

DISSERTAÇÃO DE MESTRADO Nº 1056

**APRENDIZADO DE MÁQUINA APLICADO AO RECONHECIMENTO  
AUTOMÁTICO DE FALHAS EM MÁQUINAS ROTATIVAS**

**Diego Silva Caldeira Rocha**

DATA DA DEFESA: 08/06/2018

**Universidade Federal de Minas Gerais**

**Escola de Engenharia**

**Programa de Pós-Graduação em Engenharia Elétrica**

**APRENDIZADO DE MÁQUINA APLICADO AO  
RECONHECIMENTO AUTOMÁTICO DE FALHAS EM MÁQUINAS  
ROTATIVAS**

Diego Silva Caldeira Rocha

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Prof. João Antônio de Vasconcelos

Belo Horizonte - MG

Junho de 2018

---

---

# Dedicatória

---

*Dedico este trabalho aos meus pais Edivaldo Caldeira Rocha e Maria Nilza Silva Caldeira Rocha.*

---

---

# Agradecimentos

---

Renunciar sempre é mais fácil, mas a GLÓRIA da VITÓRIA após uma batalha desafiadora e sacrificante é o alcançar da PLENITUDE.

Agradeço a Deus, pois sem Ele nada seria possível. Ao meu PAI pela confiança e por estar sempre ao meu lado. À minha MÃE, por todo apoio e incentivo nesta conquista.

Minha ESPOSA e minha filha BEATRIZ pelo auxílio, carinho e compreensão e a todas as pessoas que de alguma forma me auxiliaram para alcançar este momento.

Agradeço meu orientador João Antônio de Vasconcelos pelo acolhimento, motivação, auxílio nas dificuldades e em muitas ocasiões proporcionou um ambiente amigável, eliminando os obstáculos entre aluno e professor.

Agradeço aos professores e funcionários ligados ao PPGEE que sempre me ajudaram quando solicitei.

A todos, o meu muito obrigado.

---

---

# Epígrafe

---

*“A gente só precisa acreditar...  
Acreditar que as lutas passam, que os ventos fortes se acalmam,  
que as feridas são curadas e que o amor e o perdão nos levam a  
conquistar. Nem tudo nessa vida acontece como a gente quer, mas  
tudo que está nas mãos de Deus acontece como deve acontecer, com  
perfeição e no tempo devido.”*

Autor Desconhecido

---

# Sumário

---

|  |             |
|--|-------------|
| <b>Resumo</b>  | <b>viii</b> |
| <b>Abstract</b>  | <b>ix</b>   |
| <b>Lista de Figuras</b>  | <b>xii</b>  |
| <b>Lista de Tabelas</b>  | <b>xiv</b>  |
| <b>Lista de Símbolos</b>   | <b>xv</b>   |
| <b>Lista de Siglas e Abreviações</b>   | <b>xvi</b>  |
| <b>1 Introdução</b>  | <b>1</b>    |
| 1.1 Justificativas e Motivações . . . . .  | 1           |
| 1.2 Problema de Diagnóstico de Falhas em Máquinas Rotativas . . . . .  | 2           |
| 1.3 Objetivo . . . . .   | 4           |
| 1.4 Organização do Trabalho . . . . .  | 4           |
| 1.5 Publicação e Palestra . . . . .  | 5           |
| <b>2 Base de dados</b>   | <b>6</b>    |
| 2.1 Mafaulda ( <i>Machinery Fault Database</i> ) . . . . .   | 6           |
| 2.1.1 Defeitos Simulados . . . . .   | 8           |
| 2.1.2 Cenários de Simulação . . . . .  | 11          |
| 2.2 Conclusão . . . . .  | 12          |
| <b>3 Metodologia</b>   | <b>14</b>   |
| 3.1 Métodos de Extração de Características . . . . .   | 14          |
| 3.1.1 Método de Extração de Características Baseado em RMS . . . . .   | 14          |
| 3.1.2 Método de Extração de Características Baseado na Transformada Wavelet de Haar e Dimensão Fractal . . . . . | 15          |
| 3.1.3 Método de Extração de Características Baseado na Transformada de Fourier com Dados Estatísticos . . . . .  | 18          |
| 3.2 Classificação . . . . .  | 23          |
| 3.2.1 <i>K-Nearest Neighbors (KNN)</i> . . . . .   | 25          |
| 3.2.2 <i>Support Vector Machine (SVM)</i> . . . . .  | 26          |

|          |  |           |
|----------|--|-----------|
| 3.2.3    | <i>Extreme Gradient Boosting</i> (XGBoost) . . . . .                     | 32        |
| 3.3      | Métricas de desempenho . . . . .   | 35        |
| 3.4      | Resumo . . . . .   | 37        |
| 3.5      | Software e Computador Utilizados . . . . .                               | 39        |
| 3.6      | Conclusão . . . . .  | 39        |
| <b>4</b> | <b>Resultados das Classificações das Falhas em Máquinas Rotativas</b>    | <b>40</b> |
| 4.1      | Resultados e Análise . . . . .   | 41        |
| 4.2      | Conclusão . . . . .  | 48        |
| <b>5</b> | <b>Resultados das Classificações das Falhas nos Mancais de Rolamento</b> | <b>49</b> |
| 5.1      | Resultados e Análise . . . . .   | 49        |
| 5.2      | Conclusão . . . . .  | 55        |
| <b>6</b> | <b>Conclusões</b>  | <b>56</b> |
| 6.1      | Sugestões para Trabalhos Futuros . . . . .                               | 57        |
|          | <b>Referências Bibliográficas</b>  | <b>62</b> |

---

---

# Resumo

---

As máquinas rotativas como motores, geradores e motobombas são equipamentos comumente utilizados em quase todos os processos industriais. A análise de vibrações mecânicas tem sido uma importante técnica adotada nas empresas para avaliações do estado de operação das máquinas industriais. Este trabalho utiliza-se de uma base de dados de sinais de vibrações mecânicas para classificar automaticamente falhas em máquinas rotativas. São apresentados três modelos de extração de características de sinais de vibrações mecânicas: (i) RMS (Root Means Squares), (ii) Wavelet de Haar e dimensão fractal e (iii) FFT (Fast Fourier Transform) com dados estatísticos. Por fim, é utilizado o conceito de aprendizado de máquina com os classificadores KNN (K-vizinhos mais próximos), SVM (Máquina de Vetores de Suporte) e XGBoost (*Extreme Gradient Boosting*) para diagnosticar as falhas. Os resultados demonstram eficiência de todas as técnicas, sendo que na abordagem Wavelet de Haar e dimensão fractal combinados com o XGBoost obteve os melhores resultados. Foi possível atingir uma acurácia de 98.7% (MAUC (*Multi-class Extension of AUC*)=0.9704) nas falhas em máquinas rotativas e 99.36% de acurácia (MAUC=0.9965) na detecção de problemas nos mancais de rolamento. Além disso, obteve resultados intraclasse notáveis, se mostrando bastante promissor para o objeto dessa dissertação.

**Palavras-chave:** KNN, SVM, XGBoost, Diagnóstico de falhas em máquinas rotativas, Extração de características, Vibrações mecânicas

---

# Abstract

---

Rotary machines such as motors, generators and pumps are commonly used in almost all industrial processes. The analysis of mechanical vibrations has been an important technique adopted in companies to evaluate the state of operation of industrial machines. This work uses a database of mechanical vibration signals to automatically classify faults in rotary machines. Three models of extraction of characteristics of mechanical vibration signals are presented: (i) RMS (Root Means Squares), (ii) Haar Wavelet and fractal dimension and (iii) FFT (Fast Fourier transform) with statistical data. Finally, the machine learning concept is used with the classifiers KNN (K-Nearest Neighbors), SVM (Support Vector Machine) and XGBoost (Extreme Gradient Boosting) to diagnose faults. The results demonstrate the efficiency of all the techniques, although wavelet approach and fractal dimension combined with XGBoost, presenting the best results. It was possible to reach an accuracy of 98.7% (MAUC (Multi-class Extension of AUC) = 0.9704) on rotating machine failures and 99.36% of accuracy (MAUC = 0.9965) for bearing problems. In addition, it obtained remarkable intraclass results and was very promising for the subject of this dissertation.

**Keywords:** KNN, SVM, XGBoost, Fault diagnosis on rotating machines, Extraction of characteristics, Mechanical vibrations.

---

# Lista de Figuras

---

|     |   |    |
|-----|---|----|
| 2.1 | Bancada experimental de aquisição de dados (MAFAULDA, 2017), Figura retirada e adaptada de (Marins et al., 2017). . . . .   | 7  |
| 2.2 | Diagrama de aquisição de dados da base (MAFAULDA, 2017), Figura retirada e adaptada de (S. Martins, 2016). . . . .  | 8  |
| 2.3 | Desalinhamento horizontal, Figura retirada de (Lima et al., 2013). . . . .  | 9  |
| 2.4 | Desalinhamento vertical, Figura retirada de (Lima et al., 2013). . . . .  | 9  |
| 2.5 | Disco usado no cenário de desbalanceamento. O furo central é onde o disco é fixado ao eixo e os orifícios periféricos são usados para desequilibrar com cargas extras, Figura retirada e adaptada de (Lima et al., 2013). . . . . | 10 |
| 2.6 | Mancal de rolamento. . . . .  | 10 |
| 2.7 | Estrutura de um rolamento. . . . .  | 11 |
| 3.1 | Trecho de sinal no domínio do tempo de um acelerômetro utilizado. . . . .   | 16 |
| 3.2 | Transformada Wavelet de Haar de um acelerômetro, escala à esquerda e detalhe à direita. . . . .   | 17 |
| 3.3 | Exemplificação do algoritmo de contagem de caixas. . . . .  | 18 |
| 3.4 | Sinal no espectro da frequência de um acelerômetro, através da FFT. . . . .   | 21 |

|      |   |    |
|------|---|----|
| 3.5  | Diagrama básico de um aprendizado de máquina para um classificador. Figura retirada e adaptada de (Lorena e de Carvalho, 2007). . . . . | 24 |
| 3.6  | Conjunto de treinamento binário e três diferentes hipóteses. Figura retirada e adaptada de (Lorena e de Carvalho, 2007). . . . .        | 25 |
| 3.7  | Curvas de Classificação para o KNN a) $k = 1$ , b) $k = 100$ . Figura retirada e adaptada de (James et al., 2013). . . . .              | 27 |
| 3.8  | Exemplo de superfície de separação gerada pelo SVM. Figura retirada e adaptada de (Abe, 2005). . . . .                                  | 27 |
| 3.9  | Conjunto de dados não linearmente separáveis. Figura retirada e adaptada de (Lorena e de Carvalho, 2007). . . . .                       | 30 |
| 3.10 | Mapeamento do espaço de entrada via função kernel. Figura retirada e adaptada de (Sinhorigno, 2007). . . . .                            | 31 |
| 3.11 | a) Gráfico bidimensional de dados de classificação e b) Resultado da árvore de decisão. . . . .   | 32 |
| 3.12 | Diagrama em blocos da metodologia utilizada. . . . .  | 38 |
| 4.1  | Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do KNN. . . . .                                  | 41 |
| 4.2  | Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do SVM. . . . .                                  | 44 |
| 4.3  | Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do XGBoost. . . . .                              | 44 |
| 4.4  | <i>Boxplot</i> condensado como valores de ACC. . . . .  | 45 |
| 4.5  | <i>Boxplot</i> condensado como valores de MAUC. . . . .   | 45 |
| 4.6  | Diferença dos valores médios da ACC. . . . .  | 47 |
| 4.7  | Diferença dos valores médios da MAUC. . . . .   | 47 |

|     |   |    |
|-----|---|----|
| 5.1 | Gráficos dos valores de ACC à esquerda e MAUC à direita, para os método nas execuções do KNN. . . . .     | 51 |
| 5.2 | Gráficos dos valores de ACC à esquerda e MAUC à direita, para os método nas execuções do SVM. . . . .     | 51 |
| 5.3 | Gráficos dos valores de ACC à esquerda e MAUC à direita, para os método nas execuções do XGBoost. . . . . | 51 |
| 5.4 | <i>Boxplot</i> condensado como valores de ACC. . . . .  | 52 |
| 5.5 | <i>Boxplot</i> condensado como valores de MAUC. . . . .   | 53 |
| 5.6 | Diferença dos valores médios da ACC. . . . .  | 54 |
| 5.7 | Diferença dos valores médios da MAUC. . . . .   | 54 |

---

---

# Lista de Tabelas

---

|     |  |    |
|-----|--|----|
| 2.1 | Especificações técnicas do motor. . . . .  | 7  |
| 2.2 | Quantidade de cenários por classes. . . . .  | 12 |
| 2.3 | Quantidade de cenários por subclasses. . . . .   | 12 |
| 4.1 | Siglas dos modelos de classificação. . . . .   | 40 |
| 4.2 | Resultados intraclasses de ACC para as classes (I), (II) e (III). . . . .  | 42 |
| 4.3 | Resultados intraclasses de ACC para as classes (IV), (V) e (VI). . . . .   | 43 |
| 4.4 | Resultados condensados de médias e desvios de ACC e MAUC. . . . .  | 44 |
| 4.5 | Matriz de confusão obtida pelo modelo classificação XGBoost com preditores extraído através método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2) (ACC = 0.9951 e MAUC= 0.9879). . . . . | 46 |
| 4.6 | Ranqueamento dos modelos baseado no teste de Tukey do ACC. . . . .   | 47 |
| 4.7 | Ranqueamento dos modelos baseado no teste de Tukey do MAUC. . . . .  | 48 |
| 5.1 | Resultados intraclasses de ACC para as subclasses (VII), (VIII) e (IX). . . . .  | 50 |

|     |   |    |
|-----|---|----|
| 5.2 | Resultados condensados de médias e desvios de ACC e MAUC. . . . .   | 52 |
| 5.3 | Matriz de confusão obtida pelo modelo classificação XGBoost com preditores extraídos através do método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (ACC = MAUC=1). 52 |    |
| 5.4 | Ranqueamento dos modelos baseado no teste de Tukey do ACC. . . . .  | 54 |
| 5.5 | Ranqueamento dos modelos baseado no teste de Tukey do MAUC. . . . .   | 55 |

---

---

## Lista de Símbolos

---

- A* Matriz de características da extração da subseção 3.1.1;
- En* Energia de vibração dos detalhes;
- Et* Entropia dos detalhes;
- Df* Dimensão fractal dos detalhes;
- B* Matriz de características da extração da subseção 3.1.2;
- I<sub>d</sub>* Índice da frequência com maior magnitude;
- S<sub>I</sub>* Valores da magnitude dos sinais da FFT;
- f<sub>s</sub>* Frequência de amostragem;
- Rf* Frequência de rotação;
- He* Entropia do sinal no domínio do tempo;
- Me* Média do sinal no domínio do tempo;
- Ku* Curtose do sinal no domínio do tempo;
- C* Matriz de características da extração da subseção 3.1.3;
- Esp* Especificidade;
- Sen* Sensibilidade.

---

## Lista de Siglas e Abreviações

---

|          |  |
|----------|--|
| ACC      | <i>Acurácia;</i>                                 |
| AUC      | <i>Area Under ROC;</i>                           |
| DFT      | <i>Discret Fourier Transform;</i>                |
| FFT      | <i>Fast Fourier Transform;</i>                   |
| KNN      | <i>K-Nearest Neighbors;</i>                      |
| Mafaulda | <i>Machinery Fault Database;</i>                 |
| MAUC     | <i>Multi-class Extension of AUC;</i>             |
| MFS      | <i>SpectraQuest's Machinery Fault Simulator;</i> |
| MLP      | <i>Multilayer Perceptron;</i>                    |
| RBF      | <i>Radial Basis Function;</i>                    |
| RMS      | <i>Root Mean Square;</i>                         |
| ROC      | <i>Receiver Operating Characteristic Curve;</i>  |
| SVM      | <i>Support Vector Machine;</i>                   |
| XGBoost  | <i>Extreme Gradient Boosting.</i>                |

# Introdução

---

Este capítulo aborda as principais justificativas e motivações para o desenvolvimento deste trabalho. Além disso, os objetivos, a organização do texto e as publicações também são apresentados.

## 1.1 Justificativas e Motivações

A revolução industrial ocorreu em meados do século XVIII no Reino Unido, momento em que foram assinados vários tratados com outros países para escoar seus produtos manufaturados. A expansão das encomendas fez com que aparecessem as máquinas rotativas para atender ao aumento da demanda.

O uso contínuo das máquinas geram desgastes em suas peças, o que faz necessário ter algum plano de manutenção funcionando. A manutenção corretiva é realizada quando o equipamento deixa de cumprir sua função principal em virtude de uma falha, já a manutenção preventiva é realizada de forma programada. A manutenção preditiva é aquela cujo objetivo é manter os equipamentos funcionando ininterruptamente em condições normais de operação através de um sistema supervisorio. Uma boa manutenção preditiva reduz as manutenções preventivas e corretivas.

As máquinas rotativas como motores, geradores e motobombas são equipamentos comumente utilizados em quase todos os processos industriais. A análise de sinais de sensores (acelerômetros, tacômetros e microfones) como técnica de manutenção preditiva e preventiva desses equipamentos permite que os especialistas verifiquem, de forma taxativa, seu estado de funcionamento. A análise de vibrações mecânicas tem sido uma importante técnica adotada nas empresas para avaliação do estado de operação das máquinas industriais. Essa análise permite desenvolver métodos de antecipação à falha e realizar diagnósticos de operação das máquinas rotativas (Pestana-Viana et al., 2016).

O estado de operação das máquinas rotativas podem ser fiscalizados através de um sistemas de monitoramento industriais que podem ser classificados como do tipo *off-line*, onde os dados dos sensores são coletados e posteriormente analisados, ou tipo *on-line*, quando as informações são adquiridas em tempo real permitindo uma análise imediata das máquinas (Alexandre et al., 2017).

Portanto, é importante a existência de um sistema automático que seja capaz de emitir alarmes eficientes quanto ao diagnostico de falhas em máquinas rotativas.

## **1.2 Problema de Diagnóstico de Falhas em Máquinas Rotativas**

O procedimento de diagnóstico se baseia em detecção e classificação de falhas. O diagnóstico de falhas em máquinas rotativas constitui no mapeamento dos dados obtidos no espaço de medições ou no espaço de características (Jardine et al., 2006). Já o prognóstico tem a finalidade de além de reconhecer o estado de operação das máquinas, estima o tempo de vida útil (Jardine et al., 2006).

As falhas em máquinas podem ser divididas em dois tipos: abruptas e incipientes. As falhas abruptas resultam em grandes desvios nas condições normais de operação, e ocorrem em curto espaço de tempo (Benbouzid, 2000). Este tipo de falha é mais fácil de detecção, visto que normalmente retira a máquina de operação (Benbouzid, 2000). As falhas incipientes afetam gradualmente o funcionamento normal do processo e por isso, são mais difíceis de serem detectadas (Benbouzid, 2000).

As falhas em máquinas rotativas podem ser agrupadas como: mecânicas e elétricas (Singh et al., 2003). As falhas mecânicas podem ocorrer no rotor: excentricidade (estática ou dinâmica) e desalinhamento. No estator encontram-se também problemas de excentricidade. Bem como, as falhas nos rolamentos, podem causar excentricidade do rotor e representam grande parte das falhas em máquinas rotativas (Singh et al., 2003). Já as falhas elétricas, destacam-se: falhas nos enrolamentos e falhas no núcleo de chapas (Singh et al., 2003).

A literatura é repleta de estudos sobre Aprendizado de Máquina para classificação de falhas incipientes mecânicas, utilizando-se de dados de vibrações mecânicas.

Os autores (Nguyen et al., 2007) apresentam um classificador baseado em *K-Nearest Neighbors* (KNN) para classificar falhas em diferentes motores de indução. A extração é feita a partir de inferências estatísticas no domínio do tempo, foi alcançado uma acurácia de 99,9%.

