

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO DE MATEMÁTICA  
INSTITUTO TERCIO PACITTI DE APLICAÇÕES  
E PESQUISAS COMPUTACIONAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

CIRO ESTEVES LIMA SOBRAL

**DIAGNÓSTICO DE PROBLEMAS EM MOTORES  
A DIESEL ATRAVÉS DA ANÁLISE DE ÓLEO  
LUBRIFICANTE USANDO LÓGICA NEBULOSA**

Rio de Janeiro  
2015



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO DE MATEMÁTICA  
INSTITUTO TÉRCIO PACITTI DE APLICAÇÕES  
E PESQUISAS COMPUTACIONAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

CIRO ESTEVES LIMA SOBRAL

**DIAGNÓSTICO DE PROBLEMAS EM MOTORES  
A DIESEL ATRAVÉS DA ANÁLISE DE ÓLEO  
LUBRIFICANTE USANDO LÓGICA NEBULOSA**

Dissertação de Mestrado submetida ao  
Corpo Docente do Departamento de Ci-  
ência da Computação do Instituto de Ma-  
temática, e Instituto Tércio Pacitti de  
Aplicações e Pesquisas Computacionais da  
Universidade Federal do Rio de Janeiro,  
como parte dos requisitos necessários para  
obtenção do título de Mestre em Informá-  
tica.

Orientador: Josefino Cabral Melo Lima

Co-orientador: Adriano Joaquim de Oliveira Cruz

Rio de Janeiro  
2015

S677 Sobral, Ciro Esteves Lima

Diagnóstico de Problemas em Motores a Diesel através da  
Análise de Óleo Lubrificante usando Lógica Nebulosa / Ciro Esteves Lima  
Sobral. – 2015.

135 f.: il.

Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Programa de Pós-Graduação em Informática, Rio de Janeiro, 2015.

Orientador: Josefino Cabral Melo Lima.

Co-orientador: Adriano Joaquim de Oliveira Cruz.

1. Lógica Nebulosa. 2. Lógica Nebulosa Tipo-2. 3. Lógica Nebulosa Tipo-2 Intervalar. 4. Análise de Óleo Lubrificante. 5. Modelagem. – Teses. I. Lima, Josefino Cabral Melo (Orient.). II. Cruz, Adriano Joaquim de Oliveira (Co-orient.). III. Universidade Federal do Rio de Janeiro, Instituto de Matemática, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Programa de Pós-Graduação em Informática. IV. Título

CDD

CIRO ESTEVES LIMA SOBRAL

**Diagnóstico de Problemas em Motores a Diesel através  
da Análise de Óleo Lubrificante usando Lógica Nebulosa**

Dissertação de Mestrado submetida ao Corpo Docente do Departamento de Ciência da Computação do Instituto de Matemática, e Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para obtenção do título de Mestre em Informática.

Aprovado em: Rio de Janeiro, 17 de março de 2015.

Cabral Lima

Prof. Docteur. Josefino Cabral Melo Lima (Orientador)

Adriano Joaquim de Oliveira Cruz

Prof. Ph.D. Adriano Joaquim de Oliveira Cruz (Co-orientador)

Antonio Carlos Gay Thomé

Prof. Ph.D. Antonio Carlos Gay Thomé (Co-orientador)

Jonas Knopman

Prof. D.Sc. Jonas Knopman

Priscila Machado Vieira Lima

Profª. Ph.D. Priscila Machado Vieira Lima

Silvana Rossetto

Profª. D.Sc. Silvana Rossetto

Leila Cristina V. de Andrade

Profª. D.Sc. Leila Cristina Vasconcelos de Andrade

Rio de Janeiro  
2015



## AGRADECIMENTOS

Agradeço antes de tudo a Deus e a minha família, meu irmão Thales e meus pais, Ivanildo e Tereza, por terem me dado além do apoio emocional e financeiro (sempre bem-vindo), a força para enfrentar desafios, a coragem para não temer o desconhecido e a humildade para saber ouvir.

Um agradecimento especial à minha namorada, Jamille, pelas incontáveis horas de Skype e pelos períodos em que estivemos juntos, que com seu apoio e seu incentivo nunca me deixou desanimar.

Agradeço também aos meus amigos, Fernanda, Jacy, Pedro, João, Sérgio e Leonardo, entre outros que passaram pela minha vida e que de alguma forma me deram força e coragem nessa caminhada.

Quero agradecer ao meu orientador, Adriano Cruz, pela coragem de enfrentar essa empreitada de me orientar num problema novo para nós dois.

Agradeço ao CAPES e ao IM-NCE pelo apoio financeiro à pesquisa. Devo agradecer, também, aos colegas do Laboratório de Inteligência Computacional (LabIC) pelas opiniões e pelo companheirismo.





## RESUMO

Sobral, Ciro Esteves Lima. **Diagnóstico de Problemas em Motores a Diesel através da Análise de Óleo Lubrificante usando Lógica Nebulosa**. 2015. 116 f. Dissertação (Mestrado em Informática) - PPGI, Instituto de Matemática, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

Sistemas especialistas tem sido utilizados para o problema de diagnóstico por muitos anos, especialmente na área médica. Esse tipo de sistema tem como objetivo detectar falhas através do resultado de testes simples que possam fornecer parâmetros para a detecção do problema. O sucesso na aplicação desse sistema depende de quão preciso é o modelo usado. No entanto, a modelagem depende da extração de conhecimento de especialistas. Este processo pode ocorrer de duas maneiras distintas: ter um especialista disponível e disposto a transformar seu conhecimento em regras ou códigos de computador; ou através da aplicação de um método de aprendizado de máquina supervisionado. Os métodos de aprendizado são problemáticos quando há dados imprecisos ou quando não há consenso entre os especialistas quanto a classificação de cada amostra. Apresentamos aqui uma comparação de lógica fuzzy tradicional, com lógica nebulosa tipo-2 intervalar para a classificação de problemas em motores diesel através do resultado da análise de óleo. Para isso, foi utilizado um método *neuro-fuzzy* bem conhecido, o ANFIS, para a modelagem do sistema e desenvolvemos uma estrutura para usar um sistema de inferência nebuloso tipo-2 intervalar no MATLAB®. Esse sistema pretende realizar a identificação de possíveis casos de falha mesmo com imprecisão nos dados e no conhecimento dos especialistas.

**Palavras-chave:** Lógica Nebulosa, Lógica Nebulosa Tipo-2, Lógica Nebulosa Tipo-2 Intervalar, Análise de Óleo Lubrificante, Modelagem.



# ABSTRACT

Sobral, Ciro Esteves Lima. **Diagnóstico de Problemas em Motores a Diesel através da Análise de Óleo Lubrificante usando Lógica Nebulosa**. 2015. 116 f. Dissertação (Mestrado em Informática) - PPGI, Instituto de Matemática, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

Expert systems have been used for diagnostic problem for many years, especially for medical applications. This kind of system aims to detect failures through the result of simple tests used as parameters for detection of the problem. The success of this systems depends on how precise is the model used. However, the modeling depends on the extraction of knowledge from specialists. This process can occur in two distinct ways: having an expert available and willing to turn his knowledge into rules or computer codes; or by applying a supervised machine learning method. There are problems with the last method when there is imprecise data or when there is no consensus among experts for classifying each sample. Here we present a comparison of traditional fuzzy logic with interval type-2 fuzzy logic for classification of problems in diesel engines through the result of oil analysis. For this we used an neuro-fuzzy method well known, ANFIS, for system modeling and we developed a framework to use type-2 fuzzy inference system in MATLAB®. This classification system intends to carry out the identification of possible cases of failure even with uncertainty in data and in the experts knowledge.

**Keywords:** Fuzzy Logic, Type-2 Fuzzy Logic, Interval Type-2 Fuzzy Logic, Lubricant Oil Analysis, Fuzzy Modeling.



## LISTA DE FIGURAS

|      |                                                                                                         |    |
|------|---------------------------------------------------------------------------------------------------------|----|
| 1.1  | Custos de manutenção. . . . .                                                                           | 2  |
| 2.1  | Resultado de uma análise de óleo lubrificante. . . . .                                                  | 11 |
| 3.1  | Elementos de uma função de pertinência. . . . .                                                         | 25 |
| 3.2  | Esquema do Sistema de Inferência Nebuloso. . . . .                                                      | 26 |
| 3.3  | Função de pertinência de um T2FS. . . . .                                                               | 29 |
| 3.4  | Função de pertinência de um IT2FS. . . . .                                                              | 30 |
| 3.5  | Funções de pertinência superiores e inferiores de um IT2FS. . . . .                                     | 31 |
| 3.6  | Esquema do sistema de inferência nebuloso Tipo-2. . . . .                                               | 32 |
| 3.7  | Cálculo dos consequentes no sistema de inferência nebuloso tipo-2 intervalar. . . . .                   | 34 |
| 3.8  | Combinação dos consequentes no sistema de inferência nebuloso tipo-2 intervalar. . . . .                | 34 |
| 3.9  | Função de pertinência de T1FS embutido em um T2FS. . . . .                                              | 35 |
| 3.10 | Tipo-redução de um IT2FS pelo método do centroide. . . . .                                              | 37 |
| 5.1  | Método de cálculo da sensibilidade, especificidade, PPV e NPV para classificação multi-classes. . . . . | 55 |

## LISTA DE TABELAS

|      |                                                                                                                 |    |
|------|-----------------------------------------------------------------------------------------------------------------|----|
| 2.1  | Elementos encontrados na análise espectrométrica e possíveis fontes                                             | 21 |
| 5.1  | Formato da matriz de confusão . . . . .                                                                         | 50 |
| 5.2  | Matriz de pesos usada para avaliação de desempenho dos classificadores . . . . .                                | 56 |
| 6.1  | Média da taxa de acerto dos melhores classificadores . . . . .                                                  | 59 |
| 6.2  | Média da taxa de acerto dos classificadores . . . . .                                                           | 60 |
| 6.3  | Média da taxa de acerto ponderada dos melhores classificadores .                                                | 61 |
| 6.4  | Média da taxa de acerto ponderada dos classificadores . . . . .                                                 | 61 |
| A.1  | Valor máximo da taxa de acerto para o problema de Combustão .                                                   | 75 |
| A.2  | Valor médio da taxa de acerto para o problema de Combustão . .                                                  | 75 |
| A.3  | Resultado dos melhores classificadores para o problema de Combustão, avaliado pela taxa de acerto . . . . .     | 76 |
| A.4  | Valor máximo pela matriz de pesos para o problema de Combustão                                                  | 77 |
| A.5  | Valor médio pela matriz de pesos para o problema de Combustão                                                   | 77 |
| A.6  | Resultado dos melhores classificadores para o problema de Combustão, avaliado pela matriz de pesos . . . . .    | 78 |
| A.7  | Valor médio da sensibilidade para o problema de Combustão . . .                                                 | 79 |
| A.8  | Valor médio da especificidade para o problema de Combustão . .                                                  | 79 |
| A.9  | Valor médio do PPV para o problema de Combustão . . . . .                                                       | 79 |
| A.10 | Valor médio do NPV para o problema de Combustão . . . . .                                                       | 80 |
| A.11 | Valor máximo da taxa de acerto para o problema de Contaminação                                                  | 81 |
| A.12 | Valor médio da taxa de acerto para o problema de Contaminação                                                   | 81 |
| A.13 | Resultado dos melhores classificadores para o problema de Contaminação, avaliado pela taxa de acerto . . . . .  | 82 |
| A.14 | Valor máximo pela matriz de pesos para o problema de Contaminação . . . . .                                     | 83 |
| A.15 | Valor médio pela matriz de pesos para o problema de Contaminação                                                | 83 |
| A.16 | Resultado dos melhores classificadores para o problema de Contaminação, avaliado pela matriz de pesos . . . . . | 84 |
| A.17 | Valor médio da sensibilidade para o problema de Contaminação .                                                  | 85 |
| A.18 | Valor médio da especificidade para o problema de Contaminação .                                                 | 85 |
| A.19 | Valor médio do PPV para o problema de Contaminação . . . . .                                                    | 85 |
| A.20 | Valor médio do NPV para o problema de Contaminação . . . . .                                                    | 86 |
| A.21 | Valor máximo da taxa de acerto para o problema de Corrosão . .                                                  | 87 |

|      |                                                                                                                              |     |
|------|------------------------------------------------------------------------------------------------------------------------------|-----|
| A.22 | Valor médio da taxa de acerto para o problema de Corrosão . . .                                                              | 87  |
| A.23 | Resultado dos melhores classificadores para o problema de Corrosão, avaliado pela taxa de acerto . . . . .                   | 88  |
| A.24 | Valor máximo pela matriz de pesos para o problema de Corrosão                                                                | 89  |
| A.25 | Valor médio pela matriz de pesos para o problema de Corrosão .                                                               | 89  |
| A.26 | Resultado dos melhores classificadores para o problema de Corrosão, avaliado pela matriz de pesos . . . . .                  | 90  |
| A.27 | Valor médio da sensibilidade para o problema de Corrosão . . . .                                                             | 91  |
| A.28 | Valor médio da especificidade para o problema de Corrosão . . . .                                                            | 91  |
| A.29 | Valor médio do PPV para o problema de Corrosão . . . . .                                                                     | 91  |
| A.30 | Valor médio do NPV para o problema de Corrosão . . . . .                                                                     | 92  |
| A.31 | Valor máximo da taxa de acerto para o problema de Combustão, usando SMOTE . . . . .                                          | 93  |
| A.32 | Valor médio da taxa de acerto para o problema de Combustão, usando SMOTE . . . . .                                           | 93  |
| A.33 | Resultado dos melhores classificadores para o problema de Combustão, usando SMOTE, avaliado pela taxa de acerto . . . . .    | 94  |
| A.34 | Valor máximo pela matriz de pesos para o problema de Combustão, usando SMOTE . . . . .                                       | 95  |
| A.35 | Valor médio pela matriz de pesos para o problema de Combustão, usando SMOTE . . . . .                                        | 95  |
| A.36 | Resultado dos melhores classificadores para o problema de Combustão, usando SMOTE, avaliado pela matriz de pesos . . . . .   | 96  |
| A.37 | Valor médio da sensibilidade para o problema de Combustão, usando SMOTE . . . . .                                            | 97  |
| A.38 | Valor médio da especificidade para o problema de Combustão, usando SMOTE . . . . .                                           | 97  |
| A.39 | Valor médio do PPV para o problema de Combustão, usando SMOTE . . . . .                                                      | 98  |
| A.40 | Valor médio do NPV para o problema de Combustão, usando SMOTE . . . . .                                                      | 98  |
| A.41 | Valor máximo da taxa de acerto para o problema de Contaminação, usando SMOTE . . . . .                                       | 99  |
| A.42 | Valor médio da taxa de acerto para o problema de Contaminação, usando SMOTE . . . . .                                        | 99  |
| A.43 | Resultado dos melhores classificadores para o problema de Contaminação, usando SMOTE, avaliado pela taxa de acerto . . . . . | 100 |
| A.44 | Valor máximo pela matriz de pesos para o problema de Contaminação, usando SMOTE . . . . .                                    | 101 |

|      |                                                                                                                             |     |
|------|-----------------------------------------------------------------------------------------------------------------------------|-----|
| A.45 | Valor médio pela matriz de pesos para o problema de Contaminação, usando SMOTE . . . . .                                    | 101 |
| A.46 | Resultado dos melhores classificadores para o problema de Contaminação, usando SMOTE, avaliado pela matriz de pesos . . . . | 102 |
| A.47 | Valor médio da sensibilidade para o problema de Contaminação, usando SMOTE . . . . .                                        | 103 |
| A.48 | Valor médio da especificidade para o problema de Contaminação, usando SMOTE . . . . .                                       | 103 |
| A.49 | Valor médio do PPV para o problema de Contaminação, usando SMOTE . . . . .                                                  | 104 |
| A.50 | Valor médio do NPV para o problema de Contaminação, usando SMOTE . . . . .                                                  | 104 |
| A.51 | Valor máximo da taxa de acerto para o problema de Corrosão, usando SMOTE . . . . .                                          | 105 |
| A.52 | Valor médio da taxa de acerto para o problema de Corrosão, usando SMOTE . . . . .                                           | 105 |
| A.53 | Resultado dos melhores classificadores para o problema de Corrosão, usando SMOTE, avaliado pela taxa de acerto . . . . .    | 106 |
| A.54 | Valor máximo pela matriz de pesos para o problema de Corrosão, usando SMOTE . . . . .                                       | 107 |
| A.55 | Valor médio pela matriz de pesos para o problema de Corrosão, usando SMOTE . . . . .                                        | 107 |
| A.56 | Resultado dos melhores classificadores para o problema de Corrosão, usando SMOTE, avaliado pela matriz de pesos . . . . .   | 108 |
| A.57 | Valor médio da sensibilidade para o problema de Corrosão, usando SMOTE . . . . .                                            | 109 |
| A.58 | Valor médio da especificidade para o problema de Corrosão, usando SMOTE . . . . .                                           | 109 |
| A.59 | Valor médio do PPV para o problema de Corrosão, usando SMOTE                                                                | 110 |
| A.60 | Valor médio do NPV para o problema de Corrosão, usando SMOTE                                                                | 110 |



# SUMÁRIO

|          |                                       |           |
|----------|---------------------------------------|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>                     | <b>1</b>  |
| 1.1      | Motivação                             | 2         |
| 1.2      | Hipótese                              | 4         |
| 1.3      | Justificativa                         | 4         |
| 1.4      | Objetivo                              | 5         |
| 1.5      | Trabalhos Relacionados                | 5         |
| 1.6      | Estrutura da dissertação              | 6         |
| <b>2</b> | <b>ANÁLISE DO LUBRIFICANTE</b>        | <b>9</b>  |
| 2.1      | Funções do Lubrificante               | 12        |
| 2.2      | Problemas no Lubrificante             | 13        |
| 2.3      | Testes Aplicados ao Lubrificante      | 14        |
| 2.4      | Propriedades do Lubrificante          | 16        |
| <b>3</b> | <b>LÓGICA NEBULOSA</b>                | <b>23</b> |
| 3.1      | Conjuntos Nebulosos                   | 24        |
| 3.1.1    | Sistema de Inferência Nebuloso        | 24        |
| 3.2      | Conjuntos Nebulosos Tipo-2            | 28        |
| 3.2.1    | Sistema de Inferência Nebuloso Tipo-2 | 32        |
| <b>4</b> | <b>MODELO PROPOSTO</b>                | <b>39</b> |
| 4.1      | Descrição dos Dados                   | 41        |
| 4.2      | Descrição do <i>Framework</i>         | 43        |
| 4.2.1    | Funções                               | 45        |
| <b>5</b> | <b>METODOLOGIA</b>                    | <b>49</b> |
| 5.1      | Seleção de atributos                  | 51        |
| 5.2      | Descrição dos testes sem uso do SMOTE | 52        |
| 5.3      | Descrição dos testes com uso do SMOTE | 53        |
| 5.4      | Definição das métricas de avaliação   | 53        |
| <b>6</b> | <b>RESULTADOS E DISCUSSÃO</b>         | <b>59</b> |
| <b>7</b> | <b>CONCLUSÕES</b>                     | <b>65</b> |
| 7.1      | Contribuições                         | 66        |
| 7.2      | Trabalhos Futuros                     | 67        |
| 7.3      | Dificuldades Encontradas              | 68        |

|                                                    |     |
|----------------------------------------------------|-----|
| REFERÊNCIAS . . . . .                              | 71  |
| APÊNDICE A – RESULTADOS . . . . .                  | 75  |
| APÊNDICE B – ARTIGO PUBLICADO NO SMC2014 . . . . . | 111 |



# 1 INTRODUÇÃO

O trabalho proposto nesta dissertação aborda o problema do diagnóstico computadorizado de problemas em motores a diesel, através do resultado da análise de amostras de óleo lubrificante usado. O diagnóstico pode ser realizado com o auxílio de técnicas de classificação, que são usadas com sucesso em diversas áreas, como identificação de caracteres em imagens (MORI; SUEN; YAMAMOTO, 1995), análise de expressões faciais (FASEL; LUETTIN, 2003) e detecção de doenças através de exames de imagem (ADLASSNIG, 1982; PHUONG, 1995; GINNEKEN et al., 2001).

Óleos lubrificantes são compostos por uma combinação de substâncias de diversas origens. Eles têm a função de diminuir o atrito entre superfícies em movimento, formando uma camada intermediária que evita o contato direto entre as partes. A análise de óleos lubrificantes usados consiste na aplicação de um conjunto de testes aplicados ao óleo, visando obter informações sobre a condição do lubrificante e sobre o funcionamento da máquina.

Os dados utilizados nesta pesquisa consistem em resultados de análises de óleo lubrificante provenientes de motores de combustão interna à diesel. O resultado da análise de óleo traz uma série de parâmetros que serão usados como entrada para o sistema de classificação e as classificações informadas por especialistas, que serão usadas como a saída esperada do sistema. No entanto, alguns desses resultados possuem ruídos nos dados, dados faltantes e dados imprecisos. Todas essas fontes de incerteza dificultam a classificação automatizada.

Dentre diversas técnicas de classificações como árvores de decisão, redes neurais e SVM (*Support Vector Machines*) decidiu-se usar um sistema de aprendizado supervisionado através de lógica nebulosa. Tal escolha se justifica pela capacidade de modelagem de incertezas dos sistema de lógica nebulosa. Os sistemas de inferência nebulosos usam um conjunto de regras do tipo "se-então" para definir a pertinência de um elemento a um determinado grupo. Essas regras permitem modelar intuitivamente o conhecimento de especialistas para detectar os problemas de forma similar a dos mesmos.

Mais especificamente, o presente trabalho propõe a criação de classificadores usando lógica nebulosa tipo-2 intervalar a partir de classificadores baseados em lógica nebulosa tradicional. Então, pretende-se comparar estes classificadores com a intenção de avaliar se lógica nebulosa tipo-2 intervalar é melhor do que a tradicional para modelar o problema abordado.

## 1.1 Motivação

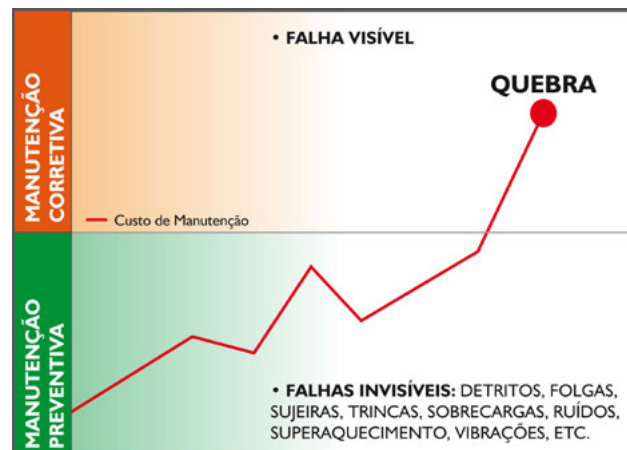


Figura 1.1: Custos de manutenção.

Equipamentos podem apresentar falhas devido ao desgaste natural das peças, uso em condições de funcionamento desfavoráveis ou provenientes de mau uso. Falhas em equipamentos trazem prejuízos pelos custos de reparo e interrupção da produção, o que gera perdas por lucros cessantes. Quanto mais tarde for detectado um problema, maiores são os custos para seu reparo (THWEATT; BELL, 2006), como exemplificado na Figura 1.1. A manutenção preditiva é uma ferramenta que pode ajudar a diminuir o número de falhas e o tempo de manutenção. No entanto, a definição de sua frequência depende de indicadores confiáveis e modelos de diagnóstico (CARNERO, 2005).

Técnicas de monitoramento da condição do motor através de análise de óleo lubrificante são meios efetivos e largamente usados para diagnóstico de desgaste anormal e mal funcionamento dos motores (RAMEZANI; MEMARIANI, 2011). Através da análise do lubrificante usado em um motor de combustão interna é possível detectar problemas como desgaste excessivo das peças, problemas na combustão e contaminação do óleo antes que o mesmo apresente risco iminente de falha, prevenindo danos e custos maiores (THWEATT; BELL, 2006). Por esse motivo, a análise do lubrificante é de suma importância para a manutenção preditiva.

A análise do óleo lubrificante fornece informações de forma similar a um exame de sangue. Os óleos lubrificantes usados possuem diversos compostos químicos e partículas resultantes da combustão, do desgaste das peças e provenientes de contaminação (AKINTUNDE, 2008; THWEATT; BELL, 2006). No entanto, para compreender os possíveis problemas detectados na análise, esses dados devem ser interpretados por um especialista que realize o diagnóstico. Esse diagnóstico é trabalhoso e por esse motivo, métodos computadorizados de interpretação do resultado da análise dos lubrificantes se tornam uma alternativa viável (SALA et al., 2005; MACIAN et al., 2006).

Realizar a extração desse conhecimento de especialistas não é uma tarefa trivial. Decidiu-se usar a técnica de aprendizado supervisionado para modelagem do sistema. Essa técnica consiste na geração/treinamento dos parâmetros do sistema a partir de um conjunto de treino com os dados de entrada e as correspondentes saídas esperadas. Entretanto a geração do modelo a partir dos dados pode causar sobreajuste (*overfitting*), que é quando o sistema perde a capacidade de generalização, não respondendo corretamente a dados que não foram usados no processo de treinamento. Ainda essa técnica pode levar a resultados incorretos caso haja inconsistência nos dados, como ruído nos dados, amostras classificadas incorretamente ou dados faltantes, assim como podem levar. Situações onde essas inconsistências podem ocorrer serão explicadas na seção 4.1.

## 1.2 Hipótese

Devido à incerteza presente nos dados disponíveis para realizar a pesquisa, acredita-se que lógica nebulosa tipo-2 seja mais apropriado para a modelagem do problema do que sistemas de inferência nebulosos tradicionais.

## 1.3 Justificativa

Sistemas baseados em lógica nebulosa são conhecidos por ser uma solução útil no tratamento de dados incertos (RAJU et al., 2011). No entanto, mesmo esses sistemas têm suas limitações e não conseguem tratar incertezas diretamente (MENDEL; KARNIK, 1998; MENDEL, 2001). Para superar estas limitações, alguns autores como MENDEL; KARNIK (1998); MENDEL (2001); MELIN; CASTILLO (2013) propõem usar sistemas baseados em lógica nebulosa tipo-2 para modelar melhor as incertezas, pois estes permitem modelar o conhecimento de vários especialistas,

mesmo não havendo consenso entre os mesmos. Como a classificação das amostras de óleos lubrificantes usada nessa pesquisa foi dada por especialistas, essa situação pode ocorrer no caso de amostras semelhantes serem avaliadas por critérios distintos.

Os sistemas de classificação baseados em lógica nebulosa permitem tratar imprecisão nos dados e com isso conseguir melhores resultados no funcionamento do sistema de classificação. No caso da classificação de amostras de óleos lubrificantes as imprecisões podem ser causadas por falhas na medição e pelos métodos de avaliação do óleo lubrificante, que como será tratado mais a frente, nem sempre usam métodos rígidos de medição.

## **1.4 Objetivo**

O presente trabalho busca investigar o uso de lógica nebulosa para o problema de classificação da condição de motores, através do resultado da análise de óleo lubrificante usado. Ainda, pretende-se estabelecer uma comparação entre os classificadores gerados usando lógica nebulosa tipo-2 e os de lógica nebulosa tipo-1.

