INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA INFORMÁTICA




# Deteção de convulsões epiléticas a partir de eletroencefalogramas

**HÉLDER JOSÉ PEDROSA DE ABREU MARTINS**
Outubro de 2016

POLITÉCNICO
DO PORTO

Instituto Superior de Engenharia do Porto

# Deteção de convulsões epiléticas a partir de eletroencefalogramas

**Hélder José Pedrosa de Abreu Martins**

**Dissertação para obtenção do Grau de Mestre em Engenharia Informática, Área de Especialização em Sistemas Computacionais**

**Orientador: Elsa Maria de Carvalho Ferreira Gomes**

**Júri**:

Presidente:

[Nome do Presidente, Categoria, Escola]

Vogais:

[Nome do Vogal1, Categoria, Escola]

[Nome do Vogal2, Categoria, Escola] (até 4 vogais)

Porto, Outubro de 2016

# À minha família

# Resumo

A epilepsia atinge cerca de 1% da população mundial, e é caracterizada pela ocorrência de crises espontâneas.

Pretende-se detetar (e prever) convulsões analisando os dados obtidos a partir de eletroencefalogramas (EEG).

Para que os sistemas de deteção/previsão baseados em EEG funcionem de forma eficaz, os algoritmos computacionais devem identificar com segurança os períodos de maior probabilidade de ocorrência de convulsão. Se estes estados cerebrais de alerta fossem detetados, os pacientes poderiam evitar atividades potencialmente perigosas, como conduzir ou natação, e poderia ser administrada medicação somente quando necessário para evitar crises iminentes, reduzindo os efeitos colaterais globais.

O objetivo deste trabalho é produzir um método/algoritmo capaz de classificar sinais resultantes do EEG, para deteção de convulsões epiléticas.

Existem várias técnicas utilizadas para processamento dos sinais EEG e para os classificar.

Pretende-se que sejam aplicadas neste trabalho técnicas utilizadas previamente, em trabalhos desenvolvidos anteriormente, em dados de sons cardíacos para deteção de patologia cardíaca.

**Palavras-chave**: Classificação, *data mining*, *motifs*.

# Abstract

Epilepsy is present in about 1% of world's population and it is characterized by the occurrence of spontaneous seizures.

The goal is to detect (and predict) seizures by analyzing data captured from electroencephalograms (EEG).

In order for systems that detect/predict seizures base in EEG data work properly is necessary that computational algorithms reliably detect the periods of increased probability of seizure occurrence. If these cerebral states of warning can be identified, the patients could avoid potential dangerous activities like driving or swimming and medications could be administrated only when needed to prevent imminent seizures, reducing the overall side effects.

The objective of this work is to develop a method/algorithm able to classify signs received from EEG data for detection of epilepsy seizures.

There are many techniques used for processing and classifying the signals of EEG data.

One of the goals of this work is to use techniques already used, in previous works with cardiac sounds for detection of cardiac pathologies.

**Keywords**: Classification, data mining, motifs.

# Agradecimentos

Quero agradecer a todas as pessoas que direta ou indiretamente contribuíram para a elaboração desta dissertação.

Um forte agradecimento à orientadora desta dissertação, a professora Elsa Gomes por todo o auxílio prestado, pela dedicação, disponibilidade e também pelos valiosos conselhos que foi dando ao longo de todo este período.

Agradeço à minha família por todo o apoio, paciência e suporte que sempre me deram ao longo de toda esta etapa, contribuindo decisivamente para que fosse possível terminar com sucesso.

Agradeço também a todos os meus amigos que sempre me apoiaram e a todos os colegas do ISEP pelo companheirismo, entreajuda e partilha de conhecimentos.

A todos, o meu sincero obrigado.

# Índice


**1 Introdução ..... 1**

1.1 Contexto ..... 1

1.2 Problema ..... 2

1.3 Análise de valor ..... 2

1.4 Avaliação de resultados ..... 2

1.5 Estrutura do documento ..... 3

**2 Contexto ..... 5**

2.1 Contextualização ..... 5

2.1.1 Deteção de convulsões ..... 6

2.1.2 Previsão de convulsões ..... 7

2.1.3 Definição do problema ..... 7

2.2 Análise de valor ..... 7

2.2.1 Proposta de valor ..... 7

2.2.2 Qual o valor? ..... 8

2.2.3 Negociação ..... 10

2.2.4 Modelo de negócio de Canvas ..... 11

2.2.5 Análise de criação de valor ..... 12

**3 Estado da arte ..... 15**

3.1 Algoritmos de classificação ..... 15

3.1.1 Random Forests ..... 15

3.1.2 J48 ..... 15

3.1.3 Support Vector Machines ..... 16

3.1.4 Regressão Logística ..... 16

3.1.5 Medidas de avaliação ..... 17

3.2 Abordagens existentes ..... 18

3.2.1 Métodos no domínio do tempo ..... 22

3.2.2 Métodos no domínio da frequência ..... 25

3.2.3 Métodos no domínio de wavelet ..... 25

3.2.4 Método de decomposição empírica ..... 26

3.2.5 Método de decomposição em valores singulares ..... 27

3.2.6 Método de análise de componentes principais ..... 27

3.2.7 Método de análise de componentes independentes ..... 27

3.3 Multiresolution Motif Discovery ..... 28

3.4 Avaliação das soluções ..... 29

3.4.1 Medidas de avaliação ..... 29

3.4.2 Hipóteses e metodologias de avaliação ..... 29

3.4.3 Teste estatístico ..... 30

3.5 Tecnologias utilizadas ..... 30

3.5.1 Java ..... 30

3.5.2 NetBeans ..... 30
3.5.3 *Weka* ..... 31

**4 Solução ..... 33**

4.1 Design da solução ..... 33
4.1.1 Dataset ..... 33
4.1.2 Processo ..... 34

4.2 Implementação ..... 36
4.2.1 Dataset ..... 36
4.2.2 Transformar *dataset* e pré-processamento ..... 37
4.2.3 Extração de atributos ..... 38
4.2.4 Reamostragem e execução algoritmos de classificação ..... 39
4.2.5 Validação dos resultados ..... 40

**5 Experiências e avaliação ..... 41**

5.1 Medidas de avaliação ..... 41

5.2 Hipóteses e metodologias de avaliação ..... 44

5.3 Avaliação Area Under the ROC Curve ..... 45

5.4 Avaliação Area Under the ROC Curve com reamostragem ..... 51
5.4.1 Paciente 1 ..... 52
5.4.2 Paciente 2 ..... 57
5.4.3 Cão 1 ..... 63
5.4.4 Cão 2 ..... 69

5.5 Avaliação *Accuracy* e *F-Measure* ..... 75
5.5.1 Paciente 1 ..... 75
5.5.2 Paciente 2 ..... 78
5.5.3 Cão 1 ..... 82
5.5.4 Cão 2 ..... 86

5.6 Discussão da metodologia de avaliação ..... 90

5.7 Teste estatístico ..... 91

**6 Conclusão ..... 95**

6.1 Trabalho futuro ..... 96

# Lista de Figuras

Figura 1 – Sistema de implantação e registo de um EEG (Kaggle.com, 2016) .....6
Figura 2 – EEG com uma convulsão (Alotaiby et al., 2014) .....23
Figura 3 – Utilização da técnica da janela para previsão de convulsões (Mormann, 2008) .....24
Figura 4 – M*otif* numa série temporal de um EEG (Castro and Azevedo, 2010) .....28
Figura 5 – Diagrama de atividades do processo de desenvolvimento da solução .....35
Figura 6 – Exemplo de um registo de EEG com 16 canais (Data | Kaggle, 2016) .....39
Figura 7 – Matriz confusão paras as classes preictal e interictal .....42
Figura 8 – Exemplo espaço ROC .....43
Figura 9 – Exemplo de curvas ROC .....44
Figura 10 – Gráfico comparativo das curvas ROC para os 4 algoritmos para o Paciente 1 .....47
Figura 11 - Gráfico comparativo das curvas ROC para os 4 algoritmos para o Paciente 2 .....48
Figura 12 - Gráfico comparativo das curvas ROC para os 4 algoritmos para o Cão 1 .....49
Figura 13 - Gráfico comparativo das curvas ROC para os 4 algoritmos para o Cão 2 .....50
Figura 14 – Gráfico com os melhores resultados para AUC sem reamostragem .....51
Figura 15 – Gráfico com os melhores resultados para AUC com reamostragem (Paciente 1) ..56
Figura 16 – Curvas de ROC para os 4 algoritmos para o Paciente 1 com reamostragem .....57
Figura 17 – Gráfico com os melhores resultados para AUC com reamostragem (Paciente 2) ..62
Figura 18 – Curvas ROC para os 4 algoritmos para o Paciente 2 com reamostragem .....63
Figura 19 – Gráfico com os melhores resultados para AUC com reamostragem (Cão 1) .....68
Figura 20 – Curvas ROC para os 4 algoritmos para o Cão 1 com reamostragem .....69
Figura 21 – Gráfico com os melhores resultados para AUC com reamostragem (Cão 2) .....73
Figura 22 – Curvas ROC para os 4 algoritmos para o Cão 2 com reamostragem .....74
Figura 23 – Matriz confusão (5ª reamostragem - J48 - parâmetros 30 10 10) .....75
Figura 24 – Matriz confusão (5ª reamostragem - *Random Forest* - parâmetros 30 10 10) .....76
Figura 25 – Matriz confusão (5ª reamostragem - SVM - SMO - parâmetros 40 30 20) .....77
Figura 26 – Matriz confusão (4ª reamostragem - Regressão Logística - parâmetros 40 10 10) .78
Figura 27 – Matriz confusão (2ª reamostragem - J48 - parâmetros 40 10 10) .....79
Figura 28 – Matriz confusão (2ª reamostragem – *Random Forest* - parâmetros 30 10 10) .....80
Figura 29 – Matriz confusão (sem reamostragem – SVM - SMO - parâmetros 30 20 10) .....81
Figura 30 – Matriz confusão (5ª reamostragem – Regressão Logística - parâmetros 40 20 10) 82
Figura 31 – Matriz confusão (5ª reamostragem - J48 - parâmetros 40 20 10) .....83
Figura 32 – Matriz confusão (5ª reamostragem - *Random Forest* - parâmetros 40 20 10) .....84
Figura 33 – Matriz confusão (5ª reamostragem - SVM - SMO - parâmetros 40 20 10) .....85
Figura 34 – Matriz confusão (5ª reamostragem - Regressão Logística - parâmetros 40 30 20) .86
Figura 35 – Matriz confusão (5ª reamostragem - J48 - parâmetros 40 20 10) .....87
Figura 36 – Matriz confusão (5ª reamostragem - *Random Forest* - parâmetros 40 20 10) .....88
Figura 37 – Matriz confusão (5ª reamostragem - SVM - SMO - parâmetros 40 20 10) .....89
Figura 38 – Matriz confusão (5ª reamostragem - Regressão Logística - parâmetros 40 30 20) .90
Figura 39 – Resultado do *t-test* – Paciente 1 .....91
Figura 40 - Resultado do *t-test* – Paciente 2 .....92

Figura 41 - Resultado do *t-test* – Cão 1 ..... 93
Figura 42 - Resultado do *t-test* – Cão 2 ..... 94

# Lista de Tabelas

Tabela 1 - Benefícios e Sacrifícios em cada valor temporal .....10
Tabela 2- Modelo de Canvas .....12
Tabela 3 – Resumo das técnicas analisadas .....18
Tabela 4 – Resultados do Paciente 1 para AUC sem reamostragem .....46
Tabela 5 – Resultados do Paciente 2 para AUC sem reamostragem .....47
Tabela 6 – Resultados do Cão 1 para AUC sem reamostragem .....48
Tabela 7 – Resultados do Cão 2 para AUC sem reamostragem .....50
Tabela 8 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 30 10 10) ....52
Tabela 9 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 30 20 10) ....53
Tabela 10 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 10 10) ..53
Tabela 11 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 20 10) ..54
Tabela 12 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 20 20) ..54
Tabela 13 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 30 10) ..55
Tabela 14 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 30 20) ..55
Tabela 15 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 60 20 10) ..56
Tabela 16 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 30 10 10) ..58
Tabela 17 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 30 20 10) ..58
Tabela 18 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 10 10) ..59
Tabela 19 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 20 10) ..59
Tabela 20 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 20 20) ..60
Tabela 21 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 30 10) ..60
Tabela 22 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 30 20) ..61
Tabela 23 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 60 20 10) ..61
Tabela 24 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 30 10 10) .....64
Tabela 25 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 30 20 10) .....64
Tabela 26 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 10 10) .....65
Tabela 27 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 20 10) .....65
Tabela 28 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 20 20) .....66
Tabela 29 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 30 10) .....66
Tabela 30 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 30 20) .....66
Tabela 31 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 60 20 10) .....67
Tabela 32 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 30 10 10) .....69
Tabela 33 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 30 20 10) .....70
Tabela 34 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 10 10) .....70
Tabela 35 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 20 10) .....71
Tabela 36 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 20 20) .....71
Tabela 37 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 30 10) .....72
Tabela 38 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 30 20) .....72
Tabela 39 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 60 20 10) .....73
Tabela 40 – Resultados do Paciente 1 para *accuracy* e *f-measure* para J48 .....75

Tabela 41 – Resultados do Paciente 1 para *accuracy* e *f-measure* para *Random Forest* .....76
Tabela 42 – Resultados do Paciente 1 para *accuracy* e *f-measure* para SVM - SMO .....77
Tabela 43 – Resultados do Paciente 1 para *accuracy* e *f-measure* para Regressão Logística....78
Tabela 44 – Resultados do Paciente 2 para *accuracy* e *f-measure* para J48.....79
Tabela 45 – Resultados do Paciente 2 para *accuracy* e *f-measure* para *Random Forest* .....80
Tabela 46 – Resultados do Paciente 2 para *accuracy* e *f-measure* para SVM - SMO .....81
Tabela 47 – Resultados do Paciente 2 para *accuracy* e *f-measure* para Regressão Logística....82
Tabela 48 – Resultados do Cão 1 para *accuracy* e *f-measure* para J48 .....83
Tabela 49 – Resultados do Cão 1 para *accuracy* e *f-measure* para *Random Forest* .....84
Tabela 50 – Resultados do Cão 1 para *accuracy* e *f-measure* para SVM - SMO.....85
Tabela 51 – Resultados do Cão 1 para *accuracy* e *f-measure* para Regressão Logística .....86
Tabela 52 – Resultados do Cão 2 para *accuracy* e *f-measure* para J48 .....87
Tabela 53 – Resultados do Cão 2 para *accuracy* e *f-measure* para *Random Forest* .....88
Tabela 54 – Resultados do Cão 2 para *accuracy* e *f-measure* para SVM - SMO.....89
Tabela 55 – Resultados do Cão 2 para *accuracy* e *f-measure* para Regressão Logística .....90

# Acrónimos e Símbolos

## Lista de Acrónimos

| | |
|---|---|
| **AHP** | *Analytic hierarchy process* |
| **AUC** | *Area Under the ROC Curve* |
| **CSV** | Comma-separeted values |
| **EEG** | Eletroencefalograma |
| **FN** | Falso negativo |
| **FP** | Falso positivo |
| **ICA** | Análise de componentes independentes |
| **IDE** | *Integrated Development Environment* |
| **iSAX** | *indexable Symbolic Aggregate approXimation* |
| **JVM** | *Java Virtual Machine* |
| **MrMotif** | *Multiresolution Motif Discovery* |
| **ROC** | *Receiver operating characteristic* |
| **SVM** | *Support Vector Machines* |
| **TFP** | taxa de falsos positivos |
| **TVP** | Taxa de verdadeiros positivos |
| **VN** | Verdadeiro negativo |
| **VP** | Verdadeiro positivo |

# 1 Introdução

## 1.1 Contexto

A área da saúde é considerada uma das mais relevantes na sociedade e como tal exige uma pesquisa constante, com o objetivo de alcançar novos tipos de tratamento ou aperfeiçoamento dos existentes.

Segundo a Sociedade Americana de Epilepsia, a epilepsia é uma doença cerebral que provoca convulsões recorrentes num determinado paciente, é diagnosticada quando ocorrem duas ou mais convulsões não provocadas. (American Epilepsy Society, 2016)

O Eletroencefalograma (EEG) é um exame que regista as correntes elétricas espontâneas desenvolvidas no cérebro, através de elétrodos aplicados no couro cabeludo, permitindo estudar e avaliar a atividade elétrica cerebral.

Através da captura dos sinais de um EEG, aliado à implementação de um algoritmo adequado, é possível desenvolver um sistema de previsão de convulsões que poderão desempenhar um importante avanço, pois poderão ser desencadeadas ações de forma a minimizar o impacto das convulsões na vida dos pacientes.

A base de trabalho para a realização desta dissertação são gravações de EEG que foram realizados em pessoas e em cães com epilepsia. Estes registos terão de ser trabalhados de forma a conseguir potenciar as frequências ideais, de seguida tentar descobrir padrões que permitam identificar potenciais convulsões e por fim efetuar uma análise e avaliação dos resultados obtidos.

