

# Diagnóstico de falhas em máquinas rotativas utilizando Random Forest

Dionisio Henrique Carvalho de Sá S6 Martins\*, Denys Pestana-Viana\*, Thiago de Moura Prego\*,†, Amaro Azevedo de Lima\*,†, Rafael Zambrano-López† e Sergio Lima Netto†.

**Resumo**—Este artigo apresenta uma forma de realizar o diagnóstico de falhas em máquinas rotativas através da análise de sinais de vibração. O presente trabalho realiza a classificação de 9 tipos de defeitos que acontecem em motores: desbalanceamento; desalinhamento paralelo horizontal; desalinhamento paralelo vertical; defeito na gaiola, na esfera e na pista externa do mancal não invertido; defeito na gaiola, na esfera e na pista externa do mancal invertido. A separação dos conjuntos de treinamento e teste é feita através do método *k-fold*. O algoritmo usado para fazer a classificação é o *Random Forest* que atingiu uma acurácia de 97,48%.

**Palavras-Chave**—Diagnóstico de falhas, *k-fold*, *Random Forest*.

**Abstract**—This article introduces a way to realize the fail diagnosis in rotation machines analysing vibrations signals. The present work realizes the classification of 9 types of faults that happen in motors: unbalance; horizontal parallel misalignment; vertical parallel misalignment; defect in the cage, in the roller and in the outer track of non inverter bearing; defect in the cage, in the roller and in the outer track of inverter bearing. The separation of train and test sets is made by the *k-fold* method. The algorithm used to make the classification is the *Random Forest* that reached an accuracy of 97,48%.

**Keywords**—Fail Diagnosis, *k-fold*, *Random Forest*.

## I. INTRODUÇÃO

As falhas em mancais são as principais causas de paradas em máquinas rotativas. Estas paradas causam perdas na produção. Como o ramo industrial a cada dia é mais competitivo há a necessidade de se evitar que ocorram interrupções no processo produtivo por falta de um programa de manutenção adequado. A manutenção preditiva tem se tornado mais comum nas empresas. Conforme a NBR 5462 de 1994 [1], a manutenção preditiva é aquela que tem a finalidade de manter o equipamento sempre em condições normais de operação, através da aplicação de técnicas de análise que usam meios de supervisão ou de amostragem, com o objetivo de diminuir as manutenções corretivas e preventivas.

A análise de vibrações nos processos industriais modernos faz-se necessária, através da manutenção preditiva, com o objetivo de garantir elevado grau de confiabilidade das máquinas rotativas. Análise esta, em particular, permite desenvolver métodos de antecipação à falha e realizar diagnósticos de comportamento da máquina [2], [3].

\* Programa de Pós-graduação em Engenharia Elétrica, CEFET-RJ, Rio de Janeiro, Brasil. † Programa de Engenharia Elétrica, COPPE/UFRJ, Rio de Janeiro, Brasil, e-mails: dionisiohenrique@ig.com.br, denys.pestana@gmail.com, thiago.prego@cefet-rj.br, amaro.lima@smt.ufrj.br, rafael.lopez@smt.ufrj.br, sergioln@smt.ufrj.br

Há casos em que o procedimento da inspeção visual das características dos sinais de vibração no domínio da frequência é suficiente para identificar falhas. No entanto é necessário que este processo de realização de diagnóstico seja feito de maneira rápida e automática, por isso utilizam-se os algoritmos de classificação, dentre eles destacam-se para o reconhecimento de padrões com múltiplas classes o *Random Forest* e as Redes Neurais Artificiais [2], [4], [3].

Este artigo está organizado da seguinte forma: a Seção II apresenta a descrição do sistema utilizado para realizar os experimentos analisados neste trabalho; na Seção III são abordados os procedimentos para realizar a extração de características; na Seção IV é introduzido o método de separação dos conjuntos de treinamento e teste através da validação cruzada; na Seção V é explicado o funcionamento do algoritmo de classificação utilizado neste artigo; na Seção VI são dados detalhes sobre a base de dados usada na realização dos experimentos; na Seção VII são fornecidos detalhes a respeito do procedimento experimental para obter o diagnóstico de falhas do equipamento; na Seção VIII são feitas as considerações finais com base nos resultados obtidos.