Em (Loureiro, 2009), o autor mostra duas estratégias para seleção de características relevantes para classificação de falhas em motobombas: (i) a utilização de heurísticas com base no conhecimento especializado, (ii) algoritmo de seleção e técnicas de processamento de sinais que são utilizados para obter atributos descritivos a partir dos sinais de vibração. Ao final utiliza-se o classificador KNN em seus resultados, foi obtida uma acurácia de 66,44%.

Em (Lima et al., 2013), os autores apresentam uma técnica de extração de característica baseada na Transformada Discreta de Fourier (DFT, do inglês *Discret Fourier Transform*) para estimar a frequência de rotação do motor, junto com dados estatísticos no domínio do tempo. Foi utilizada a rede neural artificial multicamadas perceptron (MLP, do inglês *multilayer perceptron*) como classificador, o que resultou em 93% de acurácia.

Em (Ramalho et al., 2014), os autores apresentam uma técnica de reconhecimento de padrões com transformada Wavelet de Haar e dimensão fractal para extração de características dos sinais de acelerômetros. Esses atributos são utilizados como entrada em uma rede neural multicamadas perceptron para classificar as falhas em motores, a acurácia média foi de 84%.

Em (Martins et al., 2016), os autores propõem uma modificação no método de (Lima et al., 2013) com a utilização do classificador *Random Forest* e agregam novas subclasses, resultando em uma acurácia de 91,42%. Já (Marins et al., 2017), os autores acrescentam a redução de dimensionalidade no pré-processamento através de matriz de similaridade, a acurácia foi de 98.5%.

Nota-se aqui a diversidade de abordagens tanto nas metodologia de extração de características quanto na aplicação de aprendizado de máquina no reconhecimento automático de falhas em máquinas rotativas. Estas abordagens e interesse da comunidade científica mostram que esta área de pesquisa é relevante e se encontra ainda aberta para novos estudos e proposição de novas metodologias.

## 1.3 Objetivo

O objetivo geral deste trabalho é o desenvolvimento de uma metodologia de análise de sinais de vibrações mecânicas para fins de detecção de falhas incipientes em máquinas rotativas.

Para que o objetivo principal seja alcançado, os seguintes objetivos específicos foram estabelecidos:

- Levantamento de técnicas de extrações de características em máquinas rotativas com base em dados de vibrações mecânicas;
- Implementação dos extratores de características;
- Implementação de um sistema automático de classificação.

## 1.4 Organização do Trabalho

A dissertação está organizada da seguinte forma:

- O **Capítulo 2** detalha as especificações da base de dados utilizada;
- O **Capítulo 3** apresenta os métodos utilizados nesse trabalho para extrair as características da base de dados, as métricas de desempenho empregadas e a metodologia formulada para classificar as falhas em máquinas rotativas;
- O **Capítulo 4** apresenta os resultados obtidos na classificação das falhas em máquinas rotativas;
- O **Capítulo 5** apresenta os resultados obtidos das classificação das falhas em mancais de rolamento;
- O **Capítulo 6** é dedicado às conclusões desse estudo e exhibe as sugestões para trabalhos futuros.

## 1.5 Publicação e Palestra

O seguinte trabalho científico foi aceito para publicação e apresentado durante a elaboração da dissertação:

- ROCHA, D. S. C.; DE VASCONCELOS, J. A.; DO CARMO, A. C. Aprendizado de Máquina Aplicada ao Reconhecimento Automático de Falhas em Motores Elétricos. *Computer On The Beach* (IX COTB). Florianópolis-SC, Março de 2018.

Ainda foi realizada a palestra pelo autor no seguinte evento:

- **Congresso:** XXXI Encontro Nacional dos Estudantes de Computação (XXXI ENE-COMP). **Tema:** Métodos de extração de características de sinais para classificação de falhas em máquinas rotativas. Ouro Preto-MG, Dezembro de 2017.

## Base de dados

---

Este trabalho teve o desafio de conseguir base de dados com medições recentes e que abrangesse a maioria das falhas ocorridas em máquinas rotativas. As empresas restringem base de dados, para evitar a espionagem industrial e em muitas Universidades tem seus projetos de pesquisas em parcerias com outras indústrias. Optou-se em utilizar uma base acadêmica densa, com boas mensurações e publicações, desenvolvida pelo Laboratório de Sinais, Multimídias e Telecomunicação da UFRJ/COPPE. Este capítulo detalha as especificações e defeitos induzidos pela base.

### 2.1 Mafaulda (*Machinery Fault Database*)

A base de dados referenciada em (MAFAULDA, 2017) é composta por séries temporais multivariadas, adquirida por sensores em MFS (*SpectraQuest's Machinery Fault Simulator*) (Ganeriwala et al., 1998). O MFS é uma máquina que simula a dinâmica de motores com dois rolamentos de suporte de eixo e permite o estudo de múltiplas falhas, como massa desequilibrada, desalinhamento de eixo e problemas de rolamento (Marins et al., 2017).

A Tabela 2.1 apresenta as especificações técnicas do motor elétrico de indução. O simulador faz a coleta de dados de oito sensores: tacômetro, acelerômetro triaxial mancal interno (axial, radial e tangencial), acelerômetro triaxial mancal externo (axial, radial e tangencial) e microfone.

O tacômetro faz a medida da rotação do sistema (Marins et al., 2017). Os acelerômetros são posicionados nos mancais nas direções radial, tangencial e axial, a fim de se obter uma melhor análise das vibrações. Quando ocorrem defeitos nos rolamentos, sugere-se o uso do conjunto de acelerômetros em cada mancal (White, 2008). O Microfone captura o som durante a operação do sistema (Marins et al., 2017).

Os sinais são adquiridos com uma taxa de amostragem de 51,2 kHz e são apresentados seis tipos de falhas (classes) em motores. A Figura 2.1 exibe a bancada de simulação da Mafaulda e a Figura 2.2 apresenta o diagrama de aquisição de dados.

Tabela 2.1: Especificações técnicas do motor.

| Características          | Valores      |
|--------------------------|--------------|
| Motor                    | 1/4 CV DC    |
| Alcance de Frequência    | 700-3600 rpm |
| Peso do sistema          | 22 kg        |
| Diâmetro do eixo         | 16 mm        |
| Comprimento do eixo      | 520 mm       |
| Rotor                    | 15.24 cm     |
| Distância dos rolamentos | 390 mm       |
| Número de esfera         | 8            |

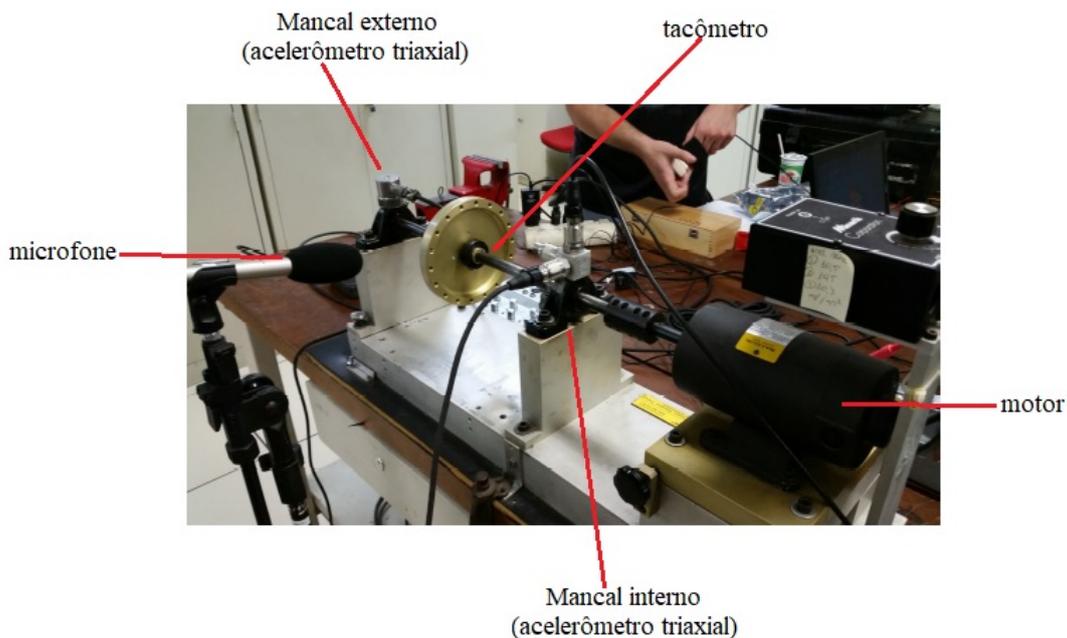


Figura 2.1: Bancada experimental de aquisição de dados (MAFAULDA, 2017), Figura retirada e adaptada de (Marins et al., 2017).

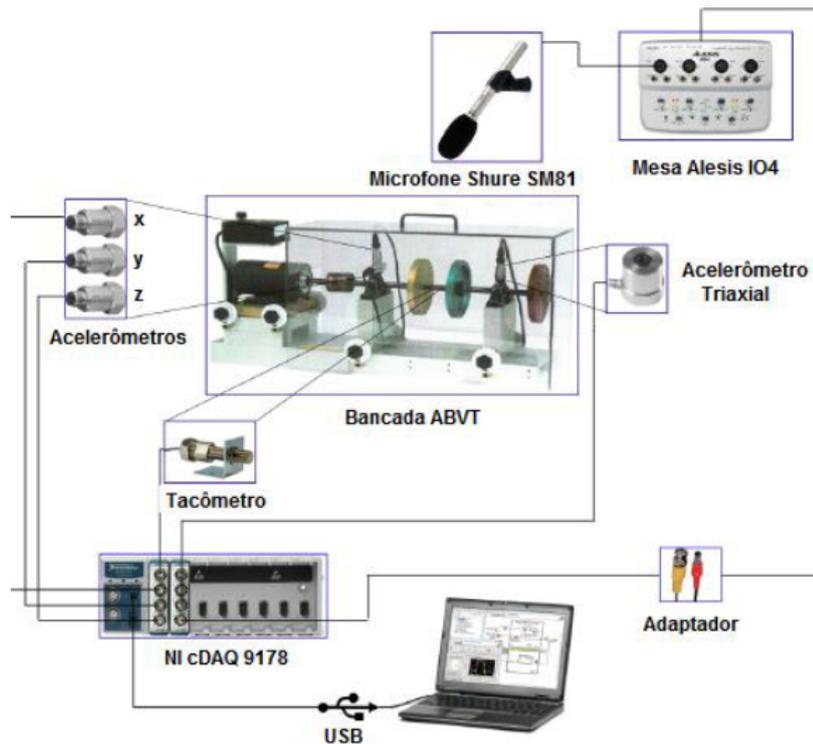


Figura 2.2: Diagrama de aquisição de dados da base (MAFAULDA, 2017), Figura retirada e adaptada de (S. Martins, 2016).

### 2.1.1 Defeitos Simulados

A base de dados Mafaulda tenta abranger ao máximo, as possíveis falhas encontradas nos motores de indução ou em outras máquinas rotativas. Além da coleta de informações do sistema em perfeito funcionamento, são estimulados os seguintes defeitos: desalinhamento, desbalanceamento dos eixos e mancais defeituosos, que serão explicados a seguir.

#### Desalinhamento

Em condições normais os eixos radiais das máquinas motriz e motora necessitam estarem alinhados e concêntricos. No desalinhamento é induzido descolamentos nas direções (horizontal e vertical), conforme as Figuras 2.3 e 2.4, esse desalinhamento foi implementado movendo o eixo do rotor lateralmente ou perpendicularmente (Lima et al., 2013).

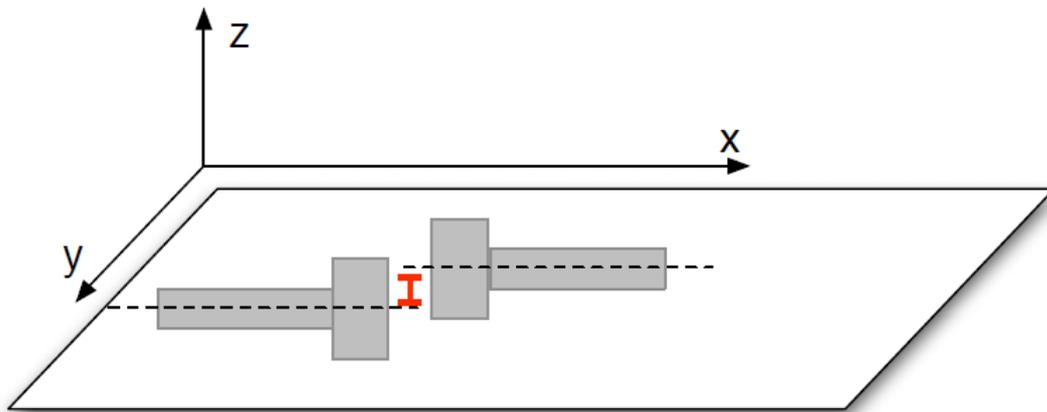


Figura 2.3: Desalinhamento horizontal, Figura retirada de (Lima et al., 2013).

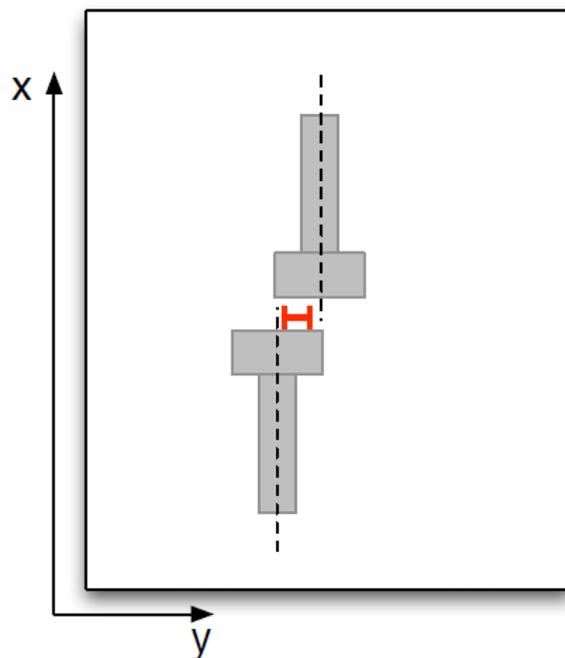


Figura 2.4: Desalinhamento vertical, Figura retirada de (Lima et al., 2013).

### Desbalanceamento

É desejável que em situação típica o rotor possua simetria axial perfeita, contudo, em eventuais desgastes ou incidentes, essa simetria pode ser avariada. No desbalanceamento, a carga não é distribuída de forma igual na direção angular. Essa falha

é provocada posicionando uma carga extra em um dos orifícios periféricos do disco mostrado na Figura 2.5 (Lima et al., 2013).

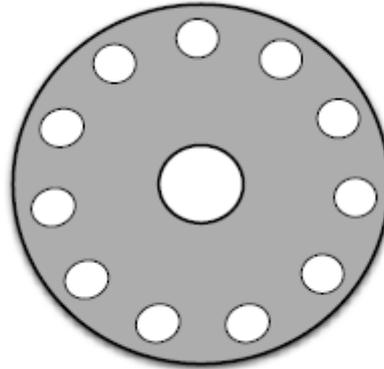


Figura 2.5: Disco usado no cenário de desbalanceamento. O furo central é onde o disco é fixado ao eixo e os orifícios periféricos são usados para desequilibrar com cargas extras, Figura retirada e adaptada de (Lima et al., 2013).

### Mancais defeituosos

Os mancais conforme o diagrama apresentado na Figura 2.6, são responsáveis pela sustentação dos eixos e é um dos elementos mais relevantes das máquinas. Os rolamentos são estruturas complexas, conforme Figura 2.7. São os elementos mais suscetíveis à ocorrência de problemas. Cerca de 37% das falhas são ocasionadas pelos mancais de rolamento (Fujimoto, 2005). Devido à sua complexidade, inúmeros defeitos podem surgir. A base de dados utilizada neste trabalho simula três problemas: defeito na gaiola, nas esferas e na parte externa (anel externo ou pista).

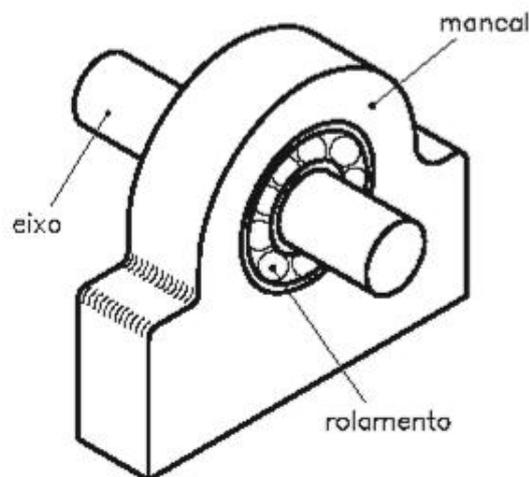


Figura 2.6: Mancal de rolamento.

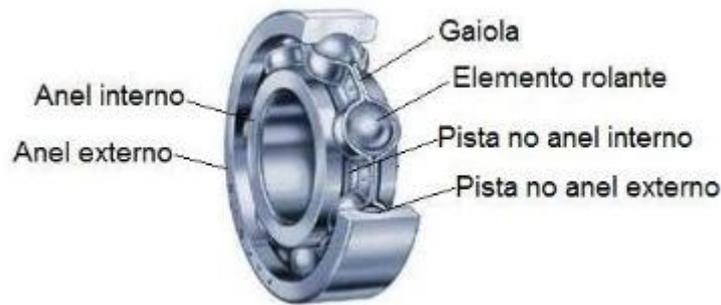


Figura 2.7: Estrutura de um rolamento.

### 2.1.2 Cenários de Simulação

Em (MAFAULDA, 2017) são descritos os testes experimentais e especificações das falhas. Para cada classe apresenta vários cenários de alteração da rotação do motor. Cada um desses cenários tem 250.000 amostras diferentes coletadas de forma temporal, resultando 13Gb de dados. A seguir serão transcritos os cenários de simulação para cada classe.

- **Normal:** esta classe representa o sistema em operação normal, sem falhas, com variação de rotação de 737 a 3686 rpm;
- **Desbalanceamento:** foram acoplados no rotor diferentes cargas no intervalo de 6 a 35g. Para cada valor de carga inferior a 30g, a frequência de rotação assumida é de 737 a 3686 rpm. Para cargas iguais ou superiores a 30g, no entanto, a vibração resultante torna impraticável para o sistema alcançar rotações acima de 3300 rpm, limitando o número de frequências de rotação distintas;
- **Desalinhamento paralelo horizontal:** esta classe foi provocada deslocando o eixo do motor horizontalmente de 0.5mm, 1.0mm e 2.0mm, com variação de rotação de 737 a 3686 rpm;
- **Desalinhamento paralelo vertical:** esta falta foi induzida com deslocamento do eixo vertical do motor entre intervalo de 0.51 a 1.9mm, com variação de rotação de 737 a 3686 rpm;
- **Mancal interno:** localizado pela Figura 2.1, existem três tipos de rolamentos defeituosos (esfera, gaiola e parte externa), na ocorrência desta falha cada um destes foram colocados de cada vez no mancal de rolamento interno.

- **Mancal externo:** identificado pela Figura 2.1, os três tipos de rolamentos defeituosos, são inseridos de cada vez no mancal de rolamento externo.

As falhas no rolamento (mancal interno e externo) são praticamente imperceptíveis quando não há desequilíbrio. Assim, foram adicionadas três massas de 6g, 20g e 35g para induzir um efeito detectável, com diferentes frequências de rotação de 737 a 3686 rpm (MAFAULDA, 2017).

A Tabela 2.2 apresenta a quantidade de cenários para cada classe.

Tabela 2.2: Quantidade de cenários por classes.

| Classes                                  | Qtde de cenários |
|--|------------------|
| Normal (I)                               | 49               |
| Desbalanceamento (II)                    | 333              |
| Desalinhamento paralelo horizontal (III) | 197              |
| Desalinhamento paralelo vertical (IV)    | 301              |
| Mancal interno (V)                       | 558              |
| Mancal externo (VI)                      | 513              |
| Total                                    | 1951             |

Como foi descrito as classes mancais interno e externo, são compostas de subclasses de defeitos nos rolamentos exibido pela Tabela 2.3.