## 1.2 Problema

A epilepsia atinge cerca de 1% da população mundial e é caracterizada pela ocorrência de crises espontâneas.

Pretende-se detetar (e prever) convulsões analisando dados obtidos a partir de eletroencefalogramas.

Para que os sistemas de deteção/previsão baseados em EEG funcionem de forma eficaz, os algoritmos computacionais devem identificar com segurança os períodos de maior probabilidade de ocorrência de convulsão. Se estes estados cerebrais de alerta fossem detetados, os pacientes poderiam evitar atividades potencialmente perigosas, como conduzir ou nadar e poderia ser administrada medicação somente quando necessário para evitar crises iminentes, reduzindo os efeitos colaterais globais (Kaggle.com, 2016).

## 1.3 Análise de valor

Devido ao facto da epilepsia afetar cerca de 1% da população mundial, existe uma crescente preocupação por parte da comunidade científica em tentar perceber e identificar as origens das convulsões epiléticas. Segundo Carney et al (Carney et al., 2011), apesar da existência de bastantes algoritmos previsão de convulsões já publicados, ainda não está claramente comprovada a utilidade clínica dos mesmos.

O objetivo deste projeto é contribuir para o desenvolvimento de técnicas e/ou metodologias que possam de alguma forma acrescentar valor e qualidade à previsão e deteção de convulsões epitéticas através da análise de eletroencefalogramas de forma a melhor a vida dos portadores da doença.

## 1.4 Avaliação de resultados

No decorrer das experiências a efetuar com o objetivo de avaliar a capacidade de deteção/previsão dos modelos será usado o método de validação cruzada (*cross-validation*) com 10 subconjuntos (*folds*). Este método consiste em dividir o conjunto de dados de entrada em 10 subconjuntos e construir 10 submodelos, onde em cada um são utilizados 9 conjuntos para treino e um conjunto para teste. Em cada submodelo, o conjunto de teste usado será diferente, assim como os dados de treino que não terão presentes os dados de teste. Para cada submodelo é então calculada a respetiva medida de avaliação como, por exemplo a taxa

de acerto (*accuracy*). No caso da taxa de acerto, o resultado da avaliação será a média das taxas e quanto maior for a taxa de acerto, melhor será o modelo (Wong, 2015).

Para além da taxa de acerto serão usadas outras medidas de avaliação dos modelos tais como *F-measure* e *Area Under the ROC Curve* (AUC), para comparar os modelos de classificação.

Para suportar as afirmações relativas à comparação de modelos serão usados testes estatísticos. Será utilizado o *t-test* para amostras emparelhadas.

## 1.5 Estrutura do documento

No primeiro capítulo desta dissertação é possível visualizar a introdução, que disponibiliza o contexto, a descrição e valorização do problema, a forma como poderão ser avaliados os resultados e também a estrutura do documento.

No segundo capítulo é realizada contextualização do problema e a análise de valor mais detalhada.

Durante o terceiro capítulo é possível observar algumas formas de avaliação das soluções anteriormente propostas. Também é possível visualizar as abordagens já realizadas por outros autores e a descrição de algumas ferramentas que irão ser utilizadas.

No quarto capítulo está presente o *design* da solução proposta, ou seja, o tipo de abordagem realizada para alcançar a solução e também a descrição detalhada da implementação do modelo.

O quinto capítulo fica destinado às experiencias e avaliações resultantes do modelo proposto.

No sexto capítulo estão presentes as conclusões finais a todo o trabalho desenvolvido, bem como, algumas propostas para trabalho futuro.

Por fim, encontra-se a bibliografia, onde poderão ser consultadas todas as fontes que servem de suporte a esta dissertação.

# 2 Contexto

## 2.1 Contextualização

Segundo Nunes de Oliveira e Rosado (Nunes de Oliveira and Rosado, 2004), uma crise epitética é um fenómeno paroxístico, de causa primária ou secundariamente encefálica causada por uma descarga neuronal anormal e excessiva podendo ter várias formas de apresentação clínica.

Um dos aspetos mais assustadores para os doentes com epilepsia é o facto de poderem ocorrer convulsões de uma forma repentina e totalmente inesperada. Estes pacientes durante a maior parte do tempo fazem uma vida normal, sem apresentar sinais da doença durante o intervalo entre as convulsões. Este intervalo entre as convulsões é denominado por intervalo interictal (Fisher et al., 2000).

Tem sido cada vez mais referido que existem 4 estados temporais que permitem classificar a atividade cerebral: preictal (ocorre antes das convulsões), interictal (período entre as convulsões), ictal (convulsões) e pós-ictal (ocorre após as convulsões). O desafio para a previsão de convulsões tem sido diferenciar o estado preictal dos restantes, mais especificamente do interictal.

Na Figura 1 é possível ver um cão com um sistema de deteção de atividade cerebral implantado, que permite efetuar a captura dos sinais de um eletroencefalograma. Na zona mais inferior da Figura 1 é possível visualizar a atividade de um EEG, a primeira zona assinalada (i) apresenta um falso alarme, pois apenas é enviado um aviso, a segunda zona (ii) também representa um falso alarme pois continuam a ser apenas avisos, enquanto a terceira zona (iii) já apresenta um alarme real e a ocorrência de uma convulsão (Kaggle.com, 2016).

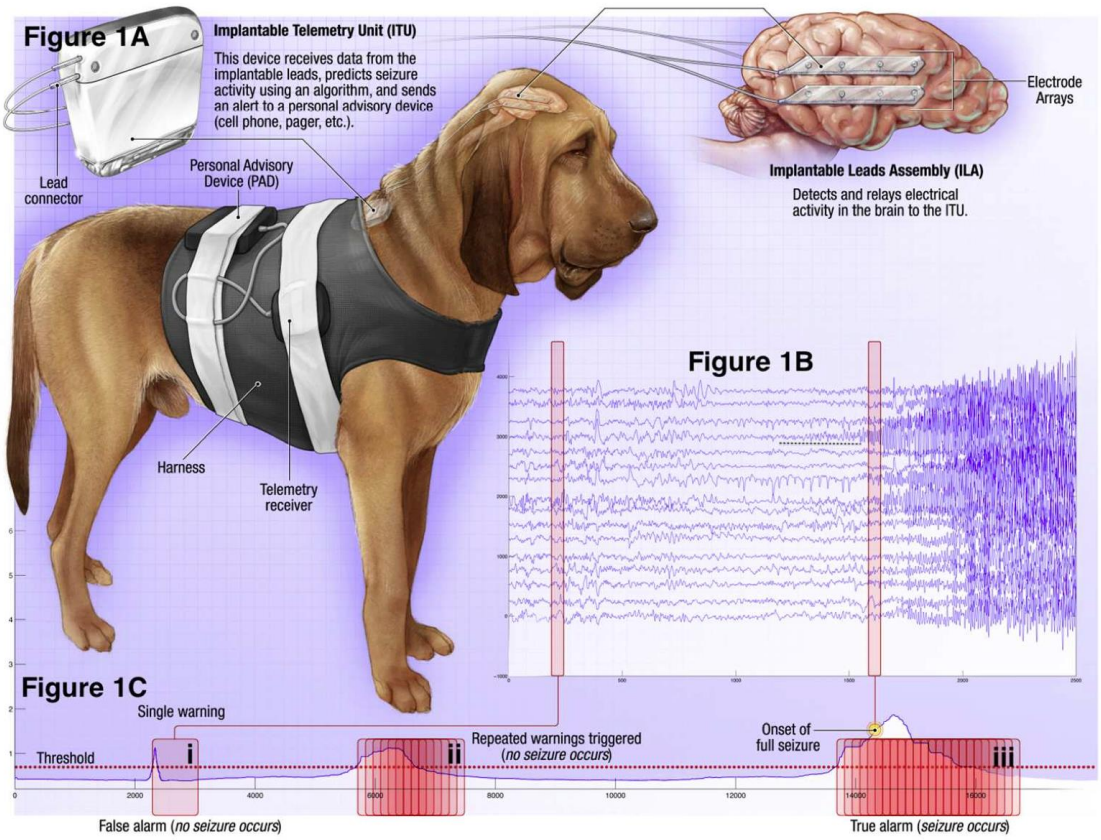


Figura 1 – Sistema de implantação e registo de um EEG (Kaggle.com, 2016)

### 2.1.1 Deteção de convulsões

A deteção de convulsões através da interpretação visual de EEG é tida como um processo entediante e que consome bastante tempo. Um EEG, normalmente contém gravações referentes a vários dias e tem a necessidade de ser analisado na totalidade por um especialista da área, para que seja possível identificar a atividade epitética.

Segundo Frei (Frei, 2013), através da análise de sinais biológicos gravados em pacientes com epilepsia, nomeadamente recorrendo à utilização de EEG, é possível detetar automaticamente as convulsões recorrendo a um algoritmo de deteção e a um sistema de classificação.

De uma forma geral, é possível afirmar que o objetivo é analisar um conjunto de sinais e transformar esse conteúdo num indicador capaz de mostrar se o paciente está ou passou por um estado de convulsão (Gajic et al., 2014).

### 2.1.2 Previsão de convulsões

A previsão de convulsões consiste na antecipação da ocorrência de convulsões epiléticas. Como já foi referido, habitualmente é monitorizada a atividade elétrica do cérebro através de eletroencefalogramas. Um método que fosse capaz de monitorizar constantemente e prever com um grau de precisão elevado a ocorrência de convulsões poderia melhorar significativamente a qualidade de vida destes pacientes e proporcionar novas formas de combater a doença, nomeadamente simples sistemas de alarme, medicação apenas quando existir sinais de potencial convulsão e estimulação elétrica ou outra (Stacey and Litt, 2008).
A principal questão que se coloca é saber se é possível distinguir um estado de pré-convulsão de um estado interictal. Segundo Mormann el al., foram realizados estudos científicos que identificam um estado de pré-convulsão que se traduz num aumento do fluxo sanguíneo cerebral, disponibilidade de oxigénio, e nível de oxigénio no sangue dependente (BOLD), bem como, alterações da frequência cardíaca (Mormann et al., 2007).

### 2.1.3 Definição do problema

Esta dissertação tem com objetivo produzir um método capaz de classificar sinais resultantes do EEG, para deteção de convulsões epiléticas.
Já foram desenvolvidas variadas técnicas para o processamento e classificação dos sinais EEG.
Nesta tese serão aplicadas técnicas utilizadas previamente, em trabalhos anteriormente realizados, nomeadamente em dados de sons cardíacos para deteção de patologia cardíaca (Gomes et al., 2014) (Gomes et al., 2013).
Pretende-se ainda que sejam utilizadas técnicas (estatísticas e de *data mining*) disponíveis em ferramentas como o *Weka*.

## 2.2 Análise de valor

### 2.2.1 Proposta de valor

Uma proposta de valor bem definida tem a capacidade de acrescentar valor para o cliente e em simultâneo gerar beneficio para quem a fornece, ou seja, quanto maior for a proposta de valor melhor será a experiência para o cliente e maior será o benefício. (Barnes et al., 2009)

É possível afirmar que uma proposta de valor deverá destacar as mais-valias que irá disponibilizar para quem usufruir dos produtos e/ou serviços, ou seja, explicar convenientemente que tipo de valor vai acrescentar. Outos aspetos que deverão constar são a clara definição do produto/serviço que será disponibilizado, em que é que ele se distingue dos restantes e qual o público-alvo.

Para que seja possível realizar uma boa proposta de valor é também necessário compreender as perceções que os clientes têm em relação aos produtos/serviços que estão a ser disponibilizados, bem como ao preço. A relação entre os benefícios que irá alcançar através da aquisição do produto/serviço (por exemplo: qualidade, personalização, assistência técnica) terá sempre de ser superior aos sacrifícios que terá de fazer para o obter (por exemplo: custos).

### 2.2.2 Qual o valor?

O termo "valor para o cliente" pode ser definido como uma perceção pessoal decorrente de uma associação entre um cliente e uma organização e poderá ocorrer devido à redução de sacrifício, aumento de benefício, uma combinação ponderada entre o sacrifício e beneficio ou a uma conjugação de todos ou parte dos elementos anteriormente referenciados (Woodall, 2003).

É possível identificar cinco formas de caracterizar o "valor para o cliente":

- Valor líquido: é a relação entre os benefícios e os sacrifícios interpretada pelo consumidor em que os benefícios terão sempre de ser maiores do que os sacrifícios.
- Experiência do utilizador: benefícios alcançados através de relatos de utilizadores realizados de uma forma independente, sem que seja necessário identificarem qualquer tipo de sacrifício associado.
- Atributos percebidos: características identificadas no produto.
- Preço: significa que o preço deverá ser o mais baixo possível tendo em consideração a posição em relação ao mercado. Está normalmente associado a uma redução do sacrifício.
- Relação preço/objetivos: esta caracterização é feita com o rácio entre os atributos percebidos e o preço, ou seja, o consumidor terá de concluir é a oferta entre estes atributos é honesta e justa.

O valor para o cliente pode ser interpretado com uma perspetiva longitudinal de valor e que engloba quatro valores temporais:

- Pré-compra: é caracterizada pelos desejos e ideias preconcebidas do consumidor em relação ao valor.
- Transação: valor para o cliente que é definido no momento da transação, aquisição ou troca.
- Pós-compra: altura em que são apresentados os resultados em função das escolhas dos clientes.
- Após utilização: reflete a experiência de utilização por parte do cliente e a altura em que faz a venda ou troca.

A previsão das convulsões epiléticas poderá ter um papel muito importante na vida dos doentes associados a esta patologia, pois este possível avanço permitiria que o doente passasse a ter uma vida considerada normal.

Estes doentes vivem o dia-a-dia em constante sobressalto pois nunca sabem quando aparecerá a próxima convulsão, mas com um sistema de notificação, poderiam estar mais relaxados, pois saberiam sempre quando é que uma convulsão iria acontecer.

Quando o paciente está indicado, por parte de um profissional de saúde para tomar medicação, esta poderia ser apenas administrada apenas quando o paciente obtivesse o aviso que a convulsão iria surgir e desta forma evitar uma sobrecarga de medicação.

Caso a medicação não seja uma opção para o paciente, poderiam existir outro tipo de abordagens, nomeadamente a estimulação elétrica do cérebro a fim de contrariar o efeito provocado pelas convulsões.

Na Tabela 1 é possível visualizar os benefícios e sacrifícios que os utilizadores deste sistema teriam em cada momento temporal.

Tabela 1 - Benefícios e Sacrifícios em cada valor temporal

| | Benefícios | Sacrifícios |
|---|---|---|
| Pré-compra | - Previsão/deteção convulsões<br>- Alertas automáticos<br>- Personalização | - Preço |
| Transação | - Previsão/deteção convulsões<br>- Alertas automáticos<br>- Personalização | - Custo da aquisição |
| Pós-compra | - Previsão/deteção convulsões<br>- Alertas automáticos<br>- Personalização | - Tempo de aprendizagem de manuseamento<br>- Tempo de formação ao paciente |
| Utilização | - Previsão/deteção convulsões<br>- Alertas automáticos<br>- Personalização | |

### 2.2.3 Negociação

Após uma análise a variadas definições de negociações, Aldo de Moor e Hans Weigand (de Moor and Weigand, 2004) concluíram que existem elementos em comum, tais como o facto de todas as negociações envolverem dois ou mais intervenientes e cada um com os seus objetivos e que poderão não ser totalmente compatíveis. Destacam também que as negociações não são estáticas, ou seja, há várias formas de se alcançarem os objetivos de cada um dos intervenientes.

Habitualmente são definidos dois tipos de negociação:

- Distributiva (ganha - perde): Este tipo de negociação normalmente existe em compras isoladas ou pouco frequentes. Não existe nenhuma solução para que ambas as partes possam ganhar, logo se uma parte ganha, é inevitável que a outra parte perca na mesma proporção. Nestes casos existe um ponto denominado como ponto de resistência que é o ponto que representa o ponto máximo de perdas e caso as essas perdas ultrapassem o ponto de resistência então a negociação é quebrada.
- Integrativa (ganha - ganha): quando esta negociação é aplicada, existem variadas soluções para os problemas apresentados. Os intervenientes cooperam de forma a

garantir que todos alcançam os objetivos definidos e normalmente as partes criam laços que lhes permite fazer novas cooperações no futuro.

Segundo Aldo de Moor e Hans Weigand, cada processo de negociação tem um ciclo de vida que consiste num número de etapas e dependendo do modelo utilizado estas etapas podem variar consideravelmente (de Moor and Weigand, 2004).

A maior parte dos modelos de negociação estão de acordo no facto de existir pelo menos algum tipo de (1) preparação da negociação, (2) condução da negociação e (3) implementação dos resultados, muitas vezes incluindo a renegociação (de Moor and Weigand, 2004).