## II. DESCRIÇÃO DO SISTEMA

Foi utilizada neste projeto a bancada experimental Aligment/Balance Vibration Trainer (ABVT) representada pela Figura 1, fabricada pela empresa Spectra Quest Inc. As principais características desta bancada são: possui um motor de corrente contínua com potência de 0,25 CV, a sua faixa de velocidade varia no intervalo de [11,67 ; 63] Hz, o comprimento do seu eixo é de 520 mm e o diâmetro do seu eixo é de 16 mm, o diâmetro do rotor possui 15,24 mm e a distância da separação dos mancais é de 390 mm.

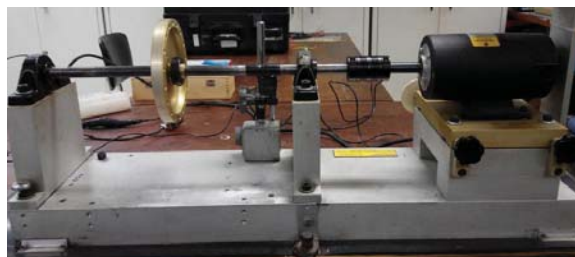


Fig. 1. Bancada Experimental usada nas simulações.

Os defeitos analisados através desta bancada foram os seguintes: desalinhamento paralelo horizontal, desalinhamento

paralelo vertical, desbalanceamento, defeito na gaiola do mancal não invertido, defeito na esfera do mancal não invertido, defeito na pista externa do mancal não invertido, defeito na gaiola do mancal invertido, defeito na esfera do mancal invertido e defeito na pista externa do mancal invertido. O mancal não invertido está entre a carga e o rotor enquanto que o mancal invertido é aquele que a carga fica entre o mancal e o rotor.

O defeito de desalinhamento paralelo horizontal é implementado por meio de um parafuso que regula a quantidade de desalinhamento, isto é feito com o auxílio de um paquímetro digital.

O desalinhamento paralelo vertical é implementado na bancada através da inserção de calços na base do motor. O defeito de desbalanceamento é inserido na bancada através da fixação de parafusos e arruelas no rotor. Estes parafusos e arruelas são pesados em balanças de precisão a fim de reproduzir o defeito de maneira eficaz.

Os defeitos nos mancais não invertido e invertido são inseridos na bancada através da substituição de mancais normais por mancais com defeitos na gaiola, na esfera ou na pista externa.

### III. EXTRAÇÃO DE CARACTERÍSTICAS

A extração de características é uma etapa fundamental para o reconhecimento de falhas em máquinas rotativas, pois ao realizar esta etapa corretamente consegue-se reduzir consideravelmente o número de amostras dos sinais a serem analisados.

Os sinais de vibração são captados por acelerômetros piezoelétricos nos mancais não invertido e invertido nas direções axial, radial e tangencial. Estes são processados com a finalidade de extrair as características dos sinais que sejam capazes de diferenciar os tipos de falhas. Além disso, é desejável que os algoritmos de extração das características permitam computações não dispendiosas.

Este processamento é feito através dos cálculos da transformada de Fourier, da curtose e da entropia do sinal.

A transformada de Fourier foi utilizada porque cada tipo de defeito altera as vibrações do motor em determinadas faixas de frequência. Para cada sinal proveniente dos acelerômetros nas três direções em cada mancal foram obtidas três características que correspondem ao valor da amplitude do espectro nas frequências  $R_f$  (frequência de rotação do motor),  $2R_f$ , e  $3R_f$ . Isto dá origem a 18 características.

Além destas, utiliza-se também como característica a frequência de rotação do motor, obtida através do tacômetro.

No trabalho de [5] são mostradas figuras que representam a inserção de defeitos na bancada, as especificações dos acelerômetros e do tacômetro utilizados nos experimentos presentes neste artigo.

#### A. Curtose

A curtose  $K$  pode ser entendida como o grau de achatamento da função distribuição de probabilidade de uma variável aleatória. Ela também é conhecida como o momento central de quarta ordem. A curtose pode ser obtida através da seguinte equação:

$$K = \frac{E[x(n) - \mu]^4}{\sigma^4}, \quad (1)$$

em que  $x(n)$  é uma variável aleatória discreta,  $\mu$  é a média,  $\sigma$  é o desvio padrão e  $E[\cdot]$  é o operador da esperança matemática [6].

A curtose é muito utilizada para encontrar e reconhecer defeitos que produzam sinais de vibração impulsivos, pois esta técnica é bastante robusta mesmo quando há a presença de ruídos nos sinais [7].