Tabela 2.3: Quantidade de cenários por subclasses.

| Classes               | Subclasses (defeitos) | Qtde de cenários |
|-----------------------|-----------------------|------------------|
| <i>Mancal interno</i> | esferas (VII)         | 186              |
|                       | gaiola (VIII)         | 188              |
|                       | parte externa (IX)    | 184              |
| <i>Mancal externo</i> | esferas (VII)         | 137              |
|                       | gaiola (VIII)         | 188              |
|                       | parte externa (IX)    | 188              |

## 2.2 Conclusão

Esse capítulo descreveu como foram simulados as faltas da base de dados de vibrações mecânica utilizada neste trabalho e as diversificações das classes e subclasses. Foi

---

possível averiguar a relevância da base (MAFAULDA, 2017), que se encontra em domínio público contribuindo significativamente para o aprimoramento da comunidade científica.

# Metodologia

---

Este capítulo apresenta a metodologia utilizada para criação de um sistema automático de diagnóstico de falhas em máquinas rotativas. Serão exibidas as estratégias de extração de características, a forma de implementação dos classificadores, as métricas de avaliação selecionadas. Posteriormente, apresenta um resumo da metodologia empregada. Por fim são apresentados as configurações do computador e software empregados.

## 3.1 Métodos de Extração de Características

Os dados adquiridos da base (MAFAULDA, 2017) são temporais e volumosos, necessita-se de métodos de extração de características que possam ser utilizados no diagnóstico de falhas em máquinas rotativas. A extração de características ao ser realizado de forma correta consegue-se reduzir, consideravelmente, o número de amostras dos sinais a serem analisados.

Nas subseções (3.1.1, 3.1.2 e 3.1.3), descrevem três métodos encontrados na literatura para extrair informações dos sinais de vibrações mecânicas.

### 3.1.1 Método de Extração de Características Baseado em RMS

O RMS (*root mean square*) ou raiz média quadrática de um sinal, apresentado pela Equação (3.1) é um recurso de análise no tempo, corresponde à medida do conteúdo de energia da assinatura de vibração, amplamente utilizado na literatura em máquinas rotativas (Tsytkin, 2013), (Guo et al., 2018). Esse recurso é clássico para análise de sinais e apresenta bons resultados para a rastreabilidade do nível geral de ruído. (Al-Ghamd e Mba, 2006). Pode ser muito eficaz na detecção de uma importante saída do equilíbrio em sistemas rotativos (Al-Ghamd e Mba, 2006).

$$RMS = \sqrt{1/Z * \sum_{i=1}^Z x_i^2} \quad (3.1)$$

em que  $i = 1, 2, \dots, Z$ , sendo  $Z$  o número de amostras.

Essa extração de características foi realizada para cada cenário, referente a Tabela 2.2, tanto para os seis sinais dos acelerômetros, quanto para os sinais dos tacômetro e microfone. O vetor de características resultante da extração possui oito valores de  $RMS$ , sendo utilizados na construção da matriz  $A$ , dado pela Equação (3.2).

$$A = \{RMS_1, RMS_2, RMS_3, RMS_4, RMS_5, RMS_6, RMS_7, RMS_8\}_{1951 \times 8} \quad (3.2)$$

### 3.1.2 Método de Extração de Características Baseado na Transformada Wavelet de Haar e Dimensão Fractal

Uma forma de transformada de sinais é a DWT (*Discrete Wavelet Transform*), que promove a decomposição do sinal em coeficientes de escala e detalhe. Ou seja, dado um sinal  $x$  unidimensional, periódico e de frequência discreta, coletado em uma determinada frequência de amostragem ao longo de certo tempo, através da Wavelet de Haar (Mallat, 1999), extraem-se a aproximação, apresentada pela Equação (3.3) e detalhe, apresentada pela Equação (3.4) do primeiro nível de decomposição.

$$a_i = \frac{x(2i-1) + x(2i)}{2} \quad (3.3)$$

$$d_i = x(2i-1) - x(2i) \quad (3.4)$$

em que  $i = 1, 2, \dots, N/2$ , sendo  $N$  o número de amostras.

As decomposições seguintes utilizam as mesmas Equações (3.3) e (3.4), calculadas sobre a componente de aproximação do nível anterior. Esse procedimento ocorre até um critério de parada. A Figura 3.1 apresenta a amplitude de um trecho de sinal na forma temporal de um acelerômetro e na Figura 3.2 exibe o primeiro nível de decomposição, através da transformada Wavelet de Haar.

A Transformada Wavelet de Haar é amplamente utilizada para extrair características para realização de diagnóstico de falhas em máquinas rotativas (Yan et al., 2014), (Asfani et al., 2012).

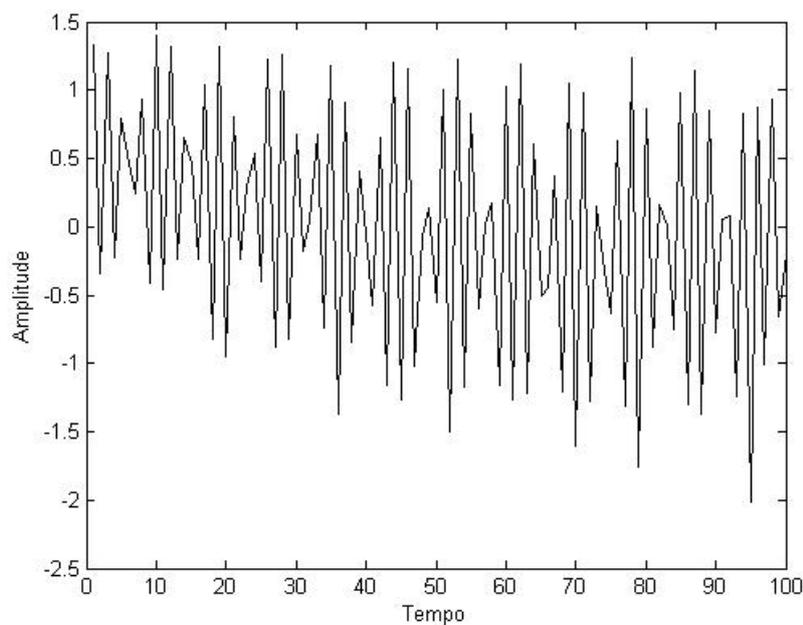


Figura 3.1: Trecho de sinal no domínio do tempo de um acelerômetro utilizado.

O método de extração de características baseado na transformada Wavelet de Haar (subseção 3.1.2) e dimensão fractal. Foi proposto por (Ramalho et al., 2014), através do conjunto de detalhes  $d_i$ , dado pela Equação (3.4), pode-se obter a extração da energia ( $En$ ), da entropia ( $Et$ ) e da dimensão fractal ( $Df$ ). É utilizado o primeiro nível de decomposição do sinal.

A energia ( $En$ ), apresentada pela Equação (3.5), fornece uma estimativa indireta da intensidade de vibração através do nível de variação do conjunto de detalhes.

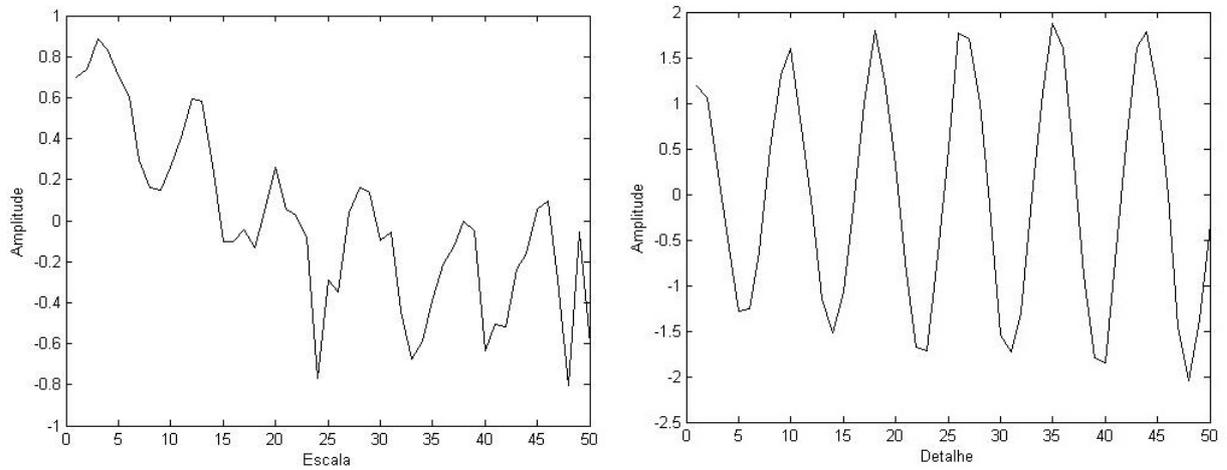


Figura 3.2: Transformada Wavelet de Haar de um acelerômetro, escala à esquerda e detalhe à direita.

$$En = 1/D * \sum_{i=1}^D d_i^2 \quad (3.5)$$

onde  $D$  é o número de detalhes dos sinais.

A entropia ( $Et$ ), exibida pela Equação (3.6), fornece uma estimativa do grau de desorganização dos sinais. Essa medida é realizada pela função densidade de probabilidade ( $P$ ). uma das formas de construção da função  $P$  é através de um histograma dos sinais.

$$Et = - \sum_{i=1}^D P(d_i) \log P(d_i) \quad (3.6)$$

A dimensão fractal ( $Df$ ), dada pela Equação (3.7), é uma métrica que estima o comprimento do desdobramento de um sinal, utilizando o algoritmo de contagem de caixas.

A Figura 3.3 exemplifica o algoritmo de contagem de caixas. O gráfico bidimensional é fragmentado em caixas (quadros), dessa forma o algoritmo faz a contagem dos quadros que possui algum trecho do sinal.



Figura 3.3: Exemplificação do algoritmo de contagem de caixas.

$$Df = \log N_b / \log R \quad (3.7)$$

em que  $N_b$  é o número total de caixas (*boxes*) e  $R$  é o número de caixas necessárias para cobrir os elementos não nulos do sinal (Falconer, 2004).

Embora essa estratégia tenha sido proposta para análise de vibrações (Ramalho et al., 2014), nesse experimento, tanto para os seis sinais dos acelerômetros, quanto para os sinais do tacômetro e microfone, calculado a partir dos oito sinais de cada cenário exibido pela Tabela 2.2. O vetor de características resultante da extração possui oito valores de  $En$ ,  $Et$  e  $Df$ , sendo utilizados na construção da matriz  $B$ , dado pela Equação (3.8), totalizando 24 dimensões.

$$B = \{En_1, Et_1, Df_1, \dots, En_8, Et_8, Df_8\}_{1951 \times 24} \quad (3.8)$$

### 3.1.3 Método de Extração de Características Baseado na Transformada de Fourier com Dados Estatísticos

A Transformada de Fourier é largamente utilizada para extrair características para efetuar o diagnóstico de falhas em máquinas rotativas (Mehala, 2010), (Amirat et al.,

2010), (Pestana-Viana et al., 2016). A Transformada de Fourier converte o sinal no domínio do tempo para frequência. Dado um sinal  $s(t)$  no domínio do tempo a Equação (3.9) define sua transformada.

$$F\{s(t)\} := F(f) = \int_{-\infty}^{\infty} s(t)e^{-i\omega t} dt \quad (3.9)$$

em que  $\omega$  é a frequência angular.

Que pode ser escrita, utilizando seno e cosseno pela Equação (3.10).

$$F(f) = \int_{-\infty}^{\infty} s(t) \cos(2\pi ft) dt - i \int_{-\infty}^{\infty} s(t) \sin(2\pi ft) dt \quad (3.10)$$

A transformada se baseia em sinais contínuos e periódicos que podem ser vistos como um conjunto de somatório de funções seno e cosseno. Para o caso de sinais discretos, que é o caso em questão abordado neste trabalho, existe sua modificação para DFT (*Discrete Fourier Transform*) (Rader, 1968), exibido pela Equação (3.11).

$$f_n = \sum_{k=0}^{N-1} f_k e^{-inx_k} \quad (3.11)$$

em que  $f_k = f(x_k), n = 0, \dots, N - 1$  (Kreyszig, 2010).

Em (Cochran et al., 1967) os autores apresentam um algoritmo eficiente para o cálculo da DFT, denominado FFT (*Fast Fourier Transform*). A FFT divide o somatório em duas partes relacionadas aos índices pares (real) e ímpares (complexo). Assim sendo, um sinal de tamanho  $N$  pode ser reescrito como a soma de duas Transformadas de Fourier, cada uma com tamanho igual a  $M = N/2$ , ou seja,  $N = 2M$ . A função exponencial pode ser reescrita pela Equação (3.12) (Kreyszig, 2010).

$$W_N^2 = W_{2M}^2 = (e^{-2\pi i/N})^2 = e^{-4\pi i/(2M)} = e^{-2\pi i/M} = W_M \quad (3.12)$$

Os componentes da DFT (Equação (3.11)), podem ser reescritos como:

$$f_n = \sum_{k=0}^{N-1} W_N^{kn} f_k \quad (3.13)$$

A FFT possui dois somatórios de  $M = N/2$ , apresentada pela Equação(3.14)(Kreyszig, 2010).

$$f_n = \sum_{k=0}^{M-1} W_N^{2kn} f_{2k} + \sum_{k=0}^{M-1} W_N^{(2k+1)n} f_{2k+1} \quad (3.14)$$

Conforme demonstrado pela Equação (3.12),  $W_N^2 = W_M$  e colocado o termo  $W_M^n$  no segundo somatório. A FFT pode ser reescrita pela Equação (3.15) (Kreyszig, 2010).

$$f_n = \sum_{k=0}^{M-1} W_N^{kn} f_{par,k} + W_M^n \sum_{k=0}^{M-1} W_M^{kn} f_{impar,k} \quad (3.15)$$

Nesta pesquisa, optou-se pela FFT, sendo a mais indicada para elaboração do método de extração de características descrito nesta subseção, pois apresenta menor esforço computacional (Shirado et al., 2015).

A Figura 3.1 exibe a amplitude de um trecho de sinal na forma temporal e na Figura 3.4 exibe sua transformação no espectro de frequência através da FFT.

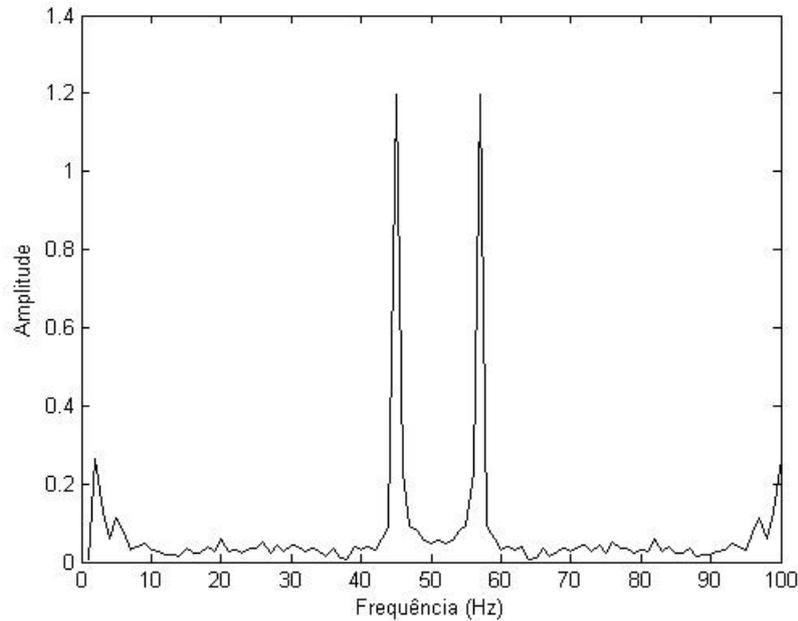


Figura 3.4: Sinal no espectro da frequência de um acelerômetro, através da FFT.

O método de extração de características baseado na FFT com dados estatísticos, foi proposto em (Marins et al., 2017) e se divide em duas etapas: a primeira etapa, no domínio da frequência, a segunda etapa, no domínio do tempo.

A base (MAFAULDA, 2017) apresenta uma informação relevante: o valor de rotação (fornecida pelo tacômetro), dado que permite o cálculo da frequência de rotação do motor (Lima et al., 2013).

Inicialmente, é realizada a FFT para todos os sinais dos sensores em cada cenário apresentado pela Tabela 2.2. Com o resultado da transformada do sinal do tacômetro, com a Equação (3.16), se extrai o índice da frequência ( $I_d$ ) com maior magnitude. Posteriormente, ao valor desta magnitude é atribuído o valor zero.

$$I_d = \arg \max |S_I(t)| \quad (3.16)$$

Sendo  $S_I(t)$ , valores da magnitude dos sinais da FFT.

Conforme a Equação (3.17), se extrai a frequência  $f_i$ .

$$f_t = I_d * f_s / N \quad (3.17)$$

em que  $f_s$  é a frequência de amostragem de  $N$  amostras.

Esse processo, exibido pelas Equações (3.16) e (3.17), é repetido por mais três vezes resultando em quatro frequências, alterando o valor do  $I_d$ , e pela Equação (3.18), é extraída a frequência de rotação  $Rf$ .

$$Rf = \min\{f_{t1}, f_{t2}, f_{t3}, f_{t4}\} \quad (3.18)$$

O valor resultante da frequência de rotação é o primeiro atributo extraído. Em seguida, são extraídos a magnitude da FFT nos outros sinais pelo valor fundamental  $Rf$  e nos dois primeiros harmônicos  $2Rf$  e  $3Rf$  (Lima et al., 2013). Esse passo inicial de extração é executado em todos os sinais (seis acelerômetros e o microfone). O vetor de características da extração possui além do valor  $Rf$ , sete valores de  $Rf_i$ ,  $2Rf_i$  e  $3Rf_i$  de cada cenário, exibido pela Tabela 2.2, resultando em 22 características:  $\{Rf, Rf_2, 2Rf_2, 3Rf_2, \dots, Rf_8, 2Rf_8, 3Rf_8\}$ .

No passo seguinte, no domínio do tempo, são retirados os seguintes atributos dos valores medidos de cada sensor: a entropia ( $He$ ), a média ( $Me$ ) e a curtose ( $Ku$ ). Estes valores são descrito a seguir.

Entropia ( $He$ ), dada pela Equação (3.19), a qual estima o grau de desorganização dos sinais no domínio do tempo.

$$He = - \sum_{i=1}^N P(x_i) \log P(x_i) \quad (3.19)$$

a função densidade de probabilidade ( $P$ ) é construída através de um histograma dos sinais temporais.

A média ( $Me$ ) é o momento de primeira ordem de sinal para as  $N$  amostras, dada pela Equação (3.20).

$$Me = \frac{\sum_{i=1}^N(x_i)}{N} \quad (3.20)$$

A curtose ( $Ku$ ) é a medida estatística de quarta ordem de um sinal, dada pela Equação (3.21).

$$Ku = \frac{E[(X - M_X)^4]}{E[(X - M_X)^2]^2} \quad (3.21)$$

onde  $M_X$  é a média da variável aleatória  $X$ .

O vetor de características da extração do segundo passo, possui oito valores de  $He$ ,  $Me$  e  $Ku$  de cada cenário, da Tabela 2.2, com as medidas dos oitos sensores, resultando em 24 características:  $\{He_1, Me_1, Ku_1, \dots, Me_8, He_8, Ku_8\}$ .

Ao final desse método de extração obtém-se a matriz de características:

$$C = \{Rf, Rf_2, 2Rf_2, 3Rf_2, \dots, Rf_8, 2Rf_8, 3Rf_8, He_1, Me_1, Ku_1, \dots, Me_8, He_8, Ku_8\}_{1951 \times 46} \quad (3.22)$$

totalizando 46 dimensões.

## 3.2 Classificação

Para se fazer a classificação utiliza-se o aprendizado de máquina, que consiste no princípio de inferência denominado indução, para se obter conclusões genéricas através de um universo limitado de exemplos. No aprendizado de máquina "supervisionado", há a figura do professor que possui o conhecimento e, através de entradas e saídas preestabelecidas, este é capaz de rotular a saída para novas entradas apresentadas. Já no aprendizado "não supervisionado", não existe a figura do professor e o algoritmo,

de forma contínua, tende a agrupar as entradas para fornecer suas respectivas saídas (James et al., 2013).