No contexto em que este documento se insere é possível tentar prever que uma negociação do tipo Integrativa (ganha - ganha) terá uma maior importância, pois será importante manter e ir melhorando o relacionamento entre as partes envolvidas na negociação. É de crer que durante a negociação sejam claramente definidos os objetivos de ambas as partes. É muito importante ter um profundo conhecimento dos assuntos que serão abordados, é necessário ter bem claro onde se poderá ceder, onde é imprescindível manter os objetivos e tentar perceber o que a outra parte irá propor, de forma a poder efetuar contra propostas.

### 2.2.4 Modelo de negócio de Canvas

A ideia de negócio que poderá ser associada ao tema desta dissertação é a comercialização de um produto que engloba um componente de *hardware* conjugado com um *software*. A proposta passa por um dispositivo capaz de obter dados das correntes elétricas desenvolvidas no encéfalo em um paciente com epilepsia. Este dispositivo para além de captar esses dados teria de os analisar, interpretar e classificar, recorrendo a metodologias utilizadas nesta dissertação.

Em termos de modelo de negócio, este passaria pela comercialização indireta ao consumidor final, ou seja, a venda seria sugerida pelo profissional de saúde em ambientes médicos e hospitalares.

Na Tabela 2 é possível visualizar o Modelo de Canvas proposto de forma a poder criar um modelo de negócio através do tema desta dissertação.

Tabela 2- Modelo de Canvas

<table>
<tr><th>Key Partners</th><th>Key Activities</th><th>Value Propositions</th><th>Customer Relationships</th><th>Customer Segments</th></tr>
<tr><td rowspan="3">- Hospitais<br>- Profissionais de saúde</td><td>- Desenvolvimento de software<br>- Interligação com hardware<br>- Informar os profissionais de saúde</td><td rowspan="3">- Melhorias na vida de pacientes com epilepsia<br>- Monitorização de pacientes<br>- Redução de medicação</td><td>- Campanhas de marketing<br>- Suporte técnico através dos profissionais de saúde e com linhas dedicadas</td><td rowspan="3">- Pacientes com epilepsia</td></tr>
<tr><th>Key Resources</th><th>Channels</th></tr>
<tr><td>-Fornecedor de hardware</td><td>- Venda ao paciente através do médico</td></tr>
<tr><th colspan="3">Cost Structure</th><th colspan="2">Revenue Streams</th></tr>
<tr><td colspan="3">-Desenvolvimento e manutenção da aplicação<br>-Custo com o hardware<br>-Assistência ao cliente<br>-Salários</td><td colspan="2">- Compra do equipamento<br>- Compra da manutenção anual</td></tr>
</table>

### 2.2.5 Análise de criação de valor

Para ser possível quantificar o valor criado é necessário recorrer a métodos analíticos. Para o efeito os métodos mais utilizados são os seguintes:

- Teoria dos conjuntos aproximativos: não requer informações adicionais sobre os dados, pode trabalhar sobre dados imprecisos ou incertos. É capaz de descobrir factos importantes ocultos e é capaz de os expressar em linguagem natural (Liou and Tzeng, 2010).
- Teoria dos conjuntos nebulosos: esta teoria foi desenvolvida Zadeh em 1965 e define que uma verdade é expressa em valores linguísticos (verdade, muito verdade, pouco verdade, …) ao invés dos sistemas lógicos binários que apenas podem ser definidos através de dois valores (verdadeiro ou falso) (Gomide et al., 1995). Nos últimos anos

tem vindo a ser provado que esta teoria é útil para lidar com dados imprecisos (Bray et al., 2015).

- Teoria dos jogos: Segundo Sartini el al., a teoria dos jogos é uma teoria matemática criada para se modelar fenómenos que podem ser observados quando dois ou mais agentes de decisão interagem entre si. Ela fornece a linguagem para a descrição de processos de decisão conscientes e objetivos envolvendo mais do que um indivíduo.

Esta teoria tem inúmeras utilizações, e para ser aplicada é necessário a existência de dois ou mais "jogadores". Cada jogador tem a suas estratégias e terão de ser identificados todos os cenários possíveis para cada situação. Pode-se dizer então que um jogo tem um conjunto finito de jogadores, cada jogador possui um conjunto finito de opções (estratégias) e cada opção fica associada a um ganho (*payoff*).

Dependendo do cenário, o jogo poderá ter diferentes soluções. No caso de existência do "equilíbrio de Nash" significa que nenhum jogador pode melhorar a sua situação devido à estratégia que foi utilizada pelo adversário, ou seja, está a fazer o melhor possível perante o cenário apresentado. Existe também a possibilidade do jogo ter apenas um vencedor possível e nesses casos não se verifica o "equilíbrio de Nash". (Sartini et al., 2004)

- Apoio Multicritério: Os métodos multicritério utilizam modelações matemáticas para ajudar a tomada de decisão quando existem diferentes objetivos a atingir e com diferentes critérios de escolha. É possível destacar os seguintes:
    - AHP (*Analytic hierarchy process*): é uma técnica baseada em matemática e psicologia utiliza uma forma estruturada desenvolvida com a finalidade de ajudar o tratamento de decisões complexas. Este método ajuda a escolher a melhor opção dentro das alternativas possíveis apresentadas pelo problema, mas não define qual a opção correta (Boas and de Lima, 2010).
    - MACBETH (*Measuring attractiveness by a categorical based evaluation technique*): é uma técnica interativa de análise de decisão que permite fazer uma representação numérica sobre o que é considerado mais importante nas escolhas (Bana e Costa et al., 2011).
    - ELECTRE (*Elimination et choix traduisant la rélité*): este método utiliza as "relações de sobreclassificação (*outranking*) de tal modo que seja obtido um subconjunto k de alternativas possíveis, também chamado de mínimo subconjunto dominante"(Almeida et al., 2002). A ideia é conseguir distinguir

entre o conjunto total de alternativas, aquelas que são as preferidas na maioria dos critérios de avaliação (Mota and Almeida, 2007).

- PROMETHEE *(Preference ranking organization method for enrichment evaluations*): este método foi desenvolvido a partir do ELECTRE com a finalidade de desenvolver um método mais simples, pode-se portanto considerar uma ramificação do ELECTRE com algumas melhorias em relação ao antecessor (Morte, 2013).

# 3 Estado da arte

## 3.1 Algoritmos de classificação

Nesta secção apresentam-se alguns dos algoritmos de classificação que surgem na bibliografia e que serão utilizados neste trabalho. Apresentam-se também as medidas de avaliação mais comuns, que são utilizadas em problemas semelhantes.

### 3.1.1 Random Forests

Segundo Breiman (Breiman, 2001), *Random Forests* é um algoritmo que cria várias árvores de decisão a partir de um conjunto de dados de entrada, que posteriormente são utilizadas para classificação de um novo exemplo. Cada uma das árvores de decisão utiliza um subconjunto de atributos aleatórios a partir do conjunto original e atribui uma classificação. A classificação que será escolhida é a que tiver o maior número de votos.

Este algoritmo é bastante rápido, é executado de forma eficiente em grandes bases de dados, consegue lidar com milhares de variáveis de entrada e consegue estimar quais as variáveis mais importantes na classificação.

Tem um método eficaz para estimar os dados em falta e mantém a taxa de acerto mesmo quando uma grande parte dos dados está em falta (Breiman, 2001) (Random Forests, 2016).

### 3.1.2 J48

Para efetuar a classificação de um novo item, o algoritmo de árvore de decisão J48 começa por criar uma árvore de decisão baseado nos valores dos atributos fornecidos pela amostra e desta forma é possível identificar as diferentes instâncias com mais clareza. Quanto mais

identificativas forem as amostras melhor serão as classificações obtidas pelo algoritmo, ou seja, se a amostra tiver pouca ou nenhuma ambiguidade na classificação o algoritmo terá uma maior facilidade na atribuição da classe/categoria corretamente, ou seja, para cada instância que entre nessa categoria será atribuído o mesmo valor e o ramo terá o valor do maior número de instâncias. Para os outros casos será encontrado um atributo com o maior ganho. Desta forma segue o processo até se obter uma classificação clara ou até se ficar sem atributos. No caso de se esgotarem os atributos ou de não ser possível atribuir um valor inequívoco, o ramo terá o valor da maioria dos itens do ramo (Classification Methods, 2016).

### 3.1.3 Support Vector Machines

O algoritmo *Support vector machines* (SVM) é um algoritmo de classificação muito popular, não só pela sua forte fundamentação teórica como pelos resultados quando comparado com outros classificadores, principalmente perante uma grande dimensão de dados (Liu, 2007).
Segundo Hearst et al (Hearst et al., 1998), o principal objetivo do algoritmo *Support vector machines (SVM)* é definir um híper-plano que separe dois grupos de classes existentes e deixar a melhor margem de separação entre eles. Desta forma, pode ser aplicado em reconhecimento de padrões, estimativa de regressão e previsão de séries temporais.
O *SVM* permite aplicar técnicas lineares de classificação em dados não lineares, é aplicada uma transformação para que todos os dados sejam binários e seja possível efetuar uma separação através do híper-plano (Hearst et al., 1998).
O SMO (sequência mínima de otimização) surgiu em 1998 com o objetivo de resolver o problema da programação quadrática, pois até esse momento apenas era possível processar todo um determinado problema de uma só vez e com este algoritmo é possível fazer o processamento em partes mais pequenas, logo ficou possível fazer o processamento de uma quantidade muito maior de informação (Platt, 1998).

### 3.1.4 Regressão Logística

O modelo de Regressão Logística tem como objetivo a previsão de valores. Este modelo tem como base matemática o logaritmo natural de uma probabilidade. Utiliza uma variável binária, logo tem a possibilidade de assumir dois valores (por exemplo 0 e 1) (Peng et al., 2002).

### 3.1.5 Medidas de avaliação

Nesta subsecção apresentam-se as medidas mais frequentemente utilizadas na avaliação dos algoritmos de classificação na área dos EEG.

$$Sensibilidade = \frac{VP}{VP + FN} \quad (1)$$

$$Especificidade = \frac{VN}{VN + FP} \quad (2)$$

$$Taxa\ de\ acerto = \frac{VP + VN}{VN + FP + VP + FN} \quad (3)$$

$$Precisão = \frac{VP}{VP + FP} \quad (4)$$

O verdadeiro positivo (VP) corresponde ao número de convulsões detetadas no intervalo tanto pelos algoritmos como através da visualização de médicos experientes, os falsos negativos (FN) correspondem a todas as convulsões que não são detetadas pelos algoritmos no intervalo, mas são detetadas pelos médicos, o verdadeiro negativo (VN) corresponde ao número de não convulsões no intervalo que são avaliadas tanto pelo algoritmo como pelo médico e o falso positivo (FP) corresponde a não convulsões detetadas pelo algoritmo e que não foram detetadas pelo médico no período (Alotaiby et al., 2014).

Uma medida muito utilizada é a medida F (*F-Measure*) que é a média harmónica ponderada da taxa de acerto com a taxa de verdadeiros positivos.

A análise das curvas ROC (*receiver operating characteristic*) é uma alternativa de análise a classificadores em problemas binários. O gráfico ROC é um gráfico bidimensional num espaço denominado ROC, com dois eixos que representa as medidas de TFP (taxa de falsos positivos) e TVP (taxa de verdadeiros positivos). A área abaixo da curva ROC designa-se por AUC (*Area Under ROC Curve*) e varia entre 0 e 1. Valores mais próximos de 1 são considerados melhores (Gama et al., 2012).

## 3.2 Abordagens existentes

Nesta secção referem-se algumas das técnicas encontradas na bibliografia. Na Tabela 3 apresenta-se um resumo das mesmas. Na coluna Domínio/Tipo está identificado se a técnica se aplica a deteção ou previsão de convulsões e também que tipo de abordagem é que a técnica teve (por exemplo: tempo, *wavelet*, etc). Na coluna *Dataset* apresenta-se a origem dos dados, bem como o tipo de canal utilizado (canal único ou multicanal). Na coluna Algoritmo é possível ver o(s) algoritmo(s) utilizados durante a aplicação das técnicas. Na coluna Avaliação encontram-se as características que os diferentes autores utilizaram para avaliar as diferentes soluções. Na coluna Comentários estão algumas palavras-chave ou comentários que poderão ser associados às técnicas em questão.

Tabela 3 – Resumo das técnicas analisadas

| Referência | Domínio / Tipo | Dataset | Algoritmo | Avaliação | Comentários |
|---|---|---|---|---|---|
| (Runarsson and Sigurdsson, 2005) | Deteção convulsões (Tempo) | Dados próprios (Canal único) | SVM | Sensibilidade: 90% | Análise de histograma entre variação mínima e máxima |
| (Yoo et al., 2013) | Deteção convulsões (Tempo) | Base de dados MIT (EEG couro cabeludo) (Multicanal) | SVM | Taxa de acerto: 84.4% | Energias de sub-bandas |
| (Dalton et al., 2012) | Deteção convulsões (Tempo) | Dataset com 21 convulsões (Canal único) | Template mathching | Sensibilidade: 91 %, Especificidade:84 % | Assinatura da convulsão |
| (Zandi et al., 2010) e (Shahidi Zandi et al., 2013) | Previsão convulsões (Tempo) | 561h de EEG de couro cabeludo com 86 convulsões em 20 pacientes (Multicanal) | SVM não linear com classif. Gaussiana | Sensibilidade: 88.34%, taxa de falsas previsões: 0.155 h−1, tempo de previsão médio: 22.5 min | Índice combinado |
| (Aarabi and He, 2012) | Previsão convulsões (Tempo) | 316h de dados iEEG com 49 convulsões em 11 pacientes da base de dados de Freiburg (Canal único) | abordagem própria | Sensibilidade média: 79.9%, 90.2% com taxa de previsão falsa média: 0.17 e 0.11/h | Corelação de tamanho, nível de ruido |

| Referência | Domínio / Tipo | Dataset | Algoritmo | Avaliação | Comentários |
|---|---|---|---|---|---|
| (Schelter et al., 2011) | Previsão convulsões (Tempo) | 8 pacientes com epilepsia focal (Canal único) | Abordagem própria | Sensibilidade: 60% | Limite probabilístico |
| (Wang et al., 2010) | Previsão convulsões (Tempo) | Dados de 5 pacientes (Canal único) | KNN | Taxa de acerto médio: 70% | |
| (Bedeeuzzaman et al., 2014) | Previsão convulsões (Tempo) | 21 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) (Canal único) | Classificador linear binário | Sensibilidade: 100%, taxa falsos positivos: 0 (em 12 pacientes), tempo de previsão médio: 51-96m | |
| (Howbert et al., 2014) | Previsão convulsões (Tempo) | 3 Cães (16 canais iEEG) | Regressão Logística com treino | Sensibilidade<br>Taxa falsos positivos | 10 folder Cross validation |
| (Li et al., 2013) | Previsão convulsões (Tempo) | 21 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) (Multicanal) | Abordagem própria | Taxa de acerto: 75.8%, taxa de falsas previsões: 0.09 | |
| (Chisci et al., 2010) | Previsão convulsões (Tempo) | 21 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) (Multicanal) | SVM | Sensibilidade: 100%, tempo previsão médio: 60 min | |
| (Park et al., 2011) | Previsão convulsões (Tempo, bipolar e potência espetral) | 18 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) (Multicanal) | SVM | Sensibilidade: 93,8%, | Resampling, bipolar |
| (Rana et al., 2012) | Deteção convulsões (Frequência) | EEG e ECoG series de tempo, Entre: 0.5 a 100 Hz (Multicanal) | Abordagem própria | Taxa de acerto de deteção: 100% | Índice de inclinação |

| Referência | Domínio / Tipo | Dataset | Algoritmo | Avaliação | Comentários |
|---|---|---|---|---|---|
| (Khamis et al., 2013) | Deteção convulsões (Frequência) | 12 pacientes com 6 gravações de dados (R1 ao R6) (Canal único) | Powell's direction set method | Sensibilidade 91%, falsos positivos por hora: 0.02 | Intervalo de frequências |
| (Acharya et al., 2012) | Deteção convulsões (Frequência) | Dados próprios (Canal único) | SVM, FSC, PNN, KNN, NBC, DT e GMM | Taxa de acerto médio: 98,1% | |
| (Zhou et al., 2013) | Deteção convulsões (Wavelet) | 21 pacientes da base de dados de Freiburg (Multicanal) | Bayesian linear discriminant analysis (BLDA) | Sensibilidade 96.25%, taxa de falsas deteções: 0.13/h, tempo atraso médio: 13.8s | Índice de lacunaridade e flutuação em escalas waveltet |
| (Liu et al., 2012) | Deteção convulsões (Wavelet) | 21 pacientes com epilepsia com 509h (Multicanal) | SVM | Sensibilidade 94.46%, Especificidade: 95.26%, taxa de falsas deteções: 0.58/h | Energia relativa, amplitude relativa, coeficiente de variação, índice de flutuação |
| (Panda et al., 2010) | Deteção convulsões (Wavelet) | 500 Intervalos EEG de 5 atividades cerebrais diferentes (100 sinais por intervalo) (Canal único) | SVM | Taxa de acerto: 91,2% | Energia, entropia, desvio padrão |
| (Khan et al., 2012) | Deteção convulsões (Wavelet) | Dados próprios de 5 pacientes (Canal único) | LDA Simples | Sensibilidade: 83%, Especificidade: 100%, Taxa de acerto: 92%, Taxa de acerto geral: 87% | Energia e coeficiente de variação normalizado |
| (Hung et al., 2010) | Previsão convulsões (Wavelet) | 11 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) | chanel interaction metric | Taxa de acerto: 87%, Taxa falsa previsão: 0.24/h, tempo médio aviso: 27 min antes do ictal | Características Chaos, características wavelet, dimensão correlação |
| (Chiang et al., 2011) | Previsão convulsões (Wavelet) | 8 pacientes base dados Freiburg 1 paciente Hospital da Universidade Nacional Taiwan (Canal único) | SVM | Sensibilidade: 74.2% e 52.2% | Wavelet |