Para cada sinal de vibração foi calculada a curtose, resultando em outras seis características.

#### B. Entropia

A entropia  $H$  pode ser definida como uma medida capaz de aferir quão imprevisível é uma variável aleatória [4],[8]. Esta medida pode ser calculada através da equação abaixo:

$$H = - \sum_{i=1}^C p_i(x = a_i) \cdot \log(p_i(x = a_i)), \quad (2)$$

onde  $x(n)$  é uma variável aleatória discreta,  $a_i$  são os possíveis valores de  $x(n)$ ,  $p_i(x = a_i)$  são as probabilidades de  $x(n) = a_i$ , onde  $p_i(x)$  é Função Densidade de Probabilidade (PDF, do inglês *probability density function*) do sinal, onde é obtido através de um estimador de densidade *Kernel*, aplicando o método da janela de Parzen-Rosenblat [9] and  $C$  é o número de pontos associado à PDF.

A entropia foi utilizada por [3] para problemas envolvendo falha de mancal e de eixo, obtendo bons resultados de classificação. Assim como a curtose, que fornece a medida de impulsividade do sinal, a entropia também pode gerar características discriminativas entre classes.

Para que a equação da entropia fique mais compacta pode-se definir a função abaixo:

$$f(p) = -p \cdot \log(p), \text{ para } 0 < p < 1, \quad (3)$$

combinando as Equações 2 e 3, a entropia pode ser representada pela Equação 4, logo:

$$H(X) = \sum_{\langle i \rangle} f(p(X = a_i)), \quad (4)$$

pode-se generalizar a definição de entropia de variáveis aleatórias discretas, para se obter a entropia de variáveis aleatórias contínuas ou de um vetor. Esta generalização é chamada de Entropia diferencial representada pela seguinte equação:

$$H_d(X) = \int_{\langle i \rangle} f(p_x(\xi)) d\xi, \quad (5)$$

onde  $p_x$  é a densidade de probabilidade da variável aleatória  $X$ .

Para cada sinal de vibração foi calculada a entropia, resultando em mais seis características.

### C. Vetor de características

O vetor de característica usado no experimento 1 apresenta 16 características obtidas a partir do mancal não invertido. Sendo 10 delas extraídas através da transformada rápida de Fourier, 3 extraídas por meio do cálculo da curtose e outras 3 extraídas por meio do cálculo da entropia extraídas do mancal não invertido. Já os vetores de características utilizados nos experimentos 2 e 3 deste trabalho tem 31 características obtidas através dos mancais invertido e não invertido. Sendo 19 delas extraídas através da transformada rápida de Fourier, 6 extraídas por meio do cálculo da curtose e outras 6 extraídas por meio do cálculo da entropia.

### IV. SEPARAÇÃO DO CONJUNTO DE DADOS USANDO *k-fold*

Quando as amostras a serem analisadas estão em grandes quantidades e são representativas utiliza-se o método de separação de dados chamado de *Hold Out* que consiste em separar o conjunto de amostra em 1/3 de amostras teste e 2/3 de amostras de treinamento. Porém neste trabalho a quantidade de amostras é limitada logo é necessário utilizar a validação cruzada para obter resultados mais confiáveis.

A validação cruzada é uma técnica estatística utilizada para estimar a performance de um algoritmo baseado num modelo de predição.

Na área de reconhecimento de padrões, são bastantes usados métodos de validação cruzada como: *k-fold*, *leave one out* e *jackknife test* para mostrar a eficácia do classificador independente do conjunto de teste escolhido.

Neste trabalho foi utilizado o método *k-fold* usando 10 *folds*. Seguindo esta metodologia, o conjunto original de dados é dividido de forma aleatória em 10 partes com tamanhos aproximadamente iguais. Dentre estas 10 partes são escolhidas de forma aleatória 3, para fazerem parte do conjunto de teste e as demais são usadas para o conjunto de treinamento. Este processo é repetido 120 vezes a fim de que todas as combinações possíveis sejam feitas para compor os conjuntos de teste e de treinamento. Por fim é feita uma média dos resultados da classificação das 120 combinações para obter um único resultado.

### V. RANDOM FOREST

*Random Forest* é um método de classificação *ensemble*, ou seja, ele combina a decisão de um conjunto de classificadores através de um processo de votação, a fim de classificar um exemplo não conhecido. Um classificador *ensemble* é geralmente mais eficaz do que qualquer um dos indivíduos que o formam [10].