Na Figura 3.5, a base de dados  $X$  possui  $m$  atributos ou características e  $n$  amostras, ou seja,  $X = (x_{i,j}, \dots, x_{n,m})$ . Para cada amostra  $x_i$ , haverá uma variável de saída  $y_i$  que representa a classificação (classes) (D. Silva e Scarpel, 2007). No conjunto rotulado em  $(x_i, y_i)$ , em que  $x_i$  é uma coleta e  $y_i$  é seu rótulo, o processo de indução de um classificador  $f(x)$  é denominado treinamento. O classificador pode ser visto em uma função  $f$ : para cada  $x$  há sua predição  $y$  (Lorena e de Carvalho, 2007). Sendo que o conjunto  $\mathbb{F}$  representa todos os possíveis classificadores de um determinado algoritmo de aprendizagem.

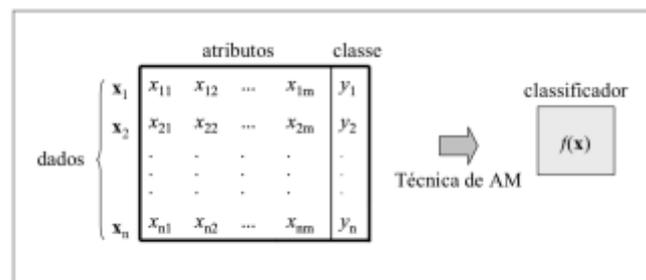


Figura 3.5: Diagrama básico de um aprendizado de máquina para um classificador. Figura retirada e adaptada de (Lorena e de Carvalho, 2007).

O algoritmo de aprendizado de máquina utiliza um conjunto de treinamento  $\mathbb{T}$ , composto por  $n$  pares (amostras e rótulos) (Baranauskas e Monard, 2000). Como exemplo a Figura 3.6 apresenta três hipóteses com objetivo de conseguir uma fronteira de classificação para separar os triângulos dos círculos. A Figura 3.6 (a) apresenta três funções capazes de fazer a classificação correta das amostras. Já a Figura 3.6 (b) exibe somente uma função flexível, em que duas amostras são classificadas incorretas. Enfim, a Figura 3.6 (c) apresenta como resultado uma função linear, em que inúmeras amostras são rotuladas de classes erradas.

A classificação é a última etapa deste trabalho, são utilizados três tipos de classificadores KNN (K-vizinhos mais próximos), SVM (Máquina de Vetores de Suporte) e XGBoost (*Extreme Gradient Boosting*), e cada um desses, foram combinados com os vetores de características resultantes nos métodos descritos nas subseções 3.1.1, 3.1.2 e 3.1.3. Foram realizadas 30 execuções de cada modelo de classificação (Rezende et al., 2017). Em cada execução, o conjunto de dados são divididos randomicamente 70% utilizados para treinamento e 30% para teste (Platt et al., 2000), e também subdivide

aleatoriamente, o conjunto de treinamento disponível, utilizando a técnica de validação cruzada, com o número de partição estabelecido, *k-fold* ( $k=10$ ). Os dados de partição usados para o aprendizado formam o conjunto de aprendizagem e os demais partição formam um conjunto de validação. Ao final de cada modelo, se tem 30 resultados de desempenho. O Algoritmo 1 exemplifica a estratégia padronizada de classificação. Nas Subseções 3.2.1, 3.2.2 e 3.2.3 serão apresentados os classificadores utilizados.

---

**Algoritmo 1:** Classificação
 

---

**Entrada:** Características extraídas( $A_{1951 \times 8} | B_{1951 \times 24} | C_{1951 \times 46}$ )

**Saída:** acurácia, MAUC, matriz de confusão

```

1 início
2   para  $i = 1$  até 30 faça
3     ALEATORIZA AS AMOSTRAS DE ENTRADA
4     treino ← 70% dos dados
5     teste ← 30% dos dados
6     VALIDAÇÃO CRUZADA K-FOLD=10
7     para  $k = 1$  até 10 faça
8       modelo ← treinClassif(treino(k))
9       acc ← predi(modelo, teste(k))
10    fim
11     $\{ACC_i, MAUC_i, MatConf\} \leftarrow modelo(treino)$ 
12  fim
13 fim
  
```

---

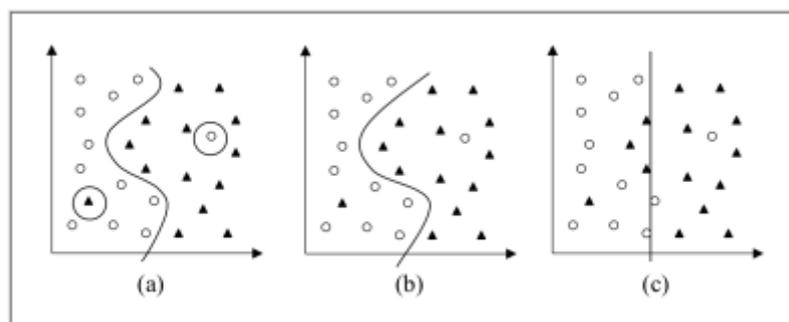


Figura 3.6: Conjunto de treinamento binário e três diferentes hipóteses. Figura retirada e adaptada de (Lorena e de Carvalho, 2007).

### 3.2.1 *K-Nearest Neighbors (KNN)*

O KNN é um classificador clássico descrito em (Cover e Hart, 1967), onde o aprendizado é baseado na memória. O conjunto de treinamento é formado por vetores

$m$ -dimensionais e cada elemento deste conjunto representa um ponto no espaço  $m$ -dimensional (M. Silva, 2005).

O kNN possui dois passos, o primeiro denominado treino em que se armazenam todas as amostras de treinamento rotuladas. No segundo passo, ou também chamada de fase de classificação, primeiramente calcula-se as distâncias de uma nova amostra a ser classificada com todos os padrões do conjunto de treinamento. Em seguida, o algoritmo considera os  $k$  padrões do conjunto de treinamento com as menores distâncias ao padrão a ser classificado. Por último, o novo padrão é classificado como pertencente à classe da maioria dos  $k$  padrões mais próximos do conjunto de treinamento (Reis Gade et al., 2017). O

As métricas de distância mais utilizada no KNN são Euclidiana, Manhattan e Minkowski, foram adotados, para execução deste trabalho a distância Euclidiana e se manteve o valor de  $K = 1$ . A Figura 3.7 apresenta uma comparação dos limites de decisão do KNN, curvas obtidas usando  $K = 1$  e  $K = 100$ . Com  $K = 1$ , o limite de decisão é excessivamente flexível, enquanto com  $K = 100$  é menos ajustável.

Neste trabalho, optou-se em utilizar o classificador KNN, pois nos resultados: método de extração de características baseado em RMS (Subseção 3.1.1) e método de extração de características Baseado na Transformada Wavelet de Haar e dimensão fractal (Subseção 3.1.2). Os atributos resultantes foram plotados em gráficos bidimensionais combinados em pares. Nos gráficos foi perceptível em alguns casos o agrupamento das classes e subclasses. A existências de aglomerações é um bom indicativo para o uso do KNN.

### 3.2.2 *Support Vector Machine (SVM)*

Máquinas de Vetores Suporte ou SVM é um classificador não-paramétrico. É uma técnica de aprendizado que possui minimização do risco estrutural. Esta técnica busca minimizar o erro com relação ao conjunto de treinamento, assim como o erro com relação ao conjunto de teste, isto é, conjunto de amostras não empregadas no treinamento do classificador (Adreola, 2009). O objetivo do SVM consiste em obter um equilíbrio entre esses erros, minimizando o excesso de ajustes com respeito às amostras de treinamento e aumentando conseqüentemente a capacidade de generalização (Vapnik, 2013).

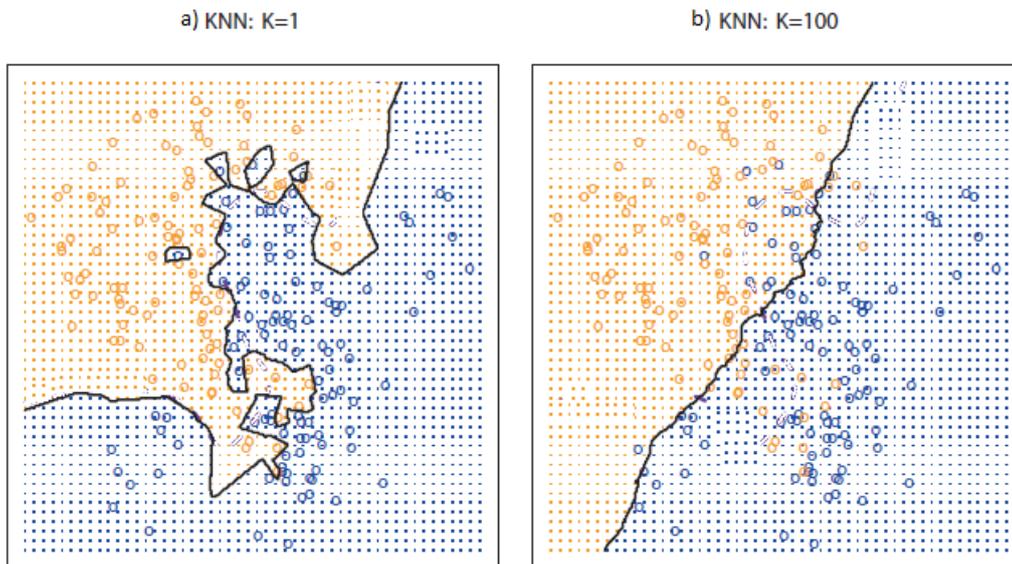


Figura 3.7: Curvas de Classificação para o KNN a)  $k = 1$ , b)  $k = 100$ . Figura retirada e adaptada de (James et al., 2013).

Assumindo que as amostras de treinamento de duas classes são linearmente separáveis, a função de decisão mais adequada é aquela em que a distância entre os conjuntos das amostras de treinamento é maximizada. Neste contexto, a função de decisão que maximiza esta separação é denominada de ótima, apresentada pela Figura 3.8.

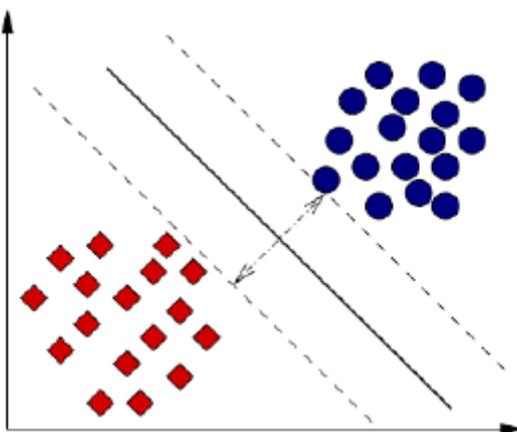


Figura 3.8: Exemplo de superfície de separação gerada pelo SVM. Figura retirada e adaptada de (Abe, 2005).

Um hiperplano é apresentada pela seguinte equação:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (3.23)$$

em que  $\mathbf{x}$  é o vetor de entrada,  $\mathbf{w}$  é o vetor de pesos e  $b \in \mathbb{R}$  (Haykin, 2009). Essa equação divide o espaço dos dados  $\mathbf{x}$  em duas regiões:  $\mathbf{w} \cdot \mathbf{x} + b > 0$  e  $\mathbf{w} \cdot \mathbf{x} + b < 0$ . Uma função  $f(x)$  pode então ser empregada na obtenção das classificações, conforme exibido nas equações a seguir:

$$f(x) = \begin{cases} \mathbf{x}_i \cdot \mathbf{w} + b \geq +1 \text{ para } y_i = +1 \\ \mathbf{x}_i \cdot \mathbf{w} + b \leq -1 \text{ para } y_i = -1 \end{cases} \quad (3.24)$$

onde  $f(x)$  mapeia duas classes  $\{-1, +1\}$ .

A maximização da margem de separação dos dados em relação a  $\mathbf{w} \cdot \mathbf{x} + b = 0$ , pode ser obtida pela minimização de  $\|\mathbf{w}\|^2$  (Burgess, 1998). Dessa forma, recorre-se ao seguinte problema de otimização primal:

$$\text{Minimizar: } \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.25)$$

$$\text{sujeito a: } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad \forall i = 1, 2, \dots, n \quad (3.26)$$

onde  $n$  é a quantidade de padrões do conjunto de treinamento.

O problema de otimização obtido é quadrático, cuja solução possui uma ampla e estabelecida teoria matemática (Smola et al., 2000). A função objetivo que será minimizada é convexa e os pontos que satisfazem as restrições formam um conjunto convexo, esse problema possui um único mínimo global. Problemas assim podem ser solucionados com a introdução de uma função Lagrangeana, que engloba as restrições à função objetivo, associadas a parâmetros denominados multiplicadores de Lagrange  $\alpha_i$ , Equação (3.27) (Burgess, 1998).

$$L(\mathbf{w}, \mathbf{b}, \alpha) = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) \quad (3.27)$$

A função Lagrangeana deve ser minimizada, o que implica em maximizar as variáveis  $\alpha_i$  e minimizar  $w$  e  $b$  (Borges, 1998). Fazendo o uso da derivada parcial tem-se um ponto de sela exibido pela Equação (3.28).

$$\frac{\partial L}{\partial b} = 0 \text{ e } \frac{\partial L}{\partial w} = 0 \quad (3.28)$$

A solução da derivação resulta nas Equações (3.29 e 3.30) (Borges, 1998).

$$\mathbf{w} = \sum_{i=1} \alpha_i y_i \mathbf{x}_i \quad (3.29)$$

$$\sum_{i=1} \alpha_i y_i = 0 \quad (3.30)$$

Substituindo as Equações (3.29 e 3.30) na Equação (3.27), resulta na Equação (3.31) (Borges, 1998).

$$L_D = \sum_{i=1} \alpha_i - \frac{1}{2} \sum_{i,j=1} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{y}_j \quad (3.31)$$

Ao maximizar a equação (3.31) tem-se como resultado o seguinte problema dual:

$$\text{Maximizar: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (3.32)$$

$$\text{sujeito a: } \begin{cases} \alpha_i \geq 0, \forall i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (3.33)$$

onde  $\alpha$  é o vetor que representa os multiplicadores de Lagrange e  $\alpha^*$  é a solução ótima para o problema dual.

Seja  $\alpha^*$  a solução do problema dual e  $w^*$  e  $b^*$  as soluções da forma primal. Obtido o valor de  $\alpha^*$ ,  $w^*$  pode ser determinado pela Equação (3.29). O parâmetro  $b^*$  é definido por  $\alpha^*$  e por  $\alpha^*$  e por condições de Karush-Kuhn-Tucker, com  $\alpha_i^*(y_i(w^* \cdot x_i + b^*) - 1) = 0 \forall i = 1, 2, \dots, n$  (Burges, 1998).

Os SVMs lineares são eficientes para separação de dados linearmente separáveis. Porém há casos que não é possível fazer a separação satisfatória. Um exemplo se mostra na Figura 3.9 em que a fronteira curva seria melhor para separação das classes (Lorena e de Carvalho, 2007). Para solucionar esta classe de problemas o SVM utiliza funções Kernel.

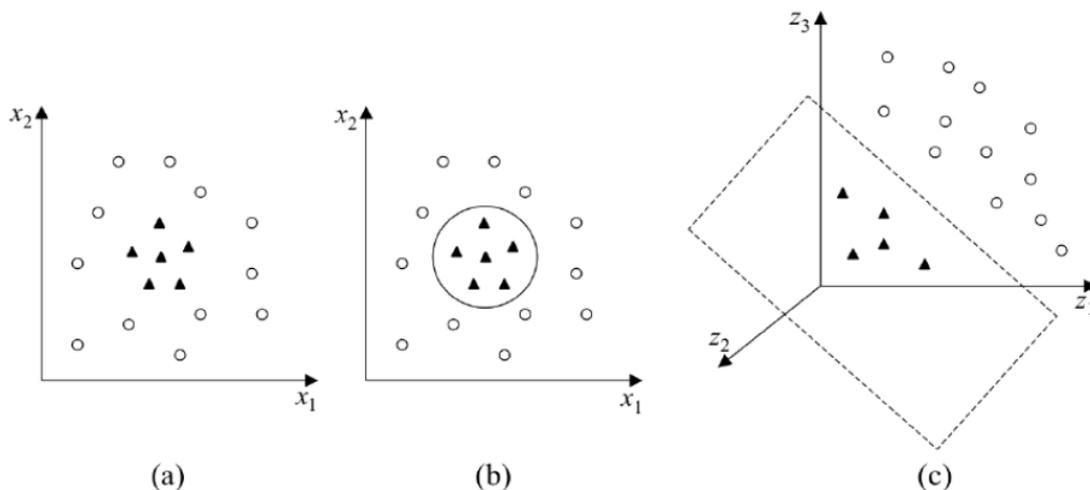


Figura 3.9: Conjunto de dados não linearmente separáveis. Figura retirada e adaptada de (Lorena e de Carvalho, 2007).

As representações Kernel trabalha com a projeção dos dados em um espaço de características com alta dimensão para permitir a classificação em espaços não linearmente separáveis. Trata-se, em primeira instância, de uma estratégia de pré-processamento que envolve mudar a representação dos dados da seguinte forma:

$$\mathbf{X} = (x_1, \dots, x_n) \Rightarrow \Phi(x) = (\Phi_1(x), \dots, \Phi_n(x)) \quad (3.34)$$

nesse momento ocorre um mapeamento do espaço de entrada  $\mathbf{X}$  em um novo espaço  $Z = \{\Phi(x) | x \in \mathbf{X}\}$  chamado espaço de características em  $\Phi_i$  são funções kernel. A Figura

3.10 mostra um mapeamento de um espaço de entrada linearmente inseparável, para um espaço de características de maior dimensão, em que os dados podem ser separados linearmente (Sinhorigno, 2007).

Os SVMs com função Kernel tem a seguinte função objetivo  $\mathbf{x}$ :

$$\text{Maximizar: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j Q(\mathbf{x}_i, \mathbf{x}_j) \quad (3.35)$$

$$\text{sujeito a: } \begin{cases} \alpha_i \geq 0 \geq C, \forall i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (3.36)$$

onde  $C$  é um parâmetro que visa ponderar a minimização do erro no conjunto de treinamento e  $Q(\mathbf{x}_i, \mathbf{x}_j)$  representa a função Kernel.

A literatura é repleta de funções Kernel como: polinomial, perceptron, sigmoide ou *Radial Basis Function* (RBF). Neste trabalho, optou-se pela utilização da RBF devido a existência dos agrupamentos descritos na subseção anterior, pois eles podem ser melhores identificados com uma superfície gaussiana, bem como os autores de Marins et al. (2017) apresentou bons resultados aplicando-o.

O Kernel RBF é definido como:

$$Q(\mathbf{x}_i, \mathbf{x}_j) = Q_{RBF}(x, x') = \exp[-\gamma \|x - x'\|^2] \quad (3.37)$$

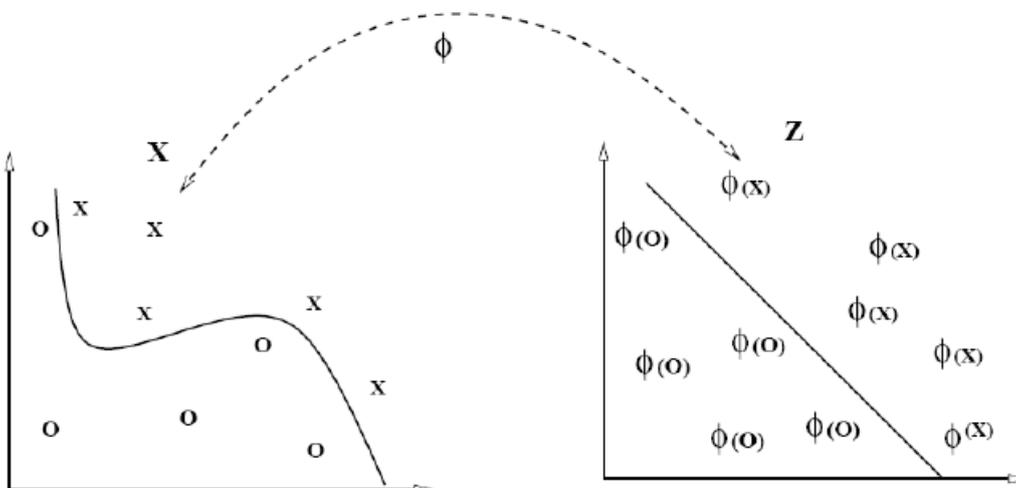


Figura 3.10: Mapeamento do espaço de entrada via função kernel. Figura retirada e adaptada de (Sinhorigno, 2007).

em que  $\gamma$  é um parâmetro que define a propagação do Kernel que define o raio da superfície gaussiana. O kernel calcula o produto interno entre dois vetores projetados.