## **1.5 Trabalhos Relacionados**

Lógica nebulosa pode ser usada para detectar problemas em motores através da análise de lubrificantes, como proposto por SALA et al. (2005), que usou um modelo similar ao Mamdani, aplicando um conjunto elaborado de regras para transformações de valores e lógica nebulosa (tipo-1) intervalar para suprimir os efeitos da ausência de dados. Esse mesmo sistema foi testado em MACIAN et al. (2006). Porém os dados da pesquisa não foram divulgados e os resultados apresentados no artigo não permitem comparação.



Sistemas baseados em lógica nebulosa e neuro-fuzzy foram desenvolvidos para diagnóstico em diversos equipamentos (CALADO et al., 2001). No meio automotivo, existem várias aplicações de sistemas nebulosos para detecção de falhas (MOHAMMADPOUR; FRANCHEK; GRIGORIADIS, 2012). Alguns autores já abordaram o diagnóstico de problemas em motores a Diesel usando lógica nebulosa (RAMEZANI; MEMARIANI, 2011; SALA et al., 2005; MACIAN et al., 2006). No entanto, a comparação com ambos os trabalhos não é possível pois os resultados, quando apresentados, não permitem comparação com o deste trabalho. Acredita-se que isso se justifica pela necessidade de proteção do conhecimento intelectual devido à competitividade das empresas que trabalham no setor automotivo.

Outros trabalhos usando lógica nebulosa tipo-2 para os problemas de classificação de tipos de veículos a partir de informações como distancia entre eixos, altura e comprimento do veículo, detecção de câncer através de mamografia, entre outros foram listados por MELIN; CASTILLO (2013).

## 1.6 Estrutura da dissertação

Esta dissertação está dividida em 7 capítulos e apêndices, e está disposta na seguinte ordem: No Capítulo 2, são apresentados conceitos e abordagens sobre óleo lubrificante. No Capítulo 3 é feita uma revisão rápida sobre conjuntos nebulosos, sistemas de inferência nebulosos e apresentados seus correspondentes em lógica nebulosa tipo-2 intervalar. No Capítulo 4, são apresentados os dados usados na pesquisa e as definições sobre o sistema usado para classificação. No Capítulo 5, são abordados aspectos metodológicos da dissertação, como a descrição dos passos dos testes e das métricas usadas para avaliação do sistema. No Capítulo 6, apresenta-se os resultados dos classificadores usando lógica nebulosa tradicional, lógica nebulosa tipo-2 intervalar, e a comparação entre ambos. No Capítulo 7, apresenta-se as con-

tribuições, sugestões para trabalhos futuros e descreve-se as dificuldades encontradas na execução deste trabalho. Por fim, nos apêndices, é possível encontrar todas as tabelas de resultados e o artigo publicado no SMC 2014.



## 2 ANÁLISE DO LUBRIFICANTE

Óleos lubrificantes são compostos por uma combinação de substâncias de diversas origens. Em geral, os lubrificantes são compostos por uma base de origem mineral (derivados de petróleo) ou sintética (produzidos em laboratório), podendo ser também uma combinação de ambos, chamada de semissintética. A essa base, são acrescentados aditivos que servem para conferir propriedades que os óleos não possuam naturalmente (SENAI-ES, 1997; BARRETT, 2007?)

Eles têm a função de diminuir o atrito entre superfícies em movimento, formando uma camada intermediária que evita o contato direto entre as partes. O bom funcionamento de motores de combustão interna depende de lubrificação adequada, pois o contato das partes móveis com as partes fixas pode causar desde aumento de consumo de combustível até a quebra prematura das peças. Os óleos lubrificantes podem ainda exercer as funções de controle de desgaste, de temperatura, de corrosão, transmissão de força, amortecimento de choques, remoção de contaminantes e vedação (SENAI-ES, 1997; BARRETT, 2007?; Supreme Lubrificantes, 2011?; Dep. Técnico da Lubrikol Lubrificantes, 2013)

O óleo lubrificante percorre as partes do motor de forma similar ao sangue nos animais. Para tal, amostras de lubrificantes usados podem ser recolhidas e examinadas para identificar o estado do motor como em um exame de sangue. A análise de óleos lubrificantes usados consiste na aplicação de um conjunto de testes aplicados ao óleo, visando obter informações sobre a condição do lubrificante e sobre o funcionamento da máquina. Uma avaliação criteriosa desses resultados permite o diagnóstico de falhas nos componentes do motor, a identificação de irregularidades no funcionamento, bem como traçar um perfil de desgaste do equipamento e de

seus componentes, através das partículas de contaminação e das partículas de desgaste detectadas (SENAI-ES, 1997; CARNERO, 2005; THWEATT; BELL, 2006; MACIAN et al., 2006; AKINTUNDE, 2008).

O estudo dos resultados da análise do lubrificante permite a identificação de possíveis falhas no equipamento. Em geral são definidos valores limite para cada elemento da análise. No entanto, uma vez que esses valores tenham sido excedidos, não é possível indicar se a máquina apresenta risco iminente de falha, apenas que foi encontrado um risco em potencial para o desempenho ou a durabilidade da mesma. Outro ponto importante, é que não existem valores limites universalmente aceitos. Investigadores e condições de operação distintos podem sugerir limites diferentes (BOOSER, 1993; THWEATT; BELL, 2006).

A Figura 2.1 traz um exemplo de relatório de uma análise de óleo lubrificante do motor de uma moto. Nesse exemplo é possível observar os valores de diversos atributos do óleo lubrificante e a análise do resultado dado por um especialista, onde indica a existência de contaminantes como silício e presença de combustível, em quantidades acima dos limites aceitos, que causaram diminuição na viscosidade.

A análise do óleo lubrificante deve ser vista por três aspectos: a condição do lubrificante, a presença de contaminantes e o desgaste da máquina (CARNERO, 2005; THWEATT; BELL, 2006; BARRETT, 2007?).

A condição do lubrificante revela se o mesmo ainda mantém as propriedades que permitam seu uso ou se deve ser trocado. No entanto, apenas a substituição do óleo contaminado pode não ser o suficiente para a resolução do problema, visto que o sistema pode continuar contaminado provocando desgastes até a falha completa de um ou mais de seus componentes mecânicos (Dep. Técnico da Lubrikol Lubrificantes, 2013).



## Lubricant Analysis Report

877-458-3315

|        |          |          |   |   |
|--------|----------|----------|---|---|
| 0      | 1        | 2        | 3 | 4 |
| NORMAL | ABNORMAL | CRITICAL |   |   |

Overall report severity based on comments.

| Comments                      |                          | Data flagged for observation only; Silicon is at a MODERATE LEVEL; SILICON sources can be abrasives (dirt, Alumina Silica), seals and gasket material, lube additive or lube supplement, and/or environmental contaminant; FUEL DILUTION is at a MINOR LEVEL; FUEL DILUTION has caused viscosity to decrease slightly below grade; FUEL DILUTION possibly caused by excessive idling; CADMIUM sources include ELECTROPLATED COATING, SILVER SOLDER, or as a specialty alloy with a low coefficient of friction; Lubricant and filter change acknowledged; |           |           |             |            |               |               |          |                    |                          |                  |             |                           |            |           |           |         |                       |           |         |        |             |      |              |  |  |                  |  |  |  |  |  |  |
|-------------------------------|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------|-------------|------------|---------------|---------------|----------|--------------------|--------------------------|------------------|-------------|---------------------------|------------|-----------|-----------|---------|-----------------------|-----------|---------|--------|-------------|------|--------------|--|--|------------------|--|--|--|--|--|--|
| Sample #                      | Wear Metals (ppm)        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |           |           |             |            |               |               |          |                    | Contaminant Metals (ppm) |                  |             | Multi-Source Metals (ppm) |            |           |           |         | Additive Metals (ppm) |           |         |        |             |      |              |  |  |                  |  |  |  |  |  |  |
|                               | Iron                     | Chromium                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Nickel    | Aluminum  | Copper      | Lead       | Tin           | Cadmium       | Silver   | Vanadium           | Silicon                  | Sodium           | Potassium   | Titanium                  | Molybdenum | Antimony  | Manganese | Lithium | Boron                 | Magnesium | Calcium | Barium | Phosphorous | Zinc |              |  |  |                  |  |  |  |  |  |  |
|                               | 1                        | 17                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 0         | 4         | 14          | 12         | 0             | 0             | 1        | 0                  | 0                        | 41               | 3           | 1                         | 0          | 112       | 0         | 0       | 0                     | 1         | 6       | 2064   | 17          | 733  | 823          |  |  |                  |  |  |  |  |  |  |
| Sample Information            |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |           |           |             |            |               |               |          |                    |                          |                  |             |                           |            |           |           |         |                       |           |         |        |             |      | Contaminants |  |  | Fluid Properties |  |  |  |  |  |  |
| Sample #                      | Date Sampled             | Date Received                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Lube Time | Unit Time | Lube Change | Lube Added | Filter Change | Fuel Dilution | Soot     | Water              | Viscosity 40 °C          | Viscosity 100 °C | Acid Number | Base Number               | Oxidation  | Nitration |           |         |                       |           |         |        |             |      |              |  |  |                  |  |  |  |  |  |  |
| 1                             | 09-Apr-2011              | 19-Apr-2011                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 388       | 388       | Yes         |            | Yes           | 2.3 - GC      | <.1      | <.1 - FTIR         |                          | 12.1             | 6.03        | 5                         | 12         |           |           |         |                       |           |         |        |             |      |              |  |  |                  |  |  |  |  |  |  |
| Particle Count (particles/mL) |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |           |           |             |            |               |               |          | Additional Testing |                          |                  |             |                           |            |           |           |         |                       |           |         |        |             |      |              |  |  |                  |  |  |  |  |  |  |
| Sample #                      | ISO Code Based On 4/6/14 | > 4 µm                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | > 6 µm    | > 10 µm   | > 14 µm     | > 21 µm    | > 38 µm       | > 70 µm       | > 100 µm | Test Method        |                          |                  |             |                           |            |           |           |         |                       |           |         |        |             |      |              |  |  |                  |  |  |  |  |  |  |
| 1                             |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |           |           |             |            |               |               |          |                    |                          |                  |             |                           |            |           |           |         |                       |           |         |        |             |      |              |  |  |                  |  |  |  |  |  |  |

Figura 2.1: Resultado de uma análise de óleo lubrificante.

Os contaminantes entram no sistema vindos do ambiente, em forma de poeira, água e contaminações no processo e são a principal causa de degradação da máquina (BARRETT, 2007?; Supreme Lubrificantes, 2011?) e sua análise se justifica, pois uma máquina com problemas gera partículas de desgaste em ritmo exponencial (BARRETT, 2007?).

Por estes motivos, a análise sistemática do óleo lubrificante é uma alternativa para fornecer informações sobre o estado de máquinas e motores, visando sua adequada manutenção (Dep. Técnico da Lubrikol Lubrificantes, 2013). Com isso aumenta-se a confiabilidade, a disponibilidade e a vida útil dos equipamentos, enquanto minimiza-se os custos com trocas e reparos, uma vez que permite detectar

problemas antes da falha de um componente (THWEATT; BELL, 2006; BARRETT, 2007?; Dep. Técnico da Lubrikol Lubrificantes, 2013; PERKINELMER, 2012; Supreme Lubrificantes, 2011?).

## **2.1 Funções do Lubrificante**

Óleos lubrificantes são usados nas mais diversas máquinas, para atuar em uma grande variedade de ambientes e por isso devem realizar diferentes funções. As principais funções dos lubrificantes são as seguintes (Dep. Técnico da Lubrikol Lubrificantes, 2013; Supreme Lubrificantes, 2011?):

### **Lubrificação**

Cria um filme entre as partes móveis, separando as superfícies de contato, reduzindo a fricção entre elas e evitando a perda de energia (BARRETT, 2007?; SENAI-ES, 1997).

### **Arrefecimento**

Absorve o calor gerado pela fricção das superfícies (BARRETT, 2007?; SENAI-ES, 1997).

### **Limpeza**

Retira as partículas sólidas, tanto contaminantes quanto as resultantes do atrito das áreas de contato (BARRETT, 2007?).

### **Proteção**

Cobre os componentes proporcionando uma barreira contra oxidação, que pode levar à corrosão, e contra a ação de ácidos (BARRETT, 2007?; SENAI-ES, 1997).

### **Vedação**

Forma uma barreira para impedir a entrada de contaminantes, gases e fluidos indesejados (BARRETT, 2007?; SENAI-ES, 1997).

Por exercer tantas funções em contato direto com as partes internas da máquina o lubrificante se torna um bom meio para obter impressões sobre o estado operacional do equipamento (Dep. Técnico da Lubrikol Lubrificantes, 2013).

## **2.2 Problemas no Lubrificante**

A partir do uso de lubrificante, algumas situações podem ocorrer e se tornarem problemáticas para o bom funcionamento do lubrificante. De acordo com BARRETT (2007?) os seguintes problemas são os mais comuns:

### **Contaminação**

Partículas advindas de fontes externas (entrada de água ou poeira, por exemplo) ou internas (devidas ao funcionamento da máquina) que se misturam ao lubrificante e modificam suas propriedades físico-químicas.

### **Degradação**

Causada pela oxidação, que acontece quando as moléculas de oxigênio se combinam com os hidrocarbonetos. Esses componentes se transformam então em ácidos carboxílicos.

### **Depleção de Aditivos**

Os aditivos são consumidos ou modificados quimicamente enquanto realizam sua função e com isso, alguma das características do lubrificante pode ser alterada.



## 2.3 Testes Aplicados ao Lubrificante

Diversos testes podem ser aplicados ao lubrificante, para investigar aspectos do óleo lubrificante. Para algumas propriedades, existe mais de um teste possível de ser aplicado. Existem diversas razões para a escolha do método de avaliação do óleo lubrificante e fica a cargo de quem vai aplicar os testes a escolha dos métodos mais adequados. Para o escopo desse trabalho vale ressaltar que alguns possuem resultados mais precisos e outros menos, o que é uma fonte de imprecisão.

*Análises físico-químicas* entre outras, permitem fazer avaliação do seu desempenho e encontrar partículas e substâncias que podem atrapalhar o bom funcionamento do lubrificante e até danificar o equipamento (SENAI-ES, 1997).

*Exames de ferrografia e espectrometria* são utilizados para identificar no lubrificante elementos que podem estar presentes por serem partículas de desgaste, componentes de aditivos ou contaminantes (SOUZA, 2000). Os elementos de interesse ao estudo das partículas de desgaste geralmente são metais como: ferro, cromo, níquel, alumínio, cobre, chumbo, estanho, cádmio, prata, titânio e vanádio. Alguns aditivos também possuem metais e semimetais em sua composição como: magnésio, boro, cálcio, bário, fósforo, molibdênio e zinco. Há também alguns elementos que podem ser caracterizados como contaminantes como: silício, sódio, boro e potássio (ConocoPhillips Lubricants, 2007).

Outras análises identificam compostos químicos não metálicos que prejudicam o funcionamento normal do lubrificante. Essas incluem testes físico-químicos e análise por infravermelho. Os resultados permitem avaliar os níveis de glicol, sulfatos, oxidação, fuligem e contaminação por água (SOUZA, 2000).

## Teste de viscosidade

O óleo é medido com o uso do *viscosímetro* e a viscosidade é medida para 40 °C ou 100 °C. O viscosímetro é composto por um tubo onde o lubrificante é aquecido até a temperatura de teste e então escorre por gravidade até outro recipiente e então é medido o tempo desde o início até o fim do escoamento. O valor é medido em cSt (centiskote) (BARRETT, 2007?).

## Espectrometria

Para identificação dos elementos que compõem o óleo lubrificante são utilizados, em geral, dois tipos de espectrômetros:

Um é o *espectrômetro de emissão atômica por plasma induzido*, que permite determinar a concentração de cada um dos elementos da amostra. Dessa forma, é possível obter valores quantitativos da presença de metais de desgaste, contaminantes e aditivos no lubrificante (BARRETT, 2007?).

O outro é o *espectrômetro por infravermelho*, que mede o espectro da passagem de luz infravermelha pela amostra. O valor da absorvância<sup>1</sup> encontrada para os diferentes comprimentos de onda determina quais os compostos encontrados na amostra. Esse equipamento permite quantificar a fuligem, a oxidação, a nitração, a sulfatação, a água e o glicol presentes no lubrificante (BARRETT, 2007?; GARRY, 2007).

## Acidez e Basicidade

O *teste de acidez* é realizado através da medição da quantidade necessária de uma base, em geral hidróxido de potássio, para neutralizar um grama de óleo. Já o teste de basicidade segue o mesmo princípio, apenas substituindo a base por um ácido, em geral ácido clorídrico (Dep. Técnico da Lubrikol Lubrificantes, 2013).

---

<sup>1</sup>Também chamada de absorvância ou absorvência, é a capacidade intrínseca dos materiais em absorver radiações em frequência específica. Usualmente, tal propriedade é empregada na análise de soluções em química analítica.

### Presença de Água

Existem alguns testes para verificar a presença de água no motor. Dentre eles, destacam-se dois: um para análise qualitativa e outro quantitativa. O primeiro é o *teste por crepitação*, que é feito com uma chapa aquecida, onde a amostra do lubrificante é despejada e caso ocorra a crepitação (som da água em ebulição), é indicada a presença de água. Já o segundo é feito através da *destilação da água* presente no lubrificante.

### Presença de Combustível

Pode-se detectar tanto a presença quanto a quantidade de combustível presente a partir do vapor gerado através do aquecimento do lubrificante.

Um método qualitativo chamado de *ponto de fulgor*, consiste em aquecer o óleo, monitorando sua temperatura, com uma chama posta em contato com os vapores saídos do óleo. No momento em que ocorrer a queima, é marcada a temperatura. Os lubrificantes saem de fábrica com uma especificação para este teste, caso o ponto de fulgor seja menor do que o original, é possível que tenha havido contaminação por combustível (Dep. Técnico da Lubrikol Lubrificantes, 2013).

Para se obter a quantidade de combustível, os vapores da amostra aquecida são passados por um sensor que absorve os hidrocarbonetos presentes no combustível, indicando o percentual de combustível presente (BARRETT, 2007?).

## 2.4 Propriedades do Lubrificante

As seguintes propriedades observadas no lubrificantes mudam a maneira do óleo atuar no sistema.

## Viscosidade

A propriedade mais importante dos lubrificantes é a viscosidade e é fundamental que seja mantida dentro das especificações para o bom funcionamento do motor.

Os efeitos da mudança na viscosidade são o aumento dos custos operacionais, superaquecimento do motor e aumento no desgaste. No caso de um óleo com menor viscosidade que o ideal, pode haver o rompimento do filme lubrificante, causando atrito entre as superfícies, o que leva ao superaquecimento e ao desgaste prematuro. No caso de aumento da viscosidade, o óleo atua resistindo ao movimento, o que causa maior necessidade de força para exercer o movimento, consequentemente maior estresse sobre as peças e ainda há a possibilidade de problemas no filtro de óleo e formação de depósitos (ConocoPhillips Lubrificants, 2007).

As causas da diminuição de viscosidade podem indicar contaminação por água, combustível ou adição de lubrificante de menor viscosidade. Seu aumento pode indicar a presença de contaminantes sólidos, a oxidação do lubrificante, combustão incompleta (problemas na mistura ar/combustível), temperatura de operação alta ou adição de lubrificante com maior viscosidade (BARRETT, 2007?; ConocoPhillips Lubrificants, 2007).

## Acidez e Basicidade Total

O valor da acidez total (TAN) é menor em um óleo virgem, e aumenta de acordo com seu uso . Por esse motivo a acidez é usada como um indicativo de degradação do óleo lubrificante. A formação de ácidos carboxílicos é um subproduto da oxidação. Por esse motivo, os lubrificantes saem da fábrica com uma reserva alcalina, que serve para combater a formação de ácidos e é medida pela basicidade total (TBN) (BARRETT, 2007?; Dep. Técnico da Lubrikol Lubrificantes, 2013).

## **Partículas Sólidas**

Os elementos químicos encontrados na análise do lubrificante, são em grande parte, metais usados na composição do motor e alguns outros elementos usados como aditivos. Por esse motivo estes são encontrados, em grande parte, em estado sólido.

A presença de partículas sólidas dificulta a lubrificação, cria depósitos, aumenta o desgaste e causa diminuição na vida útil do motor (ConocoPhillips Lubricants, 2007). Na Tabela 1 é apresentada uma lista dos elementos mais comuns e suas possíveis fontes em máquinas de maneira geral.

## **Oxidação**

Quando em altas temperaturas, o óleo e outros componentes expostos ao oxigênio oxidam e formam uma variedade de subprodutos prejudiciais. Por isso, a maioria dos lubrificantes contém aditivos para inibir ou retardar o processo de oxidação. Uma das causas da oxidação é o superaquecimento e a demora na troca do óleo lubrificante. A oxidação causa aumento na viscosidade e aumento na corrosão das partes de metal (BARRETT, 2007?; ConocoPhillips Lubricants, 2007).

## **Nitração**

Resultado da reação dos componentes do óleo com óxidos de nitrogênio (BARRETT, 2007?). O produto dessa reação é altamente ácido, cria depósitos e acelera a oxidação do óleo lubrificante. Como resultado, causa aumento no desgaste de cilindros e no comando de válvulas (ConocoPhillips Lubricants, 2007).

## **Sulfatação**

Vários óxidos de enxofre e água reagem e levam a formação de ácidos, como o ácido sulfúrico. Motores à Diesel apresentam esse problema muitas vezes pela presença de enxofre no combustível (BARRETT, 2007?).

## **Fuligem**

Formada a partir da queima incompleta de combustível, a fuligem é um bom indicador da eficiência de combustão de um motor (ConocoPhillips Lubrificants, 2007; Dep. Técnico da Lubrikol Lubrificantes, 2013). Sua presença, que acaba sendo depositada no lubrificante, exaure os aditivos e pode entupir o filtro de óleo (Dep. Técnico da Lubrikol Lubrificantes, 2013).

A presença de fuligem está associada ao uso de combustível de baixa qualidade, a mistura ar/combustível incorreta, a problemas nos bicos injetores e a baixa compressão, que pode ser causada por pistões e anéis desgastados. Uma vez misturado ao lubrificante causa o aumento da viscosidade (ConocoPhillips Lubrificants, 2007).

## **Glicol**

É um tipo de composto orgânico usado como anticongelante no sistema de arrefecimento. Sua presença no lubrificante indica contaminação por vazamento do sistema de refrigeração. A presença de glicol interfere na quantificação de água na espectrometria por infravermelho. Entretanto sua presença, corroborada pela presença de sódio e boro, tem maior impacto no desempenho do motor (GARRY, 2007).

## **Diluição por Água**

Os lubrificantes, de maneira geral, não possuem água em sua composição e sua presença pode acarretar em problemas para o funcionamento do motor. A água geralmente é evaporada quando os motores operam em temperatura normal, porém quando a temperatura de operação é baixa, tanto a condensação da umidade atmosférica quanto a água formada na queima do combustível pode contaminar o lubrificante. A presença de água pode indicar contaminação por fontes externas, como o sistema de refrigeração (ConocoPhillips Lubrificants, 2007; SENAI-ES, 1997).

A presença de água no motor causa aumento da viscosidade do lubrificante, com consequente diminuição na lubrificação, formação de ácidos (sendo um dos principais componentes do ácido sulfúrico) e corrosão (ConocoPhillips Lubricants, 2007; SENAI-ES, 1997).

### **Diluição por Combustível**

As causas da diluição por combustível são períodos estendidos em ponto morto, condição de uso intermitente (anda e para no trânsito, por exemplo), e outras relacionadas ao ajuste do motor, como mistura ar/combustível incorreta, *timing* incorreto do motor. Em geral, quando essa medida ultrapassa 2,5 a 5% devem ser tomadas ações corretivas (ConocoPhillips Lubricants, 2007; SENAI-ES, 1997).

A diluição do lubrificante por combustível não queimado reduz a eficácia do lubrificante através da redução da viscosidade, risco de fogo/explosão, aumento no gasto de combustível e diminuição no desempenho do motor (ConocoPhillips Lubricants, 2007; SENAI-ES, 1997).

Tabela 2.1: Elementos encontrados na análise espectrométrica e possíveis fontes

| <b>Elemento</b> | <b>Símbolo</b> | <b>Fonte possível</b>                                               |
|-----------------|----------------|---------------------------------------------------------------------|
| Ferro           | Fe             | Eixos, engrenagens, bloco, anéis de pistão, camisa do cilindro      |
| Cobre           | Cu             | Rolamentos, buchas, arruelas de pressão                             |
| Chumbo          | Pb             | Rolamentos                                                          |
| Estanho         | Sn             | Ligas de rolamentos, gaiolas de rolamento, Solda                    |
| Alumínio        | Al             | Bombas, arruelas de pressão, pistões, turbo-compressores            |
| Crômio          | Cr             | Rolamentos, anéis de pistão, camisa de cilindro                     |
| Níquel          | Ni             | Bombas, válvulas                                                    |
| Prata           | Ag             | Rolamentos                                                          |
| Magnésio        | Mg             | Detergente aditivo, aditivo de refrigeração                         |
| Silício         | Si             | Sujeira, aditivo antiespumante                                      |
| Boro            | B              | Aditivo anti-corrosão do sistema de refrigeração                    |
| Sódio           | Na             | Detergente aditivo, vazamento de líquido do sistema de refrigeração |
| Bário           | Ba             | Antioxidante e inibidores de corrosão                               |
| Cálcio          | Ca             | Aditivo detergente                                                  |
| Fósforo         | P              | Aditivo anti-desgaste e de pressão extrema                          |
| Potássio        | K              | Vazamento de líquido do sistema refrigeração                        |
| Molibdênio      | Mo             | Anéis e aditivo de pressão extrema                                  |
| Zinco           | Zn             | Aditivo anti-desgaste                                               |
| Vanádio         | V              | Lâminas de turbina                                                  |

Fonte: (BARRETT, 2007?; ConocoPhillips Lubricants, 2007)





### 3 LÓGICA NEBULOSA

Para diversos problemas, existem duas formas de tratamento: Usando conhecimento objetivo, representado por métodos quantitativos e precisos, como os que são usados o tempo todo na modelagem matemática; ou conhecimento subjetivo, que representa a informação linguística que é impossível de quantificar usando métodos matemáticos tradicionais (MENDEL, 2001).

Os métodos quantitativos tiveram grande êxito na resolução de problemas devido à sua precisão e rigor, que são bem úteis ao método científico. A criação dos computadores ainda resultou em uma rápida expansão do uso de métodos quantitativos.

No entanto, os computadores não são completamente capazes de realizar julgamentos como um ser humano, o que dificulta seu uso para áreas ligadas às ciências humanas. Tal fato é uma manifestação do princípio da incompatibilidade (ZADEH, 1975). Esse princípio diz que a alta complexidade é incompatível com precisão absoluta.

Uma forma de representar o conhecimento subjetivo é através dos conjuntos nebulosos, que buscam modelar a incerteza contida nas palavras que definem um determinado grupo. A lógica nebulosa (lógica difusa ou lógica *fuzzy*) é uma forma de lógica multi-valorada que lida com valores imprecisos, ao contrário do que ocorre com a lógica booleana que opera sobre conjuntos binários aceitando apenas os valores *verdadeiro* ou *falso*.

