNDICES INDIVIDUAIS

Marcador	Algoritmo	Acurácia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE
DYNAMIC_GAIN_ILV_ACOPLAM	SVM	0,790909091	0,782539683	0,914285714	0,841428571	35	22
HF_DG_ILV_ACOPLAM	SVM	0,789393939	0,798412698	0,885714286	0,837857143	35	21
DYNAMIC_GAIN_RCC	SVM	0,764052288	0,757655678	0,928787879	0,832887402	56	33
LF_DG_RCC	SVM	0,718954248	0,720757919	0,931818182	0,809336589	57	32
HF_DG_RCC	SVM	0,704575163	0,735734266	0,854545455	0,787162837	56	32
LF_DG_ILV_ACOPLAM	SVM	0,690909091	0,683333333	0,914285714	0,780588235	35	23
QUARTIL_3	SVM	0,684189723	0,671318681	0,855128205	0,747449821	62	52
IRM_RCC	SVM	0,68245614	0,779090909	0,709090909	0,729423793	55	36
MEDIANA	SVM	0,675494071	0,65702381	0,871794872	0,745777863	62	52
IRM_ILV_ACOPLAM	SVM	0,672727273	0,700974026	0,885714286	0,775873016	35	20
MEDIA	SVM	0,666403162	0,652261905	0,855128205	0,736091253	62	52
QUARTIL_1	SVM	0,65770751	0,64584439	0,838461538	0,727493797	62	52
MAXIMO	SVM	0,652173913	0,637301587	0,833333333	0,721041921	60	52
DYNAMIC_GAIN_ABR	SVM	0,606410256	0,606410256	1	0,754486216	37	24
LF_HF_RATIO_DBP_WELCH	SVM	0,606410256	0,613253968	0,657142857	0,631160506	33	31
MINIMO	SVM	0,605533597	0,615796703	0,7	0,653802198	60	52
HF_RRI_WELCH	SVM	0,602380952	0,604084967	0,843939394	0,69957672	57	46
LF_HF_RATIO_SBP_WELCH	SVM	0,6	0,614285714	0,79047619	0,679487179	33	27
DYNAMIC_GAIN_SBP_ACOPLAM	SVM	0,596969697	0,607878788	0,971428571	0,747024424	35	22
HF_DG_ABR	SVM	0,596153846	0,596153846	1	0,746867168	37	25
FAIXA_DINAMICA	SVM	0,585714286	0,598327228	0,718181818	0,645427741	56	48
RMSSD	SVM	0,585714286	0,583247863	0,810606061	0,674471756	57	49
SD_1	SVM	0,585714286	0,583247863	0,810606061	0,674471756	57	49
IRM_SBP_ACOPLAM	SVM	0,585454545	0,585454545	1	0,737908497	31	22
IRM_ABR	SVM	0,582051282	0,575050505	0,885714286	0,695571035	34	28
LF_HF_RATIO_RRI_WELCH	SVM	0,565367965	0,565583567	0,859090909	0,680597701	59	47
COEFICIENTE_VARIANCIA	SVM	0,564285714	0,570075188	0,983333333	0,721260997	58	43
LF_DG_ABR	SVM	0,56043956	0,562615163	0,785714286	0,647936508	37	31
HF_DG_SBP_ACOPLAM	SVM	0,553030303	0,58030303	0,914285714	0,706322669	34	22
LF_DG_SBP_ACOPLAM	SVM	0,527272727	0,555411255	0,857142857	0,669340017	35	24
SDNN	SVM	0,525238095	0,58341593	0,677272727	0,604494548	58	43
LF_RRI_WELCH	SVM	0,524761905	0,564548872	0,781818182	0,652600733	59	44
pNN50	SVM	0,521637427	0,521637427	1	0,685385878	48	44
LF_SBP_WELCH	SVM	0,515384615	0,534188034	0,914285714	0,671315789	35	29
SD_2	SVM	0,485714286	0,557619048	0,677272727	0,585901296	58	43
HF_SBP_WELCH	SVM	0,476923077	0,482051282	0,685714286	0,545	35	30
LF_DBP_WELCH	SVM	0,458974359	0,495940171	0,8	0,607368421	35	28
HF_DBP_WELCH	SVM	0,426923077	0,462692308	0,50952381	0,444444444	33	30

Figura D.1. Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para *support vector machines* (SVM). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias.