Neste trabalho, os parâmetros  $C$  e  $\gamma$  Equações (3.36 e 3.37) são ajustados através da técnica *grid search*, que escolhe, dentre vários parâmetros, a melhor combinação.

### 3.2.3 Extreme Gradient Boosting (XGBoost)

O *Extreme Gradient Boosting* (XGBoost) é um algoritmo de classificação, o qual é considerado uma melhoria do algoritmo *gradient boosting* de (Friedman, 2001), cujas modificações estão descritas em (Chen e Guestrin, 2016). O XGBoost é recente na literatura, apresenta bons resultados e em seu modelo utiliza-se do princípio de árvore de decisão, como exemplificado pela Figura 3.11.

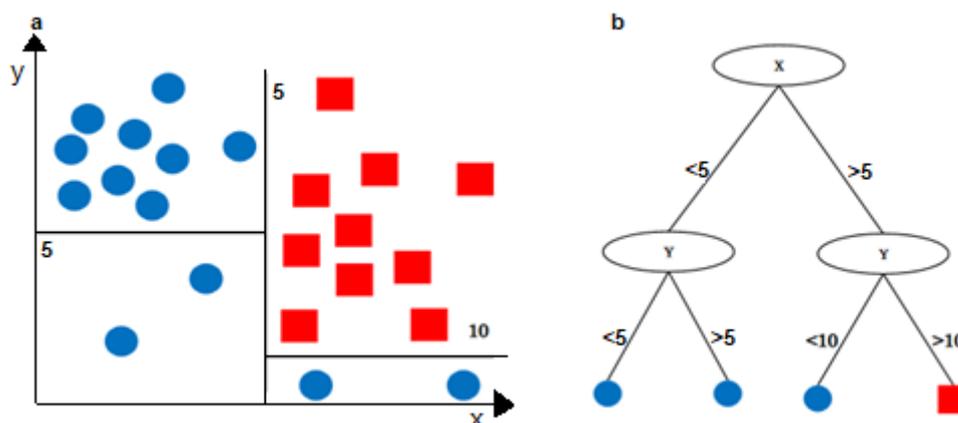


Figura 3.11: a) Gráfico bidimensional de dados de classificação e b) Resultado da árvore de decisão.

Os métodos baseados em árvores de decisão são simples e úteis para interpretação, no entanto, eles geralmente não são competitivos em relação às melhores abordagens de aprendizagem supervisionadas clássicas, em termos de precisão (James et al., 2013). Uma alternativa de melhoria é a combinação de várias árvores de decisão, tendo assim uma floresta, que resulta em um classificador com alto poder preditivo. Essa é a estratégia adotada pelo classificador *XGBoost*.

Dado que  $X$  é o conjunto de entrada de treinamento e  $Y$  é a sua saída. O XGBoost é um modelo iterativo que a cada passo tende a melhorar o modelo de árvore de decisão anterior, isso pode ser descrito na seguinte forma:

$$h(x) = \sum_{b=1}^B f_b(x), f_b \in F \quad (3.38)$$

em que  $h(x)$  é a predição,  $B$  é o número de árvores,  $f$  é uma função no espaço funcional  $F$  e  $F$  é o conjunto de todas as possíveis florestas. Dessa forma é necessário possuir a função objetivo que pode ser descrita da seguinte forma:

$$obj(\theta) = \sum_i^n l(y_i, h(x)) + \sum_{b=1}^B \Omega(f_b) \quad (3.39)$$

em que  $\theta$  são os parâmetros de otimização,  $l$  é a função de perda,  $\Omega$  é a função de regularização e  $n$  é o número de amostras.

Conforme a modelagem introduzida, pode-se explicar os passos do treinamento. Considerando a seguinte função objetivo modificada:

$$obj = \sum_{i=1}^n l(y_i, h(x)[t]) + \sum_{i=1}^t \Omega(f_i) \quad (3.40)$$

em que  $t$  é o número de passos de treinamento,  $f_i$  são funções de aprendizado. É necessário enfatizar que cada uma das funções de aprendizado,  $f_i$  contém a estrutura da árvore de decisão e as pontuações dos nós das folhas.

Sabendo que o algoritmo é iterativo que a cada iteração, tenta melhorar o modelo de árvore de decisão anterior. Atribuindo o valor da predição  $h(x)$  nos passos  $t$ , tem se:

$$h(x)[0] = 0 \quad (3.41)$$

$$h(x)[1] = f_1(x_i) = h(x)[0] + f_1(x_i) \quad (3.42)$$

$$h(x)[2] = f_1(x_i) + f_2(x_i) = h(x)[1] + f_2(x_i) \quad (3.43)$$

...

$$h(x)[t] = \sum_{k=1}^t f_k(x_i) = h(x)[t-1] + f_t(x_i) \quad (3.44)$$

A árvore de decisão que é inserida em cada etapa, tem sua estrutura criada pela importância das características através de uma avaliação combinada e que otimiza a função objetivo (Cho, 2017):

$$obj^{(t)} = \sum_{i=1}^n l(y_i, h(x)[t]) + \sum_{i=1}^t \Omega(f_i) \quad (3.45)$$

$$= \sum_{i=1}^n l(y_i, h(x)[t-1] + f_t(x_i)) + \Omega(f_t) \quad (3.46)$$

Com a inserção do erro médio quadrático na função perda, ela terá a seguinte formulação:

$$obj^{(t)} = \sum_{i=1}^n (y_i - (h(x)[t-1] + f_t(x_i)))^2 + \sum_{i=1}^t \Omega(f_i) \quad (3.47)$$

$$obj^{(t)} = \sum_{i=1}^n [2(h(x)[t-1] - y_i)f_t(x_i) + f_t(x_i)^2] + \Omega(f_t) \quad (3.48)$$

$$(3.49)$$

O Algoritmo 2 apresenta o XGBoost para classificação.

---

**Algoritmo 2: XGBoosting**


---

**Entrada:** conjunto de treinamento  $(x_i, y_i)_{i=1}^n$ , função de perda diferenciável  $l(y_i, h(x))$ , número de iterações  $B$

1 **início**

2  $h(x)[0] = \arg_{\Omega} \min \sum_{i=1}^n l(y_i, h(x)).$

3 **para**  $b = 1$  até  $B$  **faça**

4     A) CALCULAR PSEUDO-RESÍDUOS:  $r_{ib} \leftarrow -\left[\frac{\partial l(y_i, h(x))}{\partial h(x)}\right]_{h(x)=h(x)[b-1]}$  PARA  $i = 1, \dots, n$ ;

5     B) AJUSTAR UM APRENDIZ BASE  $f_b(x_i)$  PARA PSEUDO-RESIDUAIS, ISTO É, TREINÁ-LO USANDO UM CONJUNTO DE TREINO  $(x_i, r_{ib})_{i=1}^n$ ;

6     C) CALCULAR O MULTIPLICADOR  $\Omega_b$  RESOLVENDO O SEGUINTE PROBLEMA DE OTIMIZAÇÃO UNIDIMENSIONAL:  $\Omega_b = \arg_{\Omega} \min \sum_{i=1}^n l(y_i, h(x)) + \Omega f_b(x_i)$ ;

7     D) ATUALIZAR O MODELO:  $h(x) = h(x)[b-1] + \Omega_b f_b(x_i)$ ;

8 **fim**

**Saída:**  $h(x)[B]$

9 **fim**

---

Neste trabalho, optou-se pela utilização do XGBoost devido aos bons resultados recentes em várias publicações (Georganos et al., 2018), (Dong et al., 2018). Além disso, as melhores performances encontrados nas publicações da base de dados que foi descrita no capítulo 2, utilizaram classificadores baseados em árvore de decisão (Marins et al., 2017), (Martins et al., 2016).

Os parâmetros do XGBoost, profundidade máxima e taxa de aprendizagem foram escolhidos através da técnica de *grid search*.

### 3.3 Métricas de desempenho

Esta seção exhibe e explana os formalismos adotados para aferir as performances dos classificadores durante os experimentos.

Com o uso da matriz de confusão<sup>1</sup> é possível extrair pela Equação (3.50) a acurácia (ACC). É uma métrica clássica e muito utilizada na literatura, contudo, tem problemas para análise de desempenho na existência de classes desbalanceadas (Pereira et al., 2014). No entanto, ela será utilizada neste trabalho para fins de comparação dos resultados obtidos com outras publicações. Uma das formas de validação da acurácia

é o uso da sensibilidade e especificidade exibidas pelas Equações (3.51 e 3.52). Os valores de sensibilidade e especificidade próximo a 1 evidência uma boa classificação pela acurácia.

$$ACC = \frac{(TP + TN)}{(TP + FN + FP + FN)} \quad (3.50)$$

$$Sen = \frac{TP}{(TP + FN)} \quad (3.51)$$

$$Esp = \frac{TN}{(TP + FN)} \quad (3.52)$$

sendo: ( $TP$ ) número de verdadeiro positivos, ( $TN$ ) número de verdadeiro negativos, ( $FP$ ) número de falso positivos e ( $FN$ ) número de falso negativos.

Existe outra métrica muito utilizada na literatura que evidência a classificação não homogênea em classes desbalanceadas. A métrica de desempenho área da curva ROC (*Receiver Operating Characteristic Curve*) ou AUC (*Area Under ROC*), pode ser demonstrado por: dado um conjunto de instâncias  $\mathbf{S} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , onde  $x_i \in \mathbb{R}^m$  sendo o vetor de características, e  $y_i \in \{0, 1\}$  rótulos das classes. O classificador  $U(x_i) \rightarrow \mathbb{R}$  produz resultados numéricos que indicam a confiança de que  $x_i$  pertence a classe (0 ou 1). A AUC de um classificador pode ser calculado através da Equação (3.53) (Tang et al., 2011).

$$AUC = \frac{\sum_{x_i \in \text{Classe}(0); x_j \in \text{Classe}(1)} \mathbf{S}(x_i, x_j)}{n_1 * n_0} \quad (3.53)$$

<sup>1</sup>Matriz de confusão fornece o número de previsões corretas e incorretas feitas pelo modelo em comparação com os resultados reais (valores-alvo) dos dados. (Gupta, 2015).

onde  $n_1$  e  $n_0$  denotam o número de instâncias das classes  $\{0,1\}$  e  $\mathbb{S}(x_i, x_j)$  é definido pela Equação 3.54 (Tang et al., 2011). A AUC é utilizado originalmente para classificação binária.

$$\mathbb{S}(x_i, x_j) = \begin{cases} 1, & \text{se } U(x_i) > U(x_j); \\ 0.5, & \text{se } U(x_i) = U(x_j); \\ 0, & \text{se } U(x_i) < U(x_j); \end{cases} \quad (3.54)$$

Para os problemas com múltiplas classes onde  $y_i \in \{1, \dots, Cl\}$ ,  $Cl$  é o número de classes, o classificador resulta em um matriz  $M_{n \times Cl}$ . O elemento  $m_{ij}$  indica a confiança de que  $i$ -ésima instância em  $\mathbb{S}$  pertence a classe  $j$ . *Multi-class Extension of AUC* (MAUC) pode ser definido pela Equação (3.55).

$$MAUC = \frac{2}{Cl * (Cl - 1)} \sum_{i < j} \frac{a_{i,j} + a_{j,i}}{2} \quad (3.55)$$

onde  $a_{i,j}$  é a AUC entre a classe  $i$  e classe  $j$  através da matriz  $M$  (Tang et al., 2011). É válido ressaltar para o cálculo da MAUC,  $a_{i,j}$  deve ser diferente de  $a_{j,i}$  (Hand e Till, 2001). A MAUC tem sido adotado na literatura como boa métrica, pois evidência problemas de classificação em multiclases desbalanceadas.

Portanto, através desses indicadores de desempenho. Neste trabalho optou-se pelo o emprego da ACC, especificidade, sensibilidade e MAUC.

## 3.4 Resumo

Esta seção resume em linhas gerais a metodologia adotada para elaboração deste trabalho. Com a utilização da base de dados, conforme descrito no capítulo 2, foi realizado as etapas subsequentes que foram transcritas no decorrer deste capítulo. A Figura 3.12 apresenta o diagrama em blocos da metodologia utilizada para um sistema de reconhecimento automático de falhas em máquinas rotativas.

Os dados da base (MAFAULDA, 2017) foram pré-processados nas três "Extrações de Características" (Figura 3.12) : método de extração de características baseado em RMS (subseção 3.1.1), método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2), método de extração de características baseado na FFT com dados estatísticos (subseção 3.1.3).

O "Classificador 1" (Figura 3.12), identifica as classes (normal (I), desbalanceamento (II), desalinhamento paralelo horizontal (III), desalinhamento paralelo vertical (IV), mancal interno (V) e mancal externo (VI) (Tabela 2.2)) que resultam na "Saída 1" (Figura 3.12).

Devido a complexidade dos mancais de rolamento exposto na subseção 2.1.1, optou-se em utilizar o "Classificador 2" (Figura 3.12), para as falhas nos mancais: interno e externo, para diagnosticar as subclasses (esferas (VII), gaiola (VIII) e parte externa (IX) (Tabela 2.3)) que fornece a resposta na "Saída 2" (Figura 3.12). Ambos classificadores utilizam os mesmos atributos de entrada. No entanto, o "Classificador 2" (Figura 3.12) fez uso das amostras com falhas nos mancais de rolamento com a inserção dos rótulos das subclasses (esferas (VII), gaiola (VIII) e parte externa (IX) (Tabela 2.3)).

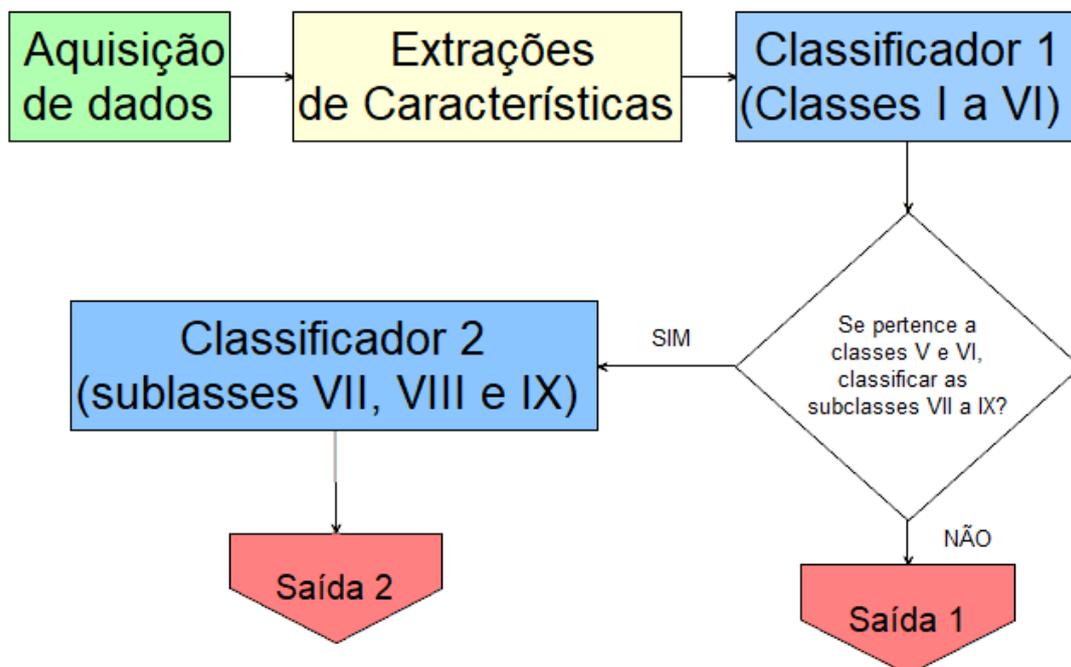


Figura 3.12: Diagrama em blocos da metodologia utilizada.

## 3.5 Software e Computador Utilizados

As extrações de características relatado nas subseções 3.1.1, 3.1.2 e 3.1.3, foram implementadas no MATLAB(2014a) e os classificadores foram desenvolvidos no ambiente RStudio (3.3). Utilizou-se um computador processador Intel(r) Core(TM) 2.5 GHz/4 Gb (RAM), Sistema Operacional Windows10 x64.

## 3.6 Conclusão

Esse capítulo descreveu a metodologia empregada para implementação de um sistema automático de reconhecimento de falhas em máquinas rotativas. Inicialmente foi exibido os formalismo adotados nas três formas de extração de características, bem como os ajustes e o algoritmos de classificação. Posteriormente, os indicadores de desempenho. Finalmente, foram expostos os softwares e o computador.

# Resultados das Classificações das Falhas em Máquinas Rotativas

Neste capítulo são apresentados os resultados obtidos pelos algoritmos de classificação (SVM, KNN e XGBoost), para classificar os dados da base (MAFAULDA, 2017) nas classes representadas pela Tabela 2.2 (normal (I), desbalanceamento (II), desalinhamento paralelo horizontal (III), desalinhamento paralelo vertical (IV)).

Os métodos descritos nas subseções: método de extração de características baseado em RMS(subseção 3.1.1), método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2), método de extração de características baseado na FFT com dados estatísticos (subseção 3.1.3), serão tratados, respectivamente, como A, B e C. As métricas utilizadas como indicadores de desempenho foram a acurácia (ACC), *Multi-class Extension of AUC* (MAUC) e produto entre sensibilidade e especificidade (*Sen X Esp*). O produto (*Sen X Esp*) próximo ao valor 1, valida uma classificação homogênea em classes binárias. Já a MAUC próximo ao valor 1, evidência a classificação homogênea entre as multiclasses. O SVM com kernel RBF será tratado como (SVM), *K-Nearest Neighbors* como KNN e o *Extreme Gradient Boosting* como XGBoost ou XG. A Tabela 4.1 sintetiza como serão tratados os modelos de classificação.

Tabela 4.1: Siglas dos modelos de classificação.

| Sigla  | Classificador e forma de extração                                     |
|--------|---|
| KNN(A) | KNN + método de extração baseado em RMS.                              |
| KNN(B) | KNN + método de extração baseado na transformada wavelet de Haar.     |
| KNN(C) | KNN + método de extração baseado em transformada de Fourier.          |
| SVM(A) | SVM + método de extração baseado em RMS.                              |
| SVM(B) | SVM + método de extração baseado na transformada wavelet de Haar.     |
| SVM(C) | SVM + método de extração baseado em transformada de Fourier.          |
| XG(A)  | XGBoost + método de extração baseado em RMS.                          |
| XG(B)  | XGBoost + método de extração baseado na transformada wavelet de Haar. |
| XG(C)  | XGBoost + método de extração baseado em transformada de Fourier.      |

## 4.1 Resultados e Análise

Cada modelo de classificação apresentou 30 matrizes de confusões com várias classes. Posteriormente, foi realizada a transformação dos resultados em uma classificação binária em cada uma das classes. A classificação binária consiste em cada classe se transformou em uma classe alvo e as demais como outras. O número total de classes foram balanceados. Por fim, foram extraídos a acurácia binária e os valores acumulados de sensibilidade e especificidade. As Tabelas 4.2 e 4.3 mostram os resultados obtidos por cada modelo em cada uma das classes, além do desvio padrão dos valores por ACC e o produto entre sensibilidade X especificidade. É importante salientar que o produto de  $(Sen \times Esp)$  próximo a 1, significa que o método foi capaz de obter uma classificação homogênea entre as classes, ou seja, o método não privilegia uma classe em relação às outras (Pereira, 2012). Todos os classificadores obtiveram um bom desempenho com uma taxa de acurácia média de 93,05%. O XG(B) se destaca, pois atingiu os melhores resultados em todas as classes alvo.