## 3.1 Conjuntos Nebulosos

Os conjuntos nebulosos foram propostos por ZADEH (1965). Estes se diferenciam dos conjuntos tradicionais, pois consideram valores intermediários entre o 1 (pertence) e o 0 (não-pertence). Esses conjuntos admitem que todos os elementos tenham um grau de pertinência em todos os conjuntos, ao contrário dos conjuntos tradicionais em que um elemento apenas "pertence" ou "não-pertence" a um determinado conjunto.

O grau de pertinência de um elemento em um conjunto é dado pela função de pertinência do conjunto, que é a representação da magnitude da pertinência de um elemento a um conjunto.

### 3.1.1 Sistema de Inferência Nebuloso

Um sistema de inferência nebuloso recebe um conjunto de entradas *crisp*, ou seja, números precisos e transforma em variáveis nebulosas, processa esses valores no motor de inferência através de lógica nebulosa, e então, calcula a saída como um valor *crisp*. O motor de inferência funciona processando as variáveis de entrada, através da aplicação dos operadores de agregação (*E*, *OU* e *NÃO*), de acordo com as instruções dos precedentes na base de regras. Quando esse processo resulta na ativação do consequente de alguma das regras, são feitos cálculos dos consequentes para composição da saída.

As variáveis nebulosas são a representação das variáveis *crisp* no sistema de inferência nebuloso, e têm uma série de conceitos associados a elas:

### O universo de discurso (U)

Conjunto de todos os valores *crisp* possíveis de uma variável nebulosa. Dentro do universo de discurso estarão definidos todos os conjuntos nebulosos (F) para uma variável;

### Grau de inclusão

Valor entre  $[0, 1]$  que representa a pertinência de um elemento em F;

### Função de pertinência ( $\mu$ )

Função que transforma as entradas *crisp* em graus de inclusão para um conjunto nebuloso. Para cada conjunto é associada uma função de pertinência  $\mu_F$ ;

### Suporte

Subconjunto de U, onde o valor de  $\mu_F$  é diferente de zero;

### Termos linguísticos

Nomes utilizados para descrever os conjuntos. Ex.: Para a variável altura, podemos ter os valores "baixo", "mediano" e "alto".

As funções de pertinência servem para modelar os conjuntos nebulosos e podem assumir várias formas. Em geral as mais usadas são as formas triangular, trapezoidal, gaussiana e sigmoide.

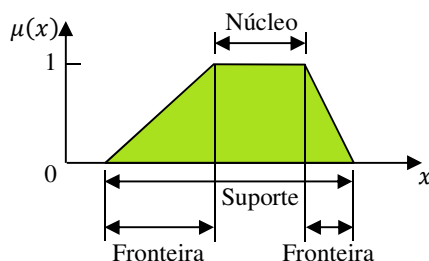


Figura 3.1: Elementos de uma função de pertinência.

A Figura 3.1 ilustra as regiões de suporte, núcleo e fronteiras de uma função de pertinência trapezoidal para um conjunto nebuloso  $A$  (ROSS, 2010).

$$A = \{(x, \mu_A(x)) \mid \forall x \in X, 0 \leq \mu_A(x) \leq 1\}$$

O suporte de uma função de pertinência é definido como a região do universo de discurso caracterizada por possuir pertinência maior que zero. Sendo assim, o suporte compreende todos os elementos de  $x$  no universo tal que  $\mu_A(x) > 0$ .

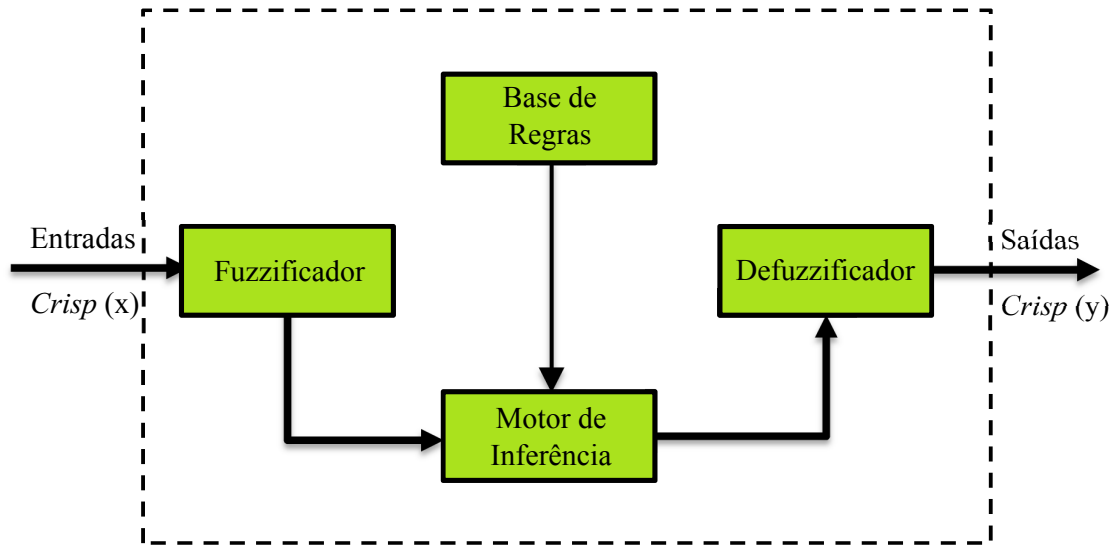


Figura 3.2: Esquema do Sistema de Inferência Nebuloso.

O núcleo de uma função de pertinência é definido como a região do universo de discurso de total pertinência no conjunto  $A$ . Sendo assim, o núcleo compreende todos os elementos de  $x$  no universo tal que  $\mu_A(x) = 1$ .

As fronteiras de uma função de pertinência são aquelas regiões do universo de discurso caracterizadas por possuírem pertinência maior que zero, no entanto não completa. Sendo assim, as fronteiras compreendem todos os elementos de  $x$  no universo tal que  $0 < \mu_A(x) < 1$ .

As partes do sistema visto na Figura 3.2 funcionam da seguinte maneira:

#### *3.1.1.1 Fuzzificação*

Calcula o grau de inclusão dos valores da entrada (*crisp*) para os conjuntos nebulosos das variáveis de entrada, através das funções de pertinência.

#### *3.1.1.2 Base de Regras*

São um conjunto de declarações constituídas de precedentes (se) e consequentes (então), ligando as variáveis de entrada às de saída por declarações do tipo "se-então".

#### *3.1.1.3 Motor de Inferência*

Componente responsável pela aplicação das regras. Funciona através da aplicação das variáveis de entrada nas regras. Então é feita a verificação das regras ativadas e depois a combinação das regras ativadas. Uma regra é ativada quando referencia pelo menos um conjunto nebuloso que possua valores dentro do seu suporte.

#### 3.1.1.4 Defuzzificação

Realiza processo inverso ao da *fuzzificação*, calculando o valor da saída a partir do grau de pertinência dos conjuntos de saída, transformando valores nebulosos para valores *crisp*.

## 3.2 Conjuntos Nebulosos Tipo-2

Frequentemente o conhecimento usado para construir sistemas nebulosos é incerto. Essa incerteza leva a regras que possuem funções de pertinências nos antecedentes e consequentes que sejam também incertas (CASTILLO; MELIN, 2008). Os conjuntos nebulosos tipo-2, a partir de agora chamados T2FS são uma generalização dos conjuntos nebulosos tradicionais, a partir de agora chamados de conjuntos nebulosos tipo-1 (T1FS) para distinção. Os T2FSs foram propostos por ZADEH (1975) e usam como funções de pertinência outros conjuntos nebulosos ao invés de um conjunto de valores *crisp* como nos T1FS.

Estendendo os conceitos de lógica nebulosa tipo-2 MENDEL; KARNIK (1998) propuseram um sistema de lógica nebuloso tipo-2. A lógica nebulosa tipo-2 é usada para minimizar os efeitos das incertezas que não são tratadas pela lógica nebulosa tipo-1.

Existem ao menos quatro fontes de incertezas que não são tratadas pelo sistema de inferência nebuloso tipo-1 (MENDEL; JOHN, 2002; CASTILLO; MELIN, 2008):

- Os termos linguísticos podem representar coisas diferentes para pessoas distintas, logo pode existir ambiguidade na definição dos conjuntos nebulosos;
- Dada uma mesma amostra, os especialistas podem ter opiniões distintas e portanto podem classificar de forma diferente, levando a uma indecisão no conhecimento usado para construção das regras;
- Pode haver a inexatidão nas medidas que servem de dados de entrada;
- Os dados usados para o ajuste das funções de pertinência podem ser tendenciosos, levando à imprecisão na escolha das funções de pertinência.

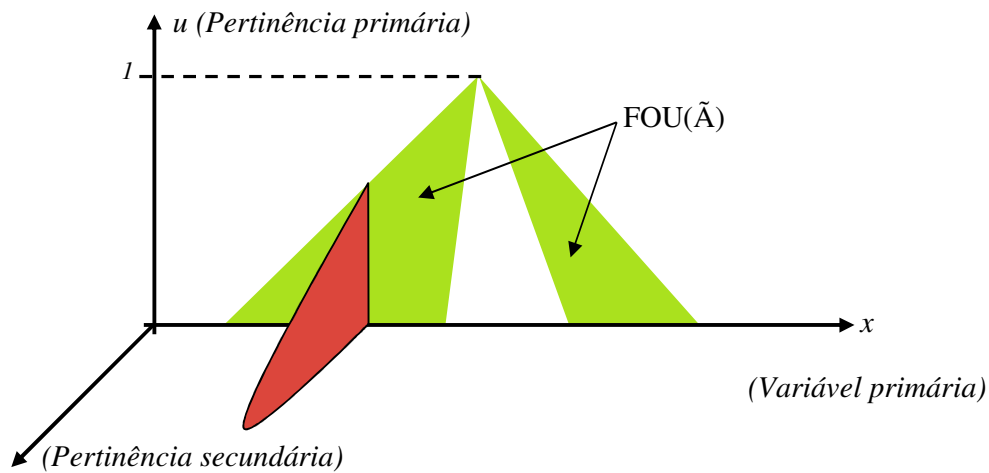


Figura 3.3: Função de pertinência de um T2FS.

Nos T1FSs o grau de pertinência é dado por um valor *crisp*, enquanto nos T2FSs, o valor da função de pertinência é um conjunto nebuloso tipo-1. A função de pertinência de um T2FS  $\tilde{A}$  (Figura 3.3) é definida por uma função  $\mu_{\tilde{A}}(x, u)$ , onde  $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$  é o valor da função de pertinência secundária de  $\tilde{A}$ ,  $u \in J_x \subseteq [0, 1]$  é o valor da função de pertinência primária e  $x \in X$  é a variável primária (MENDEL, 2001).



$$\tilde{A} = \{((x, u), \mu_{\tilde{A}}(x, u)) \mid \forall x \in X, 0 \leq \mu_{\tilde{A}}(x, u) \leq 1, \forall u \in J_x \subseteq [0, 1]\}$$

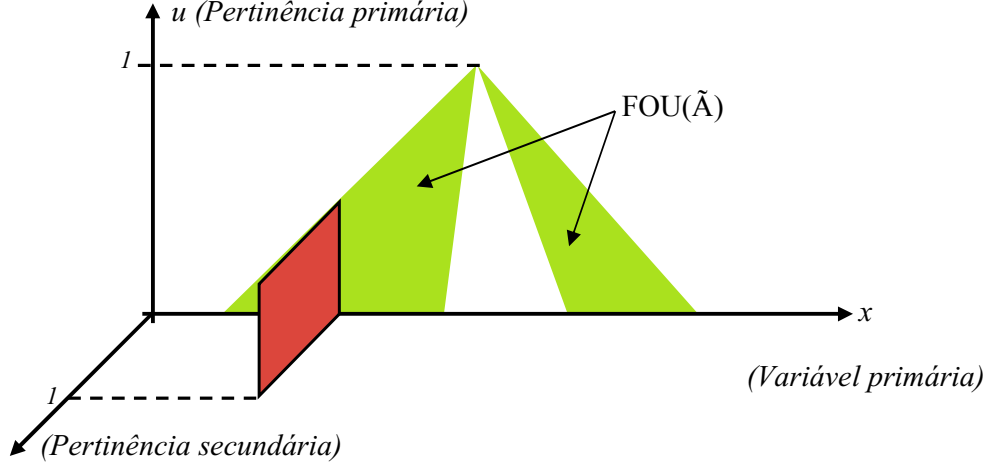


Figura 3.4: Função de pertinência de um IT2FS.

A função de pertinência secundária pode assumir qualquer forma, como triangular, trapezoidal, gaussiana ou um valor unitário. Quando a função de pertinência secundária assume apenas valores unitários ( $0$  ou  $1$ ),  $u_{\tilde{A}}(x, u) = \{0, 1\}$ , a incerteza é uniforme ao longo de todo o intervalo.

$$\tilde{A} = \{((x, u), \mu_{\tilde{A}}(x, u)) \mid \forall x \in X, \mu_{\tilde{A}}(x, u) = \{0, 1\}, \forall u \in J_x \subseteq [0, 1]\}$$

Esse caso especial permite diminuir a complexidade computacional pois para representar um conjunto nebuloso tipo-2 com valores unitários na função de pertinência secundária não é necessário uma função com 3 dimensões, bastando observar os intervalos para cada valor  $x$  da entrada, o que pode ser representado por uma função superior e uma inferior. Por isso, este caso especial de conjunto nebuloso tipo-2 é definido como conjunto nebuloso tipo-2 intervalar, a partir de agora chamado de IT2FS (KARNIK; MENDEL; LIANG, 1999).

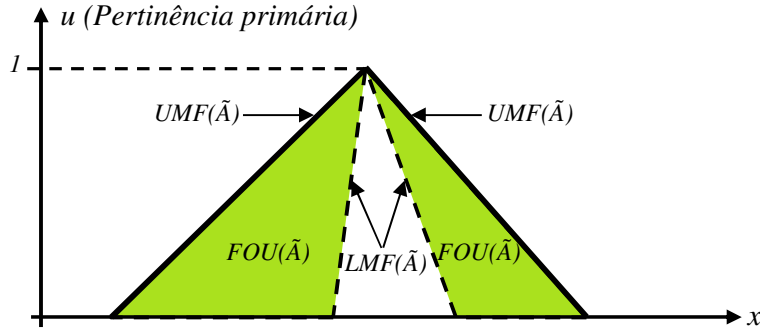


Figura 3.5: Funções de pertinência superiores e inferiores de um IT2FS.

A região marcada com domínio em  $x$  e contradomínio em  $u$  nas Figuras 3.3 e 3.4 é a mancha de incerteza (FOU, ou *Footprint of Uncertainty*). Esta região representa todos os pontos no plano  $(x, u)$  para o qual existe um valor de  $\mu_{\tilde{A}}(x, u)$ .

Pode-se definir uma função superior e uma inferior à FOU como visto na Figura 3.5. A função superior é uma função de pertinência do tipo-1 e pode ser definida como:

$$\bar{\mu}_{\tilde{A}}(x) \equiv UMF(\tilde{A}) \quad \forall x \in X$$

A função inferior também é uma função de pertinência do tipo-1 e pode ser definida como:

$$\underline{\mu}_{\tilde{A}}(x) \equiv LMF(\tilde{A}) \quad \forall x \in X$$

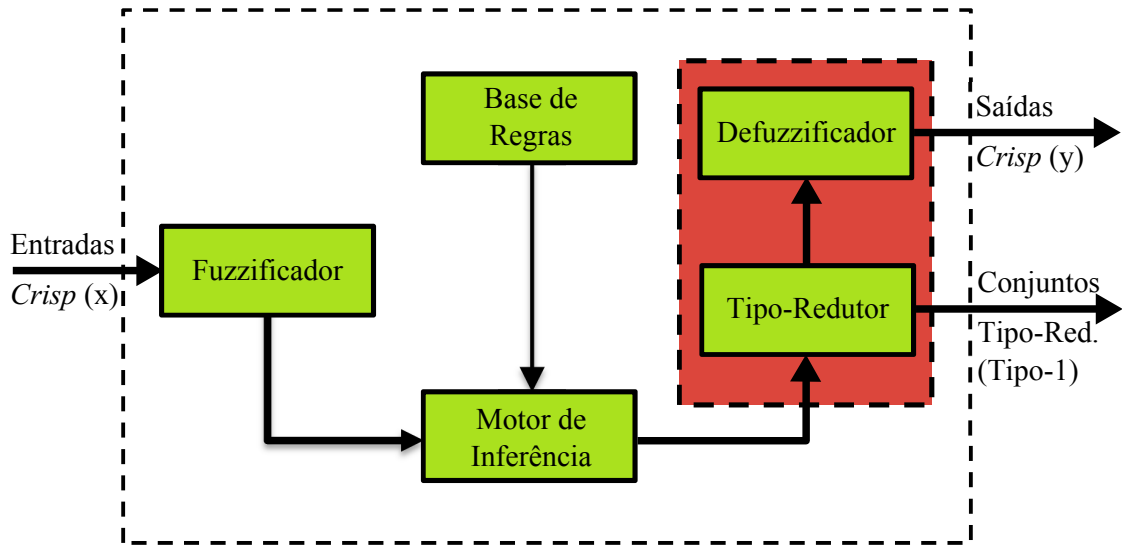


Figura 3.6: Esquema do sistema de inferência nebuloso Tipo-2.

### 3.2.1 Sistema de Inferência Nebuloso Tipo-2

Um sistema de inferência nebuloso do tipo-2 é bastante similar ao tipo-1, no entanto utiliza T2FS para modelar as variáveis de entrada e de saída. Por este motivo, o processamento da saída deixa de ser apenas a defuzzificação, que resulta em um valor *crisp* (tipo-0), ganhando um passo anterior, chamado de tipo-redução, que resulta em valores tipo-1. A seguir será explicada a forma de funcionamento dos elementos que compõem um sistema de inferência nebuloso tipo-2, como visto na Figura 3.6 CASTILLO; MELIN (2008).

#### 3.2.1.1 Fuzzificação

De forma similar ao que ocorre na fuzzificação dos sistemas tipo-1, a fuzzificação nos sistemas tipo-2 serve para mapear uma entrada *crisp* em uma variável nebulosa. No entanto, devido à característica dos T2FS a função de pertinência é

mais complexa (como visto acima). No caso dos T1FS a fuzzificação resulta em um conjunto de valores *crisp*, ou seja cada valor de  $x$  tem um valor *crisp* associado. Nos IT2FS resulta em um conjunto de intervalos, e nos T2FS resulta em um conjunto de T1FSs.

### 3.2.1.2 Base de Regras

O conjunto de regras é exatamente igual ao visto na seção 3.1.1, pois a diferença entre os sistemas de inferência nebulosos tipos 1 e 2 é o tipo de suas funções de pertinência CASTILLO; MELIN (2008).

### 3.2.1.3 Motor de Inferência

Com base nas regras contidas na base de regras, o processo de inferência é iniciado pelo cálculo dos antecedentes. Para as operações de conjunção e implicação podem ser usados os operadores mínimo ou produto. Nos sistemas baseados em IT2FS são feitos apenas dois cálculos por conjunto nebuloso (um para o valor da função inferior e outro para o valor função superior), ao contrário dos baseados em T2FS em que são calculados os antecedentes para cada um dos pontos do plano  $(x, u)$ .

Desta forma, o cálculo dos antecedentes é feito apenas para os valores das funções superiores e inferiores das funções de pertinência, como visto na Figura 3.7, onde o operador mínimo foi aplicado a às funções superiores e inferiores de  $x_1$  e  $x_2$  em  $\tilde{A}$ ,  $\tilde{B}$ ,  $\tilde{C}$ ,  $\tilde{D}$ .

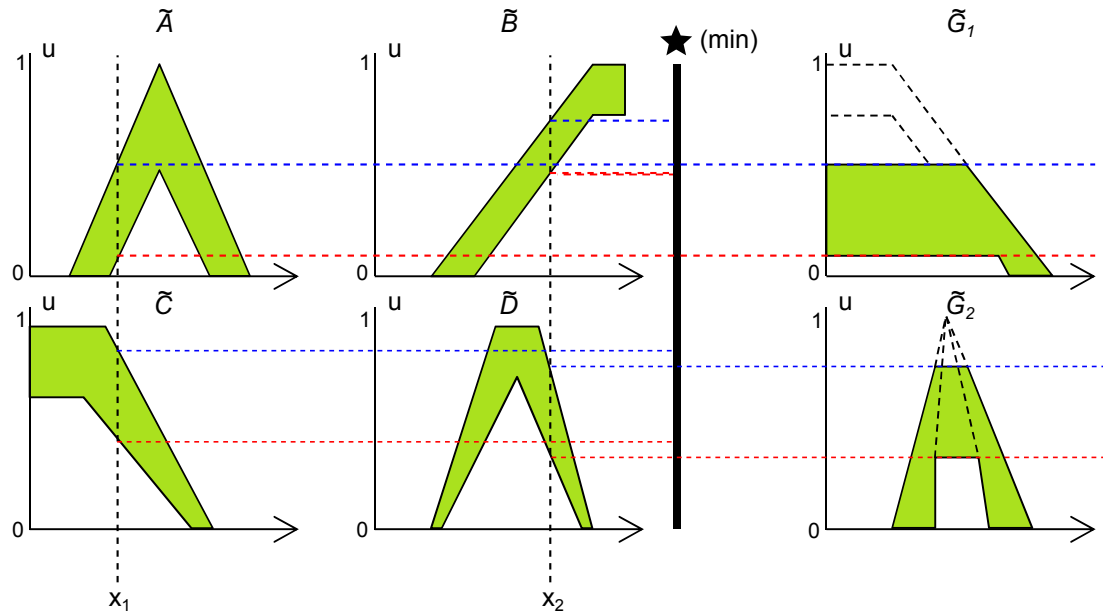


Figura 3.7: Cálculo dos consequentes no sistema de inferência nebuloso tipo-2 intervalar.

De forma similar acontece a agregação dos consequentes. Para a operação de agregação podem ser usados os operadores máximo ou soma. Como visto na Figura 3.8, onde o operador máximo foi aplicado às funções superiores do conjunto  $\tilde{G}_1$  e  $\tilde{G}_2$ .

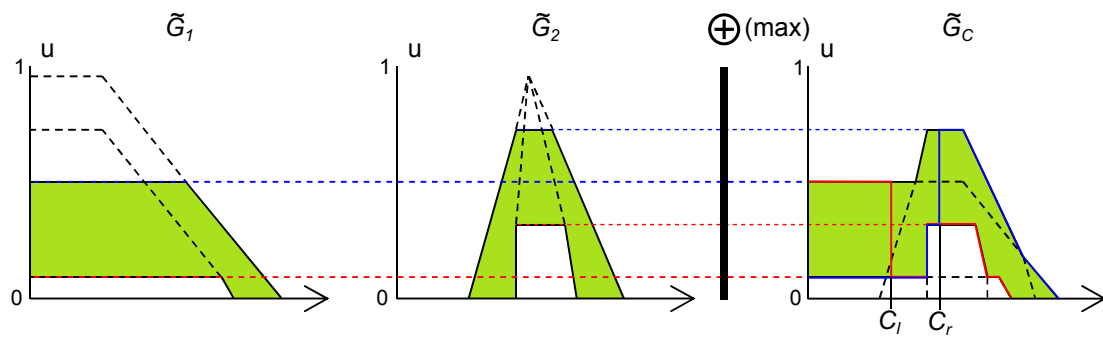


Figura 3.8: Combinação dos consequentes no sistema de inferência nebuloso tipo-2 intervalar.

### 3.2.1.4 Tipo-Redução

Os conjuntos de saída de um sistema de lógica nebulosa tipo-1 são T1FS, os conjuntos de saída de um sistema de lógica nebulosa tipo-2 são T2FS. Na defuzzificação de um sistema de inferência nebuloso tipo-1 a saída é valor *crisp*. Em um sistema de lógica nebulosa tipo-2 a operação de defuzzificação estendida no caso do tipo-2 tem como saída um T1FS. Essa operação que mapeia o conjunto de saída de um T2FS para um T1FS é chamada tipo-redução, e o conjunto resultante desse processo é chamado de conjunto tipo-reduzido. Para obter um valor *crisp* é necessário que esse conjunto tipo-reduzido seja defuzzificado. Entretanto, em algumas aplicações, o conjunto tipo-reduzido pode ser mais importante do que o valor *crisp* resultante da defuzzificação, visto que embora muitas vezes seja uma operação necessária, toda redução representa perda de informação CASTILLO; MELIN (2008).

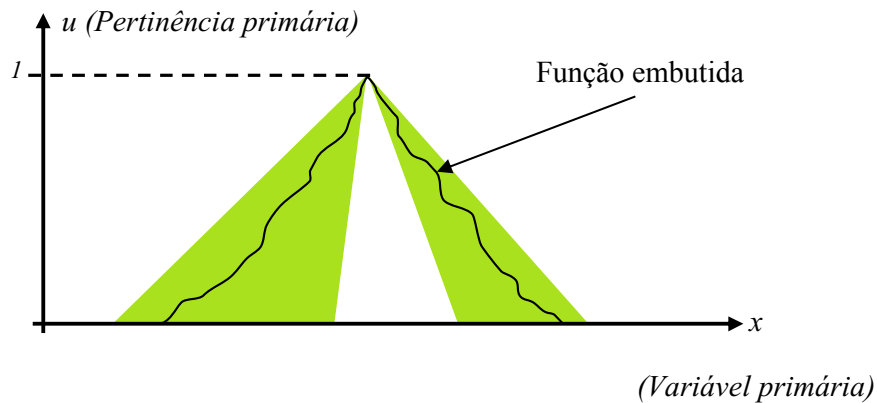


Figura 3.9: Função de pertinência de T1FS embutido em um T2FS.

Um conceito importante para compreender o processo de tipo-redução é o dos conjuntos tipo-1 embutidos (*embedded type-1 sets*). Em um universo contínuo, para cada T2FS, existem incontáveis T1FSs. Esse conceito é representado na Figura 3.9, onde é possível visualizar a representação gráfica da função de pertinência de um conjunto tipo-1 embutido.

Assim como existe mais de uma forma de defuzzificação para os sistemas nebulosos do tipo-1, existem ao menos cinco métodos de tipo-redução para os sistemas do tipo-2: Centróide, centro das somas, altura, altura modificada e centro dos conjuntos MENDEL (2001). Neste trabalho foi usado o método do centróide, através de um algoritmo bastante conhecido, chamado de Karnik-Mendel (KM) KARNIK; MENDEL (2001).

O algoritmo KM determina representa a discretização da variável primária de um IT2FS. O centróide de um IT2FS  $\tilde{A}$ ,  $c_{\tilde{A}} = [c_l, c_r]$ , pode ser calculado como a solução ótima para as seguintes funções:

$$c_l = \min_{\forall \theta_i \in [\underline{\mu}_{\tilde{A}}(x_i), \bar{\mu}_{\tilde{A}}(x_i)]} = \frac{\sum_{i=1}^N x_i \theta_i}{\sum_{i=1}^N \theta_i} = \frac{\sum_{i=1}^{k_l} x_i \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=k_l+1}^N x_i \underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{k_l} \bar{\mu}_{\tilde{A}}(x_i) + \sum_{i=k_l+1}^N \underline{\mu}_{\tilde{A}}(x_i)}$$

$$c_r = \min_{\forall \theta_i \in [\underline{\mu}_{\tilde{A}}(x_i), \bar{\mu}_{\tilde{A}}(x_i)]} = \frac{\sum_{i=1}^N x_i \theta_i}{\sum_{i=1}^N \theta_i} = \frac{\sum_{i=1}^{k_r} x_i \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=k_r+1}^N x_i \bar{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^{k_r} \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=k_r+1}^N \bar{\mu}_{\tilde{A}}(x_i)}$$

onde  $k_l$  e  $k_r$  são chamados de pontos de troca, com  $x_{k_l} \leq c_l \leq x_{k_l+1}$  e  $x_{k_r} \leq c_r \leq x_{k_r+1}$ .