| Referência | Domínio / Tipo | Dataset | Algoritmo | Avaliação | Comentários |
|---|---|---|---|---|---|
| (Soleimani-B et al., 2012) | Previsão convulsões (Wavelet e Tempo) | 21 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) (Canal único) | Neuro-fuzzy | Taxa de acerto: 99.52%, Taxa de falha de previsão: 0.1417/h | |
| (Moghim and Corne, 2011) | Previsão convulsões (Wavelet e Tempo) | 1 paciente da base de dados de Freiburg (couro cabeludo e iEEG) (Canal único) | MC-SVM e EANN | Sensibilidade e especificidade com valores variados conforme a experiência | |
| (Tafreshi et al., 2008) | Deteção convulsões (EMD) | 5 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) (Canal único) | Rede neuronal | Taxa de acerto: 95% | Média absoluta para cada IMF, características wavelet |
| (Alam and Bhuiyan, 2013) e (Alam and Bhuiyan, 2011) | Deteção convulsões (EMD) | Base de dados da Universidade de Bonn (EEG de couro cabeludo) (Canal único) | ANN | Taxa de acerto: 100% | Assimetria, variância, dimensão de correlação, maior expoente de Lyapunov, entropia aproximada de modos intrínsecos |
| (Qi et al., 2012) | Previsão convulsões (EMD) | 4 pacientes da base de dados pré-cirurgica com 80.4 h de dados de EEG com 10 convulsões (Canal único) | Abordagem própria | Sensibilidade: 100%, taxa de falsas deteções: 0.16/h, atraso médio: 10.7 a 19.4 s | IMF-Vo |
| (Vanrumste et al., 2002) | Previsão convulsões (SVD) | Simulação de sinais de EEG (Canal único) | combined IMF-VoE feature | Inspeção visual | Energia relativa residual |
| (Shahid et al., 2013) | Previsão convulsões (SVD) | Gravações em 4 pacientes pediátricos com 20 convulsões (Canal único) | SVD and dipole model approach | Inspeção visual | |
| (Xie and Krishnan, 2011) | Deteção convulsões (PCA com wavelet) | Dados de 8 pacientes (Canal único) | Adjusted random index (ARI) até 1 | Classificador empírico | Adjusted random index (ARI) |

| Referência | Domínio / Tipo | Dataset | Algoritmo | Avaliação | Comentários |
|---|---|---|---|---|---|
| (Miri and Nasrabadi, 2011) | Previsão convulsões (PCA) | 6 pacientes da base de dados de Freiburg (couro cabeludo e iEEG) (Multicanal) | Abordagem própria | Tempo de previsão: 20 ± 5 min, sensibilidade 86.6%, taxa de falsos alarmes: 0.067% | Taxa Zero-Crossing, índice estatístico, resampling |
| (Gomes et al., 2014) | Classificação batimentos cardiacos (Random Forests e Text Mining) | 312 auscultações do Real Hospital Português (Recife, Brasil) | Random Forest | Taxa de acerto | Random Forests e Text Mining, MrMotif |

É possível perceber que os diferentes autores optam por diferentes abordagens para o problema descrito. A maioria enquadra os seus métodos no domínio do tempo, mas também há outras abordagens, tais como, o domínio da frequência ou *wavelet*. Existem estudos que utilizam apenas um canal para deteção ou previsão das convulsões, mas existem outros que optam pela utilização de multicanal.

O algoritmo predominante para a avaliação dos resultados é o SVM, mas aqui as opções também são bastante diversificadas e será importante também realçar o *Random Forest* e Regressão Logística pois serão utilizados no modelo proposto. Os parâmetros mais utilizados na avaliação pelos autores analisados são a sensibilidade e a taxa de acerto.

De referir também que existe uma grande diversidade de abordagens e que alguns autores utilizam ferramentas de pré-processamento, tais como *resampling* (reamostragem) e pré-processamento bipolar.

### 3.2.1 Métodos no domínio do tempo

Segundo Carvalho et al., têm vindo a ser desenvolvidos, ao longo dos anos, vários algoritmos e técnicas de análise e processamento de sinais para a extração de informações relevantes de registos de séries temporais (Carvalho et al., 2014).

Segundo Alotaiby et al., de um modo geral os sinais obtidos através de EEG variam de paciente para paciente, portanto, os algoritmos para deteção e previsão de convulsões normalmente são específicos para cada paciente (Alotaiby et al., 2014).

A Figura 2 apresenta uma imagem com a informação de um comportamento normal, uma no estado de pré-convulsão e também uma convulsão.

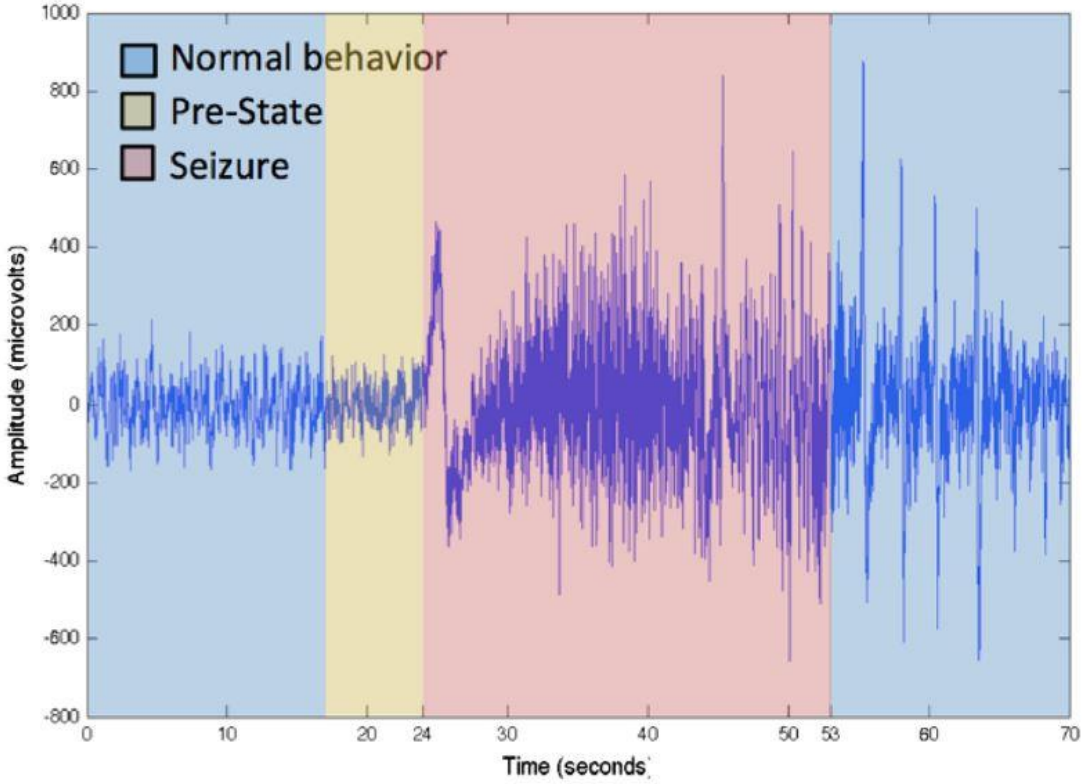


Figura 2 – EEG com uma convulsão (Alotaiby et al., 2014)

É possível visualizar na Figura 2 que existe uma diferença entre o intervalo de tempo em que está registada uma convulsão e os restantes. O objetivo dos métodos no domínio do tempo é que sejam capazes de fazer a deteção automática e de avaliar o desempenho conseguido pelas diferentes métricas, tais como sensibilidade, especificidade, taxa de acerto (*accuracy*) e falso positivo.

Uma das técnicas de previsão de convulsões mais utilizada no estudo de EEG é a que usa uma janela (intervalo) com um tamanho pré-definido e a desloca ao longo do tempo. Cada janela contém dados que são analisados. Quando é ultrapassado um determinado valor pré-definido, é acionado um alarme. Estas janelas podem conter um ou mais canais do EEG e assim que terminar a análise da área em questão, a janela move-se para a área seguinte, tal como está ilustrado na Figura 3 (Mormann, 2008).

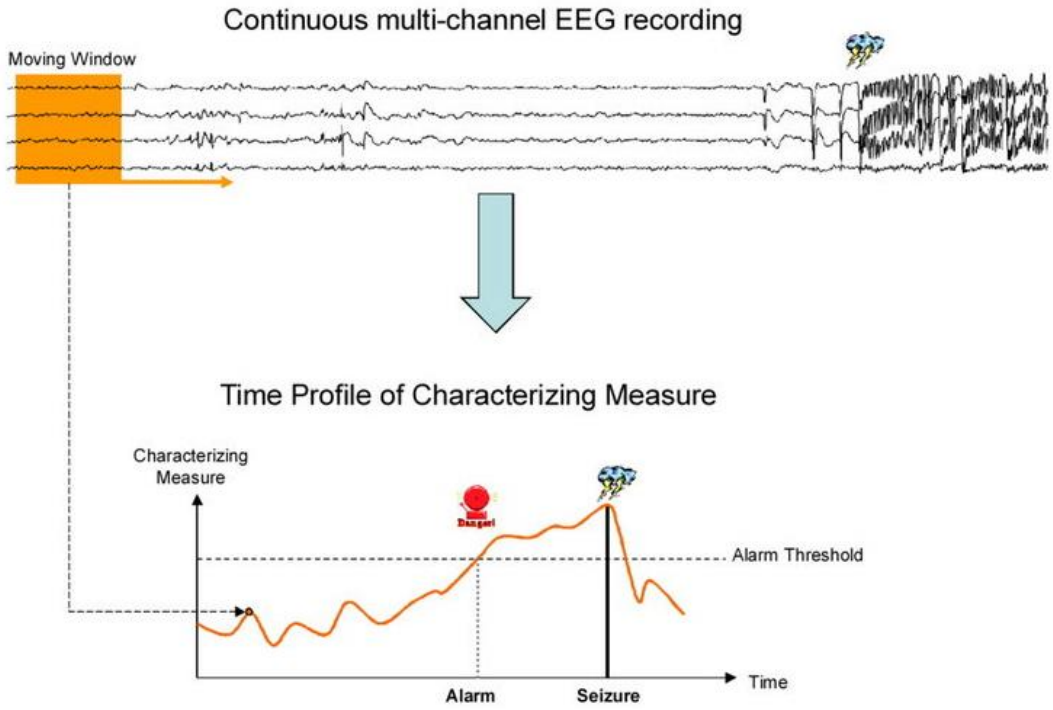


Figura 3 – Utilização da técnica da janela para previsão de convulsões (Mormann, 2008)

Caso o algoritmo detete que foi ultrapassado o limite considerado normal é emitido um alarme. É importante perceber se esta notificação corresponde a um alarme verdadeiro ou falso. Um alarme é considerado verdadeiro quando ocorre uma convulsão dentro de um período de tempo estipulado após o aviso (verdadeiro positivo), caso seja emitido o aviso e não haja a ocorrência de uma convulsão, considera-se um falso alarme (falso positivo). Os intervalos estipulados relatados na literatura estão compreendidos entre alguns minutos até algumas horas. Caso uma convulsão ocorra sem que o alarme seja acionado significa que a previsão falhou (falso negativo).

A sensibilidade de um algoritmo de previsão de convulsões habitualmente é quantificada através do número de convulsões com pelo menos um alarme dentro do horizonte de previsão anterior, dividido pelo número total de convulsões ocorridas. Uma medida de quantificação frequente encontrada na bibliografia é a taxa de falsas previsões que acontecem por hora (Mormann, 2008).

### 3.2.2 Métodos no domínio da frequência

Segundo Alotaiby et al., as técnicas no domínio da frequência têm sido utilizados para a deteção de convulsões através de EEG. A transformada de Fourier de amplitude e fase é muito utilizada neste contexto (Alotaiby et al., 2014).

Conforme Subasi e Ercelebi (Subasi and Ercelebi, 2005) relataram, este tipo de abordagem baseia-se em observações ao EEG que contenha uma forma de onda característica que encaixe essencialmente dentro de quatro bandas de frequência que são definidas como delta (<4 Hz), theta (4-8 Hz), alfa (8—13 Hz) e beta (13—30 Hz).

De uma forma geral, os sinais dos EEG são não-lineares e não-estacionários, devido a esses factos, existe uma dificuldade em caracterizar diferentes atividades dos sinais dos EEG através de determinados modelos matemáticos (Alotaiby et al., 2014). Por este motivo, Acharya et al., elaboraram um método modificado para a deteção do estado normal, pré-ictal e ictal a partir de sinais de EEG. Este modelo baseia-se em quatro características de entropia para a classificação: entropia fase 1 (S1), entropia fase 2 (S2), entropia aproximada (ApEn) e entropia amostra (SampEn). S1 e S2 são estimados a partir de espectros mais altos dos EEG e esses sinais são reveladores de estados ictal, preictal e interictal. ApEn e SampEn são métricas logarítmicas que representam a maior aproximação e correspondência entre os padrões de sinais obtidos dos EEG e os *templates*. Todas estas informações são extraídas dos EEG e posteriormente classificadas através de variados algoritmos (por exemplo, *SVM*, *fuzzy Sugeno* e *naive Bayes*). Os resultados mostraram que o algoritmo de classificação *fuzzy Sugeno* foi o que obteve a melhor taxa de acerto, 98,1% (Acharya et al., 2012).

### 3.2.3 Métodos no domínio de wavelet

Segundo Adeli et al., *wavelets* podem ser definidas como ondas pequenas de duração limitada e 0 valores médios. São funções matemáticas capazes de localizar uma função ou um conjunto de dados, tanto em tempo como em frequência.

A transformada *wavelet* é uma ferramenta eficaz no processamento de sinais devido às suas características, tais como, tempo/frequência, localização (obtendo um determinado sinal num determinado tempo e frequência, ou através da extração de determinadas localizações espaciais em diferentes escalas) e filtragem multi-taxa (diferenciar os sinais em diferentes frequências) (Adeli et al., 2003).

A transformada *wavelet* pode ser considerado como um tipo de decomposição em sub-bandas, mas com diminuição da resolução e pode ser implementada tanto em sinais analógicos, como em sinais digitais.

A habilidade de poder extrair características específicas de cada uma das sub-bandas existentes no EEG é o principal motivo para que este método seja utilizado na previsão e deteção de convulsões. Estes dados são extraídos com o objetivo de posteriormente poderem ser classificados (Alotaiby et al., 2014).

### 3.2.4 Método de decomposição empírica

Segundo Alotaiby et al., o método de decomposição empírica (EMD) é um método de decomposição de sinal, que transforma um sinal num grupo de funções de modo intrínsecas (IMFs) (Alotaiby et al., 2014).

Para a deteção e previsão de convulsões em EEG, as IMFs mostram um comportamento diferente perante a ocorrência de atividades normais e anormais nos sinais. As etapas necessárias para efetuar a extração das IMFs de um determinado sinal original, $\mathrm{x(t)}$ são as seguintes (Eftekhar et al., 2008):

1. Extrair os pontos máximos e mínimos do sinal e fazer interpolação entre eles para determinar o envelope máximo e o mínimo.
2. Calcular a média local $\mathrm{m(t)}$, utilizando estes envelopes através da equação (5) ($\mathrm{emin}$ e $\mathrm{emax}$ representam os envelopes máximo e mínimo):

$$m(t) = \frac{emin(t) + emax(t)}{2} \tag{5}$$

3. É feita a subtração desta média ao original e é utilizado o resultado como o novo sinal $\mathrm{h\,(t)\ =\ x\,(t)\ -\ m\,(t)}$.
4. Se $\mathrm{h\,(t)}$ não corresponder aos critérios de uma IMF (em que a diferença tem de ser no máximo 1 entre os números máximo e mínimo e 0 cruzamentos), então volta-se para o ponto 1, mas sendo $\mathrm{h\,(t)}$ o novo valor de entrada.
5. O $\mathrm{h\,(t)}$ é armazenado como um IMF (caso satisfaça o critério). Consequentemente, este IMF é removido do sinal original através da subtração $\mathrm{r\,(t)\ =h\,(t)\ -\ x\,(t)}$
6. Passos idênticos são realizados voltando a iniciar no passo 1.