As Figuras 4.1, 4.2 e 4.3, exibem os valores de ACC global e MAUC alcançados em cada uma das 30 execuções do algoritmo de classificação. O método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2) foi mais efetivo para o SVM e Xgboost. Já o método de extração de características baseado em RMS (subseção 3.1.1) foi melhor para o KNN.

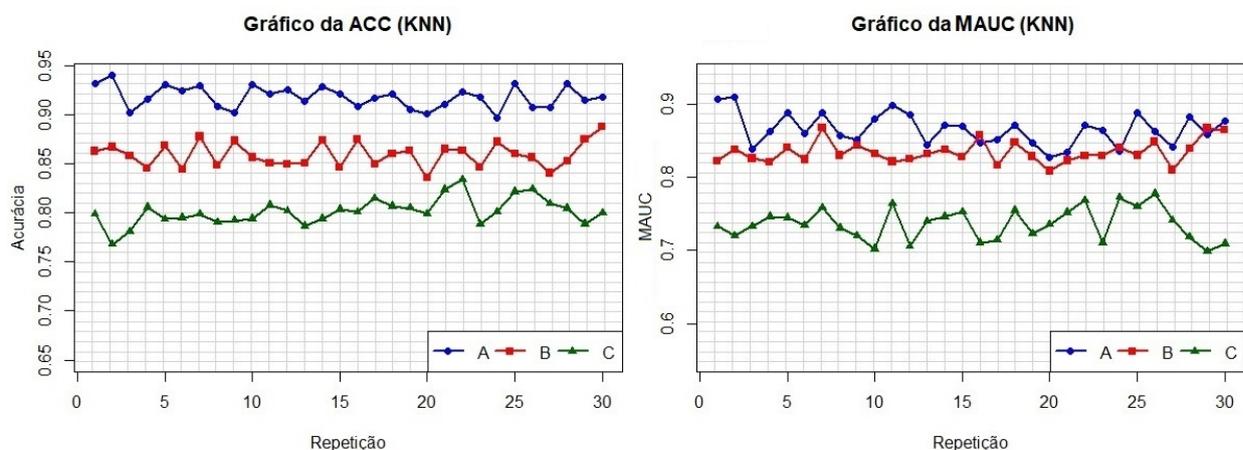


Figura 4.1: Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do KNN.

Tabela 4.2: Resultados intraclases de ACC para as classes (I), (II) e (III).

| Classe alvo                                    | Modelo       | ACC                  | Sen X Esp     |
|--|--------------|----------------------|---------------|
| <i>Normal (I)</i>                              | KNN(A)       | 0.8756 ±0.014        | 0.7523        |
|  | KNN(B)       | 0.8165 ±0.054        | 0.6365        |
|  | KNN(C)       | 0.6508 ±0.051        | 0.3107        |
|  | SVM(A)       | 0.7272 ±0.033        | 0.4552        |
|  | SVM(B)       | 0.8915 ±0.007        | 0.7838        |
|  | SVM(C)       | 0.8629 ±0.064        | 0.7267        |
|  | XG(A)        | 0.9723 ±0.034        | 0.9448        |
|  | <b>XG(B)</b> | <b>0.9900 ±0.018</b> | <b>0.9800</b> |
|  | XG(C)        | 0.9872 ±0.021        | 0.9745        |
| <i>Desbalanceamento (II)</i>                   | KNN(A)       | 0.9366 ±0.049        | 0.7726        |
|  | KNN(B)       | 0.8746 ±0.019        | 0.7584        |
|  | KNN(C)       | 0.8835 ±0.020        | 0.7747        |
|  | SVM(A)       | 0.9538 ±0.075        | 0.9086        |
|  | SVM(B)       | 0.9882 ±0.054        | 0.9764        |
|  | SVM(C)       | 0.9544 ±0.013        | 0.9096        |
|  | XG(A)        | 0.9788 ±0.010        | 0.9578        |
|  | <b>XG(B)</b> | <b>0.9919 ±0.010</b> | <b>0.9839</b> |
|  | XG(C)        | 0.9753 ±0.010        | 0.9508        |
| <i>Desalinhamento paralelo horizontal(III)</i> | KNN(A)       | 0.8839 ±0.023        | 0.8751        |
|  | KNN(B)       | 0.8673 ±0.023        | 0.7422        |
|  | KNN(C)       | 0.7570 ±0.031        | 0.5293        |
|  | SVM(A)       | 0.8380 ±0.014        | 0.6831        |
|  | SVM(B)       | 0.9650 ±0.014        | 0.9305        |
|  | SVM(C)       | 0.9372 ±0.030        | 0.8759        |
|  | XG(A)        | 0.9537 ±0.023        | 0.9084        |
|  | <b>XG(B)</b> | <b>0.9876 ±0.008</b> | <b>0.9753</b> |
|  | XG(C)        | 0.9508 ±0.016        | 0.9027        |

A Tabela 4.4 apresenta a média da acurácia global e MAUC e o desvio padrão dos resultados obtidos para cada experimento. Em negrito estão os melhores resultados. De maneira geral todos os modelos são eficientes para diagnosticar falhas, com desvios baixos e taxa de acurácia média acima de 92%. A Tabela 4.5 exhibe a matriz de confusão com a melhor resposta dentre todos os modelos, obtido pelo classificador XGBoost com método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2), indicando ACC = 0.9951 e MAUC= 0.9879.

Para visualizar a comparação dos modelos de classificação de forma simplificada, nas Figuras 4.4 e 4.5 são apresentados os gráficos do tipo boxplot obtidos a partir dos resultados apontados da distribuição dos dados da ACC e MAUC, onde se percebe a existência de simetria dos valores, bem como uma boa repartição entre os quantis, com pouca influência de *outliers*.

Tabela 4.3: Resultados intraclases de ACC para as classes (IV), (V) e (VI).

| Classe alvo                                 | Modelo       | ACC                  | Sen X Esp     |
|---|--------------|----------------------|---------------|
| <i>Desalinhamento paralelo vertical(IV)</i> | KNN(A)       | 0.9027 ±0.024        | 0.8096        |
|   | KNN(B)       | 0.9188 ±0.015        | 0.8411        |
|   | KNN(C)       | 0.8571 ±0.017        | 0.7284        |
|   | SVM(A)       | 0.9409 ±0.017        | 0.8845        |
|   | SVM(B)       | 0.9654 ±0.014        | 0.9312        |
|   | SVM(C)       | 0.9511 ±0.012        | 0.9032        |
|   | XG(A)        | 0.9448 ±0.016        | 0.8905        |
|   | <b>XG(B)</b> | <b>0.9798 ±0.008</b> | <b>0.9596</b> |
|   | XG(C)        | 0.9537 ±0.014        | 0.9081        |
| <i>Mancal interno(V)</i>                    | KNN(A)       | 0.9771 ±0.004        | 0.9547        |
|   | KNN(B)       | 0.9018 ±0.012        | 0.8103        |
|   | KNN(C)       | 0.9102 ±0.013        | 0.8275        |
|   | SVM(A)       | 0.9670 ±0.013        | 0.9348        |
|   | SVM(B)       | 0.9882 ±0.004        | 0.9765        |
|   | SVM(C)       | 0.9516 ±0.008        | 0.9055        |
|   | XG(A)        | 0.9885 ±0.004        | 0.9770        |
|   | <b>XG(B)</b> | <b>0.9963 ±0.003</b> | <b>0.9925</b> |
|   | XG(C)        | 0.9775 ±0.007        | 0.9554        |
| <i>Mancal externo(VI)</i>                   | KNN(A)       | 0.9823 ±0.007        | 0.9648        |
|   | KNN(B)       | 0.9744 ±0.007        | 0.9494        |
|   | KNN(C)       | 0.9032 ±0.014        | 0.8119        |
|   | SVM(A)       | 0.9601 ±0.014        | 0.9217        |
|   | SVM(B)       | 0.9909 ±0.006        | 0.9818        |
|   | SVM(C)       | 0.9510 ±0.013        | 0.9031        |
|   | XG(A)        | 0.9906 ±0.006        | 0.9812        |
|   | <b>XG(B)</b> | <b>0.9992 ±0.001</b> | <b>0.9984</b> |
|   | XG(C)        | 0.9723 ±0.007        | 0.9452        |

Uma vez analisados os resultados por meio da ACC e MAUC, uma análise estatística mais detalhada é necessária para comparar os melhores modelos estocásticos destacados pelas elipses nas Figuras 4.4 e 4.5. Consequentemente, primeiramente foi utilizado o teste de Shapiro-Wilk (Shapiro e Wilk, 1965) para verificar a normalidade dos dados. Shapiro-Wilk é considerado um dos mais eficazes teste de normalidade, especialmente para pequenas amostras, por exemplo, inferior a 50. Em seguida é realizada os testes de homocedasticidade e independência dos resíduos (Montgomery e Runger, 2010).

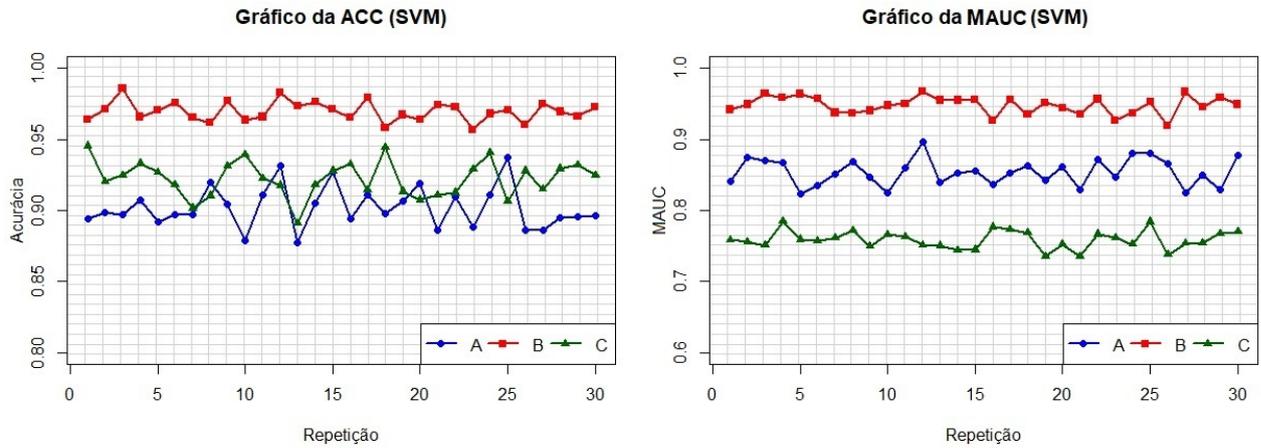


Figura 4.2: Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do SVM.

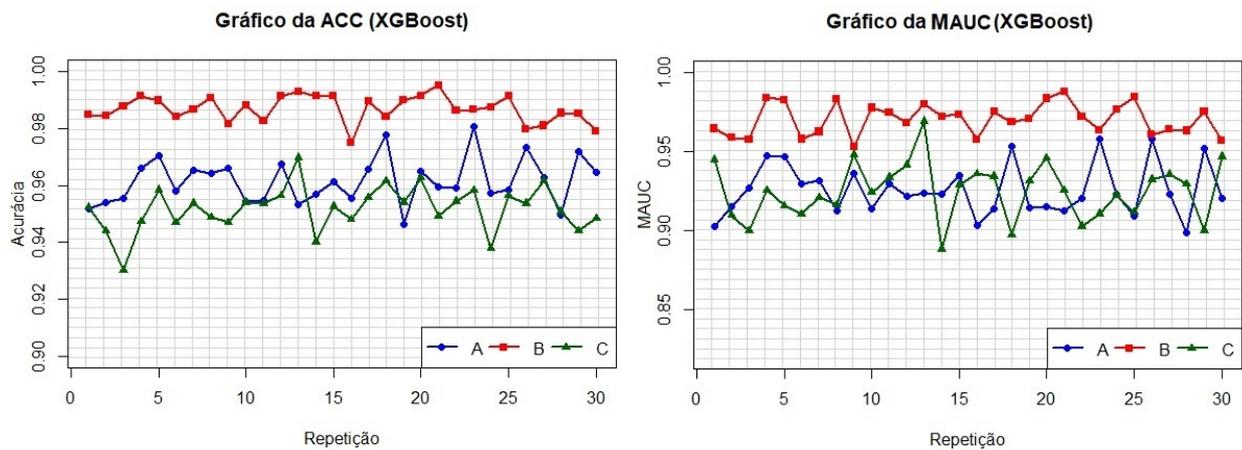
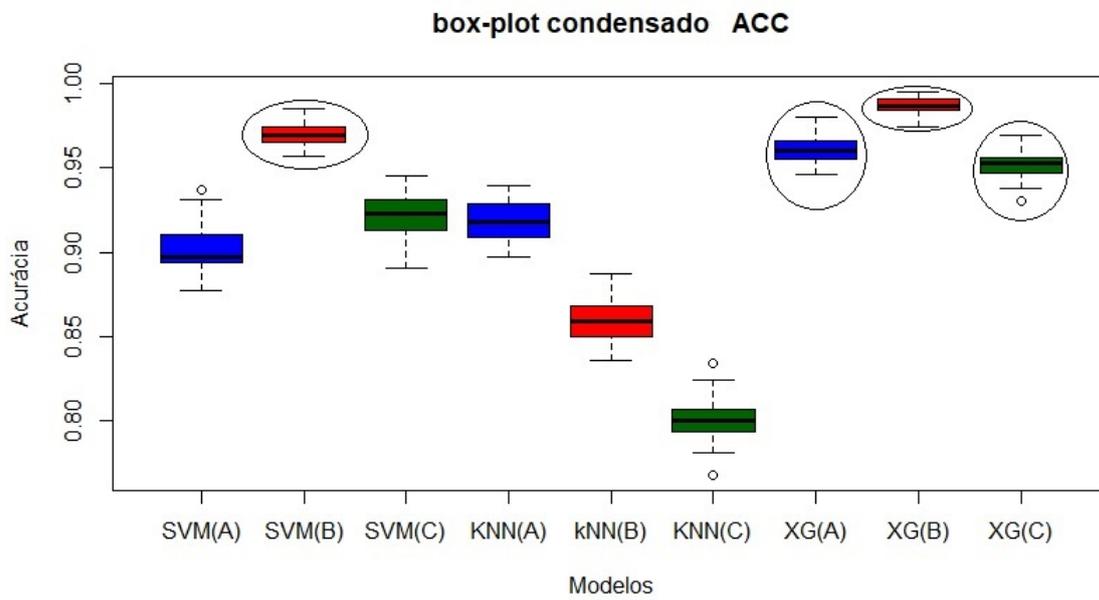
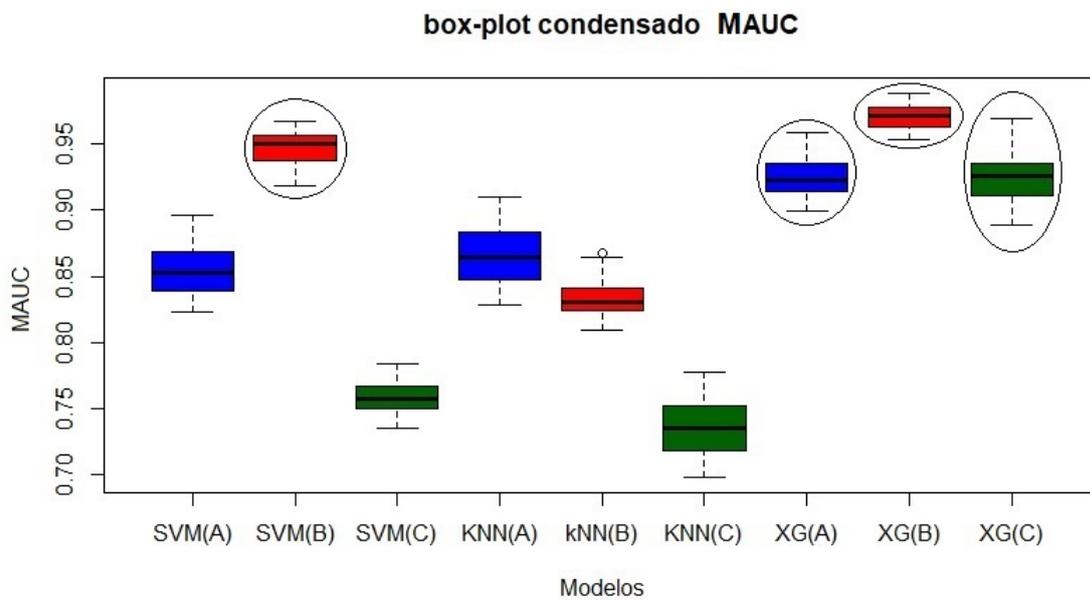


Figura 4.3: Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do XGBoost.

Tabela 4.4: Resultados condensados de médias e desvios de ACC e MAUC.

| Modelo       | ACC                   | MAUC                 |
|--------------|-----------------------|----------------------|
| KNN(A)       | 0.9180 ±0.011         | 0.8663 ±0.022        |
| KNN(B)       | 0.8592 ±0.012         | 0.8347 ±0.015        |
| KNN(C)       | 0.8010 ±0.013         | 0.7360 ±0.221        |
| SVM(A)       | 0.9018 ±0.014         | 0.8539 ±0.019        |
| SVM(B)       | 0.9697 ±0.006         | 0.9479 ±0.012        |
| SVM(C)       | 0.9221 ±0.028         | 0.7584 ±0.013        |
| XG(A)        | 0.9616 ± 0.008        | 0.9257 ± 0.016       |
| <b>XG(B)</b> | <b>0.9870 ± 0.046</b> | <b>0.9704 ±0.010</b> |
| XG(C)        | 0.9517 ±0.008         | 0.9247 ±0.020        |

Figura 4.4: *Boxplot* condensado como valores de ACC.Figura 4.5: *Boxplot* condensado como valores de MAUC.

Posteriormente, foi realizado o teste de análise de variância, também conhecida como ANOVA (Iversen e Norpoth, 1987). O teste ANOVA informa a probabilidade de existência de não semelhança entre as distribuições dos dados, ou seja, ele verifica se existem diferenças entre os modelos comparados. Conforme Equação (4.1) a hipótese nula  $L_0$  indica que os modelos são semelhantes e a hipótese alternativa  $L_i$  indica que existe alguma diferença entre eles.

$$\begin{cases} L_0 : \tau_i = 0, \forall i \\ L_i : \exists \tau_i \neq 0 \end{cases} \quad (4.1)$$

onde  $\tau_i$  é o deslocamento, da média global, do modelo  $i$ .

Para um nível de significância  $\alpha = 0.05$ , concluiu-se a existência de pelo menos 2 efeitos para os quais  $\tau_i \neq 0$ . Por fim, o teste de Tukey (Tukey, 1962) é executado para verificar se há diferença significativa ou não, entre os resultados obtidos para cada modelo selecionado. O resultado para o teste de Tukey com a diferença das médias é apresentado nas Figuras 4.6 e 4.7.

Como os resultados do teste de Tukey se baseia todos contra todos (*all Vs all*), construiu-se as Tabelas 4.6 e 4.7 com objetivo de ranquear: em que vitórias é o número de vezes que o modelo foi estatisticamente superior ( $p\text{-valor} < 0.05$ ), derrotas contabiliza as vezes que o modelo foi inferior e empate é quando os modelos comparados não possuem diferenças estatísticas.

Com nível de significância de  $\alpha = 0.05$ , em todos os casos, o classificador XG(B) obteve as melhores taxas de acerto, seguido pelo SVM(B). Os modelos XG(A) vista pela ACC é estatisticamente superior ao XG(C) e não possuem diferenças significativas pela MAUC.