O algoritmo usado para tipo-redução resulta em dois conjuntos nebulosos tipo-1 embutidos, como vistos na Figura 3.10. O algoritmo usado é iterativo e foi escolhido pois costuma convergir em um número pequeno de passos. Ainda assim, o tipo-redutor é uma das fases mais custosas computacionalmente de todo o processo do sistema de inferência nebuloso tipo-2 MENDEL (2007).

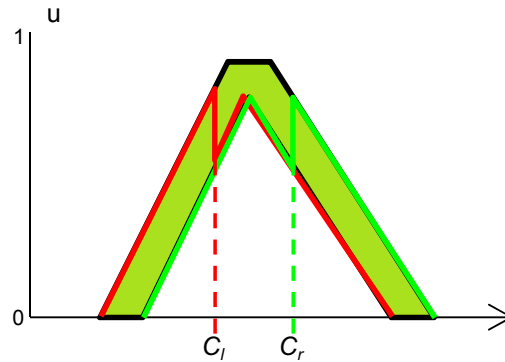


Figura 3.10: Tipo-redução de um IT2FS pelo método do centroide.

#### 3.2.1.5 Defuzzificação

Assim como a defuzzificação do sistema de inferência nebuloso tipo-1, a defuzzificação do sistema de inferência nebuloso tipo-2 transforma a saída do motor de inferência em um valor *crisp* (tipo-0). E assim como no sistema de inferência nebuloso tipo-1, o passo da defuzzificação recebe um T1FS, sendo que no sistema de inferência tipo-1 este vem do motor de inferência e nos tipo-2, vem do tipo-redutor.

Seguindo com a explicação do processamento da saída de um sistema de inferência nebuloso tipo-2 intervalar pelo método do centroide, o *defuzzificador*, obtendo os dois centroides do tipo-redutor, pode transformar esse conjunto de valores por média simples ou ponderada, obtendo como saída um valor *crisp*.





## 4 MODELO PROPOSTO

O sistema proposto usa um método de aprendizado supervisionado para definição do modelo. O aprendizado supervisionado é uma ferramenta para aquisição automática de conhecimento baseada em aprendizado indutivo.

Indução é a forma de inferência lógica que permite obter conclusões genéricas a partir de um conjunto particular de exemplos. O aprendizado indutivo se dá através da realização de inferências indutivas sobre os exemplos apresentados. Esta forma de aprendizagem se origina em um conceito específico para então generalizá-lo, partindo da parte para conhecer o todo. Não existe garantia de que as hipóteses geradas através da inferência sejam verdadeiras (REZENDE, 2003; BATISTA, 2003).

Para o aprendizado supervisionado é usado um conjunto de exemplos para treinamento do sistema. O sistema deve extrair padrões úteis ao reconhecimento de casos similares. Em geral, cada exemplo traz um conjunto de valores que representa uma ou mais classes associadas. O objetivo dessa técnica é extrair conhecimento suficiente para construção de um classificador capaz de indicar corretamente a classe de exemplos não rotulados.

Após o treinamento do sistema, geralmente é executado um processo de avaliação do classificador. Para tal, realiza-se uma série de testes no qual os parâmetros do sistema e os atributos usados para realizar a classificação podem ser ajustados (REZENDE, 2003).

Existem vários algoritmos de aprendizado supervisionado e deve ser ressaltado que não existe um algoritmo que apresente o melhor desempenho para todos os casos (REZENDE, 2003).

Decidiu-se usar um sistema neuro-*fuzzy* devido a algumas características interessantes. Ele possui características similares às funções de base radial (REZENDE, 2003), que são bastante usadas para reconhecimento de padrões (BROOMHEAD; LOWE, 1988). Outra característica desejável é a compreensibilidade do modelo gerado, visto que o conjunto de regras gerado pelo sistema é passível de ser compreendido por humanos (REZENDE, 2003).

O modelo foi criado usando um sistema neuro-*fuzzy* chamado ANFIS (JANG, 1993), usando a implementação para MATLAB através da função `anfis`. O ANFIS é um algoritmo híbrido de aprendizado, que identifica as funções de pertinência da entrada e da saída de um sistema de inferência nebuloso através da aplicação sucessiva da combinação dos métodos de mínimos quadrados e gradiente descendente. Essas aplicações sucessivas são chamadas de épocas de treinamento e são necessárias para ajustar o modelo.

Nesse sistema os consequentes das regras são funções das variáveis de entrada, diferentemente dos sistemas Mamdani em que são conjuntos nebulosos (ROSS, 2010). O ANFIS é um sistema em que as regras são criadas automaticamente, criando uma saída para cada combinação possível de conjuntos de todas as variáveis de entrada. Assim,  $k$  regras são criadas, onde  $k = \prod_{i=1}^N n_i$ , sendo  $n_i$  o número de conjuntos nebulosos para a  $i$ -ésima entrada.

Portanto, o número de regras cresce exponencialmente com a quantidade de variáveis de entrada e com o número de funções de pertinência por cada variável de entrada. Esse crescimento leva a um aumento do tempo necessário para treinamento

e também para execução do sistema. Por limitações do *hardware* usado nos testes, decidiu-se limitar o número de variáveis de entrada a 4 e a um máximo de 3 funções de pertinência por variável de entrada, o que resultou em no máximo 81 regras.

Ainda, esse fenômeno conhecido como maldição da dimensionalidade (*curse of dimensionality*), pode levar ao sobreajuste (*overfitting*), que é quando o modelo se ajusta demais aos dados de treino, perdendo a capacidade de generalização, imprescindível aos sistemas de classificação.

Para realizar a classificação usando lógica nebulosa tipo-2 intervalar, buscou-se implementações de um sistema de inferência nebuloso tipo-2 intervalar em diversas linguagens, mas como não foi possível obter, houve a necessidade de fazer sua implementação. Os detalhes dessa implementação serão discutidos mais a frente.

## 4.1 Descrição dos Dados

Nessa pesquisa foram usados dados fornecidos por uma empresa petrolífera, que pretendia desenvolver um sistema automatizado de classificação de falhas em motores através da avaliação do resultado da análise de óleos lubrificantes usados. O resultado da análise de óleo traz uma série de parâmetros que serão usados como entrada para o sistema de classificação e as classificações informadas por especialistas, que serão usadas como a saída esperada do sistema.

Os testes aplicados ao óleo lubrificante apresentados na seção 2.3 nem sempre apresentam resultados exatos ou precisos. Ainda, não há garantias de que a classificação realizada por especialistas humanos tenha seguido o mesmo critério para todas as amostras, dessa forma é possível que duas amostras semelhantes tenham re-

cebido classificações distintas. Por esses motivos, é necessário considerar a incerteza presente nesses dados.

Os dados utilizados nesta pesquisa consistem em resultados de análises de óleo lubrificante provenientes de motores de combustão interna à diesel da família OM-366, juntamente com o diagnóstico dado por especialistas para 3 problemas distintos: corrosão, combustão e contaminação.

Cada amostra contém 29 variáveis contendo informações sobre o equipamento e informações obtidas na análise do óleo lubrificante, que mostram a quantidade de elementos e compostos químicos encontrados (Quilometragem até a data do teste, Quilometragem desde a última troca, Fe, Cr, Pb, Cu, Sn, Al, Ni, Si, Na, Mg, Ca, Ba, P, Zn, Mo, Ti, Ag, B, V, K, Oxidação, Nitração, Sulfatação, Fuligem, % de água, % de combustível, % de glicol) em valores contínuos, com algumas variáveis possuindo dados imprecisos, como os dados de Oxidação, Nitração, Sulfatação, Fuligem, % de água, % de combustível, % de glicol, que trazem informações incompletas, com valores aproximados, ex.:  $< 0.1$  ou  $> 10.0$ .

Ainda, cada amostra apresenta o diagnóstico para cada um dos três problemas. Cada problema é classificado numa escala de G (verde), Y (amarelo) a R (vermelho). Sendo G o diagnóstico de motor em bom estado, Y estado de atenção e R estado crítico. Foram usadas 724 amostras, das quais 641 são classificadas como G, 64 como Y e 19 como R para o problema de corrosão; 706 classificadas como G, 14 como Y e 4 como R para o problema de combustão; e 694 classificadas como G, 8 como Y e 22 como R para o problema de contaminação.

Esse desbalanço entre as classes apresenta um problema para a construção de um classificador, pois os casos que apresentam problemas são poucos, o que dificulta a detecção de padrões. O desbalanço entre as classes poderia ser minimizado com

a geração de dados artificiais. No entanto, para este problema não é uma tarefa trivial, visto que nenhuma das variáveis apresentou distribuição normal.

Dessa forma, com os dados existentes acredita-se ser possível a construção de um classificador seguro para os casos em que não existe falha. No entanto, por haver poucos dados classificados como Y ou R, é provável que este classificador não seja tão eficaz em detectar os casos de falha e muito menos em fazer distinção entre os níveis de falha.

Para minimizar esse problema, optou-se por realizar também testes com uma implementação do SMOTE<sup>1</sup>: *Synthetic Minority Over-sampling Technique* (CHAWLA et al., 2002), visto que esse algoritmo não pressupõe o uso de nenhum tipo de distribuição estatística para a geração dos dados artificiais.

## 4.2 Descrição do *Framework*

Devido à impossibilidade de encontrar uma implementação de um sistema de inferência nebuloso tipo-2 intervalar que permitisse o teste dos classificadores, foi desenvolvido um *framework* para aplicação dos conceitos de lógica nebulosa tipo-2 intervalar, com base no *Fuzzy Logic Toolbox* do MATLAB®. A fim de aproveitar funções já existentes e facilitar o uso por quem já tenha familiaridade com o *Toolbox* original, as funções criadas obedecem a nomenclatura semelhante e a estrutura usada para armazenar os parâmetros do sistema é bastante similar à original. As operações sobre os conjuntos nebulosos tipo-2 são implementações das operações definidas em MENDEL (2001).

---

<sup>1</sup>SMOTE é uma técnica que visa diminuir o desbalanço dos dados através da geração de amostras artificiais para as classes minoritárias, a partir de amostras das classes minoritárias e do descarte de amostras da classe majoritárias.

Na *fuzzificação* os valores da entrada são discretizados e os conjuntos nebulosos tipo-2 intervalares são definidos sobre esse espaço discreto. Os valores fuzzificados são representados por duas matrizes de 3 dimensões contendo os valores superiores e inferiores, que descrevem o intervalo, para cada função de pertinência em cada variável de entrada. As dimensões das matrizes são definidas pela quantidade de divisões do espaço de entrada, o número de variáveis de entrada e o número de funções de pertinência por variável de entrada.

Para as operações de conjunção e implicação foram definidas as funções mínimo e produto, e para agregação, as funções máximo e soma. Diferente desses operadores no tipo-1, aqui esses são aplicados ao intervalo e não somente a um ponto. Essa versão do *framework* não foi provida da função disjunção pois é possível decompor a regra original em um conjunto de regras equivalentes (MENDEL, 2001). O exemplo a seguir mostra como proceder caso seja necessário usar o conectivo "ou" nas regras:

$$\text{SE } (x_1 \text{ é } F_1 \text{ e } \dots \text{ e } x_n \text{ é } F_n) \text{ ou } (x_{n+1} \text{ é } F_{n+1} \text{ e } \dots \text{ e } x_p \text{ é } F_p) \text{ ENTÃO } y \text{ é } G$$

Essa regra pode ser expressa com as duas regras a seguir:

$$\text{SE } (x_1 \text{ é } F_1 \text{ e } \dots \text{ e } x_n \text{ é } F_n) \text{ ENTÃO } y \text{ é } G$$

$$\text{SE } (x_{n+1} \text{ é } F_{n+1} \text{ e } \dots \text{ e } x_p \text{ é } F_p) \text{ ENTÃO } y \text{ é } G$$

O motor de inferência mapeia os valores fuzzificados da entrada nos valores da variável da saída. Nessa implementação, apenas uma saída é permitida. Caso necessite de múltiplas saídas, devem-se definir múltiplos sistemas de inferência com os mesmos parâmetros para a entrada e as mesmas regras.

A tipo-redução é realizada pelo método do centróide, e a função usada para tal é uma adaptação da implementação de Nilesch N. Karnik, disponível em <http://sipi.usb.edu/~mendel/software>.

Por questões de desempenho foi necessário usar o MATLAB®Coder™ para gerar código compilável (em linguagem C), mais eficiente computacionalmente que os scripts do MATLAB®. Nos testes de desempenho realizados o tempo de execução da função compilada foi, no pior caso, 1/25 do tempo de execução da função interpretada.

#### 4.2.1 Funções

As funções criadas para implementação do sistema de inferência nebuloso tipo-2 intervalar no MATLAB® estão listadas a seguir, juntamente com uma breve descrição do funcionamento de cada uma:

##### **newit2fis**

Cria a estrutura do sistema de inferência nebuloso tipo-2 intervalar (IT2FIS), contendo os parâmetros do sistema;

##### **isit2fis**

Checa se uma estrutura contém os parâmetros de definição de um IT2FIS;

##### **it2trapmf**

Retorna os valores superiores e inferiores da função de pertinência trapezoidal tipo-2 intervalar. Permite usar centros iguais e modelar função de pertinência triangular;



**it2gaussmf**

Retorna os valores superiores e inferiores da função de pertinência gaussiana tipo-2 intervalar. Permite usar incerteza na média ou no desvio;

**mf2it2mf**

Aplica uma mancha de incerteza sobre uma função de pertinência tradicional e retorna os parâmetros da função resultante;

**fis2it2fis**

Define um IT2FIS com base num sistema de inferência nebuloso (FIS) tradicional, aplicando a função **mf2it2mf** nas funções de pertinência das variáveis de entrada e de saída do FIS;

**plotIt2Mf**

Representa graficamente uma função de pertinência tipo-2 intervalar;

**plotIt2Var**

Representa graficamente todas as funções de pertinência de uma ou mais entradas ou da saída de um IT2FIS;

**it2wtaver**

Computa os centroides da esquerda e da direita de um conjunto nebuloso tipo-2 intervalar;

**it2flsengine**

Realiza os cálculos de um IT2FIS, fuzzificando as entradas, aplicando as implicações e agregações definidas no conjunto de regras e realizando a tipo-redução e defuzzificação;

**evalit2fis**

Checa se as entradas fornecidas são válidas e fornece os dados para a função **it2flsengine** calcular os resultados.

Todas as funções citadas acima estão disponíveis no repositório remoto `git@gitlab.com:cirosobral/Type-2FuzzyFramework.git`. As funções para adicionar e remover variáveis e funções de pertinência ao sistema, assim como as funções para gerenciar as regras do sistema não precisaram ser implementadas por usar uma estrutura similar a do *Fuzzy Logic Toolbox*.



## 5 METODOLOGIA

Os testes foram conduzidos usando um Intel®, Core™i7 Q 720 @ 1.60 GHz, com 8 Gb de memória DDR3-1066, rodando Windows 7 Professional x64, Service Pack 1, e o MATLAB®8.1 R2013a, com a biblioteca de modelagem do *Fuzzy Logic Toolbox*. Para o treinamento foi usada a função `anfis` e a função `evalfis` para testes com lógica nebulosa tradicional. Para os testes com lógica nebulosa tipo-2 intervalar foi desenvolvida e usada a função `evalit2fis` do *framework* que foi desenvolvido a qual é descrita na seção 4.2.

A função `anfis` aceita como entradas os dados usados para treinamento e uma estrutura contendo o tipo das funções de pertinência usadas para entradas e saída do sistema e o número de funções de pertinência a serem usados para cada variável de entrada. Por questões de desempenho computacional, foram escolhidas funções triangulares para a entrada e valores constantes para a saída em todos os classificadores. Para o número de funções de pertinência, foram realizados testes com 2 e com 3 funções de pertinência.

Após o treinamento do sistema, a saída obtida da função `anfis` é uma estrutura contendo os parâmetros do sistema de inferência nebuloso. Esses parâmetros são usados para avaliar os dados que se pretende classificar com o sistema de inferência nebuloso, obtendo assim a classificação dada pelo modelo.

Como as saídas esperadas para o ANFIS devem ser valores numéricos, foi necessário converter os rótulos usados para classificação dos exemplos (G, Y e R). Para tal, convencionou-se os valores 1 para G, 2 para Y e 3 para R.

No caso do teste dos classificadores de lógica nebulosa tradicional, essa estrutura foi passada diretamente à função `evalfis`. Para os classificadores de lógica nebulosa tipo-2 intervalar, antes de invocar a função `evalit2fis`, foi usada a função `fis2it2fis`, também descrita na seção 4.2, que adiciona uma mancha de incerteza ao redor das funções de pertinência da entrada, criando funções de pertinência tipo-2 a partir das funções de pertinência tipo-1. Essa função aceita como parâmetro o valor do espalhamento, calculado como um percentual do suporte da função de pertinência original. Neste trabalho, foi utilizado o valor 0.1 para este parâmetro.

As saídas resultantes do uso dessas funções de avaliação são contínuas. Para classificação foi feita a aproximação dos valores da saída e considerados os valores menores que 1 como G, e os maiores que 3 como R. Então foi feito o processo inverso ao anterior, convertendo os valores numéricos em classes.

Os resultados dos classificadores foram confrontados com a classificação para um determinado problema fornecida junto aos dados e apresentados na forma de matrizes de confusão 3 x 3, como vista na tabela 5.1, para cada classificador. Para a combinação das 22 variáveis 2 a 2, foram geradas 231 matrizes, 3 a 3 foram 1540, 4 a 4 foram 7315. Cada um desses grupos foi testado com 2 e 3 funções de pertinência para cada entrada, com e sem o uso de SMOTE para os 3 problemas, totalizando 109.032 matrizes, o que tornou humanamente impossível a tarefa de avaliar todas as matrizes individualmente. Para isso, foram usadas as métricas de avaliação apresentadas na seção 5.4.

Tabela 5.1: Formato da matriz de confusão

|                      |                 |    |    |
|----------------------|-----------------|----|----|
| Classe de referência | Classe prevista |    |    |
|                      | GG              | GY | GR |
|                      | YG              | YY | YR |
|                      | RG              | RY | RR |

Visando alcançar um bom resultado, o ANFIS realiza iterativamente uma rotina de treinamento, cada uma dessas iterações é chamada de época. Os primeiros testes, para determinar a quantidade de épocas necessárias para a convergência do modelo, foram realizados com um conjunto de 4 variáveis de entrada e 3 funções de pertinência por cada entrada.

Durante a execução do treinamento com a função `anfis` é exibido o erro de cada época e a função tenta minimizar este erro. Foram feitos testes até o limite de 500 épocas de treinamento. No entanto, não foi observada diminuição significativa na taxa de erros a partir da primeira dezena de iterações. Portanto optou-se por usar o valor padrão, que é de 10 épocas.

## 5.1 Seleção de atributos

Devido ao número de variáveis presente nos dados de entrada ser maior que a quantidade de variáveis que o sistema poderia aceitar, foi necessário realizar um procedimento de seleção dessas variáveis. Esse procedimento, chamado de seleção de atributos (*feature selection*), visa encontrar o subconjunto de variáveis independentes que melhor explique a variável resposta. Seu uso diminui o tempo de treinamento do sistema e leva à geração de modelos mais compreensíveis, porém possui espaço de busca exponencial.

Foi dado prosseguimento aos testes para permitir a escolha dos atributos importantes e definição da quantidade de funções de pertinência. Foram testadas todas as possibilidades de combinação entre 2, 3 e 4 variáveis, usando o sistema ANFIS com 2 e 3 funções de pertinência por variáveis de entrada.

Para cada problema foram realizadas 6 baterias de testes, combinando as variáveis 2 a 2, 3 a 3 e 4 a 4, com 2 e 3 funções de pertinência por variável de entrada. Os testes usando dados artificiais, gerados com o SMOTE, foram conduzidos de forma distinta dos que não usaram e essa diferença será explicada nas seções 5.2 e 5.3.

Estes resultados então foram avaliados por algumas métricas distintas: a taxa de acertos, taxa de acerto ponderada e pelas métricas de sensibilidade, especificidade, PPV (*Positive Predictive Value*) e NPV (*Negative Predictive Value*), que serão explicados na Seção 5.4.

## 5.2 Descrição dos testes sem uso do SMOTE

Para este conjunto de testes, os mesmos dados usados no treinamento, foram usados no teste. Por esse motivo, foi usada validação cruzada com 5 partições para definir os conjuntos de teste e de treinamento. Esse passo foi realizado com o uso da função `cvpartition` do *Statistics Toolbox* do MATLAB.

Dessa forma, o teste de cada combinação de variáveis consiste em 5 rodadas de treinamento e teste, usando 4 partições para treinamento e 1 para teste, a cada rodada alternando a partição usada para teste. Em cada iteração, as 4 partições contendo a classificação esperada de cada exemplo foram passadas ao ANFIS como dados de treinamento. Após o ajuste das funções de pertinência da entrada e da saída realizada pela função, o sistema era usado para a classificação dos exemplos da partição restante.

As classificações realizadas pelo sistema foram então confrontadas com as classes usadas como referência, dando forma às matrizes de confusão que foram usadas para comparação dos classificadores, como descrito acima.

### 5.3 Descrição dos testes com uso do SMOTE

Neste conjunto de testes, usou-se o SMOTE para geração de dados artificiais. O uso do SMOTE envolve o descarte de parte dos dados da classe majoritária (*undersampling*) e a geração de dados artificiais (*oversampling*) para uma determinada classe, a partir de uma amostra, introduzindo perturbações vindas de  $n$  amostras vizinhas.

Nesse conjunto de testes, os dados usados para treinamento foram gerados com uso do SMOTE. Para geração desses dados, foram descartado 70% das amostras da classe majoritária e os dados artificiais gerados com base em 3 vizinhos. Assim, o conjunto de dados artificiais foi passado ao ANFIS, juntamente com as classes esperadas, e os dados originais usados para o teste. Por esse motivo não foi usada a técnica de validação cruzada.

Da mesma forma que os testes sem uso do SMOTE, as classificações foram confrontadas com as classes de referência e foram geradas as matrizes de confusão.

### 5.4 Definição das métricas de avaliação

Para avaliar quais seriam as melhores matrizes, foram usadas 6 métricas diferentes.



**Taxa de acerto** Calculada como a proporção da soma dos casos presentes na diagonal principal (identificação correta), pela soma de todos os elementos da matriz de confusão;

**Sensibilidade** Probabilidade das amostras que pertencem a uma classe serem classificadas corretamente. Indica a capacidade do classificador indicar corretamente a classe a que as amostras pertencem.

**Especificidade** Probabilidade das amostras que **não** pertencem a uma classe **não** serem classificadas naquela classe. Indica a capacidade do classificador indicar corretamente a quais classes as amostras (não) pertencem.

**Valor preditivo positivo (PPV)** Indica a porcentagem de classificações corretas entre todas as amostras classificadas em uma determinada classe. Indica a probabilidade de perante uma amostra, a classificação pertencer à classe prevista.

**Valor preditivo negativo (NPV)** Indica a porcentagem de classificação que **não** pertencem a uma determinada classe sobre todos os elementos classificados naquela classe. Indica a probabilidade de perante uma amostra, a classificação pertencer a uma classe distinta da classe avaliada.

**Taxa de acerto ponderada** Métrica definida para atribuir maior peso às classificações corretas das classes minoritárias. O valor resultante dessa métrica varia de  $]\infty; 1]$ , onde 1 representa o caso de todas as classificações corretas. O raciocínio usado na definição dessa métrica é explicado mais a frente.

As métricas de Sensibilidade (*Recall*), especificidade, PPV (*Precision*) e NPV são comumente usadas na avaliação de classificadores binários. Para o uso com 3 classes, foi usado o mesmo esquema apresentado por BELEITES; SALZER; SERGO (2013), como pode ser observado na Figura 5.1.

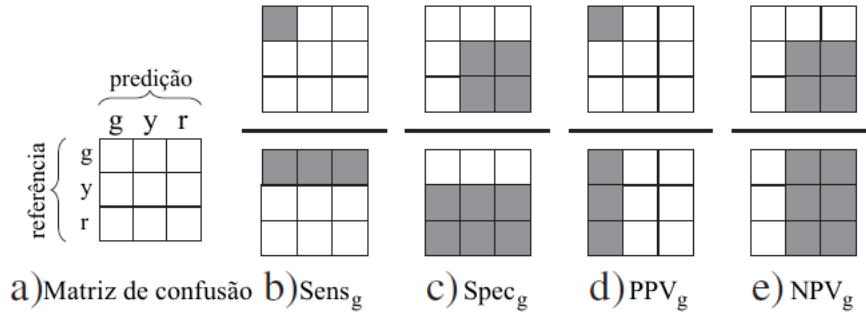


Figura 5.1: Método de cálculo da sensibilidade, especificidade, PPV e NPV para classificação multi-classes.

O cálculo da taxa de acerto obedece à equação 5.1. No entanto, a taxa de acerto foi desenvolvida para classificações binárias e portanto não é capaz de expressar todas as informações contidas nas matrizes de confusão. Essa simplificação mascara resultados desejáveis como a inexistência de falsos negativos, como no caso de exemplos das classes *Y* ou *R* serem apresentados como *G*. Por esse motivo, foram usadas métricas de sensibilidade, especificidade, PPV e NPV como visto em BELEITES; SALZER; SERGO (2013).

$$taxaDeAcerto = \frac{GG + YY + RR}{GG + GY + GR + YG + YY + YR + RG + RY + RR} \quad (5.1)$$

Ainda assim, devido a quantidade de classificadores avaliados, em diversos casos os valores para a taxa de acerto, assim como para as outras métricas tradicionais podem ser similares, impossibilitando a escolha de bons classificadores. Por esse motivo se fez necessário o uso de outra métrica para avaliação do desempenho dos classificadores.

Visando encontrar os modelos que minimizassem os casos de classificações erradas, principalmente os falsos positivos, o uso da matriz de pesos ajuda a escolher

esses classificadores. Para isso foram atribuídos pesos negativos para os casos de falha e pesos ainda mais negativos para casos de falso positivo. Dessa forma foi possível priorizar os classificadores que classificavam corretamente os casos de falhas em detrimento daqueles que classificavam corretamente apenas os casos sem falha.

Assim foi criado um sistema de classificação, usando média ponderada com pesos como vistos na tabela 5.2, que foram definidos seguindo a seguinte lógica:

Esta classificação é feita por meio da soma de todos os elementos do produto de Hadamard entre a matriz de resultados e a matriz de pesos. Esta operação é realizada para cada uma das matrizes de confusão.

Tabela 5.2: Matriz de pesos usada para avaliação de desempenho dos classificadores

|                      |   | Classe prevista |    |    |
|----------------------|---|-----------------|----|----|
|                      |   | 1               | -2 | -4 |
| Classe de referência | G | -5              | 2  | -2 |
|                      | R | -10             | -5 | 3  |

Os valores da Tabela 5.2 obedecem à seguinte ordem. Os elementos na diagonal principal, representam as classificações feitas corretamente. Portanto, receberam peso positivo. Como os dados estão desbalanceados, possuindo uma quantidade maior de casos G, foi definido que se aumentaria o peso para as classes com menor participação, em ordem crescente.