Marcador	Algoritmo	Acúrcia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE
HF_DG_RCC	ÁRVORE DECISÃO	0,751633987	0,797552448	0,822727273	0,809090909	56	32
HF_DG_ILV_ACOPLAM	ÁRVORE DECISÃO	0,751515152	0,75952381	0,885714286	0,816190476	35	21
QUARTIL_3	ÁRVORE DECISÃO	0,727667984	0,717704518	0,870512821	0,777196481	62	52
MEDIANA	ÁRVORE DECISÃO	0,718972332	0,713333333	0,803846154	0,754863655	62	52
MEDIA	ÁRVORE DECISÃO	0,684189723	0,695753823	0,791025641	0,728871511	62	52
DYNAMIC_GAIN_ILV_ACOPLAM	ÁRVORE DECISÃO	0,668181818	0,723809524	0,742857143	0,730989011	35	22
DYNAMIC_GAIN_RCC	ÁRVORE DECISÃO	0,652287582	0,699350649	0,803030303	0,746004257	56	33
MAXIMO	ÁRVORE DECISÃO	0,652173913	0,653351159	0,8	0,711426231	60	52
SD_1	ÁRVORE DECISÃO	0,642857143	0,665051615	0,689393939	0,672261999	57	49
IRM_RCC	ÁRVORE DECISÃO	0,628070175	0,849206349	0,490909091	0,613202614	55	36
LF_DG_RCC	ÁRVORE DECISÃO	0,617647059	0,705144855	0,701515152	0,697509465	57	32
RMSSD	ÁRVORE DECISÃO	0,614285714	0,636263736	0,671212121	0,644131868	57	49
LF_DG_ILV_ACOPLAM	ÁRVORE DECISÃO	0,606060606	0,636363636	0,828571429	0,711968812	35	23
QUARTIL_1	ÁRVORE DECISÃO	0,604347826	0,616922587	0,726923077	0,663170533	62	52
HF_RRI_WELCH	ÁRVORE DECISÃO	0,601904762	0,637707391	0,7	0,660952381	57	46
LF_DG_ABR	ÁRVORE DECISÃO	0,601098901	0,64968254	0,621428571	0,626648352	37	31
HF_SBP_WELCH	ÁRVORE DECISÃO	0,584615385	0,632857143	0,571428571	0,585432215	35	30
FAIXA_DINAMICA	ÁRVORE DECISÃO	0,576666667	0,617619048	0,577272727	0,579443995	56	48
HF_DG_ABR	ÁRVORE DECISÃO	0,566666667	0,618809524	0,732142857	0,664033613	37	25
IRM_ILV_ACOPLAM	ÁRVORE DECISÃO	0,563636364	0,634047619	0,714285714	0,666830424	35	20
LF_RRI_WELCH	ÁRVORE DECISÃO	0,544285714	0,580840081	0,763636364	0,6494565	59	44
HF_DG_SBP_ACOPLAM	ÁRVORE DECISÃO	0,534848485	0,574545455	0,795238095	0,650014854	34	22
COEFICIENTE_VARIANCIA	ÁRVORE DECISÃO	0,534285714	0,583408521	0,742424242	0,599802372	58	43
IRM_ABR	ÁRVORE DECISÃO	0,532051282	0,55968254	0,619047619	0,576266381	34	28
pNN50	ÁRVORE DECISÃO	0,521637427	0,521637427	1	0,685388878	48	44
LF_DG_SBP_ACOPLAM	ÁRVORE DECISÃO	0,513636364	0,576825397	0,6	0,584047619	35	24
DYNAMIC_GAIN_ABR	ÁRVORE DECISÃO	0,511538462	0,628571429	0,485714286	0,546666667	37	24
LF_HF_RATIO_RRI_WELCH	ÁRVORE DECISÃO	0,51038961	0,543343207	0,609090909	0,564710345	59	47
SDNN	ÁRVORE DECISÃO	0,504761905	0,596948052	0,465151515	0,519787942	58	43
LF_SBP_WELCH	ÁRVORE DECISÃO	0,5	0,530916306	0,685714286	0,595028011	35	29
SD_2	ÁRVORE DECISÃO	0,494761905	0,568278943	0,518181818	0,53667683	58	43
MINIMO	ÁRVORE DECISÃO	0,480632411	0,540714286	0,45	0,486853147	60	52
DYNAMIC_GAIN_SBP_ACOPLAM	ÁRVORE DECISÃO	0,471212121	0,636666667	0,571428571	0,517938831	35	22
LF_HF_RATIO_DBP_WELCH	ÁRVORE DECISÃO	0,467948718	0,477777778	0,419047619	0,437121212	33	31
IRM_SBP_ACOPLAM	ÁRVORE DECISÃO	0,454545455	0,54	0,585714286	0,54975225	31	22
LF_HF_RATIO_SBP_WELCH	ÁRVORE DECISÃO	0,416666667	0,468888889	0,457142857	0,456317016	33	27
LF_DBP_WELCH	ÁRVORE DECISÃO	0,408974359	0,451111111	0,485714286	0,465	35	28
HF_DBP_WELCH	ÁRVORE DECISÃO	0,397435897	0,443888889	0,395238095	0,374444444	33	30

**Figura D.2.** Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para árvore de decisão (AD). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias.