Tabela 4.5: Matriz de confusão obtida pelo modelo classificação XGBoost com preditores extraído através método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2) (ACC = 0.9951 e MAUC= 0.9879).

| <b>Classes</b> | (I)       | (II)       | (III)     | (IV)      | (V)        | (VI)       |
|----------------|-----------|------------|-----------|-----------|------------|------------|
| (I)            | <b>11</b> | 0          | 0         | 0         | 1          | 0          |
| (II)           | 0         | <b>102</b> | 0         | 0         | 0          | 0          |
| (III)          | 0         | 1          | <b>62</b> | 0         | 0          | 0          |
| (IV)           | 0         | 0          | 1         | <b>94</b> | 0          | 0          |
| (V)            | 0         | 0          | 0         | 0         | <b>175</b> | 0          |
| (VI)           | 0         | 0          | 0         | 0         | 0          | <b>176</b> |

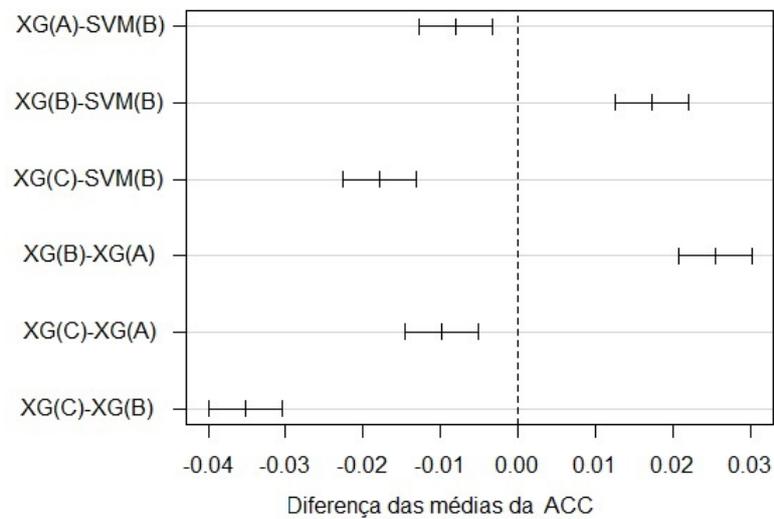


Figura 4.6: Diferença dos valores médios da ACC.

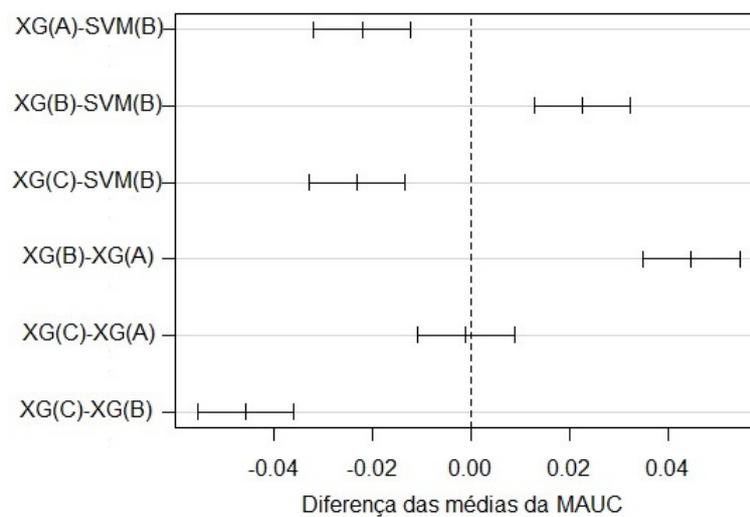


Figura 4.7: Diferença dos valores médios da MAUC.

Tabela 4.6: Ranqueamento dos modelos baseado no teste de Tukey do ACC.

| Modelo | Vitórias | Derrotas | Empates | Ranking |
|--------|----------|----------|---------|---------|
| XG(B)  | 3        | 0        | 0       | 1º      |
| SVM(B) | 2        | 1        | 0       | 2º      |
| XG(A)  | 1        | 2        | 0       | 3º      |
| XG(C)  | 0        | 3        | 0       | 4º      |

Tabela 4.7: Ranqueamento dos modelos baseado no teste de Tukey do MAUC.

| Modelo | Vitórias | Derrotas | Empates | Ranking |
|--------|----------|----------|---------|---------|
| XG(B)  | 3        | 0        | 0       | 1°      |
| SVM(B) | 2        | 1        | 0       | 2°      |
| XG(A)  | 0        | 2        | 1       | 3°      |
| XG(C)  | 0        | 2        | 1       | 3°      |

## 4.2 Conclusão

Este capítulo apresentou os resultados obtidos pelas três formas de extração de características atribuídas aos classificadores utilizados em questão. Através do teste de Tukey pode-se afirmar, com 95% de confiança, que o modelo de extração de características baseado em transformada wavelet de Haar e dimensão fractal (B) utilizada no XGBoost é significativamente melhor que os outros modelos na base de dados utilizada. Além disso, os resultados de XG(B) foram  $ACC=0.987$  e  $MAUC=0.9704$ , que são superiores aos apresentados na literatura e utilizados como base: (Lima et al., 2013), (Pestana-Viana et al., 2016), (Martins et al., 2016) e (Marins et al., 2017).

Observa-se que dos quatro melhores modelos selecionados para realização do teste de Tukey, dois utiliza-se dos métodos de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2), sendo promissor para o diagnóstico de falhas em máquinas rotativas. É válido ressaltar que o modelo XG(A), apesar de apresentar resultados inferiores a XG(B) e SVM(B), é bastante atraente para o uso em sistemas de monitoramento *on-line*, pois possui ordem de complexidade baixa no método de extração de características baseado em RMS (subseção 3.1.1)  $O(n)$ , e ainda pode ser utilizado na abordagem caixa preta na descoberta do tempo de vida útil das máquinas (Saponara, 2016). Os resultados baseados no método de extração de características baseado na FFT com dados estatísticos (subseção 3.1.3) indicam ser necessário haver uma melhoria de pré-processamento antes da realização do treinamento dos modelos.

# Resultados das Classificações das Falhas nos Mancais de Rolamento

---

Neste Capítulo são apresentados os resultados obtidos pelos algoritmos de classificação (SVM, KNN e XGBoost), para classificar os dados da base (MAFAULDA, 2017) nas subclasses representadas pela Tabela 2.3(esferas (VII), gaiola (VIII) e parte externa (IX)), referente às imperfeições dos rolamentos. É necessário enfatizar que nesta situação já existe a ocorrência de falhas nos mancais, este classificador especifica qual é o tipo problema no rolamento, sendo estes defeitos na gaiola, na pista externa e nas esferas. Os modelos de extração de características e os classificadores utilizados aqui são os mesmos apresentados na Tabela 4.1.

## 5.1 Resultados e Análise

Cada modelo de classificação apresenta 30 matrizes de confusões com várias subclasses. Posteriormente, foi realizada a transformação dos resultados em uma classificação binária em cada uma das subclasses. A classificação binária consiste em cada subclasse se transformou em uma subclasse alvo e as demais como outras. O número total de subclasses foram balanceados. Por fim, foram extraídos a acurácia binária e os valores acumulados de sensibilidade e especificidade. A Tabela 5.1 mostra os resultados obtidos por cada modelo em cada uma das subclasses, além do desvio padrão dos valores por ACC e o produto entre sensibilidade X especificidade. Todos os classificadores obtiveram excelentes desempenhos com taxa de acurácia média de 96%. O XG(B) se destaca novamente, pois atingiu os melhores resultados em todas as subclasses alvo.

Tabela 5.1: Resultados intraclasses de ACC para as subclasses (VII), (VIII) e (IX).

| Subclasse alvo            | Modelo       | ACC                   | Sen X Esp     |
|---------------------------|--------------|-----------------------|---------------|
| <i>Esferas (VII)</i>      | KNN(A)       | 0.9983 ± 0.002        | 0.9965        |
|                           | KNN(B)       | 0.9500 ± 0.016        | 0.9006        |
|                           | KNN(C)       | 0.9172 ± 0.018        | 0.8382        |
|                           | SVM(A)       | 0.9975 ± 0.002        | 0.9950        |
|                           | SVM(B)       | 0.9931 ± 0.008        | 0.9862        |
|                           | SVM(C)       | 0.9643 ± 0.014        | 0.9291        |
|                           | XG(A)        | 0.9941 ± 0.005        | 0.9881        |
|                           | <b>XG(B)</b> | <b>0.9985 ± 0.002</b> | <b>0.9970</b> |
|                           | XG(C)        | 0.9800 ± 0.011        | 0.9603        |
| <i>Gaiola (VIII)</i>      | KNN(A)       | 0.9536 ± 0.014        | 0.9085        |
|                           | KNN(B)       | 0.9259 ± 0.016        | 0.8565        |
|                           | KNN(C)       | 0.8842 ± 0.022        | 0.7792        |
|                           | SVM(A)       | 0.9485 ± 0.012        | 0.8989        |
|                           | SVM(B)       | 0.9642 ± 0.009        | 0.9296        |
|                           | SVM(C)       | 0.9337 ± 0.013        | 0.8715        |
|                           | XG(A)        | 0.9790 ± 0.009        | 0.9584        |
|                           | <b>XG(B)</b> | <b>0.9938 ± 0.004</b> | <b>0.9875</b> |
|                           | XG(C)        | 0.9627 ± 0.012        | 0.9265        |
| <i>Pista externa (IX)</i> | KNN(A)       | 0.9674 ± 0.01         | 0.9357        |
|                           | KNN(B)       | 0.9264 ± 0.013        | 0.8581        |
|                           | KNN(C)       | 0.8835 ± 0.018        | 0.7806        |
|                           | SVM(A)       | 0.9578 ± 0.011        | 0.9174        |
|                           | SVM(B)       | 0.9696 ± 0.007        | 0.9401        |
|                           | SVM(C)       | 0.9410 ± 0.012        | 0.8853        |
|                           | XG(A)        | 0.9812 ± 0.007        | 0.9626        |
|                           | <b>XG(B)</b> | <b>0.9947 ± 0.004</b> | <b>0.9894</b> |
|                           | XG(C)        | 0.9652 ± 0.009        | 0.9316        |

As Figuras 5.1, 5.2 e 5.3, exibem os valores de ACC global e MAUC alcançados em cada uma das 30 execuções do algoritmo de classificação. O método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2) foi mais efetivo para o SVM e XGBoost, já o método de extração de características baseado em RMS (subseção 3.1.1) foi melhor para o KNN.

A Tabela 5.2 apresenta a média da acurácia global e MAUC e o desvio padrão dos resultados obtidos para cada experimento. Em negrito estão os melhores resultados. De maneira geral todos os modelos são eficientes para diagnosticar defeitos nos rolamentos, com desvios baixos e taxa de acurácia média acima de 94,8%. A Tabela 5.3 exibe a matriz de confusão com a melhor resposta dentre todos os modelos, obtido pelo experimento XG(B), indicando  $ACC = MAUC = 1$ .

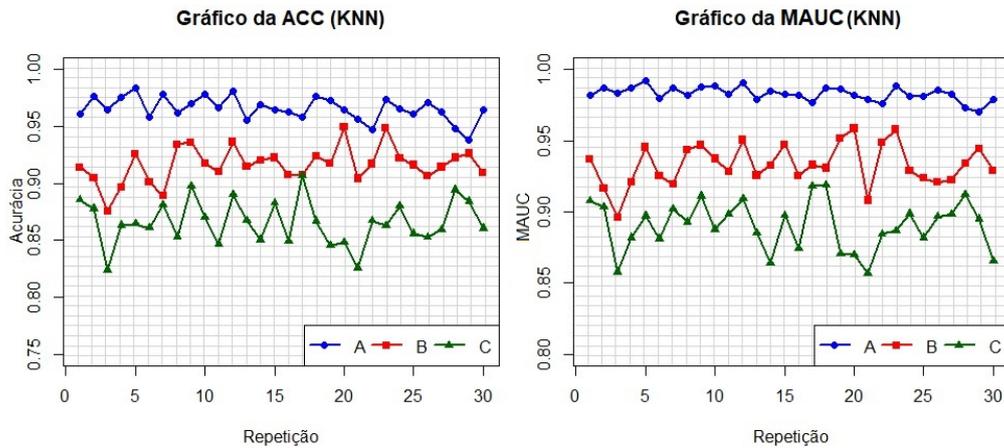


Figura 5.1: Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do KNN.

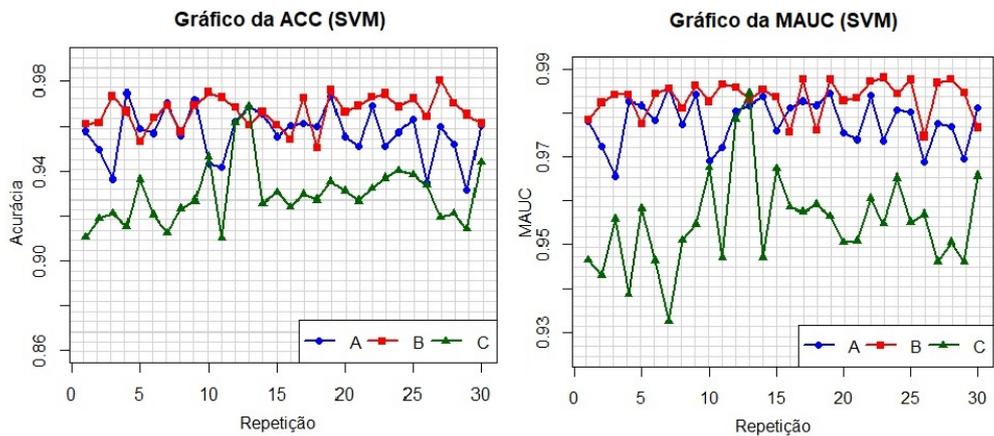


Figura 5.2: Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do SVM.

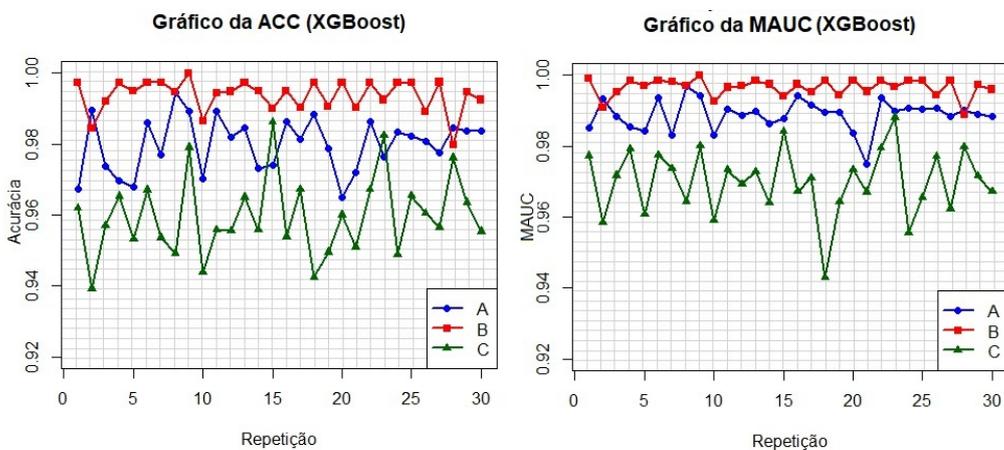


Figura 5.3: Gráficos dos valores de ACC à esquerda e MAUC à direita, para os métodos nas execuções do XGBoost.

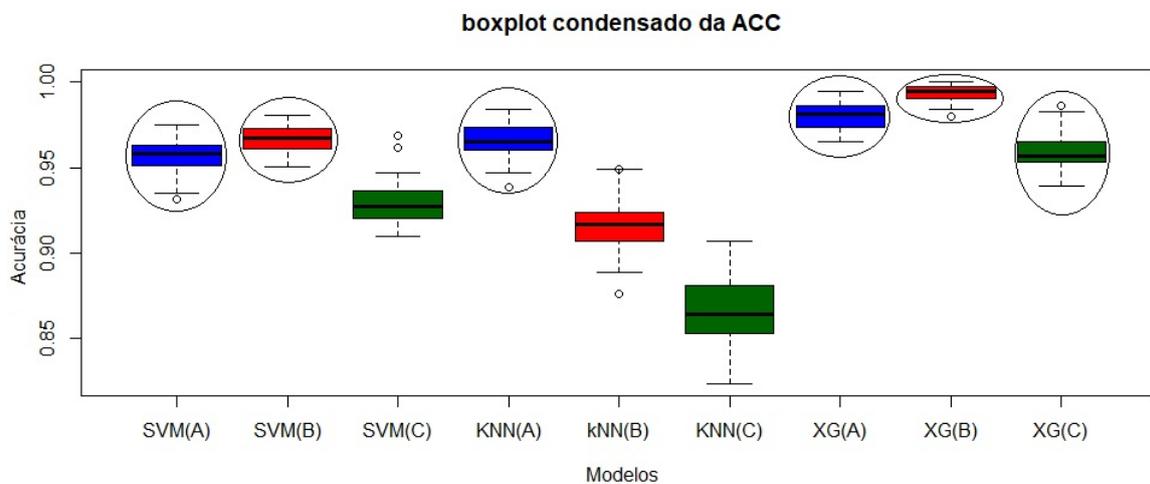
Tabela 5.2: Resultados condensados de médias e desvios de ACC e MAUC.

| Modelo       | ACC                                  | MAUC                                |
|--------------|--------------------------------------|-------------------------------------|
| KNN(A)       | 0.9654 $\pm$ 0.010                   | 0.9828 $\pm$ 0.005                  |
| KNN(B)       | 0.9162 $\pm$ 0.015                   | 0.9329 $\pm$ 0.014                  |
| KNN(C)       | 0.8656 $\pm$ 0.019                   | 0.8900 $\pm$ 0.017                  |
| SVM(A)       | 0.9569 $\pm$ 0.011                   | 0.9780 $\pm$ 0.005                  |
| SVM(B)       | 0.9665 $\pm$ 0.007                   | 0.9833 $\pm$ 0.003                  |
| SVM(C)       | 0.9293 $\pm$ 0.013                   | 0.9551 $\pm$ 0.010                  |
| XG(A)        | 0.9799 $\pm$ 0.007                   | 0.9888 $\pm$ 0.004                  |
| <b>XG(B)</b> | <b>0.9936 <math>\pm</math> 0.004</b> | <b>0.9965 <math>\pm</math>0.002</b> |
| XG(C)        | 0.9595 $\pm$ 0.011                   | 0.9698 $\pm$ 0.009                  |

Tabela 5.3: Matriz de confusão obtida pelo modelo classificação XGBoost com preditores extraídos através do método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (ACC = MAUC=1).

| Subclasses | (VII)     | (VIII)     | (IX)       |
|------------|-----------|------------|------------|
| (VII)      | <b>97</b> | 0          | 0          |
| (VIII)     | 0         | <b>113</b> | 0          |
| (IX)       | 0         | 0          | <b>111</b> |

Nas Figuras 5.4, 5.5 são apresentados os gráficos boxplot obtidos a partir dos resultados apontados da distribuição dos dados da ACC e MAUC. Os modelos exibem a existência de simetria dos valores, bem como uma boa repartição entre os quantis, com pouca influência de *outliers*.

Figura 5.4: *Boxplot* condensado como valores de ACC.