Todos os outros elementos representam classificação incorreta. Para todos eles, foram designados pesos negativos. Os elementos acima da diagonal principal, indicam casos que eram G e foram classificados como Y ou R; ou casos Y que foram classificados como R. Enquanto os casos abaixo da diagonal principal, representam casos R que foram classificados como Y ou G; ou casos Y que foram classificados como G.

Como o objetivo desse sistema é detectar casos de falha para permitir a manutenção dos equipamentos, o pior tipo de erro em classificação é quando o caso é mais grave do que o sistema informa, pois assim um problema pode passar despercebido. Dessa forma, foram definidos pesos maiores para os elementos abaixo da diagonal principal.



## 6 RESULTADOS E DISCUSSÃO

Agrupando os resultados das taxas de acerto, sensibilidade, especificidade, PPV, NPV e taxas de acerto ponderadas por problema (combustão, contaminação e corrosão), pelo tipo de dados usados para treinamento (com e sem uso de dados artificiais), pelo número de variáveis de entrada (2, 3 e 4 variáveis) e pelo número de funções de pertinência (2 e 3 funções por variável de entrada) foi possível comparar quais desses parâmetros tiveram melhor efeito na classificação e quais as variáveis de entrada mais presentes nos classificadores melhor avaliados. As tabelas resultantes desses agrupamentos foram colocadas no Apêndice A.

As tabelas presentes nesta seção trazem os dados apresentados nas tabelas do Apêndice A, agrupando todos os classificadores por grupo de teste, por problema e por tipo. Os grupos de teste são o sem SMOTE, que não fizeram uso de dados artificiais, como definido na seção 5.2 e os com SMOTE, seção 5.3. Os tipos apresentados nas tabelas, tipo-1 e tipo-2, refletem o tipo da função de pertinência na construção do classificador. Dessa forma é possível ter uma visão geral do resultado do sistema e estabelecer comparações entre os agrupamentos.

Tabela 6.1: Média da taxa de acerto dos melhores classificadores

| <b>Smote</b>           | <b>Problema</b> | <b>Tipo-1 (A)</b> | <b>Tipo-2 (B)</b> | <b>Diferença (B – A)</b> |
|------------------------|-----------------|-------------------|-------------------|--------------------------|
| Sem Smote              | Combustao       | 98,7799%          | 98,8567%          | 0,0767%                  |
|                        | Contaminacao    | 97,3297%          | 97,4294%          | 0,0998%                  |
|                        | Corrosao        | 92,6565%          | 92,4570%          | -0,1995%                 |
| <i>Total Sem Smote</i> |                 | 96,2554%          | 96,2477%          | -0,0077%                 |
| Com Smote              | Combustao       | 96,1096%          | 95,7643%          | -0,3453%                 |
|                        | Contaminacao    | 77,4171%          | 72,4217%          | -4,9954%                 |
|                        | Corrosao        | 85,5663%          | 85,9116%          | 0,3453%                  |
| <i>Total Com Smote</i> |                 | 86,3643%          | 84,6992%          | -1,6651%                 |

De uma maneira geral, observando os testes realizados sem uso do SMOTE, a taxa de acerto para os classificadores apenas com dados reais é alta, observando a média dos classificadores agregados por problemas, ficou acima de 92% para o melhor caso, e acima de 86% na média, como visto nas Tabelas 6.1 e 6.2, respectivamente. Esse fato já era esperado, devido ao desbalanço entre as classes existente nos dados usados no sistema, visto que um classificador poderia atingir taxas de acerto alta apenas classificando todos os exemplos como G.

Tabela 6.2: Média da taxa de acerto dos classificadores

| <b>Smote</b>           | <b>Problema</b> | <b>Tipo-1 (A)</b> | <b>Tipo-2 (B)</b> | <b>Diferença (B – A)</b> |
|------------------------|-----------------|-------------------|-------------------|--------------------------|
| Sem Smote              | Combustao       | 94,8762%          | 97,3872%          | 2,5109%                  |
|                        | Contaminacao    | 92,9053%          | 90,1084%          | -2,7968%                 |
|                        | Corrosao        | 86,7861%          | 88,5660%          | 1,7799%                  |
| <i>Total Sem Smote</i> |                 | 91,5225%          | 92,0205%          | 0,4980%                  |
| Com Smote              | Combustao       | 54,9666%          | 59,5741%          | 4,6075%                  |
|                        | Contaminacao    | 26,2524%          | 24,4234%          | -1,8289%                 |
|                        | Corrosao        | 61,4377%          | 57,3620%          | -4,0757%                 |
| <i>Total Com Smote</i> |                 | 47,5522%          | 47,1198%          | -0,4324%                 |

A avaliação pela taxa de acerto, no entanto, não permite concluir qual seria a melhor escolha para a quantidade de funções de pertinência para cada entrada, nem quantas variáveis de entrada deveriam ser usadas. Apesar dessa medida já permitir a distinção entre classificadores melhores dentre todos os classificadores, essa medida não representa fielmente a qualidade do classificador.

Observando os resultados presentes no Apêndice A, é possível ainda notar que o aumento do número de variáveis de entrada mostrou-se eficaz em encontrar um melhor classificador, ao passo que na observação da média entre os resultados foram similares ou inferiores. Esses resultados podem ser explicados pelo aumento do espaço de busca, que permite encontrar um classificador melhor, porém encontra também diversos classificadores com resultado abaixo da média.

Tabela 6.3: Média da taxa de acerto ponderada dos melhores classificadores

| <b>Smote</b>           | <b>Problema</b> | <b>Tipo-1 (<math>A</math>)</b> | <b>Tipo-2 (<math>B</math>)</b> | <b>Diferença (<math>B - A</math>)</b> |
|------------------------|-----------------|--------------------------------|--------------------------------|---------------------------------------|
| Sem Smote              | Combustao       | 0,943370                       | 0,950737                       | 0,007366                              |
|                        | Contaminacao    | 0,826197                       | 0,825660                       | -0,000537                             |
|                        | Corrosao        | 0,660681                       | 0,675261                       | 0,014579                              |
| <i>Total Sem Smote</i> |                 | 0,810083                       | 0,817219                       | 0,007136                              |
| Com Smote              | Combustao       | 0,902394                       | 0,889733                       | -0,012661                             |
|                        | Contaminacao    | 0,323665                       | 0,132827                       | -0,190838                             |
|                        | Corrosao        | 0,615792                       | 0,591851                       | -0,023941                             |
| <i>Total Com Smote</i> |                 | 0,613950                       | 0,538137                       | -0,075813                             |

Tabela 6.4: Média da taxa de acerto ponderada dos classificadores

| <b>Smote</b>           | <b>Problema</b> | <b>Tipo-1 (<math>A</math>)</b> | <b>Tipo-2 (<math>B</math>)</b> | <b>Diferença (<math>B - A</math>)</b> |
|------------------------|-----------------|--------------------------------|--------------------------------|---------------------------------------|
| Sem Smote              | Combustao       | 0,765992                       | 0,840135                       | 0,074143                              |
|                        | Contaminacao    | 0,561014                       | 0,465207                       | -0,095807                             |
|                        | Corrosao        | 0,304673                       | 0,359301                       | 0,054629                              |
| <i>Total Sem Smote</i> |                 | 0,543893                       | 0,554881                       | 0,010988                              |
| Com Smote              | Combustao       | -0,368487                      | -0,238961                      | 0,129525                              |
|                        | Contaminacao    | -1,289615                      | -1,356219                      | -0,066604                             |
|                        | Corrosao        | -0,182508                      | -0,314557                      | -0,132049                             |
| <i>Total Com Smote</i> |                 | -0,613537                      | -0,636579                      | -0,023043                             |

Ainda, observando o resultado dos testes sem uso de dados artificiais, o aumento do número de variáveis de entrada mostrou ser mais consistente na geração de melhores classificadores que o aumento do número de funções de pertinência, isto é, o aumento do número de variáveis de entrada causou maior melhora no resultado do classificador que o aumento do número de funções de pertinência.

Contudo, quando se observa os resultados dos testes realizados usando SMOTE, é possível observar uma maior eficácia nos classificadores com maior número de funções de pertinência. É possível que um sistema com mais funções de pertinência seja mais capaz de generalizar o conhecimento.



Novamente, na comparação dos resultados dos classificadores usando lógica nebulosa tipo-1, contra os classificadores baseados em lógica nebulosa tipo-2, é possível notar no grupo de avaliações realizado sem uso do SMOTE, que para os problemas de combustão e corrosão, os classificadores tipo-2 tiveram resultados superiores aos classificadores tipo-1 em praticamente todas as medidas. Para esses casos, os valores para as métricas de sensibilidade, especificidade, PPV e NPV foram em geral superiores, o que indica a confiabilidade dos resultados. No entanto, para o problema de contaminação, o resultado foi o inverso. Esse resultado pode ser explicado pelo fato desse ser o único problema com quantidade de casos rotulados como Y em menor quantidade que os rotulados como R. Como os classificadores Tipo-2 não foram treinados, apenas tiveram suas funções de pertinência modificadas, ampliando o suporte das mesmas, há uma tendência maior para a classificação de amostras na classe intermediária.

Ainda, para as avaliações realizadas com uso do SMOTE, os classificadores foram inferiores em praticamente todas as medidas de todos os problemas. Esse resultado era de certa forma esperado visto que nos testes sem SMOTE, os mesmos dados foram usado para treinamento e teste, usando validação cruzada. Já nos testes com SMOTE, os dados artificiais foram usados para treinamento e os reais para teste.

Interessante observar que a variável Fuligem apareça constantemente nos melhores classificadores para o problema de combustão, pois a fuligem é um dos produtos da combustão incompleta. Assim como a variável Diluição por combustível aparece nos melhores classificadores para contaminação, visto que o combustível não deveria entrar em contato com o lubrificante, sendo assim, considerado um contaminante. Da mesma forma é possível observar a variável Ferro sendo importante para a construção de um modelo de diagnóstico de corrosão, já que o bloco e grande

parte do motor é composta desse elemento. Esses resultados sugerem a correteude do modelo proposto.



## 7 CONCLUSÕES

O trabalho atingiu o objetivo de investigar o uso de lógica nebulosa tipo-2 para classificação da condição de motores, através do resultado da análise de óleo. Os resultados permitem afirmar que é possível construir, de forma simples, um sistema baseado em um modelo *neuro-fuzzy* bem conhecido, que apresenta bons resultados para o diagnóstico de casos em que não há falhas. Nos casos de falha, no entanto, o sistema não permite diferenciar qual o tipo de falha. Esse não seria um impeditivo para o uso do sistema em situações reais, visto que havendo falha o motor deve ser levado para manutenção em qualquer um dos casos.

A diferença entre os resultados dos sistemas usando lógica nebulosa tipo-1 e tipo-2 foi pequena, não permitindo afirmar que os classificadores do tipo-2 foram melhor do que os do tipo-1. Porém há indícios de que os sistemas de classificação usando lógica nebulosa tipo-2 sejam, de fato, mais robustos que os classificadores que usaram lógica nebulosa tipo-1.

O baixo número de amostras e o desbalanço de classes presente nos dados dificultaram a realização dos testes. Foi possível observar que o uso do SMOTE não foi uma boa solução para o problema do desbalanço dos dados, pois tanto os classificadores que usaram lógica nebulosa tipo-1 quanto os que usaram lógica nebulosa tipo-2 apresentaram resultados inferiores aos que usaram os dados originais.

## 7.1 Contribuições

O presente trabalho apresentou o uso do ANFIS como uma alternativa viável para modelagem do problema de classificação da condição de motores, através do resultado da análise de óleo. Esta parte do trabalho serviu como conteúdo para um artigo que foi publicado no SMC2014 (2014 *IEEE International Conference on Systems, Man, and Cybernetics*). Nesse artigo foi abordado como gerar um classificador para o problema de corrosão usando ANFIS.

Ainda, esse trabalho apresenta uma comparação entre classificadores multi-classe usando lógica nebulosa tipo-2 e lógica nebulosa tradicional. Dessa comparação pôde-se notar que foi possível alcançar melhores resultados usando lógica nebulosa tipo-2, mas que a solução da adição automatizada da mancha de incerteza não melhora os resultados em todos os casos, portanto se faz necessário uma avaliação criteriosa antes de seu uso.

Uma outra importante contribuição desse trabalho seja o desenvolvimento do *framework* de lógica nebulosa tipo-2 intervalar para MATLAB. Este *framework* pretende facilitar o uso de lógica nebulosa tipo-2 intervalar para os estudantes e profissionais que desejem usar tal ferramenta em seus trabalhos futuros. Nesse *framework*, além do sistema de inferência nebuloso tipo-2 intervalar, são encontradas funções para visualização das funções de pertinência, criação da estrutura e obtenção de parâmetros através de um sistema de lógica nebulosa tradicional.

## 7.2 Trabalhos Futuros

Há um desequilíbrio nas amostras utilizadas para treinar o sistema *neuro-fuzzy*. Há muito mais exemplos de motores em bom estado (G) do que em atenção (Y) ou condição crítica (R). Nessa pesquisa tentou-se minimizar o efeito de tais dados usando o SMOTE para geração de dados artificiais. Outra saída pode ser a de buscar um especialista para classificar uma quantidade maior de exemplos para treinamento do sistema.

É necessário encontrar outras formas de avaliar o resultado dos classificadores multi-classe para levar em consideração os casos de falha na classificação, de forma similar às métricas usadas para classificações binárias. A aplicação de tais métodos poderia levar à escolha de um melhor modelo de diagnóstico.

Nesse trabalho, a conversão do sistema nebuloso tipo-1 em tipo-2, realizada pela função `fis2it2fis`, foi realizada com o parâmetro que indica o tamanho da mancha de incerteza de 0,1. Outros valores podem ser testados para avaliar a reação do sistema.

Ainda sobre as funções de pertinência da entrada, pode-se avaliar o sistema com outras funções de pertinência que não apenas a triangular. Bem como o uso de outras técnicas como algoritmos genéticos ou redes neurais para o treinamento do sistema de inferência.

Outro ponto ainda a ser investigado é o de como o sistema de classificação proposto no trabalho se sairia em comparação a sistemas que usem outras técnicas para classificação.

### 7.3 Dificuldades Encontradas

Diversos desafios foram encontrados durante a execução desse projeto. Desde o tamanho da massa de dados disponíveis para a realização do trabalho até a falta de uma ferramenta para a realização dos testes com lógica nebulosa tipo-2 intervalar.

Para tentar resolver o problema da quantidade de dados disponíveis, foram realizados contatos com as empresas Shell, Petrobrás e Ipiranga que oferecem serviços do mesmo tipo na tentativa de obter novos dados. Nesses contatos não se obteve sucesso com a empresa que forneceu os dados originalmente, visto que eles alegaram não oferecer mais o serviço. Já outra empresa que se dispôs a cooperar, não possuía os dados com as avaliações necessárias para o treinamento e teste do sistema. Por esse motivo, optou-se por realizar a geração de dados artificiais com o SMOTE.

Com os contatos foi possível ainda obter mais conhecimento sobre análise de óleo lubrificante, pois a falta de conhecimento sobre o assunto também foi outro desafio a ser superado no processo. A análise de óleo é um procedimento bastante realizado em empresas que operam grandes frotas e maquinários pesados. No entanto, muitas informações não estão disponíveis, ou carecem de maior rigor científico, visto que os resultados não são divulgados para proteger sua vantagem competitiva.

Quanto ao problema da falta do ferramental para realizar os testes usando lógica nebulosa tipo-2 intervalar, foi tentado o contato com alguns pesquisadores que já usaram essa técnica em artigos publicados. Porém os mesmos não responderam às tentativas de contato feitas. Portanto, decidiu-se pelo desenvolvimento de um *framework* para realização dos testes. O *framework* desenvolvido foi fortemente baseado nas funções descritas em (MENDEL, 2001).

Entretanto, o algoritmo inicial era extremamente lento, impossibilitando o uso prático nos testes. Verificou-se a possibilidade de usar o MATLAB®Coder™ para ganhar desempenho no processamento. Seu uso aumenta de forma expressiva o desempenho do algoritmo, no entanto, impõe aos desenvolvedores uma série de restrições na escrita do código. Por este motivo, foi desenvolvida outra implementação do sistema de inferência nebuloso tipo-2 intervalar. Na nova versão pôde ser usado o MATLAB®Coder™ e conseguiu-se assim um desempenho satisfatório.

Além dos desafios já citados, a falta de métricas claras para avaliação de classificadores multi-classe foi outro desafio que dificultou o processo, tanto na definição do sistema quanto na apuração dos resultados.





## REFERÊNCIAS

- GUPTA, M. M.; SÁNCHEZ, E. (Ed.). **Approximate Reasoning in Decision Analysis**. [S.l.]: Elsevier/North-Holland, 1982.
- AKINTUNDE, D. O. Spectrometric Oil Analysis - An Untapped Resource for Condition Monitoring. **AU Journal of Technology**, [S.l.], v.2, n.12, p.107–114, October 2008.
- BARRETT, M. **The Practical Guide To Oil Analysis**. [S.l.]: Insight Services, 2007? n.4th ed.
- BATISTA, G. E. d. A. P. A. **Pré-processamento de Dados em Aprendizado de Máquina Supervisionado**. 2003. Tese de Doutorado — Universidade de São Paulo.
- BELEITES, C.; SALZER, R.; SERGO, V. Validation of soft classification models using partial class memberships: an extended concept of sensitivity & co. applied to grading of astrocytoma tissues. **Chemometrics and Intelligent Laboratory Systems**, [S.l.], v.122, n.0, p.12–22, 2013.
- BOOSER, E. R. **CRC Handbook of Lubrication and Tribology**. [S.l.]: CRC Press, 1993.
- BROOMHEAD, D. S.; LOWE, D. Multivariable Functional Interpolation and Adaptive Networks. **Complex Systems**, [S.l.], v.2, p.321–355, 1988.
- CALADO, J. et al. Soft Computing Approaches to Fault Diagnosis for Dynamic Systems. **European Journal of Control**, [S.l.], v.7, n.2-3, p.248–286, 2001.
- CARNERO, M. C. Selection of diagnostic techniques and instrumentation in a predictive maintenance program. A case study. **Decision Support Systems**, [S.l.], v.38, n.4, p.539–555, 2005.

- CASTILLO, O.; MELIN, P. **Type-2 Fuzzy Logic: theory and applications**. [S.l.]: Springer, 2008.
- CHAWLA, N. et al. SMOTE: synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, [S.l.], v.16, p.321–357, 2002.
- ConocoPhillips Lubricants. **Oil Analysis Technical Guide**. [S.l.]: ConocoPhillips Lubricants, 2007.
- Dep. Técnico da Lubrikol Lubrificantes. **Importância da Análise de Óleo Lubrificante**. [S.l.]: Lubrikol, 2013.
- FASEL, B.; LUETTIN, J. Automatic facial expression analysis: a survey. **Pattern Recognition**, [S.l.], v.36, n.1, p.259 – 275, 2003.
- GARRY, M. C. **FT-IR Analysis of Used Lubricant Oils – General Considerations**. [S.l.]: Thermo Fischer Scientific, 2007.
- GINNEKEN, B. V. et al. Computer-aided diagnosis in chest radiography: a survey. **IEEE Transactions on Medical Imaging**, [S.l.], v.20, p.1228–1241, 2001.
- JANG, J.-S. R. ANFIS: adaptive-network-based fuzzy inference system. **IEEE Transactions on Systems, Man, and Cybernetics**, [S.l.], v.23, n.3, p.665–685, 1993.
- KARNIK, N. N.; MENDEL, J. M. Centroid of a type-2 fuzzy set. **Information Sciences**, [S.l.], v.132, n.1-4, p.195–220, 2001.
- KARNIK, N. N.; MENDEL, J. M.; LIANG, Q. Type-2 fuzzy logic systems. **IEEE T. Fuzzy Systems**, [S.l.], v.7, n.6, p.643–658, 1999.
- MACIAN, V. et al. Fuzzy logic-based expert system for diesel engine oil analysis diagnosis. **Insight**, [S.l.], v.8, n.48, p.462–470, 2006.

- MELIN, P.; CASTILLO, O. A review on the applications of type-2 fuzzy logic in classification and pattern recognition. **Expert Syst. Appl.**, [S.l.], v.40, n.13, p.5413–5423, 2013.
- MENDEL, J. M. **Uncertain Rule-Based Fuzzy Logic Systems**: introduction and new directions. [S.l.]: Prentice Hall, 2001.
- MENDEL, J. M. Advances in type-2 fuzzy sets and systems. **Inf. Sci.**, [S.l.], v.177, n.1, p.84–110, 2007.
- MENDEL, J. M.; JOHN, R. I. Type-2 fuzzy sets made simple. **IEEE T. Fuzzy Systems**, [S.l.], v.10, n.2, p.117–127, 2002.
- MENDEL, J. M.; KARNIK, N. N. An introduction to type-2 fuzzy logic systems. In: IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE. **Anais...** [S.l.: s.n.], 1998. v.2, p.915–920.
- MOHAMMADPOUR, J.; FRANCHEK, M.; GRIGORIADIS, K. A survey on diagnostic methods for automotive engines. In: INTERNATIONAL JOURNAL OF ENGINE RESEARCH. **Anais...** SAGE, 2012. v.13, p.41–64.
- MORI, S.; SUEN, C. Y.; YAMAMOTO, K. Historical review of OCR research and development. In: O’GORMAN, L.; KASTURI, R. (Ed.). **Document image analysis**. Los Alamitos, CA, USA: IEEE Computer Society Press, 1995. p.244–273.
- PERKINELMER. **Lubricant Analysis**. [S.l.]: PerkinElmer, 2012.
- PHUONG, N. H. Fuzzy set theory and medical expert systems: survey and model. In: BARTOSEK, M.; STAUDEK, J.; WIEDERMANN, J. (Ed.). **SOFSEM ’95**: theory and practice of informatics. [S.l.]: Springer Berlin Heidelberg, 1995. p.431–436. (Lecture Notes in Computer Science, v.1012).

- RAJU, G. S. V. P. et al. Article: solving uncertain problems using anfis. **International Journal of Computer Applications**, [S.l.], v.29, n.11, p.14–21, September 2011. Full text available.
- RAMEZANI, S.; MEMARIANI, A. A Fuzzy Rule Based System for Fault Diagnosis, Using Oil Analysis Results. **International Journal of Industrial Engineering & Production Research**, [S.l.], v.22, n.2, p.91–98, June 2011.
- REZENDE, S. O. (Ed.). **Sistemas Inteligentes - Fundamentos e Aplicações**. 1 ed..ed. [S.l.]: Editora Manole Ltda., 2003.
- ROSS, T. J. **Fuzzy Logic with Engineering Applications, Third Edition**. [S.l.]: Wiley, 2010.
- SALA, A. et al. Fuzzy diagnosis module based on interval fuzzy logic: oil analysis application. In: ICINCO. **Anais...** INSTICC Press, 2005. p.85–90.
- SENAI-ES. **Lubrificação - Mecânica**. [S.l.]: SENAI-ES, 1997.
- SOUZA, M. S. M. de. Métodos Analíticos para Lubrificantes e Isolantes. **Revista QD**, [S.l.], v.382, p.20–28, Maio 2000.
- Supreme Lubrificantes. **Análise de Óleo**. 2011?
- THWEATT, D.; BELL, T. Interpreting Heavy-duty Motor Oil Analysis Reports. In: PRACTICING OIL ANALYSIS. **Anais...** [S.l.: s.n.], 2006.
- ZADEH, L. A. Fuzzy Sets. **Information and Control**, [S.l.], v.8, n.3, p.338–353, 1965.
- ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning - I. **Inf. Sci.**, [S.l.], v.8, n.3, p.199–249, 1975.