### 3.2.5 Método de decomposição em valores singulares

Segundo Shieh et al, o método de decomposição em valores singulares (SVD) é uma das ferramentas de grande utilidade da álgebra linear com diversas aplicações na codificação de imagens, estimativa de ruído e mais recentemente em marcas d'água (Shieh et al., 2006).
O SVD atua sobre uma matriz A que se decompõem em matrizes U, S e V:

$$A = USV^T \quad (6)$$

As matrizes U e V são ortogonais tais que $U^T U = I$, e $V^T V = I$. S = diag (σ1, ..., σP), onde σ1 ≥ σ2 ≥ ... ΣP ≥ 0 é a matriz de SVs de A. A matriz U contém os vetores singulares esquerdo de A, enquanto a matriz V contém os vetores singulares direita de A. Uma característica importante do SVD que pode ser utilizada na deteção de convulsões em EEG é que os SVs não são significativamente afetadas por pequenas perturbações na matriz A. Esta propriedade pode ser usada para a deteção de atividades convulsivas, tendo em consideração as ligeiras variações entre as crises (Alotaiby et al., 2014).
Shahid et al. desenvolveram um algoritmo baseado em SVD para deteção de convulsões em EEG. O SVD é aplicado de uma forma sequencial através de uma janela deslizante com 1 segundo de comprimento em dados de um EEG e os r valores singulares são obtidos e utilizados para indicar alterações repentinas nos sinais (Shahid et al., 2013).

### 3.2.6 Método de análise de componentes principais

Segundo Ghosh-Dastidar et al., o método de análise de componentes principais (PCA) é geralmente utilizado para a redução da dimensionalidade com o objetivo de reduzir a exigência computacional, e, em alguns casos, para a filtragem (Ghosh-Dastidar et al., 2008).
Este método é normalmente utilizado em conjunto com outros métodos na deteção ou previsão de convulsões (Alotaiby et al., 2014).

### 3.2.7 Método de análise de componentes independentes

Segundo Comon, a análise de componentes independentes (ICA) de um vetor aleatório consiste na pesquisa de uma transformação linear que minimiza a dependência estatística entre os seus componentes (Comon, 1994).

O principal princípio de que o ICA depende é a maximização da independência estatística entre os componentes estimados. O ICA depende, principalmente, de algumas técnicas de pré-processamento, tais como remoção média (centralização), branqueamento (*whitening*), e redução da dimensão. Tanto a redução de branqueamento como a de dimensão podem ser conseguidos com PCA ou SVD (Alotaiby et al., 2014).

## 3.3 Multiresolution Motif Discovery

No âmbito desta dissertação, a extração de atributos será realizada utilizando o algoritmo *Multiresolution Motif Discovery in Time Series* (*MrMotif*). A extração dos padrões frequentes em bases de dados temporais é uma importante tarefa de *data mining*.

Estes padrões, os *motifs*, fornecem informações uteis ao especialista e ajudam a resumir os dados na serie temporal (Castro and Azevedo, 2010).

Os *motifs* são utilizados nas mais variadas áreas, inclusivamente em séries temporais de EEG e poderá identificar um padrão que precede uma convulsão, tal como representado na Figura 4.

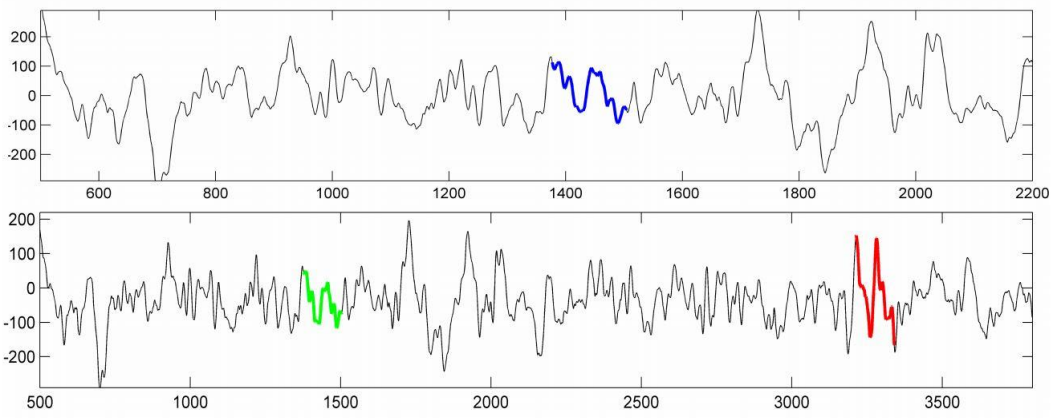


Figura 4 – M*otif* numa série temporal de um EEG (Castro and Azevedo, 2010)

O *MrMotif* baseia-se na representação *indexable Symbolic Aggregate approXimation (iSAX*), que segundo Shieh e Keogh é uma representação que suporta a indexação de grandes conjuntos de dados, tendo mostrado que consegue indexar até cem milhões de séries temporais (Shieh and Keogh, 2008). O iSAX converte uma série temporal numa sequência de símbolos (*word*).

Este algoritmo tem vindo a ser utilizado na extração de atributos para classificação de sons cardíacos (Gomes et al., 2013), (Gomes et al., 2014) e de sons urbanos (Gomes and Batista, 2015a), (Gomes and Batista, 2015b).

## 3.4 Avaliação das soluções

### 3.4.1 Medidas de avaliação

Ao longo da análise do estado da arte do tema proposto nesta dissertação, foi possível verificar que os diferentes autores tinham a preocupação de utilizar métodos para avaliação das diferentes soluções apresentadas. Foi possível notar que a maior parte dos autores utilizam os seguintes parâmetros para poder efetuar a avaliação: sensibilidade; especificidade; taxa de acerto. Informações mais detalhadas podem ser consultadas na secção 3.1 desta dissertação.

A medida *Area Under the ROC Curve* (AUC) tem características que a tornam uma candidata na classificação do problema relatado nesta dissertação, pois, segundo Gama et al, um problema de classificação binário (duas classes) poderá obter resultados interessantes através desta medida de avaliação (Gama et al., 2012).

No artigo *Classifying Heart Sounds using SAX Motifs, Random Forests and Text Mining techniques* (Gomes et al., 2014) foi utilizado a taxa de acerto como medida de avaliação e comparação, esta medida foi comparada nas diferentes classes (*normal, non-normal* e *murmur*).

### 3.4.2 Hipóteses e metodologias de avaliação

Ao longo do estudo e análise das diferentes metodologias de análise de EEG, foi possível verificar que muitos autores utilizam a comparação dos seus métodos contra a visualização médica realizada por profissionais de saúde.

No artigo Gomes et al. (Gomes et al., 2014), foi testada a hipótese nula da taxa de acerto ser igual entre as abordagens TF (*term frequency*) e TFIDF (*Term frequency - Inverse Document Frequency*), entre as abordagens TF e TFWBG (*Term frequency - Inverse Document Frequency with bi-grams*) contra a hipótese alternativa de que as soluções propostas obteriam melhores resultados (em todos os casos foi utilizado o algoritmo de classificação *Random Forest*). Estas

abordagens, TF, TFIDF e TFWBG, correspondem a diferentes metodologias utilizadas na extração de atributos de sons cardíacos.

### 3.4.3 Teste estatístico

No artigo referido anteriormente (Gomes et al., 2014), foi utilizado o teste estatístico utilizado foi o "Wilcoxon signed-rank" (teste não paramétrico para comparação de duas amostras emparelhadas).

## 3.5 Tecnologias utilizadas

### 3.5.1 Java

Java é uma linguagem de programação desenvolvida por James Gosling, juntamente com a sua equipa, na década de 1990, na empresa *Sun Microsystems*. Atualmente o Java é a linguagem mais popular em todo o mundo (Cass, 2015).

Os programas em Java são interpretados por outro programa chamado Java VM. Em vez de correr diretamente no sistema operativo nativo, o programa é interpretado pelo Java VM para o sistema operativo nativo. Isto significa que qualquer sistema que tenha o Java VM instalado, poderá correr aplicações em Java independentemente do sistema operativo em que foi desenvolvido (Java Essentials, 2016).

A linguagem de programação Java foi a escolhida para utilizar no desenvolvimento desta dissertação.

### 3.5.2 NetBeans

O NetBeans IDE é um ambiente de desenvolvimento de utilização livre, que permite escrever, compilar, depurar e executar programas. O projeto NetBeans foi criado pela *Sun Microsystems* em Junho de 2000. Todo o ambiente é desenvolvido em Java, mas suporta várias linguagens de programação (NetBeans, 2016).

Esta é a ferramenta escolhida para utilizar durante o desenvolvimento desta dissertação.

### 3.5.3 *Weka*

O *Weka* permite analisar *datasets* com algoritmos de aprendizagem. Este *software* foi desenvolvido na Nova Zelândia, mais concretamente na Universidade de Waikato. É desenvolvido em Java e corre em diferentes sistemas operativos.

O *Weka* tem uma interface uniforme para diferentes algoritmos de aprendizagem, juntamento com métodos de pré-processamento, pós-processamento e para avaliação de resultados obtidos através de esquemas de aprendizagem de determinado *dataset*. Também contém variadas ferramentas de *data mining*: classificação, regressão, *clustering*, associação de regras e seleção de atributos.

Existem várias formas de utilização desta ferramenta:

- Pode ser utilizada para aplicar um método de aprendizagem para um conjunto de dados e analisar a sua saída com o objetivo de obter mais informações acerca dos dados;
- Utilizar modelos previamente aprendidos para gerar previsões sobre novas instâncias;
- Aplicar diferentes aprendizagens e comparar a performance com o objetivo de escolher uma delas para previsão.
- É possível utilizar ficheiros ARFF ou folhas de cálculo e fazer vários tipos de personalizações de forma a poder adaptar a diferentes necessidades (Witten et al., 2011).

Esta é a ferramenta selecionada para correr os algoritmos de classificação.

# 4 Solução

Neste capítulo, apresenta-se o *design* da solução, onde se podem obter informações sobre os *datasets* utilizados e também sobre a forma como esta metodologia foi estruturada.
Ainda neste capítulo, é possível visualizar a implementação da solução proposta que contém os detalhes sobre cada um dos procedimentos efetuados com o objetivo de resolver o problema descrito na subsecção 2.1.3.

## 4.1 Design da solução

### 4.1.1 Dataset

Os *datasets* utilizados nesta dissertação foram fornecidos pela Associação Americana de Epilepsia[1], no âmbito de uma competição com o objetivo de prever convulsões através de gravações de EEG intitulada "Predict seizures in intracranial EEG recordings" (Kaggle.com, 2016). Os dados recolhidos pertencem a humanos e cães, foram obtidos a partir de EEG intracranianos e contêm dados em diferentes estados, nomeadamente, no estado preictal (estado anterior às convulsões) e no estado interictal (estado entre convulsões).
Os EEG foram gravados em cães com convulsões naturais utilizando um sistema de monitorização em regime ambulatório. Os EEG foram realizados a partir de 16 elétrodos a 400Hz. Estes *datasets* são de longa duração, pois foram obtidos durante vários meses (até um ano) e em alguns casos ocorreram centenas de convulsões.

[1] https://www.aesnet.org/

No caso dos humanos, o número de elétrodos foi variável (Paciente 1: 14; Paciente 2: 24), foram obtidos a 5000Hz e a duração foi no máximo de uma semana.
Os *datasets* do estado preictal contêm dados até uma hora antes da ocorrência de convulsão, com o objetivo de disponibilizar um intervalo de tempo suficientemente grande, que permita a emissão de notificações aos pacientes antes da ocorrência das convulsões.
De forma idêntica, os *datasets* que contêm a informação do estado interictal, também têm a duração de uma hora e foram recolhidos aleatoriamente do total das gravações. Para evitar o surgimento de contaminação dos dados, estes foram recolhidos em espaços temporais afastados das convulsões. No caso dos cães, os intervalos foram de uma semana, no caso dos humanos, o intervalo utilizado foi de quatro horas, devido ao facto de não existirem tantas horas de gravação (Kaggle.com, 2016).

### 4.1.2 Processo

A metodologia de desenvolvimento que é proposta nesta dissertação, tendo em vista a solução do problema apresentado é composta por cinco grandes etapas:

- Transformar o *dataset*;
- Pré-processamento;
- Extração de atributos;
- Reamostragem;
- Avaliação.

Todo o processo tem início na transformação do *dataset*, seguidamente são realizadas diferentes ações de pré-processamento dos dados. Com as duas primeiras fases concluídas, é possível gerar o conjunto dos atributos através da execução do *MrMotif* com diferentes parâmetros. Na fase seguinte foram testados os resultados obtidos com e sem recurso à técnica de reamostragem (*resample)* para diferentes algoritmos de classificação, nomeadamente o *Random Forest*, J48, SVM e Regressão Logística.
Na Figura 5 é possível verificar o diagrama de atividades que representa de uma forma geral o processo de desenvolvimento da solução.

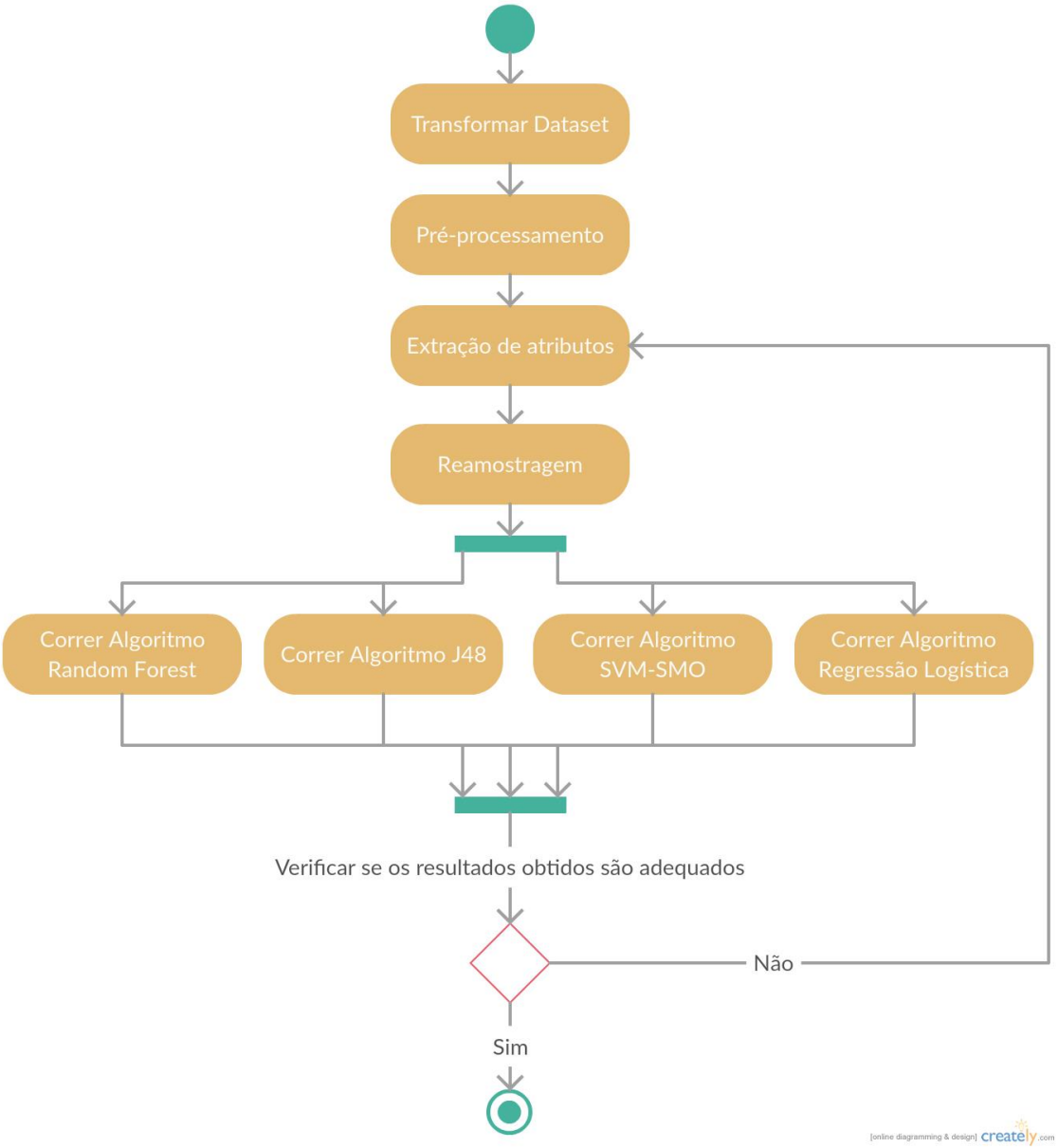


Figura 5 – Diagrama de atividades[2] do processo de desenvolvimento da solução

Na primeira etapa é efetuada toda a conversão do *dataset*, pois os dados estão gravados em formato de ficheiro Matlab[3] e as ferramentas propostas não trabalham com esse tipo de formatação. Na segunda etapa efetua-se o pré-processamento dos dados, aplicando os seguintes passos:

[2] Diagrama de atividades realizado em https://creately.com
[3] http://www.mathworks.com/products/matlab/index.html?s_tid=gn_loc_drop

- Aplicação de filtros para redução de ruido;
- Redução do tamanho da amostra (*decimate*);
- Criação do envelope de Shannon

A terceira etapa tem como objetivo a extração de atributos do *dataset* e gerar um novo conjunto de dados, composto pelos atributos extraídos. Para tal, utiliza-se o *MrMotif*. Esta fase implica um estudo dos parâmetros mais adequados aos dados de entrada, de acordo com as especificidades do *dataset*.