Marcador	Algoritmo	Acúrcia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE
HF_DG_ILV_ACOPLAM	KNN	0,807575758	0,817857143	0,885714286	0,84952381	35	21
DYNAMIC_GAIN_ILV_ACOPLAM	KNN	0,790909091	0,782539683	0,914285714	0,841428571	35	22
QUARTIL_3	KNN	0,736363636	0,74202324	0,823076923	0,771725259	62	52
DYNAMIC_GAIN_RCC	KNN	0,730718954	0,76003663	0,859090909	0,801941392	56	33
HF_DG_RCC	KNN	0,71503268	0,748787879	0,836363636	0,785981169	56	32
LF_DG_ILV_ACOPLAM	KNN	0,696969697	0,733333333	0,771428571	0,75	35	23
MEDIA	KNN	0,692094862	0,698866397	0,78974359	0,735364055	62	52
MEDIANA	KNN	0,683399209	0,685198135	0,807692308	0,733224868	62	52
IRM_SBP_ACOPLAM	KNN	0,665454545	0,701587302	0,776190476	0,73525641	31	22
LF_DG_ABR	KNN	0,662637363	0,703968254	0,675	0,68525641	37	31
MAXIMO	KNN	0,662055336	0,670606061	0,733333333	0,696871406	60	52
HF_RRI_WELCH	KNN	0,63047619	0,664891775	0,683333333	0,671939394	57	46
LF_DG_SBP_ACOPLAM	KNN	0,627272727	0,671082251	0,771428571	0,707721037	35	24
LF_DG_RCC	KNN	0,618300654	0,693223443	0,742424242	0,703109312	57	32
IRM_ILV_ACOPLAM	KNN	0,618181818	0,681507937	0,771428571	0,717641672	35	20
HF_DG_ABR	KNN	0,614102564	0,643076923	0,810714286	0,711890756	37	25
COEFICIENTE_VARIANCIA	KNN	0,604285714	0,668376068	0,642424242	0,648749482	58	43
LF_HF_RATIO_SBP_WELCH	KNN	0,6	0,645714286	0,60952381	0,622344322	33	27
QUARTIL_1	KNN	0,587747036	0,594093567	0,756410256	0,661904762	62	52
FAIXA_DINAMICA	KNN	0,585714286	0,620512821	0,627272727	0,617619048	56	48
HF_SBP_WELCH	KNN	0,553846154	0,586666667	0,571428571	0,571945701	35	30
LF_RRI_WELCH	KNN	0,543333333	0,58834303	0,677272727	0,627413755	59	44
DYNAMIC_GAIN_ABR	KNN	0,542307692	0,622222222	0,642857143	0,62384364	37	24
SD_1	KNN	0,538095238	0,581849262	0,560606061	0,567194805	57	49
HF_DG_SBP_ACOPLAM	KNN	0,536363636	0,617063492	0,652380952	0,627106227	34	22
SD_2	KNN	0,534285714	0,601631702	0,639393939	0,616638601	58	43
LF_SBP_WELCH	KNN	0,530769231	0,563059163	0,685714286	0,611145227	35	29
IRM_RCC	KNN	0,528070175	0,600692641	0,654545455	0,622055278	55	36
LF_HF_RATIO_RRI_WELCH	KNN	0,509090909	0,555119048	0,607575758	0,571938207	59	47
RMSSD	KNN	0,5	0,547863248	0,524242424	0,52954386	57	49
DYNAMIC_GAIN_SBP_ACOPLAM	KNN	0,493939394	0,583333333	0,628571429	0,603076923	35	22
SDNN	KNN	0,484285714	0,557272727	0,568181818	0,560547504	58	43
LF_HF_RATIO_DBP_WELCH	KNN	0,483333333	0,505714286	0,485714286	0,493706294	33	31
pNN50	KNN	0,478362573	0	0	0	48	44
HF_DBP_WELCH	KNN	0,476923077	0,513333333	0,60952381	0,550641026	33	30
LF_DBP_WELCH	KNN	0,475641026	0,488383838	0,6	0,526873943	35	28
IRM_ABR	KNN	0,467948718	0,514285714	0,619047619	0,559551821	34	28
MINIMO	KNN	0,445059289	0,477812178	0,566666667	0,510745947	60	52

**Figura D.3.** Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para *k-nearest neighbor* (KNN). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias.