## APÊNDICE A – RESULTADOS

### Para o problema de combustão sem uso do SMOTE

Tabela A.1: Valor máximo da taxa de acerto para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 98.62%        | 98.62%        | 0.00%            |
| <b>3 Variáveis</b> | 98.76%        | 98.62%        | -0.14%           |
| <b>4 Variáveis</b> | 98.62%        | 99.03%        | 0.41%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 98.90%        | 98.90%        | 0.00%            |
| <b>3 Variáveis</b> | 98.76%        | 98.90%        | 0.14%            |
| <b>4 Variáveis</b> | 99.03%        | 99.08%        | 0.05%            |

Tabela A.2: Valor médio da taxa de acerto para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 95.84%        | 97.59%        | 1.75%            |
| <b>3 Variáveis</b> | 95.03%        | 97.59%        | 2.57%            |
| <b>4 Variáveis</b> | 94.06%        | 97.57%        | 3.51%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 95.87%        | 97.54%        | 1.67%            |
| <b>3 Variáveis</b> | 94.75%        | 97.26%        | 2.52%            |
| <b>4 Variáveis</b> | 93.72%        | 96.76%        | 3.04%            |

Tabela A.3: Resultado dos melhores classificadores para o problema de Combustão, avaliado pela taxa de acerto

(a) Tipo-1, com 2 variáveis (Si, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 7        | 7        | 0        |
| <b>R</b> | 2        | 1        | 1        |

(b) Tipo-1, com 3 variáveis (Al, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 5        | 8        | 1        |
| <b>R</b> | 1        | 2        | 1        |

(c) Tipo-1, com 4 variáveis (Fe, Pb, Oxi, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 6        | 7        | 1        |
| <b>R</b> | 2        | 1        | 1        |

(d) Tipo-2, com 2 variáveis (Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 705      | 1        | 0        |
| <b>Y</b> | 5        | 9        | 0        |
| <b>R</b> | 1        | 3        | 0        |

(e) Tipo-2, com 3 variáveis (Pb, Mo, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 6        | 8        | 0        |
| <b>R</b> | 2        | 2        | 0        |

(f) Tipo-2, com 4 variáveis e (Fe, P, Nit, Ful) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 5        | 9        | 0        |
| <b>R</b> | 1        | 1        | 2        |

(g) Tipo-1, com 2 variáveis (Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 704      | 2        | 0        |
| <b>Y</b> | 3        | 11       | 0        |
| <b>R</b> | 0        | 3        | 1        |

(h) Tipo-1, com 3 variáveis (Fe, Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 5        | 8        | 1        |
| <b>R</b> | 0        | 3        | 1        |

(i) Tipo-1, com 4 variáveis (P, Ful, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 5        | 9        | 0        |
| <b>R</b> | 1        | 1        | 2        |

(j) Tipo-2, com 2 variáveis (P, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 704      | 2        | 0        |
| <b>Y</b> | 4        | 10       | 0        |
| <b>R</b> | 1        | 1        | 2        |

(k) Tipo-2, com 3 variáveis (Cr, Ti, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 528      | 1        | 0        |
| <b>Y</b> | 4        | 7        | 0        |
| <b>R</b> | 0        | 1        | 2        |

(l) Tipo-2, com 4 variáveis e (Cu, Ti, Ful, Agu) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 530      | 0        | 0        |
| <b>Y</b> | 3        | 7        | 0        |
| <b>R</b> | 0        | 2        | 1        |

Tabela A.4: Valor máximo pela matriz de pesos para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.9171        | 0.9268        | 0.0097           |
| <b>3 Variáveis</b> | 0.9365        | 0.9351        | -0.0014          |
| <b>4 Variáveis</b> | 0.9337        | 0.9586        | 0.0249           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.9599        | 0.9613        | 0.0014           |
| <b>3 Variáveis</b> | 0.9475        | 0.9613        | 0.0138           |
| <b>4 Variáveis</b> | 0.9655        | 0.9613        | -0.0041          |

Tabela A.5: Valor médio pela matriz de pesos para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.7859        | 0.8371        | 0.0512           |
| <b>3 Variáveis</b> | 0.7699        | 0.8437        | 0.0738           |
| <b>4 Variáveis</b> | 0.7444        | 0.8504        | 0.1060           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.7941        | 0.8426        | 0.0485           |
| <b>3 Variáveis</b> | 0.7651        | 0.8407        | 0.0756           |
| <b>4 Variáveis</b> | 0.7366        | 0.8263        | 0.0897           |



Tabela A.6: Resultado dos melhores classificadores para o problema de Combustão, avaliado pela matriz de pesos

(a) Tipo-1, com 2 variáveis (Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 705      | 1        | 0        |
| <b>Y</b> | 6        | 8        | 0        |
| <b>R</b> | 1        | 3        | 0        |

(c) Tipo-1, com 4 variáveis (Pb, Al, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 704      | 2        | 0        |
| <b>Y</b> | 5        | 9        | 0        |
| <b>R</b> | 1        | 2        | 1        |

(e) Tipo-2, com 3 variáveis (Pb, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 704      | 2        | 0        |
| <b>Y</b> | 5        | 8        | 1        |
| <b>R</b> | 0        | 3        | 1        |

(g) Tipo-1, com 2 variáveis (Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 704      | 2        | 0        |
| <b>Y</b> | 3        | 11       | 0        |
| <b>R</b> | 0        | 3        | 1        |

(i) Tipo-1, com 4 variáveis (Mg, Oxi, Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 705      | 1        | 0        |
| <b>Y</b> | 4        | 8        | 2        |
| <b>R</b> | 0        | 1        | 3        |

(k) Tipo-2, com 3 variáveis (Sul, Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 705      | 1        | 0        |
| <b>Y</b> | 5        | 8        | 1        |
| <b>R</b> | 0        | 1        | 3        |

(b) Tipo-1, com 3 variáveis (Al, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 706      | 0        | 0        |
| <b>Y</b> | 5        | 8        | 1        |
| <b>R</b> | 1        | 2        | 1        |

(d) Tipo-2, com 2 variáveis (Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 705      | 1        | 0        |
| <b>Y</b> | 5        | 9        | 0        |
| <b>R</b> | 1        | 3        | 0        |

(f) Tipo-2, com 4 variáveis e (Fe, Zn, Nit, Ful) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 705      | 1        | 0        |
| <b>Y</b> | 4        | 10       | 0        |
| <b>R</b> | 1        | 1        | 2        |

(h) Tipo-1, com 3 variáveis (Sul, Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 703      | 3        | 0        |
| <b>Y</b> | 5        | 9        | 0        |
| <b>R</b> | 0        | 2        | 2        |

(j) Tipo-2, com 2 variáveis (Zn, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 704      | 2        | 0        |
| <b>Y</b> | 4        | 10       | 0        |
| <b>R</b> | 0        | 2        | 2        |

(l) Tipo-2, com 4 variáveis e (Cu, Ti, Ful, Agu) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 530      | 0        | 0        |
| <b>Y</b> | 3        | 7        | 0        |
| <b>R</b> | 0        | 2        | 1        |

Tabela A.7: Valor médio da sensibilidade para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 36.40%        | 36.50%        | 0.10%            |
| <b>3 Variáveis</b> | 39.14%        | 39.41%        | 0.26%            |
| <b>4 Variáveis</b> | 40.38%        | 42.31%        | 1.93%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 39.27%        | 39.61%        | 0.35%            |
| <b>3 Variáveis</b> | 41.41%        | 42.12%        | 0.71%            |
| <b>4 Variáveis</b> | 41.61%        | 43.06%        | 1.45%            |

Tabela A.8: Valor médio da especificidade para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 69.98%        | 70.15%        | 0.17%            |
| <b>3 Variáveis</b> | 72.25%        | 72.16%        | -0.09%           |
| <b>4 Variáveis</b> | 73.01%        | 74.23%        | 1.22%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 72.30%        | 72.36%        | 0.07%            |
| <b>3 Variáveis</b> | 73.30%        | 74.54%        | 1.24%            |
| <b>4 Variáveis</b> | 74.49%        | 75.63%        | 1.14%            |

Tabela A.9: Valor médio do PPV para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 43.46%        | 45.20%        | 1.74%            |
| <b>3 Variáveis</b> | 48.05%        | 51.09%        | 3.05%            |
| <b>4 Variáveis</b> | 48.76%        | 54.92%        | 6.16%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 47.61%        | 47.65%        | 0.04%            |
| <b>3 Variáveis</b> | 48.85%        | 49.23%        | 0.37%            |
| <b>4 Variáveis</b> | 45.38%        | 45.13%        | -0.25%           |

Tabela A.10: Valor médio do NPV para o problema de Combustão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 80.05%        | 82.66%        | 2.61%            |
| <b>3 Variáveis</b> | 82.85%        | 85.15%        | 2.30%            |
| <b>4 Variáveis</b> | 81.10%        | 86.14%        | 5.03%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 82.73%        | 81.30%        | -1.43%           |
| <b>3 Variáveis</b> | 79.35%        | 82.60%        | 3.25%            |
| <b>4 Variáveis</b> | 78.88%        | 78.36%        | -0.52%           |

## Para o problema de contaminação sem uso do SMOTE

Tabela A.11: Valor máximo da taxa de acerto para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 96.96%        | 97.10%        | 0.14%            |
| <b>3 Variáveis</b> | 97.10%        | 97.42%        | 0.32%            |
| <b>4 Variáveis</b> | 97.51%        | 97.24%        | -0.28%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 97.38%        | 97.24%        | -0.14%           |
| <b>3 Variáveis</b> | 97.65%        | 97.79%        | 0.14%            |
| <b>4 Variáveis</b> | 97.38%        | 97.79%        | 0.41%            |

Tabela A.12: Valor médio da taxa de acerto para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 93.97%        | 95.59%        | 1.62%            |
| <b>3 Variáveis</b> | 93.05%        | 93.84%        | 0.78%            |
| <b>4 Variáveis</b> | 91.97%        | 90.56%        | -1.40%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 93.89%        | 93.02%        | -0.87%           |
| <b>3 Variáveis</b> | 92.82%        | 86.60%        | -6.22%           |
| <b>4 Variáveis</b> | 91.73%        | 81.03%        | -10.69%          |

Tabela A.13: Resultado dos melhores classificadores para o problema de Contaminação, avaliado pela taxa de acerto

(a) Tipo-1, com 2 variáveis (Ca, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 694      | 0        | 0        |
| <b>Y</b> | 7        | 1        | 0        |
| <b>R</b> | 11       | 4        | 7        |

(c) Tipo-1, com 4 variáveis (Mg, P, Agu, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 693      | 0        | 1        |
| <b>Y</b> | 5        | 3        | 0        |
| <b>R</b> | 9        | 3        | 10       |

(e) Tipo-2, com 3 variáveis (Ca, Ti, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 521      | 0        | 0        |
| <b>Y</b> | 5        | 1        | 0        |
| <b>R</b> | 7        | 2        | 7        |

(g) Tipo-1, com 2 variáveis (Al, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 693      | 1        | 0        |
| <b>Y</b> | 5        | 2        | 1        |
| <b>R</b> | 9        | 3        | 10       |

(i) Tipo-1, com 4 variáveis (Fe, Ba, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 691      | 3        | 0        |
| <b>Y</b> | 4        | 4        | 0        |
| <b>R</b> | 8        | 4        | 10       |

(k) Tipo-2, com 3 variáveis (Ti, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 520      | 0        | 0        |
| <b>Y</b> | 4        | 2        | 0        |
| <b>R</b> | 5        | 3        | 9        |

(b) Tipo-1, com 3 variáveis (Cr, Na, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 693      | 1        | 0        |
| <b>Y</b> | 6        | 2        | 0        |
| <b>R</b> | 13       | 1        | 8        |

(d) Tipo-2, com 2 variáveis (Sn, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 694      | 0        | 0        |
| <b>Y</b> | 7        | 1        | 0        |
| <b>R</b> | 13       | 1        | 8        |

(f) Tipo-2, com 4 variáveis e (Fe, Al, Ti, Dil) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 520      | 1        | 0        |
| <b>Y</b> | 3        | 3        | 0        |
| <b>R</b> | 7        | 4        | 5        |

(h) Tipo-1, com 3 variáveis (Sn, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 694      | 0        | 0        |
| <b>Y</b> | 5        | 2        | 1        |
| <b>R</b> | 8        | 3        | 11       |

(j) Tipo-2, com 2 variáveis (Zn, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 694      | 0        | 0        |
| <b>Y</b> | 7        | 1        | 0        |
| <b>R</b> | 10       | 3        | 9        |

(l) Tipo-2, com 4 variáveis e (Fe, Ba, Ti, Dil) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 520      | 1        | 0        |
| <b>Y</b> | 4        | 2        | 0        |
| <b>R</b> | 6        | 1        | 9        |

Tabela A.14: Valor máximo pela matriz de pesos para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.7638        | 0.7804        | 0.0166           |
| <b>3 Variáveis</b> | 0.8011        | 0.8085        | 0.0074           |
| <b>4 Variáveis</b> | 0.8287        | 0.8204        | -0.0083          |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.8384        | 0.8135        | -0.0249          |
| <b>3 Variáveis</b> | 0.8702        | 0.8582        | -0.0120          |
| <b>4 Variáveis</b> | 0.8550        | 0.8729        | 0.0180           |

Tabela A.15: Valor médio pela matriz de pesos para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.5709        | 0.6174        | 0.0465           |
| <b>3 Variáveis</b> | 0.5632        | 0.5816        | 0.0184           |
| <b>4 Variáveis</b> | 0.5444        | 0.5067        | -0.0377          |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.5786        | 0.5508        | -0.0278          |
| <b>3 Variáveis</b> | 0.5624        | 0.3432        | -0.2192          |
| <b>4 Variáveis</b> | 0.5467        | 0.1917        | -0.3551          |

Tabela A.16: Resultado dos melhores classificadores para o problema de Contaminação, avaliado pela matriz de pesos

(a) Tipo-1, com 2 variáveis (Al, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 693      | 1        | 0        |
| <b>Y</b> | 6        | 2        | 0        |
| <b>R</b> | 10       | 6        | 6        |

(c) Tipo-1, com 4 variáveis (Si, Ca, Agu, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 688      | 5        | 1        |
| <b>Y</b> | 3        | 4        | 1        |
| <b>R</b> | 7        | 5        | 10       |

(e) Tipo-2, com 3 variáveis (Ca, Ti, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 521      | 0        | 0        |
| <b>Y</b> | 5        | 1        | 0        |
| <b>R</b> | 7        | 2        | 7        |

(g) Tipo-1, com 2 variáveis (Si, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 692      | 2        | 0        |
| <b>Y</b> | 6        | 2        | 0        |
| <b>R</b> | 5        | 7        | 10       |

(i) Tipo-1, com 4 variáveis (Cu, Al, Sul, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 691      | 1        | 2        |
| <b>Y</b> | 6        | 0        | 2        |
| <b>R</b> | 6        | 2        | 14       |

(k) Tipo-2, com 3 variáveis (Ti, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 520      | 0        | 0        |
| <b>Y</b> | 4        | 2        | 0        |
| <b>R</b> | 5        | 3        | 9        |

(b) Tipo-1, com 3 variáveis (Si, Sul, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 690      | 4        | 0        |
| <b>Y</b> | 6        | 2        | 0        |
| <b>R</b> | 6        | 8        | 8        |

(d) Tipo-2, com 2 variáveis (Ca, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 694      | 0        | 0        |
| <b>Y</b> | 7        | 1        | 0        |
| <b>R</b> | 10       | 4        | 8        |

(f) Tipo-2, com 4 variáveis e (Al, Si, Sul, Dil) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 676      | 13       | 5        |
| <b>Y</b> | 4        | 2        | 2        |
| <b>R</b> | 2        | 7        | 13       |

(h) Tipo-1, com 3 variáveis (Si, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 692      | 2        | 0        |
| <b>Y</b> | 4        | 3        | 1        |
| <b>R</b> | 4        | 7        | 11       |

(j) Tipo-2, com 2 variáveis (Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 693      | 0        | 1        |
| <b>Y</b> | 6        | 1        | 1        |
| <b>R</b> | 8        | 4        | 10       |

(l) Tipo-2, com 4 variáveis e (Si, Ba, Ti, Dil) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 520      | 1        | 0        |
| <b>Y</b> | 4        | 2        | 0        |
| <b>R</b> | 4        | 3        | 9        |

Tabela A.17: Valor médio da sensibilidade para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 36.56%        | 36.72%        | 0.15%            |
| <b>3 Variáveis</b> | 38.24%        | 38.14%        | -0.10%           |
| <b>4 Variáveis</b> | 38.20%        | 39.03%        | 0.83%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 37.48%        | 37.17%        | -0.31%           |
| <b>3 Variáveis</b> | 38.69%        | 38.00%        | -0.69%           |
| <b>4 Variáveis</b> | 41.37%        | 38.49%        | -2.88%           |

Tabela A.18: Valor médio da especificidade para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 69.91%        | 70.20%        | 0.29%            |
| <b>3 Variáveis</b> | 72.10%        | 72.34%        | 0.24%            |
| <b>4 Variáveis</b> | 73.50%        | 74.37%        | 0.87%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 70.83%        | 70.99%        | 0.16%            |
| <b>3 Variáveis</b> | 72.58%        | 71.75%        | -0.83%           |
| <b>4 Variáveis</b> | 74.45%        | 72.48%        | -1.97%           |

Tabela A.19: Valor médio do PPV para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 39.85%        | 41.61%        | 1.76%            |
| <b>3 Variáveis</b> | 48.61%        | 44.06%        | -4.55%           |
| <b>4 Variáveis</b> | 49.35%        | 48.14%        | -1.21%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 48.99%        | 46.68%        | -2.31%           |
| <b>3 Variáveis</b> | 50.62%        | 45.73%        | -4.89%           |
| <b>4 Variáveis</b> | 51.53%        | 44.55%        | -6.99%           |



Tabela A.20: Valor médio do NPV para o problema de Contaminação

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 77.95%        | 79.68%        | 1.73%            |
| <b>3 Variáveis</b> | 81.15%        | 79.77%        | -1.38%           |
| <b>4 Variáveis</b> | 79.92%        | 78.10%        | -1.82%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 79.91%        | 77.48%        | -2.42%           |
| <b>3 Variáveis</b> | 79.49%        | 75.46%        | -4.03%           |
| <b>4 Variáveis</b> | 78.39%        | 72.92%        | -5.47%           |

## Para o problema de corrosão sem uso do SMOTE

Tabela A.21: Valor máximo da taxa de acerto para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 91.99%        | 92.27%        | 0.28%            |
| <b>3 Variáveis</b> | 92.54%        | 92.54%        | 0.00%            |
| <b>4 Variáveis</b> | 93.37%        | 92.82%        | -0.55%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 92.40%        | 91.99%        | -0.41%           |
| <b>3 Variáveis</b> | 92.82%        | 92.45%        | -0.37%           |
| <b>4 Variáveis</b> | 92.82%        | 92.68%        | -0.14%           |

Tabela A.22: Valor médio da taxa de acerto para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 87.30%        | 88.57%        | 1.27%            |
| <b>3 Variáveis</b> | 86.85%        | 88.68%        | 1.84%            |
| <b>4 Variáveis</b> | 86.45%        | 88.87%        | 2.41%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 87.19%        | 88.52%        | 1.33%            |
| <b>3 Variáveis</b> | 86.75%        | 88.46%        | 1.71%            |
| <b>4 Variáveis</b> | 86.18%        | 88.30%        | 2.12%            |

Tabela A.23: Resultado dos melhores classificadores para o problema de Corrosão, avaliado pela taxa de acerto

(a) Tipo-1, com 2 variáveis (Fe, Si) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 630      | 11       | 0        |
| <b>Y</b> | 31       | 33       | 0        |
| <b>R</b> | 1        | 15       | 3        |

(c) Tipo-1, com 4 variáveis (Fe, Cu, Ni, Ca) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 634      | 7        | 0        |
| <b>Y</b> | 30       | 34       | 0        |
| <b>R</b> | 1        | 10       | 8        |

(e) Tipo-2, com 3 variáveis (Fe, Cu, Ni) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 628      | 13       | 0        |
| <b>Y</b> | 28       | 36       | 0        |
| <b>R</b> | 0        | 13       | 6        |

(g) Tipo-1, com 2 variáveis (Fe, Si) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 629      | 11       | 1        |
| <b>Y</b> | 31       | 31       | 2        |
| <b>R</b> | 2        | 8        | 9        |

(i) Tipo-1, com 4 variáveis (Fe, Cu, Mg, P) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 631      | 10       | 0        |
| <b>Y</b> | 34       | 29       | 1        |
| <b>R</b> | 2        | 5        | 12       |

(k) Tipo-2, com 3 variáveis (Fe, Sn, Ti) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 471      | 9        | 0        |
| <b>Y</b> | 24       | 24       | 0        |
| <b>R</b> | 0        | 8        | 7        |

(b) Tipo-1, com 3 variáveis (Fe, Sn, Ni) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 630      | 11       | 0        |
| <b>Y</b> | 29       | 35       | 0        |
| <b>R</b> | 2        | 12       | 5        |

(d) Tipo-2, com 2 variáveis (Fe, Mg) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 628      | 13       | 0        |
| <b>Y</b> | 27       | 37       | 0        |
| <b>R</b> | 0        | 16       | 3        |

(f) Tipo-2, com 4 variáveis e (Fe, Pb, Sn, P) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 627      | 14       | 0        |
| <b>Y</b> | 26       | 37       | 1        |
| <b>R</b> | 1        | 10       | 8        |

(h) Tipo-1, com 3 variáveis (Fe, Cu, Ca) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 632      | 9        | 0        |
| <b>Y</b> | 33       | 30       | 1        |
| <b>R</b> | 2        | 7        | 10       |

(j) Tipo-2, com 2 variáveis (Fe, Ni) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 625      | 16       | 0        |
| <b>Y</b> | 28       | 36       | 0        |
| <b>R</b> | 2        | 12       | 5        |

(l) Tipo-2, com 4 variáveis e (Fe, Al, Nit, Dil) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 631      | 10       | 0        |
| <b>Y</b> | 28       | 33       | 3        |
| <b>R</b> | 4        | 8        | 7        |

Tabela A.24: Valor máximo pela matriz de pesos para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.6215        | 0.6492        | 0.0276           |
| <b>3 Variáveis</b> | 0.6561        | 0.6809        | 0.0249           |
| <b>4 Variáveis</b> | 0.6934        | 0.7072        | 0.0138           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.6533        | 0.6533        | 0.0000           |
| <b>3 Variáveis</b> | 0.6657        | 0.6667        | 0.0009           |
| <b>4 Variáveis</b> | 0.6740        | 0.6943        | 0.0203           |

Tabela A.25: Valor médio pela matriz de pesos para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.2719        | 0.3069        | 0.0350           |
| <b>3 Variáveis</b> | 0.3063        | 0.3642        | 0.0579           |
| <b>4 Variáveis</b> | 0.3331        | 0.4175        | 0.0843           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.2814        | 0.3177        | 0.0363           |
| <b>3 Variáveis</b> | 0.3108        | 0.3624        | 0.0516           |
| <b>4 Variáveis</b> | 0.3245        | 0.3871        | 0.0626           |

Tabela A.26: Resultado dos melhores classificadores para o problema de Corrosão, avaliado pela matriz de pesos

(a) Tipo-1, com 2 variáveis (Fe, Mg) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 626      | 15       | 0        |
| <b>Y</b> | 29       | 35       | 0        |
| <b>R</b> | 0        | 16       | 3        |

(c) Tipo-1, com 4 variáveis (Fe, Cu, Ni, Ca) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 634      | 7        | 0        |
| <b>Y</b> | 30       | 34       | 0        |
| <b>R</b> | 1        | 10       | 8        |

(e) Tipo-2, com 3 variáveis (Fe, Cu, Sn) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 626      | 15       | 0        |
| <b>Y</b> | 28       | 35       | 1        |
| <b>R</b> | 0        | 11       | 8        |

(g) Tipo-1, com 2 variáveis (Fe, Si) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 629      | 11       | 1        |
| <b>Y</b> | 31       | 31       | 2        |
| <b>R</b> | 2        | 8        | 9        |

(i) Tipo-1, com 4 variáveis (Fe, Cu, Mg, P) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 631      | 10       | 0        |
| <b>Y</b> | 34       | 29       | 1        |
| <b>R</b> | 2        | 5        | 12       |

(k) Tipo-2, com 3 variáveis (Fe, Sn, Ti) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 471      | 9        | 0        |
| <b>Y</b> | 24       | 24       | 0        |
| <b>R</b> | 0        | 8        | 7        |

(b) Tipo-1, com 3 variáveis (Fe, Sn, Si) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 629      | 11       | 1        |
| <b>Y</b> | 31       | 31       | 2        |
| <b>R</b> | 0        | 11       | 8        |

(d) Tipo-2, com 2 variáveis (Fe, Mg) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 628      | 13       | 0        |
| <b>Y</b> | 27       | 37       | 0        |
| <b>R</b> | 0        | 16       | 3        |

(f) Tipo-2, com 4 variáveis e (Fe, Cu, Mg, Mo) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 627      | 14       | 0        |
| <b>Y</b> | 28       | 33       | 3        |
| <b>R</b> | 0        | 8        | 11       |

(h) Tipo-1, com 3 variáveis (Fe, Cu, Ca) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 632      | 9        | 0        |
| <b>Y</b> | 33       | 30       | 1        |
| <b>R</b> | 2        | 7        | 10       |

(j) Tipo-2, com 2 variáveis (Fe, Si) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 624      | 16       | 1        |
| <b>Y</b> | 30       | 32       | 2        |
| <b>R</b> | 2        | 7        | 10       |

(l) Tipo-2, com 4 variáveis e (Fe, Ca, Ti, Sul) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 466      | 14       | 1        |
| <b>Y</b> | 21       | 23       | 4        |
| <b>R</b> | 0        | 4        | 10       |

Tabela A.27: Valor médio da sensibilidade para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 39.98%        | 40.27%        | 0.29%            |
| <b>3 Variáveis</b> | 43.55%        | 44.62%        | 1.06%            |
| <b>4 Variáveis</b> | 46.85%        | 49.07%        | 2.22%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 41.59%        | 42.07%        | 0.49%            |
| <b>3 Variáveis</b> | 45.08%        | 46.43%        | 1.35%            |
| <b>4 Variáveis</b> | 47.98%        | 49.31%        | 1.33%            |

Tabela A.28: Valor médio da especificidade para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 73.08%        | 73.33%        | 0.26%            |
| <b>3 Variáveis</b> | 75.31%        | 76.15%        | 0.84%            |
| <b>4 Variáveis</b> | 77.05%        | 78.52%        | 1.47%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 73.60%        | 73.78%        | 0.18%            |
| <b>3 Variáveis</b> | 75.62%        | 76.16%        | 0.54%            |
| <b>4 Variáveis</b> | 77.04%        | 77.71%        | 0.67%            |

Tabela A.29: Valor médio do PPV para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 55.79%        | 56.62%        | 0.83%            |
| <b>3 Variáveis</b> | 64.25%        | 63.63%        | -0.63%           |
| <b>4 Variáveis</b> | 66.48%        | 65.62%        | -0.86%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 56.84%        | 56.44%        | -0.40%           |
| <b>3 Variáveis</b> | 60.91%        | 60.42%        | -0.48%           |
| <b>4 Variáveis</b> | 60.46%        | 59.74%        | -0.73%           |

Tabela A.30: Valor médio do NPV para o problema de Corrosão

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 80.87%        | 81.15%        | 0.29%            |
| <b>3 Variáveis</b> | 82.60%        | 83.41%        | 0.81%            |
| <b>4 Variáveis</b> | 83.05%        | 84.20%        | 1.15%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 81.25%        | 81.82%        | 0.57%            |
| <b>3 Variáveis</b> | 82.30%        | 82.96%        | 0.66%            |
| <b>4 Variáveis</b> | 82.18%        | 82.91%        | 0.72%            |

## Para o problema de combustão usando SMOTE

Tabela A.31: Valor máximo da taxa de acerto para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 94.48%        | 93.65%        | -0.83%           |
| <b>3 Variáveis</b> | 94.75%        | 94.61%        | -0.14%           |
| <b>4 Variáveis</b> | 95.58%        | 95.58%        | 0.00%            |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 96.82%        | 96.13%        | -0.69%           |
| <b>3 Variáveis</b> | 97.24%        | 96.82%        | -0.41%           |
| <b>4 Variáveis</b> | 97.79%        | 97.79%        | 0.00%            |

Tabela A.32: Valor médio da taxa de acerto para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 38.34%        | 42.01%        | 3.67%            |
| <b>3 Variáveis</b> | 49.72%        | 57.61%        | 7.89%            |
| <b>4 Variáveis</b> | 55.36%        | 68.05%        | 12.69%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 50.07%        | 52.07%        | 2.00%            |
| <b>3 Variáveis</b> | 63.57%        | 65.67%        | 2.10%            |
| <b>4 Variáveis</b> | 72.73%        | 72.03%        | -0.70%           |



Tabela A.33: Resultado dos melhores classificadores para o problema de Combustão, usando SMOTE, avaliado pela taxa de acerto

(a) Tipo-1, com 2 variáveis (Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 669      | 37       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(c) Tipo-1, com 4 variáveis (Si, Nit, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 677      | 29       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(e) Tipo-2, com 3 variáveis (Sul, Ful, Agu) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 670      | 36       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(g) Tipo-1, com 2 variáveis (Fe, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 686      | 20       | 0        |
| <b>Y</b> | 2        | 12       | 0        |
| <b>R</b> | 0        | 1        | 3        |

(i) Tipo-1, com 4 variáveis (Fe, Pb, Ca, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 691      | 15       | 0        |
| <b>Y</b> | 0        | 14       | 0        |
| <b>R</b> | 0        | 1        | 3        |

(k) Tipo-2, com 3 variáveis (Fe, Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 685      | 21       | 0        |
| <b>Y</b> | 1        | 13       | 0        |
| <b>R</b> | 0        | 1        | 3        |

(b) Tipo-1, com 3 variáveis (Nit, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 672      | 34       | 0        |
| <b>Y</b> | 1        | 11       | 2        |
| <b>R</b> | 0        | 1        | 3        |

(d) Tipo-2, com 2 variáveis (Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 665      | 41       | 0        |
| <b>Y</b> | 1        | 10       | 3        |
| <b>R</b> | 0        | 1        | 3        |

(f) Tipo-2, com 4 variáveis e (Al, P, Sul, Ful) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 679      | 26       | 1        |
| <b>Y</b> | 3        | 10       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(h) Tipo-1, com 3 variáveis (Na, Mg, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 689      | 17       | 0        |
| <b>Y</b> | 1        | 13       | 0        |
| <b>R</b> | 0        | 2        | 2        |

(j) Tipo-2, com 2 variáveis (Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 681      | 25       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(l) Tipo-2, com 4 variáveis e (Fe, Si, Ful, Dil) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 696      | 10       | 0        |
| <b>Y</b> | 4        | 10       | 0        |
| <b>R</b> | 0        | 2        | 2        |