O método *Random Forest* constrói um grande número de árvores de decisão a partir de um subconjunto de dados proveniente do conjunto de treinamento original através do método *bagging* [10].

Este método realiza a amostragem do conjunto de treinamento com reposição, onde são escolhidos de forma aleatória elementos para constituir um novo subconjunto de treino [11].

As árvores de decisão que em conjunto formam as *Random Forest*, são classificadores instáveis, ou seja, pequenas variações no conjunto de treinamento podem gerar alterações

significativas no classificador. Para dirimir este problema, utiliza-se o método *bagging* que apresenta bons resultados quando aplicados a classificadores instáveis [12].

O procedimento de *bootstrap* no *Random Forest* é usado para formar cada árvore de decisão [10]. Cada árvore da floresta é chamada de elemento preditor. A classificação das amostras no método *Random Forest* é feita através da contabilização dos votos de cada elemento preditor para cada classe e posteriormente atribui-se ao padrão a classe que obteve mais votos.

Utilizam-se combinações das entradas em cada nó para fazer crescer cada árvore nas florestas. As florestas geradas apresentam bons resultados quando comparados com outros métodos ensembles [13]. As principais características deste algoritmo são: robustez a ruídos e *outliers*; é mais rápido que os métodos de *bagging* e *boosting*; consegue fornecer informações úteis de erro (força, correlação e importância da variável); é simples de ser implementado e fácil de ser paralelizado.

### VI. BASE DE DADOS

Os sinais de vibração utilizados neste trabalho foram obtidos a uma taxa de aquisição de 50.000 amostras durante 5 segundos resultando num total de 250.000 amostras, a frequência de amostragem utilizada foi de 50 kHz. A base de dados utilizada neste experimento tem a seguinte divisão:

- **Sinais normais:** Esta classe possui 49 amostras com velocidades de rotação diferentes. A fim de balancear a população de classes foi realizado o seguinte procedimento: a partir de cada uma das amostras foram gerados mais 7 sinais através da soma de ruído branco aos sinais originais resultando em 343 amostras destes tipo de sinal;
- **Desalinhamento paralelo horizontal:** apresenta 197 amostras divididas em quatro graus de severidade: 0,5 mm (50 amostras), 1 mm (49 amostras), 1,5 mm (49 amostras), 2 mm (49 amostras);
- **Desalinhamento paralelo vertical:** esta classe tem 301 amostras separadas em seis graus de severidade: 0,51 mm (51 amostras), 0,63 mm (50 amostras), 1,27 mm (50 amostras), 1,4 mm (50 amostras), 1,78 mm (50 amostras) e 1,9 mm (50 amostras);
- **Desbalanceamento:** possui 333 amostras separadas em sete graus de severidade: 6 g (49 amostras), 10 g (48 amostras), 15 g (48 amostras), 20 g (49 amostras), 25 g (47 amostras), 30 g (47 amostras), 35 g (45 amostras).

Os outros 6 tipos de falhas analisados são as falhas nos mancais invertidos e não invertidos. Estes tipos de falhas são mais evidentes quando associados a um desbalanceamento [5], que são:

- **Defeito na gaiola do mancal não invertido:** possui 188 amostras associadas aos seguintes valores de desbalanceamento: 0 g (49 amostras), 6 g (48 amostras), 20 g (49 amostras), 35 g (42 amostras);
- **Defeito na pista externa do mancal não invertido:** esta classe apresenta 184 amostras associadas aos seguintes valores de desbalanceamento 0 g (49 amostras), 6 g (49 amostras), 20 g (49 amostras), 35 g (37 amostras);

- **Defeito na esfera do mancal não invertido:** esta classe tem 184 amostras associadas aos seguintes valores de desbalanceamento 0 g (50 amostras), 6 g (49 amostras), 20 g (49 amostras), 35 g (38 amostras);
- **Defeito na gaiola do mancal invertido:** possui 188 amostras associadas aos seguintes valores de desbalanceamento: 0 g (49 amostras), 6 g (49 amostras), 20 g (49 amostras), 35 g (41 amostras);
- **Defeito na pista externa do mancal invertido:** esta classe apresenta 188 amostras associadas aos seguintes valores de desbalanceamento 0 g (49 amostras), 6 g (49 amostras), 20 g (49 amostras), 35 g (41 amostras);
- **Defeito na esfera do mancal invertido:** esta classe tem 137 amostras associadas aos seguintes valores de desbalanceamento 0 g (49 amostras), 6 g (43 amostras), 20 g (25 amostras), 35 g (20 amostras).