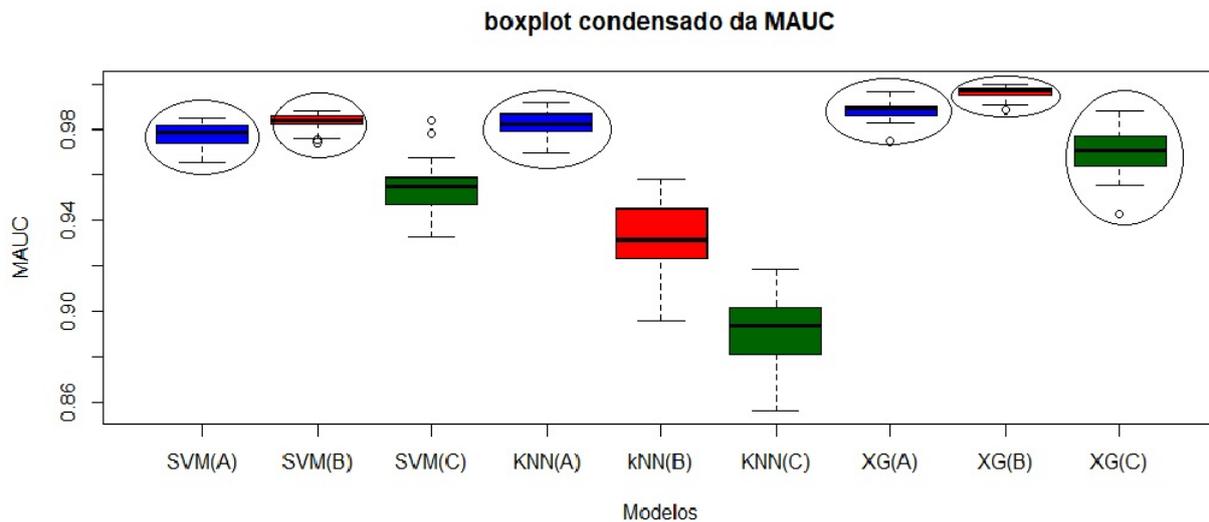


Figura 5.5: *Boxplot* condensado como valores de MAUC.

Uma vez analisados os resultados por meio da ACC e AUC, uma análise estatística mais detalhada é feita comparando os melhores modelos estocásticos destacados pelas elipses nas Figuras 5.4 e 5.5. Posteriormente, com o atendimento das premissas de normalidade, homocedasticidade e independência dos resíduos, foi realizado o teste de ANOVA. Mostrado pela Equação (4.1) onde a hipótese nula  $L_0$  indica que os modelos são semelhantes e a hipótese alternativa  $L_i$  indica que existe alguma diferença entre eles.

Para um nível de significância  $\alpha = 0.05$ , obteve-se a existência de pelo menos 2 efeitos para os quais  $\tau_i \neq 0$ . Por fim, foi realizado o teste de Tukey (Tukey, 1962). O resultado para o teste de Tukey com a diferença das médias é apresentado nas Figuras 5.6 e 5.7.

As Tabelas 5.4 e 5.5 apresentam o ranqueamento entre classificadores e extratores, conforme o capítulo anterior. Com nível de significância de  $\alpha = 0.05$ , em todos os casos, o classificador XG(B) obteve melhores taxas de acerto, seguido pelos XG(A) e SVM(B), em sequência os demais modelos KNN(A), XG(C) ou SVM(A).

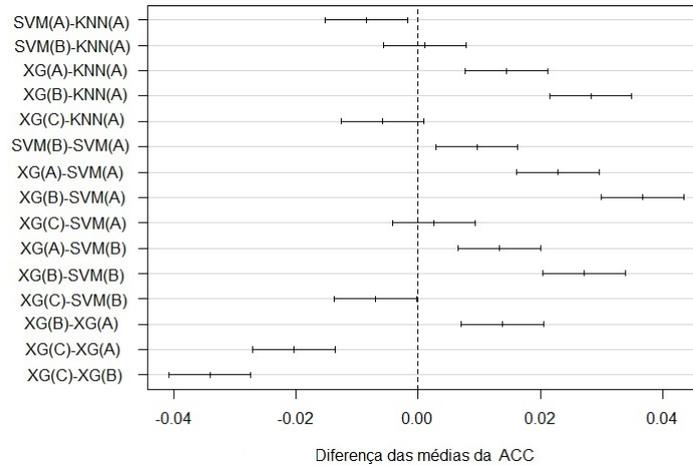


Figura 5.6: Diferença dos valores médios da ACC.

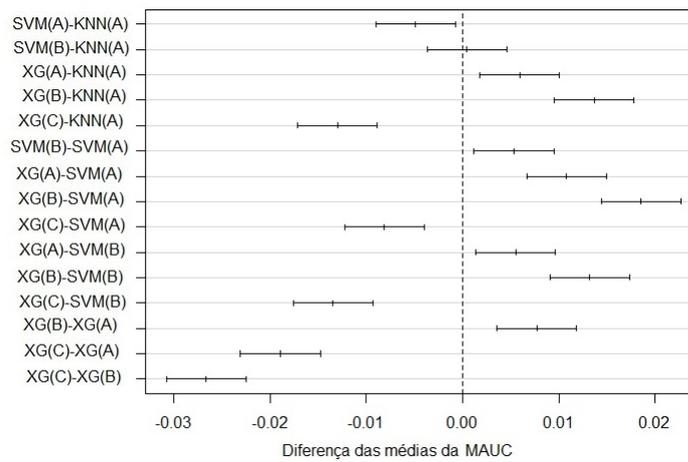


Figura 5.7: Diferença dos valores médios da MAUC.

Tabela 5.4: Ranqueamento dos modelos baseado no teste de Tukey do ACC.

| Modelo | Vitórias | Derrotas | Empates | Ranking        |
|--------|----------|----------|---------|----------------|
| XG(B)  | 5        | 0        | 0       | 1 <sup>o</sup> |
| XG(A)  | 4        | 1        | 0       | 2 <sup>o</sup> |
| SVM(B) | 2        | 2        | 1       | 3 <sup>o</sup> |
| knn(A) | 2        | 1        | 2       | 4 <sup>o</sup> |
| XG(C)  | 0        | 3        | 2       | 5 <sup>o</sup> |
| SVM(A) | 0        | 4        | 1       | 6 <sup>o</sup> |

Tabela 5.5: Ranqueamento dos modelos baseado no teste de Tukey do MAUC.

| <b>Modelo</b> | <b>Vitórias</b> | <b>Derrotas</b> | <b>Empates</b> | <b>Ranking</b> |
|---------------|-----------------|-----------------|----------------|----------------|
| XG(B)         | 5               | 0               | 0              | 1º             |
| XG(A)         | 4               | 1               | 0              | 2º             |
| SVM(B)        | 2               | 2               | 1              | 3º             |
| KNN(A)        | 2               | 2               | 1              | 3º             |
| SVM(A)        | 1               | 4               | 0              | 5º             |
| XG(C)         | 0               | 5               | 0              | 6º             |

## 5.2 Conclusão

Esse capítulo apresentou os resultados obtidos pelas três formas de extração de características atribuídas aos classificadores utilizados em questão para identificar as falhas nos mancais de rolamento. Através do teste de Tukey pode-se afirmar, com 95% de confiança, que o modelo de extração de características baseado em transformada wavelet de Haar e dimensão fractal (B) utilizada no XGBoost é significativamente melhor que os outros modelos. Além disso, os resultados de XG(B) foram ACC=0.9936 e MAUC=0.9965 e que em algumas execuções conseguiu a matriz de confusão ideal.

Observa-se que dos seis melhores modelos selecionados para realização do teste de Tukey, dois dos três superiores utiliza-se o método de extração de características baseado na transformada Wavelet de Haar e dimensão fractal (subseção 3.1.2), sendo favorável para o diagnóstico de falhas em mancais de rolamento, quanto aos outros modelos de extração são válidas as outras ressalvas descritas na seção 4.2.

## Conclusões

---

Essa dissertação empregou três técnicas de extração de características na análise de vibrações mecânicas, para o diagnóstico automático de falhas em máquinas rotativas com uso da base de dados (MAFAULDA, 2017). Os três métodos de extração de características foram combinadas com os classificadores KNN, SVM e XGBoost.

O método baseado em RMS descrito na Subseção 3.1.1, apresentou bons resultados na abordagem das falhas em mancais de rolamentos com o classificador XGBoost, com uma taxa de acurácia média de 98% (MAUC = 0.9888). O RMS também pode ser utilizado para monitoramento do tempo de vida das máquinas rotativas.

Na maioria das publicações da base (MAFAULDA, 2017), utiliza-se o método de extração baseado em transformada de Fourier com dados estatísticos, exposto na Subseção 3.1.3. Ao realizar esse método de extração de características foi possível obter resultados similares a (Lima et al., 2013), (Pestana-Viana et al., 2016) e (Martins et al., 2016). No entanto, esses resultados foram inferiores aos demais encontrados neste trabalho através dos preditores extraídos pelos métodos de extrações de características baseado em RMS (Subseção 3.1.1) e Transformada Wavelet de Haar (Subseção 3.1.2), combinados com o classificador XGBoost. O método de extração baseado em transformada de Fourier com dados estatísticos (Subseção 3.1.3) possui 46 dimensões, é necessário haver melhorias em pré-processamentos, como redução de dimensionalidade apresentado em (Marins et al., 2017).

A extração baseada na transformada Wavelet de Haar e dimensão fractal, descrito na Subseção 3.1.2 com o classificador XGBoost, conseguiu atingir uma taxa média de acurácia de 98.7% (MAUC=0.9704) nas falhas em máquinas rotativas e 99.36% (MAUC=0.9965) na detecção de problemas nos rolamentos, que são superiores aos apresentados na literatura pesquisada: (Lima et al., 2013), (Pestana-Viana et al., 2016), (Martins et al., 2016) e (Marins et al., 2017), e ainda obteve resultados intraclasse notáveis. Através do teste de Tukey pode-se afirmar que esses resultados, com 95% de confiança é significativamente melhor que os outros modelos testados. Este método

com vetor de características extraído com 24 dimensões, se mostrou bastante promissor para o objeto dessa dissertação.

Esta pesquisa apresentou boas contribuições que ainda não haviam sido obtidos com a base (MAFAULDA, 2017) em outras publicações como: Duas formas distintas de extração de característica, descritas nas Subseções 3.1.1, 3.1.2; Utilização de outros indicadores de desempenho mais abrangentes que a acurácia como MAUC e produto entre sensibilidade e especificidade (*Sen X Esp*); A análise de resultados por meio de médias, desvio padrão e testes estatísticos ANOVA e Tukey para comparar os resultados dos experimentos.

## 6.1 Sugestões para Trabalhos Futuros

Sugere-se como propostas de continuidade desse trabalho:

- O acoplamento entre técnicas de extração de características e outras transformadas wavelet (Debeuchets e Symlets);
- A hibridização entre os métodos apresentados com seleção dos melhores atributos;
- O emprego de técnicas de redução de dimensionalidade e seleção de características relevantes;
- Testar a metodologia em novas falhas;
- Aplicar outros classificadores, como: naive Bayes, redes neurais artificiais;
- Aplicar a metodologia em outras bases com máquinas rotativas como motobombas ou geradores.

---

## Referências Bibliográficas

---

- Abe, S. (2005). *Support vector machines for pattern classification*, volume 2. Springer.
- Adreola, R. (2009). *Support Vector Machines na Classificação de Imagens Hiperespectrais, 130f*. Dissertação de Mestrado, Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.
- Al-Ghamd, A. M. e Mba, D. (2006). A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical systems and signal processing*, 20(7):1537–1571.
- Alexandre, M., Carrijo, D., e AL, A. J. (2017). Arquiteturas de monitoramento e diagnóstico de transformadores de potência. *XVIII Seminário Nacional de Distribuição de Energia Elétrica*, 18.
- Amirat, Y., Choqueuse, V., e Benbouzid, M. H. (2010). Wind turbines condition monitoring and fault diagnosis using generator current amplitude demodulation. páginas 310–315.
- Asfani, D., Muhammad, A., Purnomo, M., Hiyama, T., et al. (2012). Temporary short circuit detection in induction motor winding using combination of wavelet transform and neural network. *Expert Systems with Applications*, 39(5):5367–5375.
- Baranauskas, J. A. e Monard, M. C. (2000). Reviewing some machine learning concepts and methods. *Relatórios Técnicos do ICMC/USP*, 102.
- Benbouzid, M. E. H. (2000). A review of induction motors signature analysis as a medium for faults detection. *IEEE transactions on industrial electronics*, 47(5):984–993.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.

- Chen, T. e Guestrin, C. (2016). Xgboost: A scalable tree boosting system. páginas 785–794.
- Cho, H. (2017). Speeding up gradient boosting for training and prediction. *Department of Computer Science and Engineering*, 1(1):1–23.
- Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., Maling, G., Nelson, D. E., Rader, C. M., e Welch, P. D. (1967). What is the fast fourier transform? *Proceedings of the IEEE*, 55(10):1664–1674.
- Cover, T. e Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.
- D. Silva, V. e Scarpel, R. A. (2007). Detecção de fraudes na distribuição de energia elétrica utilizando support vector machine. *Investigação Operacional*, 27(2):139–150.
- Dong, H., Xu, X., Wang, L., e Pu, F. (2018). Gaofen-3 polsar image classification via xgboost and polarimetric spatial information. *Sensors*, 18(2):611.
- Falconer, K. (2004). *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, páginas 1189–1232.
- Fujimoto, R. (2005). *Diagnóstico automático de defeitos em rolamentos baseado em Lógica Fuzzy*. 2005. 158 f. Dissertação de Mestrado, Escola Politécnica, Universidade de São Paulo, São Paulo, 2005.
- Ganeriwala, S., Patel, S., e Hartung, H. (1998). Observations concerning misalignment vibration signature. *Spectra Quest Inc*.
- Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., e Wolff, E. (2018). Very high resolution object-based land use-land cover urban classification using extreme gradient boosting. *IEEE Geoscience and Remote Sensing Letters*.
- Guo, S., Ren, Y., Shi, S., Yang, L., e Lin, J. (2018). Rotating machinery vibration analysis of the rotary-laser scanning measurement system. In *2017 International Conference on Optical Instruments and Technology: Optoelectronic Measurement Technology and Systems*, volume 10621, página 106210E. International Society for Optics and Photonics.
- Gupta, A. (2015). *Learning Apache Mahout Classification*. Packt Publishing Ltd.

- Hand, D. J. e Till, R. J. (2001). A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine learning*, 45(2):171–186.
- Haykin, S. (2009). *Neural networks and machine learning*. Pearson, página 936.
- Iversen, G. R. e Norpoth, H. (1987). *Analysis of variance*. Number 1. Sage.
- James, G., Witten, D., Hastie, T., e Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Jardine, A. K., Lin, D., e Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510.
- Kreyszig, E. (2010). *Advanced engineering mathematics*. John Wiley & Sons.
- Lima, A. d. A., Prego, T. d. M., Netto, S. L., da Silva, E. A., Gutierrez, R. H., Monteiro, U. A., Troyman, A. C., Silveira, F. J. d. C., e Vaz, L. (2013). On fault classification in rotating machines using fourier domain features and neural networks. páginas 1–4.
- Lorena, A. C. e de Carvalho, A. C. (2007). Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, 14(2):43–67.
- Loureiro, S. M. (2009). *Reconhecimento automático de padrões de defeitos em motobombas utilizando análise de sinais de vibração*. Dissertação de Mestrado, Prógrama de pós graduação em informática, Universidade Federal do Espírito Santo, Vitória, 2009.
- M. Silva, L. (2005). *Uma aplicação de Árvores de Decisão, Redes Neurais e KNN para a Identificação de Modelos ARMA não Sazonais e Sazonais*. Tese de Doutorado, PUC-Rio, Rio de Janeiro 2005.
- MAFAULDA (2017). Machinery fault database. <http://www02.smt.ufrj.br/~offshore/mfs/>. Acessado: 02/02/2018.
- Mallat, S. (1999). *A wavelet tour of signal processing*. Academic press.
- Marins, M. A., Ribeiro, F. M., Netto, S. L., e da Silva, E. A. (2017). Improved similarity-based modeling for the classification of rotating-machine failures. *Journal of the Franklin Institute*.
- Martins, D. H. d. S. S., Prego, T. d. M., e de Lima, A. A. (2016). Classificação de severidade de falhas em máquinas rotativas usando random forest e redes neurais artificiais. *Encontro Nacional de Inteligência Artificial e Computacional*, 13(1):13–24.

- Mehala, N. (2010). Condition monitoring and fault diagnosis of induction motor using motor current signature analysis. *A Ph. D Thesis submitted to the Electrical Engineering Department, National Institute of Technology, Kurushetra, India.*
- Montgomery, D. C. e Runger, G. C. (2010). *Applied Statistics and Probability for Engineers.* John Wiley & Sons.
- Nguyen, N.-T., Kwon, J.-M., e Lee, H.-H. (2007). Fault diagnosis of induction motor using decision tree with an optimal feature selection. In *Power Electronics, 2007. ICPE'07. 7th Internatonal Conference on*, páginas 729–732. IEEE.
- Pereira, M. d. A. (2012). *Classificação de Dados Híbridos Através de Algoritmos Evolucionários.* Tese de Doutorado, Prógrama de pós graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, Belo Horizonte 2012.
- Pereira, M. d. A., Júnior, C. A. D., Carrano, E. G., e De Vasconcelos, J. A. (2014). A niching genetic programming-based multi-objective algorithm for hybrid data classification. *Neurocomputing*, 133:342–357.
- Pestana-Viana, D., Zambrano-López, R., de Lima, A. A., Prego, T. d. M., Netto, S. L., e da Silva, E. A. (2016). The influence of feature vector on the classification of mechanical faults using neural networks. páginas 115–118.
- Platt, J. C., Cristianini, N., e Shawe-Taylor, J. (2000). Large margin dags for multiclass classification. In *Advances in neural information processing systems*, páginas 547–553.
- Rader, C. M. (1968). Discrete fourier transforms when the number of data samples is prime. *Proceedings of the IEEE*, 56(6):1107–1108.
- Ramalho, G. L. B., Pereira, A. H., Rebouças Filho, P. P., e Medeiros, C. (2014). Detecção de falhas em motores elétricos através da classificação de padrões de vibração utilizando uma rede neural elm/fault detection in induction motors using vibration patterns and elm neural network. *HOLOS*, 30(4):185.
- Reis Gade, L. d., de Castro, C. L., Torres, L. C. B., Coelho, F. G. F., Braga, A. P., Garcia, J. A., e Torres, F. S. (2017). Nn-clas: classificador geométrico de margem larga baseado na regra do vizinho mais próximo. *Brazilian Congresson Computational Intelligence*, 1(1):33.1–33.12.
- Rezende, T. M., de Castro, C. L., e Moreira, S. G. (2017). Análise da expressao facial em reconhecimento de sinais de libras. *VI Simpósio Brasileiro de Automação Inteligente*, 13(1):465–470.

- S. Martins, D. (2016). *Diagnóstico e prognóstico de falhas em máquinas rotativas no algoritmo random forest*. 2016. 94f. Dissertação de Mestrado, Programa de pós graduação em engenharia elétrica, CEFET/RJ, Rio de Janeiro, 2016.
- Saponara, S. (2016). Distributed measuring system for predictive diagnosis of uninterruptible power supplies in safety-critical applications. *Energies*, 9(5):327.
- Shapiro, S. S. e Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Shirado, W. H., de Abreu Moreira, M., Palma, J. G., e Júnior, S. B. (2015). Estudo comparativo entre algoritmos das transformadas discretas de fourier e wavelet. *Revista Brasileira de Computação Aplicada*, 7(3):97–107.
- Singh, G. et al. (2003). Experimental investigations on induction machine condition monitoring and fault diagnosis using digital signal processing techniques. *Electric Power Systems Research*, 65(3):197–221.
- Sinhorigno, S. (2007). *Previsão de inadimplência de transações com cartão de crédito: um estudo comparativo*. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, 2007.
- Smola, A. J., Barlett, P., Scholkopf, B., e Schuurmans, D. (2000). *Advances in large margin classifiers*. MIT press.
- Tang, K., Wang, R., e Chen, T. (2011). Towards maximizing the area under the roc curve for multi-class classification problems.
- Tsyppkin, M. (2013). Induction motor condition monitoring: Vibration analysis technique-a twice line frequency component as a diagnostic tool. In *Electric Machines & Drives Conference (IEMDC), 2013 IEEE International*, páginas 117–124. IEEE.
- Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1):1–67.
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- White, G. D. (2008). *Introduction to Machine Vibration*, volume 1. Reliabilityweb.com Press.
- Yan, R., Gao, R. X., e Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal processing*, 96:1–15.