Após a obtenção do conjunto de atributos, poderá haver a necessidade de efetuar um *resampling* (reamostragem).

Posteriormente, aplicam-se os algoritmos de classificação *Random Forest*, J48, *SVM* e Regressão Logística através do *Weka*.

Por fim, é necessário efetuar uma análise e avaliação dos resultados obtidos para cada modelo de classificação resultante da aplicação de cada um dos algoritmos anteriormente referidos.

## 4.2 Implementação

Nesta secção será apresentada e detalhada toda a sequência de procedimentos realizados para a implementação do *design* sugerido anteriormente.

É importante referir que existem alguns fatores muito importantes para obter um modelo de sucesso:

- *dataset* (conjunto de dados);
- Extração dos atributos;
- Algoritmos de classificação.

### 4.2.1 Dataset

Conforme referido na subseção 4.1.1 deste documento, os *datasets* utilizados para a realização desta dissertação são disponibilizados pela Associação Americana de Epilepsia (Kaggle.com, 2016). Foram utilizados *datasets* de dois humanos e de dois cães, com as seguintes características:

Paciente 1:

- Classe interictal: 50 ficheiros .mat;
- Classe preictal: 18 ficheiros .mat.

Paciente 2:

- Classe interictal: 42 ficheiros .mat;
- Classe preictal: 18 ficheiros .mat.

Cão 1:

- Classe interictal: 480 ficheiros .mat;
- Classe preictal: 24 ficheiros .mat.

Cão 2:

- Classe interictal: 500 ficheiros .mat;
- Classe preictal: 42 ficheiros .mat.

Cada ficheiro do *dataset* contém pastas com as séries interictais e preictais correspondentes a cada um dos humanos e também a cada um dos cães. Cada um dos ficheiros contido nestas pastas corresponde a 10 minutos de um EEG. Os ficheiros estão ordenados numericamente e armazenados em ficheiros de Matlab (.mat) com a seguinte estrutura:

- data: uma matriz de amostras de valores EEG em que cada linha corresponde aos eletrodos e cada coluna corresponde ao tempo;
- data_length_sec: o tempo de duração de cada linha;
- sampling_frequency: o número de amostras que representam 1 segundo de dados do EEG;
- channels: uma lista dos nomes dos eletrodos correspondents às linhas existentes nos dados;
- sequence: o índice do segmento de dados que contém uma hora de dados do EEG. Por exemplo: preictal_segment_6.mat contém o número 6 e representa dados obtidos do EEG entre 50 e 60 minutos de dados preictais.

### 4.2.2 Transformar *dataset* e pré-processamento

Para a transformação do *dataset* de ficheiros de matlab para ficheiros csv *(comma-separated values)* foi utilizada uma aplicação em Java. Este processo foi realizado pois csv é o formato que o *MrMotif* recebe para poder fazer a extração dos atributos.

Posteriormente, os registos são filtrados, decimados (utilizando métodos da classe jmatio-1.0.jar) e é construído o envelope de energia de Shannon normalizado. O procedimento consiste em dividir o sinal em pequenas janelas de intervalo de tempo para que o período de cada janela seja menor que metade da sua frequência, e a seguir é calculada a energia de Shannon aplicando a equação (7) para cada janela (Gomes et al., 2013).

$$E = x^2 * \log x^2 \tag{7}$$

O resultado deste processamento serve como *input* para o passo seguinte, a extração de atributos.

### 4.2.3 Extração de atributos

Para extração dos atributos, os *motifs*, é utilizada uma versão já bastante modificada de uma ferramenta, codificada em Java por Nuno Casto e Paulo Azevedo, intitulada Multiresolution Motif Discovery (*MrMotif*) (Castro and Azevedo, 2010).

Para que fosse possível executar o *MrMotf* de uma forma sistemática, foi desenvolvida uma aplicação, que permite reorganizar e dividir em pastas todos os ficheiros resultantes da conversão do *dataset* e respetivo pré-processamento, relativos ao Paciente 1, Paciente 2, Cão 1 e Cão 2.

Neste problema, pretende-se executar o *MrMotif* de modo a obter frequências dos *K* motifs principais, agregando canais os EEG para cada clip. Por exemplo, para os Cães 1 e 2, tem-se 16 linhas correspondentes aos canais de leitura do EEG, para cada registo (ver Figura 6). No caso do Paciente 1, tem-se 14 linhas e no Paciente 2, tem-se 24 linhas.

Assim, foi necessário alterar o código de modo a que seja capaz de incluir todos os canais. Em vez de fazer os cálculos para cada canal isoladamente, a escolha foi agregar todos os canais para cada clip. Na realidade o que acontece é a soma dos vetores correspondentes.

O *MrMotif* é executado mediante a passagem de alguns parâmetros:

- Tamanho do *motif*: corresponde ao tamanho da janela de discretização que contém a secção a ser analisada dentro da série temporal inicial;
- Número de *motifs* gerados: são os top-K *motifs* a detetar entre 4 e 64;
- Tamanho da palavra: número de símbolos utilizados na palavra resultante do *iSAX*;
- *Overlap*: *overlap* das janelas de discretização.

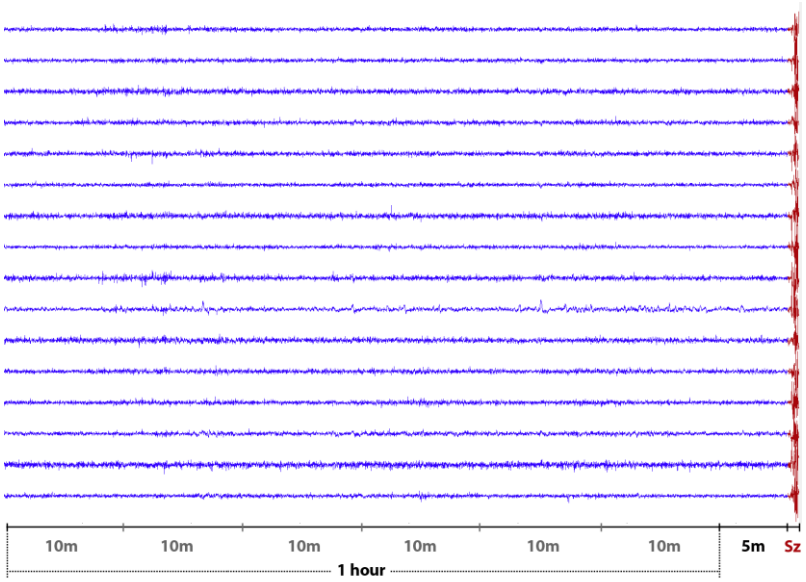


Figura 6 – Exemplo de um registo de EEG com 16 canais (Data | Kaggle, 2016)

Na realização das experiências foi utilizada uma variação no tamanho do *motif* (30, 40 e 60), no número de *motifs* gerados (10, 20 e 30) e também no tamanho do *overlap* da janela (10 e 20). O tamanho da palavra foi mantido sempre a 4, pois foi o que melhor resultado deu nos teste preliminares.

No âmbito deste problema, foi também desenvolvida uma pequena aplicação que permite agrupar cada conjunto de resultados obtidos do *MrMotif*, com o objetivo de criar um conjunto de dados com os resultados para poder ser utilizado no Weka.

### 4.2.4 Reamostragem e execução algoritmos de classificação

Após a extração de atributos descrita na subsecção 4.2.3, poderá efetuar-se um *resampling*, dado estarmos perante um problema de dados desbalanceados (Lee, 2014). Para tal, usamos o *Resample* para as instâncias, do *Weka*.

A literatura refere técnicas para melhorar o desempenho dos algoritmos na presença de dados desbalanceados: redefinição do tamanho do conjunto de dados, onde pode aumentar-se o número de objetos à classe minoritária ou eliminar-se objetos à classe maioritária; utilização de custos de classificação diferentes para as classes; usar técnicas de classificação com apenas uma classe, em que as classes são aprendidas separadamente (Gama et al., 2012).

Posteriormente serão utilizados algoritmos de classificação *Random Forest*, J48, SVM e Regressão Logística através da ferramenta *Weka* para avaliação das medidas AUC, *accuracy* e *f-measure*, bem como a matriz confusão.

### 4.2.5 Validação dos resultados

Para a avaliação dos resultados obtidos para cada modelo de classificação resultante da aplicação de cada um dos algoritmos anteriormente referidos, usamos as medidas descritas na secção 5.1. Dado estarmos perante um caso de classes desbalanceadas, usaremos outras medidas para além da taxa de acerto, assim como mostraremos a matriz de confusão. Para que um resultado seja aceitável, a taxa de acerto terá que ser superior à taxa de acerto de considerar todos os elementos na classe maioritária (Gama et al., 2012).

# 5 Experiências e avaliação

No decorrer das experiências realizadas no âmbito desta dissertação foram selecionados vários algoritmos e medidas de avaliação com o objetivo de perceber quais seriam aqueles que dariam um maior grau de confiança na utilização da metodologia proposta.

Todas as experiências foram realizadas para os *datasets* do Paciente 1, Paciente 2, Cão 1 e Cão 2. Os testes foram realizados com diferentes parâmetros para a extração de atributos e também com e sem a aplicação reamostragem (disponibilizada pelo *Weka*). Nas seguintes secções serão detalhados os resultados obtidos para cada uma das medidas avaliadas bem como para cada um dos algoritmos utilizados.

## 5.1 Medidas de avaliação

Uma forma simples de representação de resultados para problemas com duas classes é a matriz confusão. Habitualmente representa-se uma das classes como sendo positiva (+) e outra como sendo negativa (-). Após a aplicação de determinado algoritmo de classificação a matriz confusão irá apresentar o número de elementos corretamente classificados para a classe positiva (VP), o número de elementos corretamente classificados para a classe negativa (VN), o número de elementos da classe positiva que foram mal classificados (FN) e também o número de elementos da classe negativa que foram mal classificados (FP) (Gama et al., 2012). Nesta dissertação a classe positiva corresponde à classe preictal (P) e a classe negativa corresponde à classe interictal (I).

$$TVP = \frac{VP}{VP + FN} \tag{8}$$

$$TFP = \frac{VN}{VN + FP} \tag{9}$$

$$Taxa\ de\ acerto = \frac{VP + VN}{VN + FP + VP + FN} \tag{10}$$

$$Precisão = \frac{VP}{VP + FP} \tag{11}$$

$$F - Measure = \frac{2\ x\ Precisão\ x\ Sensibilidade}{Precisão +\ Sensibilidade} \tag{12}$$

Na Figura 7 é possível visualizar como é representada a matriz confusão para as classes interictal e preictal.

| | | Classe predita | |
|---|---|---|---|
| | | I | P |
| Classe verdadeira | I | VP | FN |
| | P | FP | VN |

Figura 7 – Matriz confusão paras as classes preictal e interictal

Nesta dissertação a principal medida para efetuar a avaliação dos resultados será a *Area Under the ROC Curve* (AUC), uma vez que tem sido a mais utilizada para a avaliação deste problema (Evaluation | Kaggle, 2016). Por outro lado, esta medida adequa-se a este problema dado que se trata de um problema de classificação binário (duas classes) e por existir desbalanceamento entre as classes (Gama et al., 2012).

O gráfico ROC (*Receiving Operating Characteristics*) é um gráfico bidimensional com eixos *X* e *Y* que representam respetivamente a taxa de falsos positivos (TFP) e a taxa de verdadeiros positivos (TVP) ou sensibilidade.

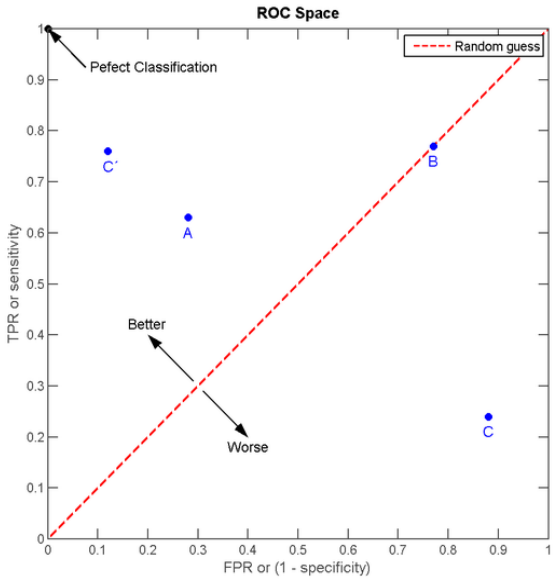


Figura 8 – Exemplo espaço ROC[4]

Na Figura 8 é possível visualizar uma linha diagonal tracejada que representa classificadores que realizam as predições aleatórias. A imagem contém resultados de quatro algoritmos de classificação, cada um deles representado por um ponto (A, B, C e D). O ponto (0,1) representa classificações perfeitas, o ponto (1,1) representa classificações sempre positivas e o ponto (0,0), classificações sempre negativas. Um classificador poderá ser considerado melhor do que outro quando a sua representação no gráfico fica por cima e à esquerda do ponto correspondente do segundo classificador.

A forma mais comum de comparar diferentes algoritmos de classificação é a criação de uma curva ROC. Na Figura 9 é possível visualizar a comparação de 3 algoritmos de classificação. Caso não exista interceção das linhas, a linha que mais se aproxime do ponto (0,1) representa o algoritmo que obtém um melhor desempenho. Caso exista interseção das linhas, significa que cada um dos algoritmos tem um melhor desempenho numa determinada região. É comum comparar o desempenho dos algoritmos através de uma medida única extraída da sua curva de ROC: a área abaixo da curva ROC (AUC).

[4] https://commons.wikimedia.org/wiki/File:ROC_space-2.png

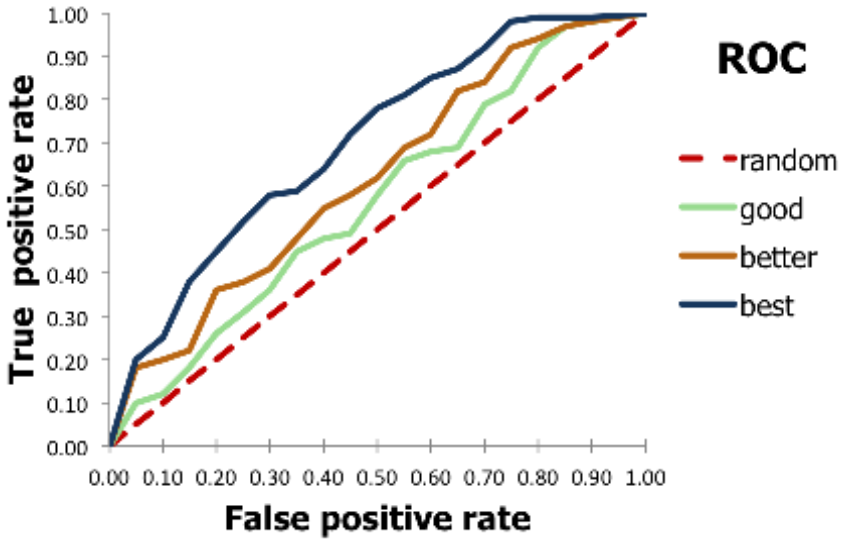


Figura 9 – Exemplo de curvas ROC[5]

A medida AUC produz valores entre 0 e 1. Os valores que estiverem mais próximos de 1 serão aqueles que são considerados melhores, logo, o algoritmo que tenha uma melhor AUC será considerado melhor.

Para além da medida AUC, também iremos analisar a *accuracy*, e a F-*Measure*. A *accuracy* neste caso corresponde à medida *Recall* e é obtida através da média ponderada de cada classe. A *F-Measure* (12) é obtida através da média harmónica ponderada da precisão e da sensibilidade ou TVP, cuja expressão é apresentada na equação (8). (Gama et al., 2012).

Nas experiências a efetuar será utilizada a metodologia de validação cruzada com a utilização de 10 subconjuntos, já abordada na secção 1.4 deste documento.

## 5.2 Hipóteses e metodologias de avaliação

Para decidir qual o melhor modelo de classificação serão comparados vários algoritmos de classificação (*Random Forest*, J48*,* SVM e Regressão Logística) para as diferentes medidas de avaliação e para os quatro casos em estudo (Paciente 1, Paciente 2, Cão 1 e Cão 2). Assim, teremos para hipótese nula a igualdade das medidas (a média ou a mediana, dependendo do

[5] https://docs.eyesopen.com/toolkits/cookbook/python/plotting/roc.html

tipo de teste a realizar) contra a hipótese alternativa em que o modelo apresentado é melhor (melhor medida).