Marcador	Algoritmo	Acurácia-Validação	Precisão-Validação	Sensibilidade-Validação	Medida-F1-Validação	DM	CONTROLE
DYNAMIC_GAIN_ILV_ACOPLAM	RL	0,790909091	0,782539683	0,914285714	0,841428571	35	22
HF_DG_ILV_ACOPLAM	RL	0,789393939	0,798412698	0,885714286	0,837857143	35	21
DYNAMIC_GAIN_RCC	RL	0,752941176	0,777482517	0,875757576	0,818141858	56	33
LF_DG_RCC	RL	0,718954248	0,736227106	0,88030303	0,799230769	57	32
HF_DG_RCC	RL	0,704575163	0,743274373	0,854545455	0,789043715	56	32
LF_DG_ILV_ACOPLAM	RL	0,692424242	0,708095238	0,857142857	0,772605042	35	23
IRM_RCC	RL	0,671929825	0,703059441	0,818181818	0,753491436	55	36
MAXIMO	RL	0,660869565	0,671266968	0,75	0,706137931	60	52
QUARTIL_3	RL	0,658102767	0,681058941	0,711538462	0,693341261	62	52
MEDIANA	RL	0,649407115	0,671336898	0,712820513	0,684995109	62	52
IRM_ILV_ACOPLAM	RL	0,636363636	0,647878788	0,942857143	0,766993464	35	20
LF_HF_RATIO_SBP_WELCH	RL	0,633333333	0,632424242	0,823809524	0,705149752	33	27
MEDIA	RL	0,623320158	0,650582751	0,680769231	0,659188477	62	52
LF_HF_RATIO_DBP_WELCH	RL	0,621794872	0,61547619	0,752380952	0,674240465	33	31
DYNAMIC_GAIN_ABR	RL	0,606410256	0,606410256	1	0,754486216	37	24
HF_DG_SBP_ACOPLAM	RL	0,606060606	0,609090909	0,933333333	0,736507937	34	22
LF_DG_ABR	RL	0,603296703	0,640606061	0,703571429	0,652345797	37	31
HF_DG_ABR	RL	0,596153846	0,596153846	1	0,746867168	37	25
QUARTIL_1	RL	0,596047431	0,61702381	0,679487179	0,640591212	62	52
DYNAMIC_GAIN_SBP_ACOPLAM	RL	0,58030303	0,60030303	0,942857143	0,732989336	35	22
FAIXA_DINAMICA	RL	0,569047619	0,577142857	0,751515152	0,652307692	56	48
SD_2	RL	0,565714286	0,589253394	0,792424242	0,67366539	58	43
SDNN	RL	0,565714286	0,589253394	0,792424242	0,67366539	58	43
LF_HF_RATIO_RRI_WELCH	RL	0,564935065	0,577296182	0,793939394	0,667128719	59	47
LF_DG_SBP_ACOPLAM	RL	0,560606061	0,604617605	0,771428571	0,671031746	35	24
RMSSD	RL	0,557575758	0,555108359	0,843939394	0,667292922	57	49
SD_1	RL	0,557575758	0,555108359	0,843939394	0,667292922	57	49
COEFICIENTE_VARIANCIA	RL	0,554285714	0,569356725	0,895454545	0,695012407	58	43
LF_RRI_WELCH	RL	0,553809524	0,57229741	0,865151515	0,686276775	59	44
LF_SBP_WELCH	RL	0,547435897	0,553449883	0,914285714	0,688194014	35	29
IRM_SBP_ACOPLAM	RL	0,530909091	0,56020202	0,90952381	0,691176471	31	22
MINIMIO	RL	0,52687747	0,545409678	0,733333333	0,619698925	60	52
pNN50	RL	0,521637427	0,521637427	1	0,685385878	48	44
HF_RRI_WELCH	RL	0,514761905	0,53477763	0,893939394	0,66834241	57	46
LF_DBP_WELCH	RL	0,508974359	0,53003108	0,885714286	0,660701754	35	28
HF_DBP_WELCH	RL	0,507692308	0,55991453	0,666666667	0,582424242	33	30
IRM_ABR	RL	0,487179487	0,507692308	0,885714286	0,643070175	34	28
HF_SBP_WELCH	RL	0,461538462	0,493406593	0,771428571	0,59647452	35	30

**Figura D.4.** Métricas de acurácia, precisão, sensibilidade e Medida F1 dos índices individuais para regressão logística (RL). Em amarelo, estão destacados os 10 (dez) melhores valores de cada métrica. A tabela está organizada por ordem de acurácia. Os dados estão em unidades unitárias.