## VII. DIAGNÓSTICO DE FALHAS

A fim de diagnosticar as falhas no motor no experimento foi utilizado o algoritmo *Random Forest*.

Para encontrar o melhor número de árvores do algoritmo, foram realizados experimentos variando o número de árvores que compõem a *Random Forest* de 1 a 100.

Aplicando o algoritmo no conjunto de treinamento a meta foi encontrar o menor número de árvores que resultasse na maior média de acerto relativo intraclasse (soma dos valores percentuais das acurácias dividida pelo número de classes) das 120 combinações produzidas pelo *k-fold*.

Uma vez escolhido o melhor número de árvores, este parâmetro é estabelecido no *Random Forest* e realiza-se o experimento nas combinações do *k-fold* que correspondem ao conjunto teste para verificar o desempenho real do método de classificação.

Estas 120 combinações associadas às 100 possibilidades de números de árvores diferentes fazem com que o algoritmo tenha que ser executado 12.000 vezes. Como este número de execuções é alto para ser feito em um computador normal os experimentos foram executados no cluster de computadores do Laboratório de Sinais, Multimídia e Telecomunicações da COPPE da UFRJ.

### A. Experimento 1

O primeiro experimento consistiu em realizar o diagnóstico de falhas utilizando um vetor com 16 características extraídas do mancal mais próximo do motor. Os defeitos analisados foram: desalinhamento horizontal, desalinhamento vertical, desbalanceamento, defeito no mancal invertido e defeito no mancal não invertido, além da classe correspondente aos sinais normais. Para este experimento o número ótimo de árvores obtido na etapa de treinamento foi de 38 árvores. Este número de árvores foi implementado no algoritmo para realizar a classificação do conjunto teste.

Para a comparação de desempenho do classificador *Random Forest*, foi implementado um classificador *Multi-Layer Perceptron* (MLP), com 16 neurônios na camada de entrada, 16 neurônios na camada oculta e com 6 neurônios na camada de saída. O algoritmo de aprendizado utilizado foi o

retropropagação, com função de ativação sigmoideal, com passo de aprendizagem de 0,01, com máximo de 1000 épocas de treinamento e erro mínimo quadrático desejado de  $10^{-5}$ . A MLP utilizada não foi parametrizada de forma ótima, mas sim apenas para se ter um efeito comparativo.

A Tabela I mostra o resultado da classificação nos *folds* correspondentes ao conjunto teste. Nesta tabela são mostrados os valores médios das acurácias e os desvios padrões para as 120 combinações do conjunto teste. Usando o *Random Forest*, a média de acerto relativo intraclasse obtida para este experimento foi de 94,30% com desvio padrão de 1,43%. O tempo de processamento deste experimento foi de 26 horas. Para a rede neural o acerto médio relativo intraclasse foi de 91,7% com desvio padrão de 3,6% e com tempo de treinamento foi de 1 hora e 10 minutos.

TABELA I

CARACTERÍSTICAS EXTRAÍDAS DO MANCAL MAIS PRÓXIMO DO MOTOR.

Classes	Classificadores	
	Random Forest (%)	RNA (%)
Normal	95,6 ± 5,9	95,2 ± 4,6
Desal. H.	93,4 ± 4,3	90,1 ± 5,2
Desal. V.	97,0 ± 2,5	94,1 ± 3,3
Desbal.	92,6 ± 2,4	92,2 ± 3,1
Def. Man. Inv.	93,9 ± 2,3	90,1 ± 2,8
Def. Man. n. Inv.	94,1 ± 1,9	91,3 ± 3,0
<b>Total (%)</b>	<b>94,3 ± 1,9</b>	<b>91,7 ± 3,6</b>

### B. Experimento 2

O segundo experimento consistiu em realizar o diagnóstico de falhas utilizando um vetor com 31 características extraídas do mancal mais próximo do motor e do mancal mais distante do motor. Os defeitos analisados foram os mesmos do experimento 1. Para o experimento 2 o número ótimo de árvores obtido na etapa de treinamento foi de 28 árvores. Este número de árvores foi implementado no algoritmo a fim de fazer a classificação do conjunto teste.