Tabela A.34: Valor máximo pela matriz de pesos para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.8508        | 0.8232        | -0.0276          |
| <b>3 Variáveis</b> | 0.8577        | 0.8550        | -0.0028          |
| <b>4 Variáveis</b> | 0.8840        | 0.8881        | 0.0041           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.9213        | 0.9006        | -0.0207          |
| <b>3 Variáveis</b> | 0.9434        | 0.9227        | -0.0207          |
| <b>4 Variáveis</b> | 0.9572        | 0.9489        | -0.0083          |

Tabela A.35: Valor médio pela matriz de pesos para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | -0.8617       | -0.7529       | 0.1089           |
| <b>3 Variáveis</b> | -0.5204       | -0.2843       | 0.2361           |
| <b>4 Variáveis</b> | -0.3496       | 0.0287        | 0.3783           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | -0.5213       | -0.4638       | 0.0575           |
| <b>3 Variáveis</b> | -0.1159       | -0.0600       | 0.0559           |
| <b>4 Variáveis</b> | 0.1580        | 0.0986        | -0.0594          |

Tabela A.36: Resultado dos melhores classificadores para o problema de Combustão, usando SMOTE, avaliado pela matriz de pesos

(a) Tipo-1, com 2 variáveis (Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 669      | 37       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(c) Tipo-1, com 4 variáveis (Si, Nit, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 677      | 29       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(e) Tipo-2, com 3 variáveis (Sul, Ful, Agu) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 670      | 36       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(g) Tipo-1, com 2 variáveis (Zn, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 682      | 24       | 0        |
| <b>Y</b> | 1        | 13       | 0        |
| <b>R</b> | 0        | 0        | 4        |

(i) Tipo-1, com 4 variáveis (Fe, Pb, Ca, Ful) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 691      | 15       | 0        |
| <b>Y</b> | 0        | 14       | 0        |
| <b>R</b> | 0        | 1        | 3        |

(k) Tipo-2, com 3 variáveis (Fe, Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 685      | 21       | 0        |
| <b>Y</b> | 1        | 13       | 0        |
| <b>R</b> | 0        | 1        | 3        |

(b) Tipo-1, com 3 variáveis (Na, Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 668      | 38       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 0        | 4        |

(d) Tipo-2, com 2 variáveis (Sul, Ful) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 665      | 41       | 0        |
| <b>Y</b> | 1        | 10       | 3        |
| <b>R</b> | 0        | 1        | 3        |

(f) Tipo-2, com 4 variáveis e (Mg, Nit, Sul, Ful) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 678      | 28       | 0        |
| <b>Y</b> | 1        | 10       | 3        |
| <b>R</b> | 0        | 0        | 4        |

(h) Tipo-1, com 3 variáveis (Zn, Ful, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 685      | 21       | 0        |
| <b>Y</b> | 0        | 14       | 0        |
| <b>R</b> | 0        | 0        | 4        |

(j) Tipo-2, com 2 variáveis (Ful, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 681      | 25       | 0        |
| <b>Y</b> | 1        | 12       | 1        |
| <b>R</b> | 0        | 1        | 3        |

(l) Tipo-2, com 4 variáveis e (Pb, Ful, Agu, Dil) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 689      | 17       | 0        |
| <b>Y</b> | 0        | 14       | 0        |
| <b>R</b> | 0        | 1        | 3        |

Tabela A.37: Valor médio da sensibilidade para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 48.15%        | 45.35%        | -2.80%           |
| <b>3 Variáveis</b> | 53.97%        | 49.44%        | -4.53%           |
| <b>4 Variáveis</b> | 58.04%        | 51.36%        | -6.68%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 53.41%        | 49.49%        | -3.91%           |
| <b>3 Variáveis</b> | 61.25%        | 53.04%        | -8.21%           |
| <b>4 Variáveis</b> | 66.85%        | 52.78%        | -14.07%          |

Tabela A.38: Valor médio da especificidade para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 76.95%        | 71.05%        | -5.90%           |
| <b>3 Variáveis</b> | 80.70%        | 71.86%        | -8.84%           |
| <b>4 Variáveis</b> | 82.74%        | 70.92%        | -11.82%          |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 80.37%        | 73.68%        | -6.69%           |
| <b>3 Variáveis</b> | 84.91%        | 73.63%        | -11.28%          |
| <b>4 Variáveis</b> | 87.93%        | 70.97%        | -16.95%          |

Tabela A.39: Valor médio do PPV para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 35.39%        | 35.13%        | -0.26%           |
| <b>3 Variáveis</b> | 36.55%        | 36.46%        | -0.09%           |
| <b>4 Variáveis</b> | 35.93%        | 35.64%        | -0.30%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 39.81%        | 35.98%        | -3.84%           |
| <b>3 Variáveis</b> | 42.40%        | 36.05%        | -6.35%           |
| <b>4 Variáveis</b> | 43.32%        | 33.94%        | -9.38%           |

Tabela A.40: Valor médio do NPV para o problema de Combustão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 64.19%        | 61.15%        | -3.05%           |
| <b>3 Variáveis</b> | 64.07%        | 59.50%        | -4.58%           |
| <b>4 Variáveis</b> | 63.21%        | 57.08%        | -6.13%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 68.45%        | 62.51%        | -5.94%           |
| <b>3 Variáveis</b> | 69.83%        | 60.16%        | -9.67%           |
| <b>4 Variáveis</b> | 70.69%        | 57.43%        | -13.26%          |

## Para o problema de contaminação usando SMOTE

Tabela A.41: Valor máximo da taxa de acerto para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 59.67%        | 45.44%        | -14.23%          |
| <b>3 Variáveis</b> | 76.80%        | 67.82%        | -8.98%           |
| <b>4 Variáveis</b> | 84.81%        | 75.55%        | -9.25%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 73.90%        | 59.12%        | -14.78%          |
| <b>3 Variáveis</b> | 82.87%        | 91.99%        | 9.12%            |
| <b>4 Variáveis</b> | 86.46%        | 94.61%        | 8.15%            |

Tabela A.42: Valor médio da taxa de acerto para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 8.20%         | 8.32%         | 0.13%            |
| <b>3 Variáveis</b> | 20.40%        | 19.03%        | -1.37%           |
| <b>4 Variáveis</b> | 33.77%        | 29.81%        | -3.96%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 16.58%        | 16.20%        | -0.38%           |
| <b>3 Variáveis</b> | 32.31%        | 29.94%        | -2.36%           |
| <b>4 Variáveis</b> | 46.26%        | 43.24%        | -3.02%           |

Tabela A.43: Resultado dos melhores classificadores para o problema de Contaminação, usando SMOTE, avaliado pela taxa de acerto

(a) Tipo-1, com 2 variáveis (Si, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 415      | 276      | 3        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 10       | 12       |

(c) Tipo-1, com 4 variáveis (Si, Ca, Agu, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 593      | 97       | 4        |
| <b>Y</b> | 1        | 6        | 1        |
| <b>R</b> | 0        | 7        | 15       |

(e) Tipo-2, com 3 variáveis (Pb, Cu, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 470      | 214      | 10       |
| <b>Y</b> | 2        | 6        | 0        |
| <b>R</b> | 1        | 6        | 15       |

(g) Tipo-1, com 2 variáveis (Si, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 516      | 174      | 4        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 8        | 14       |

(i) Tipo-1, com 4 variáveis (Si, Mg, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 603      | 86       | 5        |
| <b>Y</b> | 2        | 6        | 0        |
| <b>R</b> | 0        | 5        | 17       |

(k) Tipo-2, com 3 variáveis (Al, Si, Ba) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 656      | 31       | 7        |
| <b>Y</b> | 5        | 2        | 1        |
| <b>R</b> | 13       | 1        | 8        |

(b) Tipo-1, com 3 variáveis (Si, Ca, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 536      | 154      | 4        |
| <b>Y</b> | 1        | 6        | 1        |
| <b>R</b> | 0        | 8        | 14       |

(d) Tipo-2, com 2 variáveis (Si, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 310      | 380      | 4        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 8        | 14       |

(f) Tipo-2, com 4 variáveis e (Cu, Si, P, Dil) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 523      | 164      | 7        |
| <b>Y</b> | 0        | 7        | 1        |
| <b>R</b> | 0        | 5        | 17       |

(h) Tipo-1, com 3 variáveis (Si, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 580      | 111      | 3        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 7        | 15       |

(j) Tipo-2, com 2 variáveis (Al, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 410      | 281      | 3        |
| <b>Y</b> | 1        | 6        | 1        |
| <b>R</b> | 1        | 9        | 12       |

(l) Tipo-2, com 4 variáveis e (Si, Ba, Agu, Dil) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 670      | 24       | 0        |
| <b>Y</b> | 3        | 5        | 0        |
| <b>R</b> | 3        | 9        | 10       |

Tabela A.44: Valor máximo pela matriz de pesos para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | -0.2238       | -0.6395       | -0.4157          |
| <b>3 Variáveis</b> | 0.3025        | 0.0124        | -0.2901          |
| <b>4 Variáveis</b> | 0.5497        | 0.2831        | -0.2666          |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.2141        | -0.2459       | -0.4599          |
| <b>3 Variáveis</b> | 0.4931        | 0.5967        | 0.1036           |
| <b>4 Variáveis</b> | 0.6064        | 0.7901        | 0.1837           |

Tabela A.45: Valor médio pela matriz de pesos para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | -1.8295       | -1.8263       | 0.0031           |
| <b>3 Variáveis</b> | -1.4616       | -1.5050       | -0.0434          |
| <b>4 Variáveis</b> | -1.0565       | -1.1790       | -0.1225          |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | -1.5869       | -1.6006       | -0.0137          |
| <b>3 Variáveis</b> | -1.1100       | -1.1971       | -0.0871          |
| <b>4 Variáveis</b> | -0.6933       | -0.8294       | -0.1361          |



Tabela A.46: Resultado dos melhores classificadores para o problema de Contaminação, usando SMOTE, avaliado pela matriz de pesos

(a) Tipo-1, com 2 variáveis (Si, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 415      | 276      | 3        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 10       | 12       |

(c) Tipo-1, com 4 variáveis (Si, Ca, Agu, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 593      | 97       | 4        |
| <b>Y</b> | 1        | 6        | 1        |
| <b>R</b> | 0        | 7        | 15       |

(e) Tipo-2, com 3 variáveis (Pb, Cu, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 470      | 214      | 10       |
| <b>Y</b> | 2        | 6        | 0        |
| <b>R</b> | 1        | 6        | 15       |

(g) Tipo-1, com 2 variáveis (Si, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 516      | 174      | 4        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 8        | 14       |

(i) Tipo-1, com 4 variáveis (Si, Mg, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 603      | 86       | 5        |
| <b>Y</b> | 2        | 6        | 0        |
| <b>R</b> | 0        | 5        | 17       |

(k) Tipo-2, com 3 variáveis (Al, Si, Ba) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 656      | 31       | 7        |
| <b>Y</b> | 5        | 2        | 1        |
| <b>R</b> | 13       | 1        | 8        |

(b) Tipo-1, com 3 variáveis (Si, Ca, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 536      | 154      | 4        |
| <b>Y</b> | 1        | 6        | 1        |
| <b>R</b> | 0        | 8        | 14       |

(d) Tipo-2, com 2 variáveis (Si, Dil) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 310      | 380      | 4        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 8        | 14       |

(f) Tipo-2, com 4 variáveis e (Cu, Si, P, Dil) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 523      | 164      | 7        |
| <b>Y</b> | 0        | 7        | 1        |
| <b>R</b> | 0        | 5        | 17       |

(h) Tipo-1, com 3 variáveis (Si, Agu, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 580      | 111      | 3        |
| <b>Y</b> | 1        | 5        | 2        |
| <b>R</b> | 0        | 7        | 15       |

(j) Tipo-2, com 2 variáveis (Al, Dil) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 410      | 281      | 3        |
| <b>Y</b> | 1        | 6        | 1        |
| <b>R</b> | 1        | 9        | 12       |

(l) Tipo-2, com 4 variáveis e (Si, Ba, Agu, Dil) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 670      | 24       | 0        |
| <b>Y</b> | 3        | 5        | 0        |
| <b>R</b> | 3        | 9        | 10       |

Tabela A.47: Valor médio da sensibilidade para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 37.39%        | 38.48%        | 1.09%            |
| <b>3 Variáveis</b> | 44.12%        | 44.41%        | 0.29%            |
| <b>4 Variáveis</b> | 51.29%        | 50.82%        | -0.47%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 44.07%        | 43.99%        | -0.08%           |
| <b>3 Variáveis</b> | 51.62%        | 50.24%        | -1.39%           |
| <b>4 Variáveis</b> | 57.78%        | 54.18%        | -3.60%           |

Tabela A.48: Valor médio da especificidade para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 68.18%        | 68.13%        | -0.05%           |
| <b>3 Variáveis</b> | 71.39%        | 70.89%        | -0.50%           |
| <b>4 Variáveis</b> | 75.32%        | 74.10%        | -1.22%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 70.42%        | 70.29%        | -0.14%           |
| <b>3 Variáveis</b> | 74.88%        | 73.85%        | -1.03%           |
| <b>4 Variáveis</b> | 78.87%        | 76.76%        | -2.11%           |

Tabela A.49: Valor médio do PPV para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 34.73%        | 33.86%        | -0.88%           |
| <b>3 Variáveis</b> | 44.60%        | 42.98%        | -1.63%           |
| <b>4 Variáveis</b> | 47.96%        | 46.24%        | -1.72%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 40.83%        | 38.71%        | -2.12%           |
| <b>3 Variáveis</b> | 45.99%        | 43.26%        | -2.74%           |
| <b>4 Variáveis</b> | 46.20%        | 43.69%        | -2.52%           |

Tabela A.50: Valor médio do NPV para o problema de Contaminação, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 65.27%        | 65.77%        | 0.50%            |
| <b>3 Variáveis</b> | 67.09%        | 67.03%        | -0.06%           |
| <b>4 Variáveis</b> | 67.79%        | 67.64%        | -0.16%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 67.03%        | 67.08%        | 0.05%            |
| <b>3 Variáveis</b> | 67.80%        | 67.64%        | -0.16%           |
| <b>4 Variáveis</b> | 68.41%        | 68.16%        | -0.25%           |

## Para o problema de corrosão usando SMOTE

Tabela A.51: Valor máximo da taxa de acerto para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 82.32%        | 80.52%        | -1.80%           |
| <b>3 Variáveis</b> | 84.53%        | 83.70%        | -0.83%           |
| <b>4 Variáveis</b> | 86.33%        | 86.19%        | -0.14%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 84.81%        | 85.36%        | 0.55%            |
| <b>3 Variáveis</b> | 87.02%        | 88.54%        | 1.52%            |
| <b>4 Variáveis</b> | 88.40%        | 91.16%        | 2.76%            |

Tabela A.52: Valor médio da taxa de acerto para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 48.68%        | 45.29%        | -3.38%           |
| <b>3 Variáveis</b> | 59.78%        | 54.80%        | -4.98%           |
| <b>4 Variáveis</b> | 67.36%        | 60.68%        | -6.68%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 55.59%        | 53.78%        | -1.81%           |
| <b>3 Variáveis</b> | 65.49%        | 62.58%        | -2.91%           |
| <b>4 Variáveis</b> | 71.72%        | 67.03%        | -4.69%           |

Tabela A.53: Resultado dos melhores classificadores para o problema de Corrosão, usando SMOTE, avaliado pela taxa de acerto

(a) Tipo-1, com 2 variáveis (Fe, Pb) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 534      | 106      | 1        |
| <b>Y</b> | 7        | 48       | 9        |
| <b>R</b> | 0        | 5        | 14       |

(b) Tipo-1, com 3 variáveis (Fe, Pb, Na) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 549      | 91       | 1        |
| <b>Y</b> | 7        | 49       | 8        |
| <b>R</b> | 0        | 5        | 14       |

(c) Tipo-1, com 4 variáveis (Fe, Pb, Mg, P) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 561      | 79       | 1        |
| <b>Y</b> | 8        | 49       | 7        |
| <b>R</b> | 0        | 4        | 15       |

(d) Tipo-2, com 2 variáveis (Fe, Pb) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 521      | 119      | 1        |
| <b>Y</b> | 7        | 48       | 9        |
| <b>R</b> | 0        | 5        | 14       |

(e) Tipo-2, com 3 variáveis (Pb, Si, Agu) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 565      | 73       | 3        |
| <b>Y</b> | 28       | 29       | 7        |
| <b>R</b> | 2        | 5        | 12       |

(f) Tipo-2, com 4 variáveis e (Ca, Ba, Ful, Agu) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 605      | 31       | 5        |
| <b>Y</b> | 44       | 13       | 7        |
| <b>R</b> | 9        | 4        | 6        |

(g) Tipo-1, com 2 variáveis (Fe, Na) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 552      | 88       | 1        |
| <b>Y</b> | 8        | 48       | 8        |
| <b>R</b> | 0        | 5        | 14       |

(h) Tipo-1, com 3 variáveis (Fe, Cu, Si) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 575      | 65       | 1        |
| <b>Y</b> | 17       | 39       | 8        |
| <b>R</b> | 0        | 3        | 16       |

(i) Tipo-1, com 4 variáveis (Fe, Cu, Al, Na) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 575      | 64       | 2        |
| <b>Y</b> | 12       | 48       | 4        |
| <b>R</b> | 0        | 2        | 17       |

(j) Tipo-2, com 2 variáveis (Fe, Nit) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 560      | 81       | 0        |
| <b>Y</b> | 10       | 45       | 9        |
| <b>R</b> | 0        | 6        | 13       |

(k) Tipo-2, com 3 variáveis (Fe, Nit, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 587      | 53       | 1        |
| <b>Y</b> | 16       | 42       | 6        |
| <b>R</b> | 0        | 7        | 12       |

(l) Tipo-2, com 4 variáveis e (Fe, Al, Ca, Ti) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 613      | 27       | 1        |
| <b>Y</b> | 22       | 35       | 7        |
| <b>R</b> | 0        | 7        | 12       |

Tabela A.54: Valor máximo pela matriz de pesos para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.5221        | 0.4682        | -0.0539          |
| <b>3 Variáveis</b> | 0.5898        | 0.5428        | -0.0470          |
| <b>4 Variáveis</b> | 0.6464        | 0.6050        | -0.0414          |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 0.5925        | 0.5925        | 0.0000           |
| <b>3 Variáveis</b> | 0.6423        | 0.6492        | 0.0069           |
| <b>4 Variáveis</b> | 0.7017        | 0.6934        | -0.0083          |

Tabela A.55: Valor médio pela matriz de pesos para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | -0.5541       | -0.6473       | -0.0932          |
| <b>3 Variáveis</b> | -0.2236       | -0.3642       | -0.1405          |
| <b>4 Variáveis</b> | 0.0021        | -0.1879       | -0.1900          |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | -0.3654       | -0.4325       | -0.0671          |
| <b>3 Variáveis</b> | -0.0695       | -0.1753       | -0.1057          |
| <b>4 Variáveis</b> | 0.1155        | -0.0802       | -0.1957          |

Tabela A.56: Resultado dos melhores classificadores para o problema de Corrosão, usando SMOTE, avaliado pela matriz de pesos

(a) Tipo-1, com 2 variáveis (Fe, Pb) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 534      | 106      | 1        |
| <b>Y</b> | 7        | 48       | 9        |
| <b>R</b> | 0        | 5        | 14       |

(b) Tipo-1, com 3 variáveis (Fe, Pb, Na) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 549      | 91       | 1        |
| <b>Y</b> | 7        | 49       | 8        |
| <b>R</b> | 0        | 5        | 14       |

(c) Tipo-1, com 4 variáveis (Fe, Pb, Mg, P) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 561      | 79       | 1        |
| <b>Y</b> | 8        | 49       | 7        |
| <b>R</b> | 0        | 4        | 15       |

(d) Tipo-2, com 2 variáveis (Fe, Pb) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 521      | 119      | 1        |
| <b>Y</b> | 7        | 48       | 9        |
| <b>R</b> | 0        | 5        | 14       |

(e) Tipo-2, com 3 variáveis (Fe, Ni, Na) e 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 540      | 100      | 1        |
| <b>Y</b> | 8        | 48       | 8        |
| <b>R</b> | 0        | 5        | 14       |

(f) Tipo-2, com 4 variáveis e (Fe, Pb, Ca, Ba) 2 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 561      | 79       | 1        |
| <b>Y</b> | 10       | 47       | 7        |
| <b>R</b> | 0        | 6        | 13       |

(g) Tipo-1, com 2 variáveis (Fe, Na) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 552      | 88       | 1        |
| <b>Y</b> | 8        | 48       | 8        |
| <b>R</b> | 0        | 5        | 14       |

(h) Tipo-1, com 3 variáveis (Fe, Cu, Al) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 567      | 72       | 2        |
| <b>Y</b> | 13       | 44       | 7        |
| <b>R</b> | 0        | 2        | 17       |

(i) Tipo-1, com 4 variáveis (Fe, Cu, Al, Na) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 575      | 64       | 2        |
| <b>Y</b> | 12       | 48       | 4        |
| <b>R</b> | 0        | 2        | 17       |

(j) Tipo-2, com 2 variáveis (Fe, Nit) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 560      | 81       | 0        |
| <b>Y</b> | 10       | 45       | 9        |
| <b>R</b> | 0        | 6        | 13       |

(k) Tipo-2, com 3 variáveis (Fe, Nit, Agu) e 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 587      | 53       | 1        |
| <b>Y</b> | 16       | 42       | 6        |
| <b>R</b> | 0        | 7        | 12       |

(l) Tipo-2, com 4 variáveis e (Fe, Al, Ca, Ti) 3 funções de pertinência

|          | <b>G</b> | <b>Y</b> | <b>R</b> |
|----------|----------|----------|----------|
| <b>G</b> | 613      | 27       | 1        |
| <b>Y</b> | 22       | 35       | 7        |
| <b>R</b> | 0        | 7        | 12       |

Tabela A.57: Valor médio da sensibilidade para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 55.12%        | 54.39%        | -0.73%           |
| <b>3 Variáveis</b> | 60.61%        | 59.77%        | -0.84%           |
| <b>4 Variáveis</b> | 64.40%        | 63.37%        | -1.03%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 57.34%        | 56.64%        | -0.70%           |
| <b>3 Variáveis</b> | 63.06%        | 61.05%        | -2.01%           |
| <b>4 Variáveis</b> | 66.49%        | 62.21%        | -4.28%           |

Tabela A.58: Valor médio da especificidade para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 78.34%        | 77.58%        | -0.76%           |
| <b>3 Variáveis</b> | 81.56%        | 80.50%        | -1.06%           |
| <b>4 Variáveis</b> | 83.67%        | 82.35%        | -1.33%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 79.78%        | 79.37%        | -0.41%           |
| <b>3 Variáveis</b> | 82.56%        | 81.59%        | -0.97%           |
| <b>4 Variáveis</b> | 84.21%        | 82.10%        | -2.11%           |



Tabela A.59: Valor médio do PPV para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 50.28%        | 48.06%        | -2.22%           |
| <b>3 Variáveis</b> | 51.89%        | 50.47%        | -1.42%           |
| <b>4 Variáveis</b> | 52.79%        | 50.99%        | -1.80%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 48.42%        | 47.68%        | -0.74%           |
| <b>3 Variáveis</b> | 51.09%        | 49.89%        | -1.20%           |
| <b>4 Variáveis</b> | 52.61%        | 49.85%        | -2.76%           |

Tabela A.60: Valor médio do NPV para o problema de Corrosão, usando SMOTE

(a) Com 2 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 71.15%        | 70.65%        | -0.50%           |
| <b>3 Variáveis</b> | 72.65%        | 71.92%        | -0.72%           |
| <b>4 Variáveis</b> | 73.98%        | 72.78%        | -1.20%           |

(b) Com 3 funções de pertinência

|                    | <b>Tipo-1</b> | <b>Tipo-2</b> | <b>Diferença</b> |
|--------------------|---------------|---------------|------------------|
| <b>2 Variáveis</b> | 71.86%        | 71.47%        | -0.39%           |
| <b>3 Variáveis</b> | 73.54%        | 72.89%        | -0.65%           |
| <b>4 Variáveis</b> | 74.85%        | 73.69%        | -1.17%           |

## Diesel Engines Diagnosis Through Analysis of Lubricating Oil

Ciro E. L. Sobral

Programa de Pós-Graduação  
em Informática

Universidade Federal do Rio de Janeiro Universidade Federal do Rio de Janeiro Universidade Federal do Rio Grande do Norte  
Rio de Janeiro-RJ, Brazil Rio de Janeiro-RJ, Brazil Natal-RN, Brazil

Email: ciro.sobral@ppgi.ufrj.br

Adriano J. de O. Cruz

Programa de Pós-Graduação  
em Informática

Rio de Janeiro-RJ, Brazil

Email: adriano@nce.ufrj.br

Antonio C. G. Thomé

Programa de Pós-Graduação  
em Sistemas e Computação

Natal-RN, Brazil

Email: thome@dimap.ufrn.br

**Abstract**—Expert systems can be applied to solve complex problems in different areas of knowledge. It have been used by specialists to help diagnose problems in areas such as medicine and engineering. More recently techniques from computational intelligence area, such as neural networks and fuzzy logic were used to build expert systems. These systems can learn from examples provided by problems solved in the past. The success of the expert system depends on how good the model is defined. This paper presents the modeling of an expert system used to classify the internal combustion diesel about the level of corrosion of its metal parts on a three-level scale. The system bases its decision on the results of used lubricating oil analysis. It uses a neuro-fuzzy model, called ANFIS, trained from past diagnoses performed by specialists. The paper discusses how important parameters of the system were defined, such as which variables obtained from the analysis of the oil should be used, the number of training epochs and the number of membership functions used to map the inputs.

### I. INTRODUCTION

Equipment may fail due to natural wear and tear of its components, misuse or unfavorable operating conditions. Equipment failures cause losses due to repair costs and production downtime. A delay in finding a problem usually increases the repair costs. The later a problem is discovered, higher are the repair costs [1]. Predictive maintenance is a tool that helps decreasing the number of failures and downtime, however, depends on reliable indicators and diagnostic models.

Engine condition monitoring techniques through lubricant oil analysis are widely used and are an effective way to diagnose abnormal wear and engine malfunctioning [2]. Lubricating oils have different chemical compounds on its formulation and through the analysis of the used lubricant of an internal combustion engine it is possible to detect problems such as excessive wear of parts, combustion problems and oil contamination before it presents imminent risk of failure, preventing damage and extra costs [3]. For this reason, the analysis of the used lubricant oil plays an important role in predictive maintenance.

Lubricating oil analysis provides information similar to a blood test [3]. However, to understand the potential problems identified in the analysis, these data should be interpreted by an expert who makes the diagnosis. This diagnosis is laborious and therefore, computerized methods for interpretation of the test results become a viable alternative [4].

Systems based on fuzzy logic and neuro-fuzzy have been developed for diagnosis in many devices [5]. In the automotive environment, there are several applications of fuzzy systems for fault detection [6]. Some authors have addressed diagnosing problems in the diesel engines using fuzzy logic [2], [4], [7].

In this article we are going to discuss a new approach for fault detection in diesel engines based on the analysis of lubricating oil. We will introduce the system, which uses a neuro-fuzzy model (ANFIS) [8] to classify the engine condition for corrosion in a Green (G), Yellow (Y) and Red (R) scale, based on analysis of its lubricant. A major problem is the great number of attributes present in the results of the analysis, which hinders the use of all variables for the system modeling. To circumvent this problem it was included a data analysis step, in order to remove some irrelevant attributes. Then the tests were performed to determine which were the best set of attributes to model the system.