Para os algoritmos de classificação referidos anteriormente, foram definidos os principais parâmetros, que passamos a detalhar:

- J48:
    - confidenceFactor: 0.25;
    - minNumObj: 2;
    - numFolds: 3;
    - *seed*: 1;
    - batchSize: 100;
- Random Forest:
    - *numFeatures*: 0 (automático);
    - numIterations: 100;
    - maxDepth: 0 (ilimitado);
    - *seed*: 1;
    - batchSize: 100;
- SVM:
    - *c*: 1.0;
    - randomSeed: 1;
    - numFolds: -1;
    - *epsilon*: 1.0E-12;
    - batchSize: 100;
- Regressão Logística:
    - maxIts: -1;
    - *ridge*: 1.0E-8;
    - batchSize: 100;

## 5.3 Avaliação Area Under the ROC Curve

Nas seguintes tabelas são apresentados os melhores resultados obtidos para a medida de avaliação AUC, sendo que o valor ideal será o que mais se aproxima de 1. As experiências foram realizadas com diferentes parâmetros do *MrMotif*, tal como se pode visualizar através da coluna “Parâmetros extração atributos”, que estão detalhados na subsecção 4.2.3 deste

documento. Os dados resultantes do *MrMotif* foram executados com diferentes algoritmos de classificação (J48, *Random Forest*, SVM e Regressão Logística) e cada um é representado por uma coluna nas tabelas que se seguem. Estes dados foram obtidos sem o recurso à reamostragem.

Na Tabela 4, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC sem reamostragem para os quatro algoritmos executados.

Tabela 4 – Resultados do Paciente 1 para AUC sem reamostragem

| Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|
| **30 10 10** | 0,650 | **0,847** | 0,639 | 0,737 |
| 30 20 10 | 0,772 | **0,846** | 0,639 | 0,596 |
| 40 10 10 | 0,728 | **0,769** | 0,601 | 0,648 |
| 40 20 10 | 0,572 | **0,756** | 0,601 | 0,551 |
| 40 20 20 | 0,517 | **0,764** | 0,601 | 0,593 |
| 40 30 10 | **0,801** | 0,784 | 0,583 | 0,678 |
| 40 30 20 | 0,466 | **0,736** | 0,546 | 0,550 |
| 60 20 10 | 0,613 | **0,819** | 0,611 | 0,643 |

Como é possível observar na Tabela 4, para este caso o algoritmo de classificação que claramente obteve melhores resultados (0,847) foi o *Random Forest* e os parâmetros ideais para a extração de atributos foram os 30 10 10. De salientar que a diferença para o segundo resultado (0,846) é mínima. Este foi obtido com os parâmetros 30 20 10. O único caso em que o *Random Forest* não foi o algoritmo que obteve melhores resultados foi para os parâmetros 40 30 10 e nesse caso, foi o algoritmo J48 que conseguir obter a melhor classificação.

Na Figura 10 apresentam-se as curvas ROC para a classe preictal (P) correspondentes aos quatro algoritmos para o caso do Paciente 1.

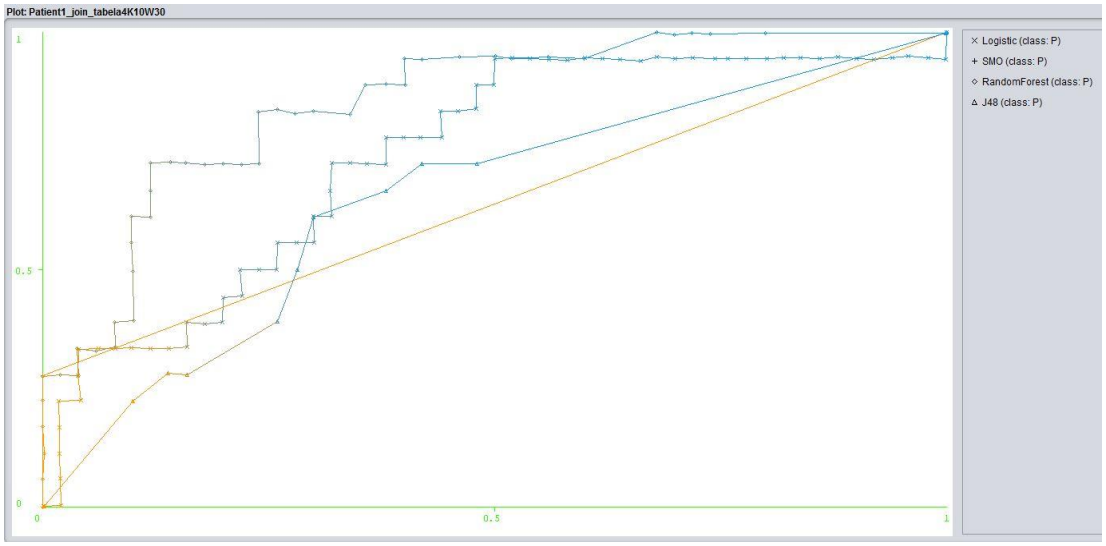


Figura 10 – Gráfico comparativo das curvas ROC para os 4 algoritmos para o Paciente 1

É possível ver na Figura 10 que o algoritmo Random Forest é o algoritmo que obtém uma curva que mais se aproxima do topo do gráfico (do valor 1), logo mais próximo dos valores ideais.

Na Tabela 5, apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC sem reamostragem para os quatro algoritmos executados.

Tabela 5 – Resultados do Paciente 2 para AUC sem reamostragem

| Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|
| 30 10 10 | 0,646 | **0,757** | 0,488 | 0,630 |
| **30 20 10** | 0,646 | **0,789** | 0,516 | 0,572 |
| 40 10 10 | 0,479 | **0,719** | 0,476 | 0,389 |
| 40 20 10 | 0,569 | **0,667** | 0,492 | 0,344 |
| 40 20 20 | 0,501 | **0,620** | 0,476 | 0,352 |
| 40 30 10 | 0,553 | **0,673** | 0,476 | 0,473 |
| 40 30 20 | 0,629 | **0,651** | 0,504 | 0,457 |
| 60 20 10 | 0,414 | **0,505** | 0,476 | 0,468 |

A Tabela 5 refere-se ao Paciente 2, o algoritmo de classificação *Random Forest* obteve os melhores resultados (0,789) em todos os casos e os parâmetros ideais para a extração de atributos foram os 30 20 10.

Na Figura 11 apresentam-se as curvas ROC para a classe preictal (P) correspondentes aos quatro algoritmos para o caso do Paciente 2.

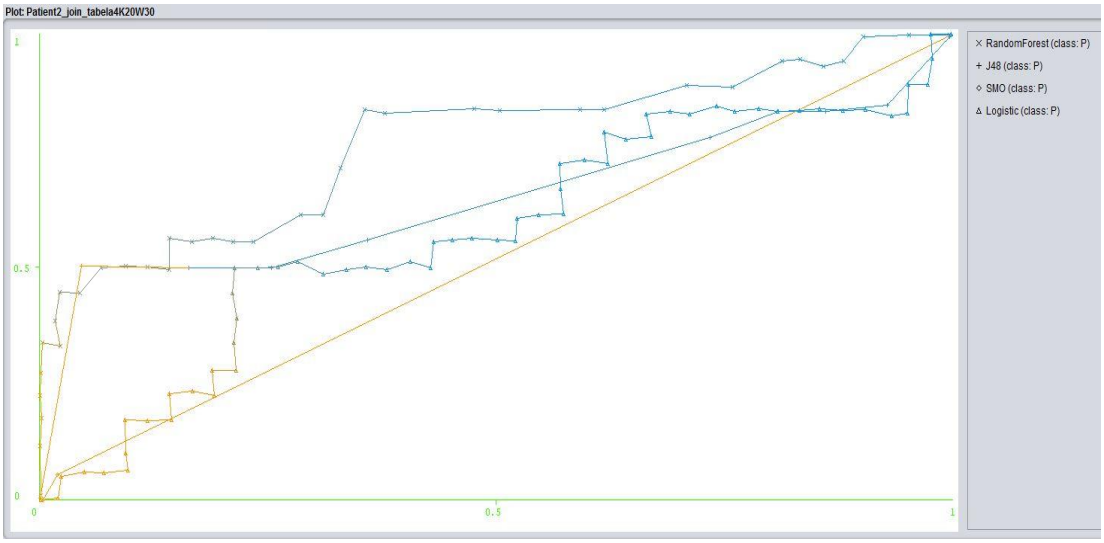


Figura 11 - Gráfico comparativo das curvas ROC para os 4 algoritmos para o Paciente 2

É possível ver na Figura 11 que o algoritmo *Random Forest* é o algoritmo que obtém uma curva que mais se aproxima do topo do gráfico durante mais tempo, logo mais próximo dos valores ideais.

Na Tabela 6 apresentam-se os resultados obtidos do Cão 1 analisando a medida AUC sem reamostragem para os quatro algoritmos executados.

Tabela 6 – Resultados do Cão 1 para AUC sem reamostragem

| Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|
| 30 10 10 | 0,450 | 0,572 | 0,500 | **0,613** |
| 30 20 10 | 0,450 | **0,613** | 0,500 | 0,653 |
| 40 10 10 | 0,455 | **0,535** | 0,500 | 0,373 |
| **40 20 10** | 0,531 | **0,631** | 0,500 | 0,396 |
| 40 20 20 | 0,461 | 0,495 | 0,500 | **0,507** |
| 40 30 10 | 0,452 | **0,555** | 0,500 | 0,270 |
| 40 30 20 | 0,465 | **0,522** | 0,500 | 0,445 |
| 60 20 10 | 0,459 | **0,566** | 0,500 | 0,512 |

Para o caso apresentado na Tabela 6, o algoritmo de classificação que obteve melhores resultados (0,631) foi o *Random Forest*, mas a variação já é maior do que a visualizada nos pacientes 1 e 2. O algoritmo SVM apresenta sempre os mesmos resultados, o que nos leva a

concluir que não é um algoritmo eficaz para aplicar a este *dataset*. Pelo que podemos observar, os parâmetros ideais para a extração de atributos foram os 40 20 10. De referir também, que, para dois parâmetros distintos (30 10 10 e 40 20 20), foi o algoritmo Regressão Logística que apresentou os melhores resultados (0,613 e 0,507 respetivamente).

Na Figura 12 apresentam-se as curvas ROC para a classe preictal (P) correspondentes aos quatro algoritmos para o caso do Cão 1.

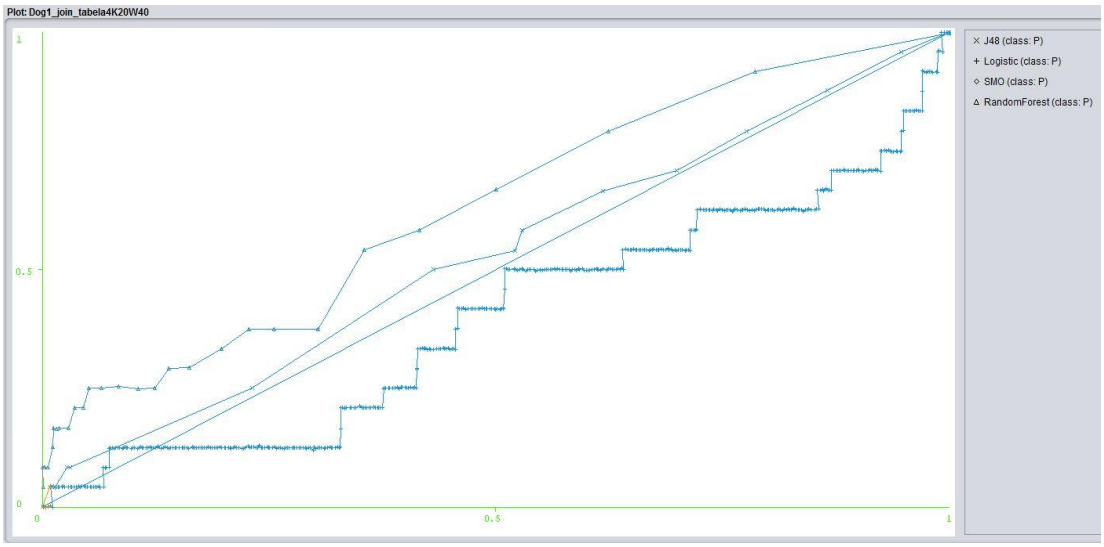


Figura 12 - Gráfico comparativo das curvas ROC para os 4 algoritmos para o Cão 1

É possível ver na Figura 12 que o algoritmo *Random Forest* é o que obtém uma curva que mais se aproxima do topo do gráfico, no entanto parece ficar longe dos valores ideais.

Na Tabela 7, apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC sem reamostragem para os quatro algoritmos executados.

Tabela 7 – Resultados do Cão 2 para AUC sem reamostragem

| Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|
| 30 10 10 | 0,530 | **0,673** | 0,500 | 0,416 |
| 30 20 10 | 0,511 | **0,529** | 0,500 | 0,386 |
| 40 10 10 | 0,532 | **0,703** | 0,500 | 0,311 |
| 40 20 10 | 0,549 | **0,659** | 0,500 | 0,236 |
| **40 20 20** | **0,565** | 0,562 | 0,500 | 0,287 |
| 40 30 10 | 0,491 | **0,654** | 0,500 | 0,171 |
| 40 30 20 | 0,563 | **0,588** | 0,500 | 0,207 |
| 60 20 10 | 0,523 | **0,669** | 0,500 | 0,509 |

É possível visualizar na Tabela 7, que o *Random Forest* continua a ser o algoritmo de classificação que mais se destaca pela positiva e no sentido inverso está o SVM que aqui também apresenta sempre os mesmos resultados, levando a crer que não é o algoritmo ideal para este caso. Os parâmetros ideais para a extração de atributos foram os 40 20 20 e a classificação máxima foi de 0,703.

Na Figura 13 apresentam-se as curvas ROC para a classe preictal (P) correspondentes aos quatro algoritmos para o caso do Cão 2.

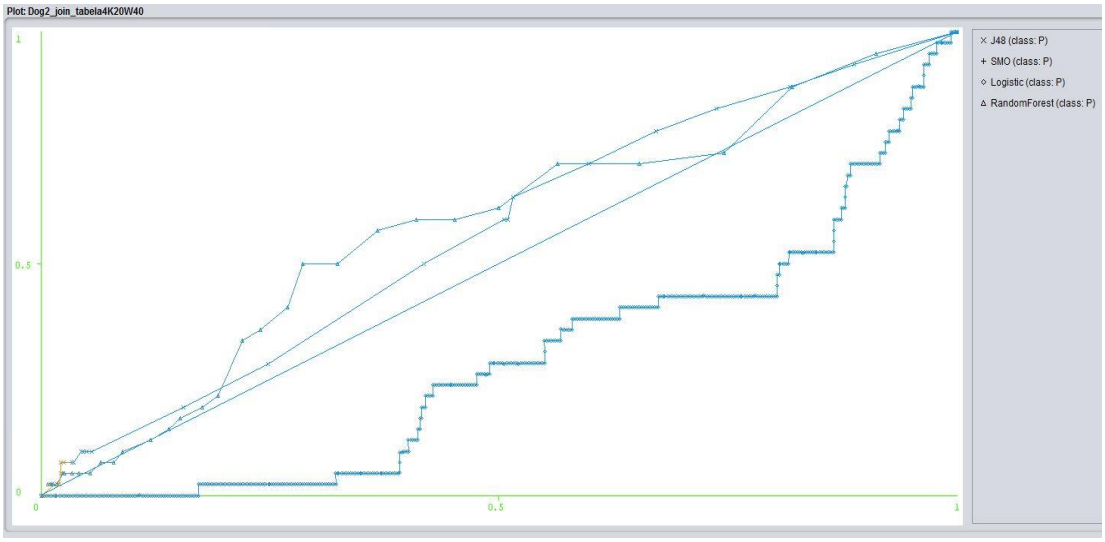


Figura 13 - Gráfico comparativo das curvas ROC para os 4 algoritmos para o Cão 2

É possível ver no gráfico da Figura 13 que o algoritmo *Random Forest* é o que obtém uma curva que mais se aproxima do topo do gráfico, no entanto, tal como na Figura 12, parece ficar longe dos valores ideais.

Na Figura 14 é possível visualizar um gráfico que apresenta um resumo dos melhores resultados de cada um dos algoritmos testados (J48, *Random Forest*, SVM e Regressão Logística) para a medida AUC sem recurso à reamostragem para o Paciente 1, Paciente 2, Cão 1 e Cão 2.