A Tabela II mostra o resultado da classificação nos *folds* correspondentes ao conjunto teste. Utilizando o *Random Forest*, a média de acerto relativo intraclasse obtida para este experimento foi de 97,02% com desvio padrão de 1,23%. O tempo de processamento deste experimento foi de 28 horas. Para a rede neural o acerto médio relativo intraclasse foi de 95,9% com desvio padrão de 1,7 % e com tempo de treinamento foi de 3 horas e 47 minutos.

TABELA II

CARACTERÍSTICAS EXTRAÍDAS DO MANCAL MAIS PRÓXIMO DO MOTOR E DO MANCAL MAIS DISTANTE DO MOTOR.

Classes	Classificadores	
	Random Forest (%)	RNA (%)
Normal	95,9 ± 5,4	98,8 ± 2,4
Desal. H.	96,2 ± 2,8	91,1 ± 2,6
Desal. V.	98,8 ± 1,1	97,3 ± 1,7
Desbal.	95,3 ± 1,8	95,7 ± 1,4
Def. Man. Inv.	98,0 ± 1,6	97,0 ± 1,2
Def. Man. n. Inv.	98,0 ± 1,2	95,5 ± 1,2
<b>Total (%)</b>	<b>97,0 ± 1,2</b>	<b>95,9 ± 1,7</b>



### C. Experimento 3

No experimento 3 foi utilizado o mesmo vetor de características do experimento 2, mas foi feita uma nova separação nas classes dos defeitos no mancal invertido e no mancal não invertido. As classes no experimento 3 foram: desalinhamento horizontal, desalinhamento vertical, desbalanceamento, defeito na gaiola do mancal não invertido, defeito na pista do mancal não invertido, defeito na esfera do mancal não invertido, defeito na gaiola do mancal invertido, defeito na pista do mancal invertido, defeito na esfera mancal invertido.

Para este experimento o número ótimo de árvores obtido na etapa de treinamento foi de 38 árvores. Este número de árvores foi implementado no algoritmo a fim de fazer a classificação do conjunto teste.

A Tabela III mostra o resultado da classificação nos *folds* correspondentes ao conjunto teste. Usando o *Random Forest*, a média de acerto relativo intraclasse obtida para este experimento foi de 97,48% com desvio padrão de 0,81%. O tempo de processamento deste experimento foi de 33 horas. Para a rede neural o acerto médio relativo intraclasse foi de 95,5% com desvio padrão de 4,5 % e com tempo de treinamento foi de 3 horas e 51 minutos.

TABELA III

CARACTERÍSTICAS EXTRAÍDAS DO MANCAL MAIS PRÓXIMO DO MOTOR E DO MANCAL MAIS DISTANTE DO MOTOR COM SEPARAÇÃO DE CLASSES DOS MANCAIS.

Classes	Classificadores	
	Random Forest (%)	RNA (%)
Normal	96,1 ± 5,8	95,2 ± 6,2
Desal. H.	96,4 ± 2,3	93,6 ± 3,0
Desal. V.	99,0 ± 1,1	96,6 ± 2,0
Desbal.	96,9 ± 2,1	96,8 ± 1,7
Def. Gaiola Man.n.Inv.	96,5 ± 2,2	94,1 ± 4,3
Def. Pista Man.n.Inv.	99,9 ± 0,4	94,6 ± 3,9
Def. Esfera Man.n.Inv.	99,1 ± 1,4	97,0 ± 3,2
Def. Gaiola M.Inv.	98,1 ± 2,1	92,6 ± 9,2
Def. Pista M.Inv.	92,8 ± 4,1	96,8 ± 9,1
Def. Esfera M.Inv.	100 ± 0	98,2 ± 1,8
<b>Total (%)</b>	<b>97,5 ± 2,2</b>	<b>95,5 ± 4,5</b>

Com a finalidade de saber o comportamento do algoritmo durante a fase de treinamento do experimento 3 foi plotado um gráfico representado pela Figura 2 que relaciona a acurácia obtida com o método *Random Forest* de acordo com a variação do número de árvores.

Analisando este gráfico verifica-se que o método atinge acurácias maiores que 99% na etapa de treinamento a partir de 4 árvores e mantém-se estável a partir de então.