This work is organized in the following way: in section II information relevant to understanding the selection of variables will be discussed; in section III information describing the data used in the study will be presented; section IV will describe how the model was developed; then in section V, some aspects of the methodology and results of the tests are presented; issues that were not covered but we intend to address in future work in section VI; and finally the section VII is reserved for conclusion.

### II. LUBRICANT OIL ANALYSIS

The analysis of used lubricating oils consists in a series of tests applied to the oil, in order to obtain information about the its condition. Also allows to have some information about the machine operation condition. It can be seen from three points of view: the condition of the lubricant, the presence of contaminants and machine wear.

The lubricant condition reveals whether it still retains the properties that allow its usage or if it should be replaced. However, the replacement of the contaminated oil may not be enough to solve an existing problem in the engine, since the system may remain contaminated, what can lead to the complete failure of one or more of its mechanical components [9].

It is important to notice that the condition of the lubricating oil is still influenced by several factors such as the operating condition, usage mode, machine load, general machine specifications and factors such as time of use and loose parts in the equipment.

Contaminants enter the system from the environment, in the form of dust, water and other materials and are a major cause of machine degradation [10], [11]. Their analysis is justified because a machine with this kind of problems generates wear particles at exponential rate [10].

A systematic analysis of the lubricating oil is a way to provide information about the status of engines and machines, aiming their proper maintenance [9]. This increases the reliability, availability, and the equipment life, while minimizing the repair costs, since it allows to detect problems before a component failure [9]–[12].

#### A. Problems on Lubricant Oil

Some circumstances may occur and become problematic for the proper functioning of the lubricant oil. According to [10] the following issues are the most common:

- Contamination: particles coming from external sources (entry of water or dust, for example) or internal (due to machine operation) that mix to the lubricant and modify its physicochemical properties.
- Degradation: caused by oxidation, which occurs when oxygen molecules combine with hydrocarbons. These components are then converted into carboxylic acids.
- Additive depletion: additives are consumed or chemically modified while performing their function and with that, some of the characteristics of the lubricant can be changed.

#### B. Lubricant Oil Properties

The most important property of a lubricant is the viscosity and it must be maintained within specifications for the proper functioning of the engine. The effects of abnormal viscosity are increased operating costs, engine overheating and increased wear. In case of an oil having its viscosity lower than ideal, there may be a disruption of the lubricant film, causing friction between the surfaces, which leads to overheating and premature wear. In case of increased viscosity, the oil acts resisting the motion, which increases the required power to move the parts, hence more stress and there is still the possibility of problems in the oil filter and deposit formation [13].

The causes of the decrease in viscosity may indicate contamination by water, fuel or addition of a lower viscosity lubricant. Its increase may indicate the presence of solid contaminants, oxidation of the lubricant, incomplete combustion (problems in the fuel/air mixture), high temperature operation or addition of lubricant with higher viscosity [10], [13].

The following properties, observed in lubricant, change the way lubricant oil act on the system, largely by changing its viscosity.

TABLE I. ELEMENTS USUALLY FOUND IN SPECTROMETRIC ANALYSIS AND POSSIBLE SOURCES. [10], [13]

| Element    | Symbol | Possible source                                         |
|------------|--------|---------------------------------------------------------|
| Iron       | Fe     | Axles, gears, block, piston rings, cylinder liner       |
| Copper     | Cu     | Bearings, bushings, thrust washers                      |
| Lead       | Pb     | Bearings                                                |
| Tin        | Sn     | Alloy bearings, bearing cages, welding                  |
| Aluminium  | Al     | Pumps, pressure washers, pistons, turbo-compressors     |
| Chromium   | Cr     | Bearings, piston rings, cylinder liner                  |
| Nickel     | Ni     | Pumps, valves                                           |
| Silver     | Ag     | Bearings                                                |
| Magnesium  | Mg     | Detergent additive, coolant additive                    |
| Silicon    | Si     | Dirt, antifoam additive                                 |
| Boron      | B      | Cooling system anti-corrosion additive                  |
| Sodium     | Na     | Detergent additive, fluid leaks from the cooling system |
| Barium     | Ba     | Antioxidant and corrosion inhibitor                     |
| Calcium    | Ca     | Detergent additive                                      |
| Phosphorus | P      | Anti-wear and extreme pressure additive                 |
| Potassium  | K      | Leaking liquid from the cooling system                  |
| Molybdenum | Mo     | Rings and extreme pressure additive                     |
| Zinc       | Zn     | Anti-wear additive                                      |
| Vanadium   | V      | Turbine blades                                          |

1) *Solid Particles*: The chemical elements found in the lubricant analysis are largely metals used in the composition of the engine and some other elements used as additives. For this reason these are found largely in the solid state.

The presence of solid particles hinders lubrication, creates deposits, increases wear, decreasing engine life [13]. Table I presents a list of the most common elements and their possible sources.

2) *Oxidation*: Oil and other components exposed to oxygen at higher temperatures oxidizes to form a variety of harmful by-products. Therefore, most of the lubricant contains additives to inhibit or slow down the oxidation process. One cause of oxidation is overheating and the extended delay in changing the lubricating oil. Oxidation causes an increase in viscosity and increase in corrosion of metal parts [10], [13].

3) *Nitration*: Nitration is the result of the reaction of the components of the oil to nitrogen oxides [10]. The product of this reaction is highly acidic, creating deposits and accelerates oxidation of the lubricating oil. As a result, cause increased wear of cylinders and camshafts [13].

4) *Sulfation*: Various sulfur oxide and water react to form acids, such as sulfuric acid. Diesel engines exhibit this problem usually due to the presence of sulfur in the fuel [10].

5) *Soot*: Formed from incomplete fuel combustion, soot is a good indicator of combustion efficiency of an engine [9], [13]. Its presence, which ends up being deposited in lubricant additives and exhausts can clog the oil filter [9].

The presence of soot is associated with the use of low quality fuel, air/fuel mixture incorrect, problems in the nozzles and low pressure which may be caused by worn piston and rings. Once mixed with the lubricant, soot causes increased viscosity [13].

6) *Dilution by Water*: Lubricants, in general, do not have water in their composition and their presence can lead to engine operation problems. The water usually evaporates when the engines work at normal operating temperatures. However when the operating temperature is below the normal, both the condensation of atmospheric humidity and the water formed by the burning of the fuel can contaminate the lubricant. The

presence of water may indicate contamination by external sources such as the cooling system [13].

The presence of water in the engine causes an increase in viscosity of the lubricant, with consequent reduction in lubrication, formation of acids (being a major component the sulfuric acid) and corrosion [13].

7) *Dilution by Fuel*: The causes of dilution by fuel are extended periods in neutral, condition of intermittent use (stop and go traffic, for example), and other related to engine tuning, like incorrect air/fuel mixture, incorrect engine timing. In general, if the measure exceeds 2.5% to 5% corrective actions should be taken [13].

The dilution of the lubricant by unburned fuel reduces the effectiveness of the lubricating oil by reducing the viscosity, fire/explosion risk, increased fuel consumption and a reduction in engine performance [13], [14].

### III. DATA DESCRIPTION

The data for this study consist of results of analyzes of lubricating oil from internal combustion engines to diesel Mercedes Benz OM-366 family.

Each sample contains 29 independent variables containing information about the equipment and the lubricating oil test results that are shown as the value of each element or chemical compounds found in the lubricant (Mileage at the test date, Mileage since the last oil change, Fe, Cr, Pb, Cu, Sn, Al, Ni, Si, Na, Mg, Ca, Ba, P, Zn, Mo, Ti, Ag, B, V, K, oxidation, nitration, sulfation, soot, % of water % of fuel, % of glycol).

Each sample presents a diagnosis for three different problems: corrosion, contamination and combustion. Each problem is classified in a Green (G), Yellow (Y) and Red (R) scale. A G classifies the motor in good condition, an Y meaning state of attention and R means critical condition. In this work, the only problem addressed is the corrosion. For this 724 samples were used, of which 641 were classified as G, Y as 64 and 19 as R.

This unbalance among the classes presents a problem for the construction of a classifier, because the cases that present problems are few, which hinders the pattern detection.

The class unbalance issue could be minimized with the use of artificial data. However, it is not an trivial task for this problem, since none of the 29 variables used follow a normal distribution.

Thus, with the data used, it can be said that it is possible to build a safe classifier for cases where there is no fault. However, because they are few data classified as Y or R, it is likely that this classifier will not be so effective in detecting the failure cases and even less in distinguishing from one another.

### IV. DEFINITION OF THE FUZZY LOGIC SYSTEM

Fuzzy logic is an extension of the traditional Boolean logic. Fuzzy logic accepts intermediate values between *false* (0) and *true* (1). Based on this assumption, Zadeh [15] proposed a definition for fuzzy sets, which differ from the traditional logic, because fuzzy sets admit that its elements may have a degree of membership that varies from 0 to 1.

Fuzzy inference systems are based on *if...then* rules. They receive a set of inputs, in general numerical, and turns them into fuzzy variables. From that transformation it is possible to make fuzzy inferences about the variables, through application of logical operators (AND, OR, NOT), according to instructions in the fuzzy rule base.

The neuro-fuzzy system used in this work is the ANFIS [8]. The consequents of the rules in the ANFIS systems are functions of the input variables, differently of the Mamdani systems in which they are fuzzy sets [16]. It is a system in which the rules are created automatically. The training algorithm creates all the rules for each possible combination of sets of all input variables. Thus,  $k$  rules are created, where  $k = \prod_{i=1}^N n_i$ , where  $n_i$  is the number of fuzzy sets for the  $i$ th entry.

The ANFIS system is trained by a hybrid learning algorithm. In the forward pass the algorithm uses the least-squares method to identify the consequent function parameters. In the backward pass the errors are propagated and the premise parameters are updated by gradient descent algorithm. Each time that these steps are applied is called a epoch.

The number of epochs can determine how well the system will model the input data. If this number is too small, the system will not extract well the characteristics of the input data and the result may be wrong. On the other hand, if the number is too big, it takes longer for the execution of the algorithm and can lead to a better result, however there is the risk of overfitting. For these reasons one should cautiously choose the system parameters.

The system parameters were defined by testing and the method used was divided into 3 steps:

- 1) Definition of the ideal number of epochs for training;
- 2) Choice of the most important attributes for modeling the problem;
- 3) Establishment of the number of membership functions for each variable;

#### A. Definition of Number of Epochs

As previously mentioned, the number of epochs is a major influence in the data modeling data by the system. In order to establish the best number of epochs a series of tests were performed. These test consisted in varying the number of epochs using the same set of input variables and determining when the algorithm stopped converging.

#### B. Choice of the Most Significant Variables

Before the choice of the most important variables, it was carried out a step of data analysis. This step allowed that some variables were discarded without loss of information in the final result. In this step, it was found that some variables presented almost zero variance, contributing with virtually no relevant information to the process. The result was that from the 29 initial variables, 7 were discarded (Mileage at the test date, Mileage since the last oil change, Ag, B, V, K, % of Glycol). This step was necessary because of the computational cost of using all input variables, even so the number of rules generated by a system with 22 variables is large.

Therefore, we chose to perform a testing round in which all possible combinations of variables were used to train the system then tested and evaluated. This step aims to find out if there are variables that appear repeatedly between the best classifiers evaluated. In this case, those variables would be good discriminators. Due to the limitations of the hardware used in the tests, it was only possible to test the combinations of up to 5 variables.

### C. Defining the Number of Membership Functions

The number of membership functions for an input variable defines how many fuzzy sets are used to map input values. A greater number of membership functions allows to define more precise rules for each piece of data. However, an increase on the number of rules does not guarantee an improvement in the system modeling. Furthermore, in ANFIS system, this greatly increases the number of rules, which can lead to the "curse of dimensionality". A smaller number of membership functions will yield faster training and execution of the system.

Each input variable may have a different number of membership functions, however, we choose for tests, carried out in this work, that all input variables would use the same number of membership functions in order to simplify the tests. We trained and tested the system with up to 3 membership functions for each input variable.

## V. EXPERIMENTS AND ANALYSIS OF RESULTS

### A. Testing environment description

The tests were conducted using a Intel(R) Core(TM) i7 CPU Q720 @ 1.60GHz 4GB DDR3, running Windows 7 x64, on MATLAB R2013a, with ANFIS modeling Fuzzy Logic Toolbox library. All tests used 5-fold cross-validation for defining sets of training and testing, which was done using the function `cvpartition` found in the Statistics Toolbox from MATLAB, which guarantees that every class will be represented in each fold.

### B. Methodology

The first tests were used to set the number of epochs needed for stopping the training. It was performed with a set of 4 variables as input and tested to the limit of 500 training epochs. However, no significant decrease was observed in the rate of errors from the first ten iterations. Therefore we chose to use the default value.

The next step was the selection of the important variables and the choice of the amount of membership functions. All possible combinations of 2, 3, 4 and 5 variables in an ANFIS system with 2 and 3 membership functions for input variables were tested.

The results of the classification were provided in the form of confusion matrices, as the example shown in table II. When combining the 22 variables 2 by 2, 231 matrices were generated, using 3 by 3, 1540 matrices, 4 by 4 were 7315 matrices and 5 by 5 were 26334. Each of these groups was tested with 2 and 3 membership functions for each input, totaling 70840 matrices.

TABLE II. EXAMPLE OF CONFUSION MATRIX.

|                 |                 |    |    |
|-----------------|-----------------|----|----|
| Reference Class | Predicted Class |    |    |
|                 | GG              | GY | GR |
|                 | YG              | YY | YR |
|                 | RG              | RY | RR |

In order to find which combination of variables provided the best results from all those matrices we defined a set of weights that was used to create a rank of all matrices in each set of testing. The Table III shows the values used.

This classification is made through the sum of all elements from the element-wise multiplication between the results matrix and the weight matrix. This operation is performed for each of the result matrices and the one with higher rank is selected.

TABLE III. WEIGHT MATRIX.

|                 |                 |    |    |
|-----------------|-----------------|----|----|
| Reference Class | Predicted Class |    |    |
|                 | 1               | -2 | -4 |
|                 | -5              | 2  | -2 |
|                 | -10             | -5 | 3  |

The values obey the following logic. The elements in the main diagonal, represent the correct classifications, therefore they receive a positive weight. Since the data have a greater number of positive classifications (G), it was defined that the weight of the classes with less samples would be increased.

All other matrix elements represent incorrect classification. For all of them negative weights were assigned. The elements above the main diagonal indicate cases that were classified as G and are Y or R; or Y cases that were classified as R. The cases below the main diagonal represent R cases that were classified as Y or G; or Y cases that were classified as G.

Since the goal of this system is to detect cases of failure to allow maintenance of the equipment, the worst kind of error in the classification is when the problem is more severe than the system reports, because it may be unnoticed. Therefore, larger weights were defined for the elements below the main diagonal.

Considering the number of tested models it is understandable to have different accuracy rates and also different rates of false positives or false negatives. Looking for models that minimize cases of wrong classifications, especially false positives, the application of the weight matrix helps to choose the best models, since it applies a negative weight for cases of failure and even higher weights for cases of false positive. Thus we seek to minimize the cases of severe errors.

Then the accuracy rate was measured for the system trained with the variables shown Table IV. The accuracy rate calculus obeys the Equation 1.

$$rate = \frac{GG + YY + RR}{GG + GY + GR + YG + YY + YR + RG + RY + RR} \quad (1)$$

### C. Most significant variables

All matrices were evaluated and those with better rankings for each test were used for comparison. The combinations with best results are shown in Table IV. All tests indicated the variable *Fe* as being very important when building a diagnostic

TABLE IV. COMBINATION OF VARIABLES WITH THE BEST RESULT ON EACH TEST SET.

| # MF | 2 Variables | 3 Variables        | 4 Variables                  | 5 Variables                      |
|------|-------------|--------------------|------------------------------|----------------------------------|
| 2 MF | Fe, Mg      | Fe, Al, Zn         | Fe, Sn, Oxidation, Sulfation | Fe, Pb, Si, Oxidation, Sulfation |
| 3 MF | Fe, Sul     | Fe, Si, % of water | Fe, Sn, Ba, Zn               | Fe, Pb, Ca, Sulfation, % of Fuel |

model from corrosion, which was as expected, since the block and the motor is largely composed of this element. However, no other variable had such performance during the tests.

#### D. Best results obtained from the classifier

The result for the main tests are shown in Table V. The selected models correctly answered more than 90% considering all classes.

TABLE V. RESULT OF ACCURACY RATE.

| # MF | 2 Variables | 3 Variables | 4 Variables | 5 Variables |
|------|-------------|-------------|-------------|-------------|
| 2 MF | 92.27%      | 93.09%      | 92.68%      | 93.09%      |
| 3 MF | 91.99%      | 92.40%      | 92.13%      | 92.54%      |

However, the accuracy rate as it was used for a binary classification, it is not able to express all the information contained in the confusion matrices. This synthesis hides desirable outcomes such as the lack of Y or R being presented as G. Therefore, we also used metrics of sensitivity and specificity, as seen in [17]. Sensitivity being the chance of testing positive among those with the condition and specificity being the chance of testing negative among those without the condition.

TABLE VI. SENSITIVITY AND SPECIFICITY, FOR 2 MEMBERSHIP FUNCTIONS.

| Measure     | 2 Variables | 3 Variables | 4 Variables | 5 Variables |
|-------------|-------------|-------------|-------------|-------------|
| Sensitivity | 57.19%      | 61.85%      | 67.09%      | 69.54%      |
| Specificity | 87.69%      | 88.35%      | 87.50%      | 86.25%      |

TABLE VII. SENSITIVITY AND SPECIFICITY, FOR 3 MEMBERSHIP FUNCTIONS.

| Measure     | 2 Variables | 3 Variables | 4 Variables | 5 Variables |
|-------------|-------------|-------------|-------------|-------------|
| Sensitivity | 64.49%      | 66.35%      | 61.31%      | 70.74%      |
| Specificity | 85.84%      | 86.00%      | 87.30%      | 87.10%      |

The best results obtained for each test are shown in Tables VIII and IX.

#### E. Result analysis

Although the classifiers had a 90% accuracy rate, as seen in V, the accuracy of class G sometimes outstripped 97%, however the accuracy for the other classes was slightly over 50%. This is largely explained by the imbalance between the size of the classes, which makes it difficult to recognize cases of engine failure, which have fewer cases.

The results provided by the accuracy rate do not establish what would be the best choice for the number of membership functions for each input, or how many input variables should be used. From the results presented in Table VI and Table VII, it is possible to see the improvement in sensitivity as the number of input variables increases.

TABLE VIII. BEST RESULT WITH 2 MEMBERSHIP FUNCTIONS.

|     |    |   |
|-----|----|---|
| 628 | 13 | 0 |
| 27  | 37 | 0 |
| 0   | 16 | 3 |

(a) 2 variables

|     |    |   |
|-----|----|---|
| 630 | 10 | 1 |
| 25  | 39 | 0 |
| 1   | 13 | 5 |

(b) 3 variables

|     |    |   |
|-----|----|---|
| 626 | 15 | 0 |
| 27  | 36 | 1 |
| 1   | 9  | 9 |

(c) 4 variables

|     |    |    |
|-----|----|----|
| 632 | 9  | 0  |
| 31  | 30 | 3  |
| 1   | 6  | 12 |

(d) 5 variables

TABLE IX. BEST RESULT WITH 3 MEMBERSHIP FUNCTIONS.

|     |    |    |
|-----|----|----|
| 623 | 18 | 0  |
| 25  | 35 | 4  |
| 0   | 8  | 11 |

(a) 2 variables

|     |    |    |
|-----|----|----|
| 628 | 12 | 1  |
| 31  | 31 | 2  |
| 1   | 8  | 10 |

(b) 3 variables

|     |    |   |
|-----|----|---|
| 626 | 14 | 1 |
| 28  | 35 | 1 |
| 0   | 13 | 6 |

(c) 4 variables

|     |    |    |
|-----|----|----|
| 625 | 16 | 0  |
| 28  | 33 | 3  |
| 1   | 6  | 12 |

(d) 5 variables

The increase in the number of variables was more consistent in creating a better classifier than the increase in the number of membership functions. Except for the case of 3 membership functions and 4 variables there was increase of the sensitivity. Furthermore, the specificity remained at about 87%.

Since there is a trade-off between the number of variables and the number of membership functions used to map input values. This trade-off between the number of variables and the number of membership functions is limited by the system processing time. Both the increase of the amount of variables and membership functions, increases the number of rules, and this way increases the time needed to train and also to run the system. So this is another issue to be taken into account when choosing best model.

## VI. FUTURE WORK

There is an unbalance in the samples used to train the neuro-fuzzy system. There are much more examples of engines in good condition (G) than in attention (Y) or critical condition (R). Although the data do not obey the normal distribution, this problem can be minimized with the use of techniques for the generation of artificial data that are not based on normal distributions, for example the SMOTE [18].

It is necessary to test other ways to assess the performance of multi-class classifiers to consider the cases of failure classification, similar to the metrics used to form binary classifications. The application of such methods could lead to the choice of a better diagnostic model.

We intend in future studies to test Principal Component Analysis [19] to try linear combinations of the input variables. It is expected to achieve better performance in the detection of cases of failure at the expense of processing time.

This study was conducted as a preliminary study of the use of fuzzy logic to model this problem. In a further step, we intend to make use of interval type-2 fuzzy logic. We intend

to use a set of rules found with this system for defining the rules of the type-2 fuzzy logic system.

## VII. CONCLUSION

This paper proposes the modeling of a system based on fuzzy logic to classify the problem of corrosion in internal combustion diesel engines. The method used for modeling the fuzzy system was based on the application of a technique called ANFIS. This technique was applied in all combinations of 2, 3, 4 and 5 variables. The resulting systems were tested to determine which would be the best combination of variables for problem classification.

This approach was used in order to determine which variables could be used as good discriminant to determine the existence of the engine problem. This strategy was proposed due to the large number of variables obtained from the analysis of the lubricating oil. The use of all variables implies a large computational cost, however, discarding variables without judgement can lead to a poor results from the classifier.

In the tests we obtained accuracy rates greater than 90%. Another important result was that the system achieved sensitivity rate of 70.74% with a specificity of 87.10%. However, to the best of our knowledge there is not another model that performed similar analysis to be used as benchmark. This makes it difficult to present comparative results.

It is important to note that it is not possible to say at the present stage to what extent the engine type and the intrinsic characteristics of the data influence the outcome of the system. That way it can not be guaranteed that it is possible to extend the results for samples of different nature.

The results shows that increasing the number of membership functions leads to better results in the classification of engine failures. The results allow us to assert that it is possible to construct a system based on a well-known neuro-fuzzy model, which presents good results in the detection of the cases with no fault. In failure cases, the system does not allow to differentiate the type of fault. This would not be an impediment to the use of the system in real situations, since in case of engine failure it must be taken to maintenance anyway.

## ACKNOWLEDGMENTS

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior), which partially supported the development of this work.

## REFERENCES

- [1] R. Smith. Asset integrity: What to do after an economic crisis, online. [Online]. Available: [http://reliabilityweb.com/index.php/articles/asset\\_integrity\\_what\\_to\\_do\\_after\\_an\\_economic\\_crisis/](http://reliabilityweb.com/index.php/articles/asset_integrity_what_to_do_after_an_economic_crisis/)
- [2] S. Ramezani and A. Memariani, "A fuzzy rule based system for fault diagnosis, using oil analysis results," *International Journal of Industrial Engineering & Production Research*, vol. 22, no. 2, pp. 91–98, June 2011. [Online]. Available: [http://ijiepr.iust.ac.ir/browse.php?a\\_code=A-10-15-1&slc\\_lang=en&sid=1](http://ijiepr.iust.ac.ir/browse.php?a_code=A-10-15-1&slc_lang=en&sid=1)
- [3] D. Thweatt and T. Bell, "Interpreting heavy-duty motor oil analysis reports," in *Practicing Oil Analysis*, Mar 2006. [Online]. Available: <http://www.machinerylubrication.com/Read/853/heavy-duty-motor-oil-analysis>
- [4] A. Sala, B. Tormos, V. Macián, and E. Royo, "Fuzzy diagnosis module based on interval fuzzy logic: oil analysis application," in *ICINCO*, J. Filipe, J. Andrade-Cetto, and J.-L. Ferrier, Eds. INSTICC Press, 2005, pp. 85–90. [Online]. Available: <http://www.icinco.org/ICINCO2005/>
- [5] J. Calado, J. Korbicz, K. Patan, R. Patton, and J. S. da Costa, "Soft computing approaches to fault diagnosis for dynamic systems," *European Journal of Control*, vol. 7, no. 2-3, pp. 248–286, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0947358001711508>
- [6] J. Mohammadpour, M. Franchek, and K. Grigoriadis, "A survey on diagnostic methods for automotive engines," in *International Journal of Engine Research*, vol. 13. SAGE, February 2012, pp. 41–64. [Online]. Available: <http://jer.sagepub.com/content/13/1/41.full.pdf>
- [7] V. Macian, B. Tormos, A. Sala, and J. C. Ramírez, "Fuzzy logic-based expert system for diesel engine oil analysis diagnosis," *Insight*, vol. 8, no. 48, pp. 462–470, 2006.
- [8] J.-S. R. Jang, "Anfis: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993. [Online]. Available: [ieeexplore.ieee.org/iel1/21/6499/00256541.pdf](http://ieeexplore.ieee.org/iel1/21/6499/00256541.pdf)
- [9] Dep. Técnico da Lubrikol Lubrificantes, "Importância da análise de Óleo lubrificante," Lubrikol, Tech. Rep., 2013. [Online]. Available: [http://www.lubrikol.com.br/admin/fotos\\_dicas/06-2013%20Importancia%20da%20An%C3%A1lise%20de%20%C3%93leo%20Lubrificante.pdf](http://www.lubrikol.com.br/admin/fotos_dicas/06-2013%20Importancia%20da%20An%C3%A1lise%20de%20%C3%93leo%20Lubrificante.pdf)
- [10] M. Barrett, *The Practical Guide To Oil Analysis*. Insight Services, [2007?], no. 4th ed.
- [11] Supreme Lubrificantes. Análise de Óleo. [Online]. Available: [http://www.supremelub.com.br/downloads/tecnicas/analise\\_de\\_oleo.pdf](http://www.supremelub.com.br/downloads/tecnicas/analise_de_oleo.pdf)
- [12] PerkinElmer, "Lubrificant analysis," PerkinElmer, Tech. Rep., 2012.
- [13] ConocoPhillips Lubrificantes, "Oil analysis technical guide," ConocoPhillips Lubrificantes, Tech. Rep., 2007.
- [14] SENAI-ES, "Lubrificação - mecânica," SENAI-ES, Tech. Rep., 1997. [Online]. Available: <http://www.abraman.org.br/docs/apostilas/Mecanica-Lubrificacao.pdf>
- [15] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [16] T. J. Ross, *Fuzzy Logic with Engineering Applications, Third Edition*. Wiley, 2010.
- [17] C. Beleites, R. Salzer, and V. Sergio, "Validation of soft classification models using partial class memberships: An extended concept of sensitivity & co. applied to grading of astrocytoma tissues," *Chemometrics and Intelligent Laboratory Systems*, vol. 122, no. 0, pp. 12–22, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169743912002419>
- [18] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. [Online]. Available: <http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/chawla02a.pdf>
- [19] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, Oct. 2002.