### VIII. CONCLUSÕES

Através dos experimentos realizados foi verificado que o algoritmo *Random Forest* é bastante eficaz para realizar o diagnóstico de múltiplas falhas em máquinas rotativas, alcançando acertos acima de 90% para todas as classes e um baixo desvio padrão. Além disso foram obtidos ótimos valores das acurácias totais para os 3 experimentos. Também foi verificado que o *Random Forest* é bastante estável a partir de um determinado número de árvores.

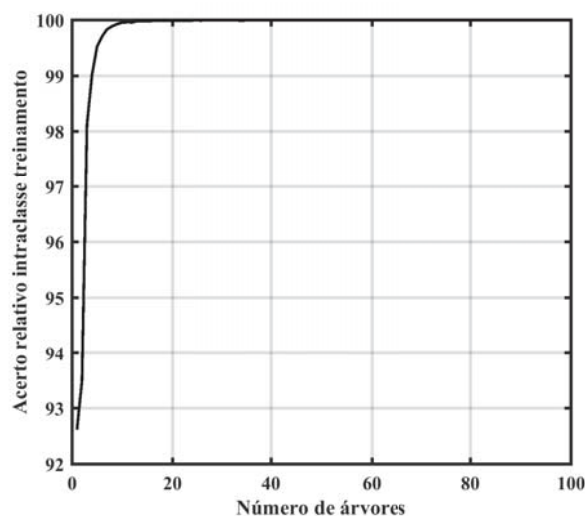


Fig. 2. Relação entre o acerto relativo intraclasse no treinamento e o nº de árvores

Sugestões para trabalhos futuros sobre este tema: realizar a classificação da severidade das classes; verificar nos vetores de características utilizados nos experimentos quais as características são mais relevantes para realizar a distinção das classes; usar valores de partições diferentes para o método *k-fold* a fim de analisar o impacto no desempenho do algoritmo.

### REFERÊNCIAS

- [1] Associação Brasileira de Normas Técnicas. “NBR 5462: Confiabilidade e manutenibilidade”. 1994.
- [2] A. A. D. Lima, T. D. M. Prego e S. L. Netto, et al., “On Fault Classification in Rotating Machines using Fourier Domain Features and Neural Networks”, *Circuits and Systems (LASCAS)*, IEEE Fourth Latin American Symposium, pp. 1–4, Fevereiro 2013.
- [3] D. Pestana-Viana, R. Zambrano-López, A. A. D. Lima, et al., “The Influence of Feature Vector on the Classification of Mechanical Faults using Neural Networks”, *Circuits and Systems (LASCAS)*, IEEE Seven Latin American Symposium, pp. 115–118, Março, 2016.
- [4] Y. Long, Y. He e L. Yuan, “Fault dictionary based switched current circuit fault diagnosis using entropy as preprocessor”, *Analog Integrated Circuits and Signal Processing*, vol. 66, pp.93–102. Janeiro, 2011.
- [5] Rafael Zambrano-López, “Classificação Automática de Defeitos em Máquinas Rotativas”, *Projeto Final de Graduação*, Universidade Federal do Rio de Janeiro, Dezembro, 2014.
- [6] W. Wang e H. Lee, “An energy kurtosis demodulation technique for signal denoising and bearing fault detection”, *Measurement Science and Technology* 24, pp.10, Janeiro 2013.
- [7] L. Saidi, J. B. Ali e F. Fnaiech, “The Use of Spectral Kurtosis as a Trend Parameter for Bearing Faults Diagnosis”, *Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, 15th International Conference, pp.394–399, Dezembro, 2014.
- [8] C.E.Shannon, “A mathematical theory of communication”, *The Bell System Technical Journal*, vol. 27, pp.379–423, 623–656., Outubro, 1948.
- [9] E. Parzen, “On estimation of a probability density function and mode”, *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, Sept., 1962.
- [10] B. S. Yang, X. Di e T. Han, “Random forests classifier for machine fault diagnosis”, *Journal of Mechanical Science and Technology*, vol. 22, pp. 1716–1725. Setembro, 2008.
- [11] L. Breiman, “Bagging Predictors”, *Machine Learning*, vol. 24, pp. 123–140, Agosto, 1996.
- [12] L. Breiman, “Random Forest”, *Machine Learning*, vol. 45, pp. 5–32, Outubro, 2001.
- [13] E. Acuña e E. Rojas, “Bagging Classifiers Based on Kernel Density Estimators”. *Proceedings of the International Conference On New Trends in Computational Statistics With Biomedical Applications*, pp. 343–350, 2001.