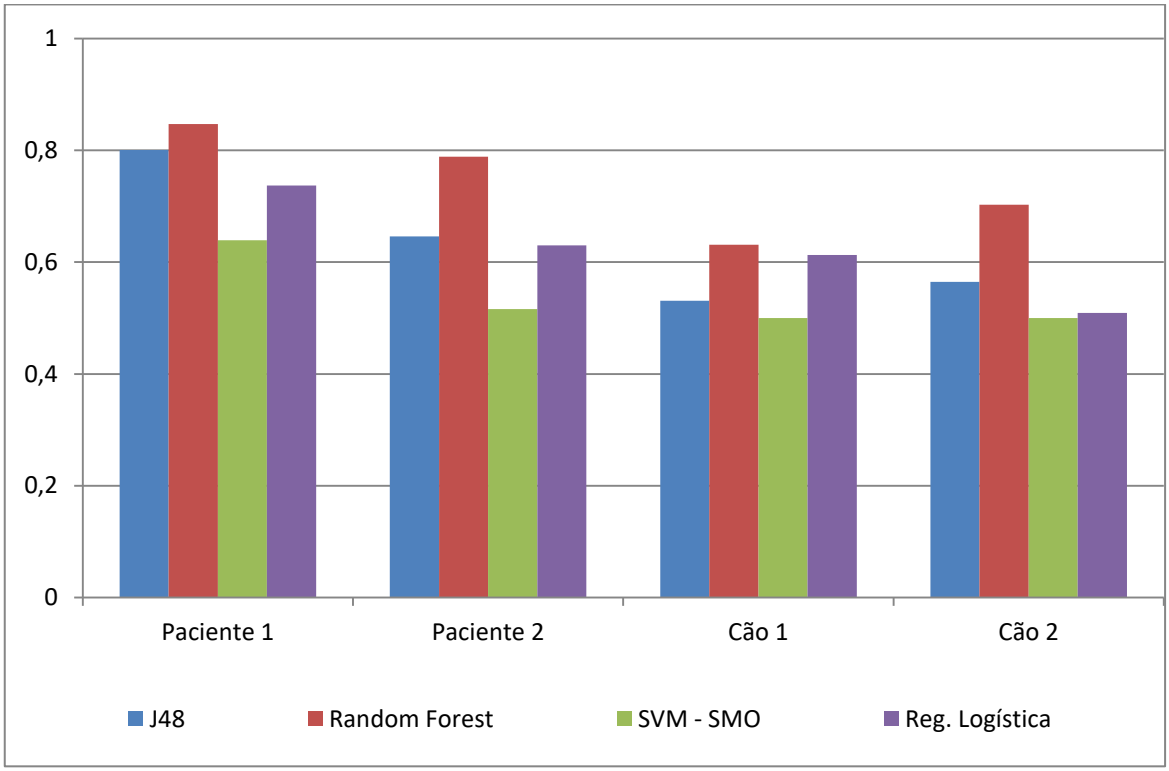


Figura 14 – Gráfico com os melhores resultados para AUC sem reamostragem

Na Figura 14 é possível verificar que em todos os casos, o algoritmo que obteve as melhores classificações foi o *Random Forest*. É possível também identificar que o SVM foi o que teve os resultados menos satisfatórios para todos os Pacientes e Cães.

## 5.4 Avaliação Area Under the ROC Curve com reamostragem

Nas seguintes tabelas estão detalhados os resultados obtidos para a medida de avaliação AUC. Todas as medidas foram obtidas após a aplicação da técnica de reamostragem (cinco aplicações) disponível no *Weka*. O número correspondente à reamostragem poderá ser visualizado na coluna "Reamostragem". As restantes colunas são idênticas às apresentadas na subsecção 5.3.

### 5.4.1 Paciente 1

Todas as medidas apresentadas nesta subsecção são referentes ao Paciente 1 para diferentes parâmetros de extração de atributos e várias aplicações da reamostragem.

Na Tabela 8, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 10 10.

Tabela 8 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 30 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 10 10 | 0,929 | 0,967 | 0,667 | 0,701 |
| 2 | 30 10 10 | 0,877 | **0,978** | 0,667 | 0,950 |
| 3 | 30 10 10 | 0,866 | 0,964 | 0,722 | 0,942 |
| 4 | 30 10 10 | 0,939 | 0,966 | 0,722 | 0,980 |
| 5 | 30 10 10 | 0,981 | 0,996 | 0,750 | 0,949 |

Para o caso apresentado na Tabela 8, o algoritmo de classificação que obteve melhores resultados (0,978) foi o *Random Forest* e foi alcançado com a 2ª aplicação da reamostragem. De referir também que tanto o algoritmo J48 (0,981), como o Regressão Logística (0,980) obtiveram resultados bastante próximos do ideal. Na Tabela 8 também é possível visualizar que o algoritmo *Random Forest* é o algoritmo que reage mais rapidamente à reamostragem, ou seja, logo na 1ª aplicação obtém resultados acima dos 0,95 mantendo-se assim para todas as aplicações.

Na Tabela 9, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 20 10.

Tabela 9 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 30 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 20 10 | 0,881 | 0,975 | 0,667 | 0,775 |
| 2 | 30 20 10 | 0,886 | 0,972 | 0,667 | 0,907 |
| 3 | 30 20 10 | 0,787 | 0,917 | 0,722 | 0,907 |
| 4 | 30 20 10 | 0,939 | 0,986 | 0,722 | 0,950 |
| 5 | 30 20 10 | 0,964 | **0,997** | 0,750 | 0,935 |

Para os parâmetros 30 20 10 apresentados na Tabela 9, o algoritmo de classificação que obteve melhores resultados (0,997) foi o *Random Forest* e foi alcançado com a 5ª aplicação da reamostragem. Tanto aqui, como na Tabela 8, também se verificam valores acima dos 0,95 para os algoritmos J48 e Regressão Logística.

Na Tabela 10, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 10 10.

Tabela 10 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 10 10 | 0,708 | 0,917 | 0,657 | 0,662 |
| 2 | 40 10 10 | 0,780 | 0,889 | 0,667 | 0,814 |
| 3 | 40 10 10 | 0,844 | 0,867 | 0,722 | 0,771 |
| 4 | 40 10 10 | 0,919 | 0,954 | 0,722 | 0,958 |
| 5 | 40 10 10 | 0,990 | **0,998** | 0,750 | 0,985 |

Tal como nas duas tabelas anteriores, para os valores da Tabela 10 também foi o algoritmo *Random Forest* que obteve melhores resultados (0,998), este foi alcançado na 5ª aplicação da reamostragem. Também é possível visualizar valores interessantes para os algoritmos J48 (0,990) e Regressão Logística (0,985), todos na 5ª aplicação da reamostragem.

Na Tabela 11, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 10.

Tabela 11 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 10 | 0,871 | 0,946 | 0,657 | 0,749 |
| 2 | 40 20 10 | 0,867 | 0,934 | 0,667 | 0,931 |
| 3 | 40 20 10 | 0,979 | 0,902 | 0,722 | 0,825 |
| 4 | 40 20 10 | 0,912 | 0,953 | 0,722 | 0,902 |
| 5 | 40 20 10 | 0,983 | **0,996** | 0,750 | 0,934 |

Para os valores da Tabela 11 também foi o algoritmo *Random Forest* que obteve melhores resultados (0,996), este foi alcançado na 5ª aplicação da reamostragem.

Na Tabela 12Tabela 13, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 20.

Tabela 12 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 20 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 20 | 0,845 | 0,961 | 0,657 | 0,705 |
| 2 | 40 20 20 | 0,907 | 0,978 | 0,667 | 0,909 |
| 3 | 40 20 20 | 0,741 | 0,904 | 0,694 | 0,878 |
| 4 | 40 20 20 | 0,903 | 0,958 | 0,722 | 0,910 |
| 5 | 40 20 20 | 0,962 | 0,995 | 0,722 | **0,997** |

Para os parâmetros apresentados na Tabela 12, foi o algoritmo Regressão Logística que obteve o melhor resultado (0,997), este foi alcançado na 5ª aplicação da reamostragem. De salientar que o *Random Forest* ficou muito próximo (0,955), mas mostra valores muito mais consistentes do que os restantes algoritmos.

Na Tabela 13, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 10.

Tabela 13 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 30 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 10 | 0,920 | 0,933 | 0,629 | 0,762 |
| 2 | 40 30 10 | 0,869 | 0,948 | 0,667 | 0,840 |
| 3 | 40 30 10 | 0,941 | 0,888 | 0,702 | 0,875 |
| 4 | 40 30 10 | 0,938 | 0,958 | 0,712 | 0,938 |
| 5 | 40 30 10 | 0,981 | **0,986** | 0,740 | 0,984 |

A Tabela 13 mostra que o algoritmo *Random Forest* volta a obter a melhor classificação (0,986) e que esta também foi obtida na 5ª aplicação da reamostragem. Apenas o algoritmo SVM fica mais distante dos valores ideais.

Na Tabela 14, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 20.

Tabela 14 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 40 30 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 20 | 0,859 | 0,952 | 0,629 | 0,739 |
| 2 | 40 30 20 | 0,907 | 0,986 | 0,667 | 0,960 |
| 3 | 40 30 20 | 0,823 | 0,902 | 0,674 | 0,876 |
| 4 | 40 30 20 | 0,949 | 0,963 | 0,712 | 0,883 |
| 5 | 40 30 20 | 0,973 | **0,993** | 0,796 | 0,932 |

É possível verificar que para os parâmetros da Tabela 14, o algoritmo *Random Forest* foi o que obteve a melhor classificação (0,993) e que esta também foi obtida na 5ª aplicação da reamostragem. Tal como na tabela anterior, apenas o algoritmo SVM fica mais distante dos valores ideais.

Na Tabela 15, apresentam-se os resultados obtidos do Paciente 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 60 20 10.

Tabela 15 – Resultados do Paciente 1 para AUC com reamostragem (parâmetros 60 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 60 20 10 | 0,863 | 0,961 | 0,667 | 0,838 |
| 2 | 60 20 10 | 0,861 | 0,946 | 0,667 | 0,949 |
| 3 | 60 20 10 | 0,902 | 0,906 | 0,722 | 0,964 |
| 4 | 60 20 10 | 0,907 | 0,963 | 0,722 | 0,992 |
| 5 | 60 20 10 | 0,978 | **1,000** | 0,750 | 0,982 |

Finalmente, para os dados da Tabela 15 é possível verificar que o algoritmo *Random Forest* não só é o que obtém o melhor resultado, mas também é o responsável por atingir o valor máximo de classificação, ou seja, AUC igual a 1. Com estes parâmetros os resultados foram de uma forma geral bastante satisfatórios, ficando apenas aquém do esperado no caso do SVM.

Na Figura 15 é possível visualizar um gráfico que apresenta um resumo dos melhores resultados de cada um dos algoritmos testados (J48, *Random Forest*, SVM e Regressão Logística) para a medida AUC com recurso às várias reamostragens para o Paciente 1.

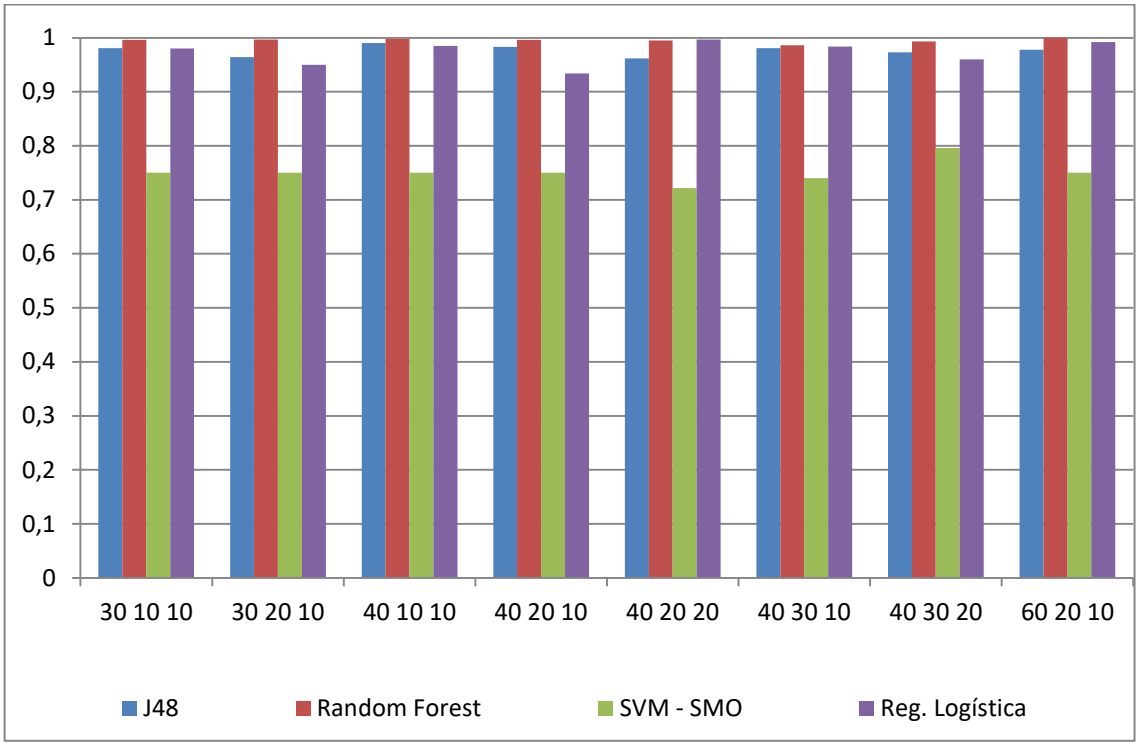


Figura 15 – Gráfico com os melhores resultados para AUC com reamostragem (Paciente 1)

É possível verificar que independentemente dos parâmetros utilizados para a extração dos atributos, no Paciente 1, o algoritmo que obteve as melhores classificações foi o *Random Forest*, mas a diferença para o J48 e para o Regressão Logística não é muito evidente. É possível também identificar que o SVM está muito longe dos resultados obtidos pelos restantes algoritmos.

Na Figura 16 apresentam-se as curvas ROC para a classe preictal (P), correspondentes aos quatro algoritmos utilizados para o caso do Paciente 1. Os parâmetros utilizados para a visualização das curvas ROC na Figura 16 foram 60 20 10 com cinco aplicações da reamostragem. Estes foram os parâmetros que levaram à obtenção do melhor resultado para o Paciente 1 (*Random Forest* 1,000).

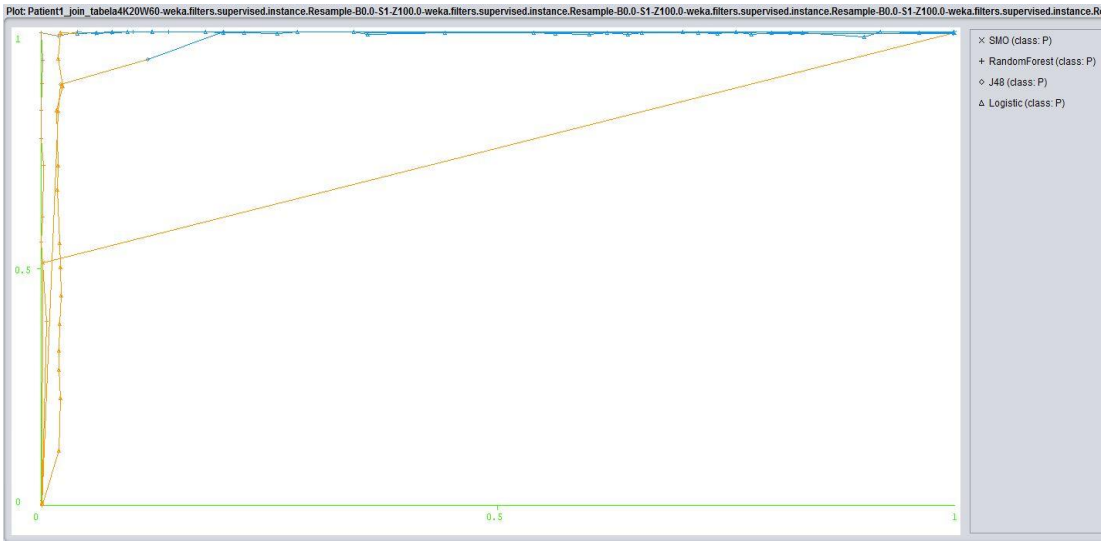


Figura 16 – Curvas de ROC para os 4 algoritmos para o Paciente 1 com reamostragem

No gráfico da Figura 16 é possível visualizar que para a classe preictal, o algoritmo *Random Forest* é aquele que apresenta uma melhor curva de ROC, pois a sua trajetória é na maioria dos casos superior às restantes linhas. Também é claro que os resultados para o SVM ficam muito longe dos restantes algoritmos.

### 5.4.2 Paciente 2

Todas as medidas apresentadas nesta subsecção são referentes ao Paciente 2 para diferentes parâmetros de extração de atributos e várias aplicações da reamostragem.

Na Tabela 16 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 10 10.

Tabela 16 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 30 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 10 10 | 0,936 | 0,941 | 0,500 | 0,759 |
| 2 | 30 10 10 | 0,948 | **0,996** | 0,500 | 0,859 |
| 3 | 30 10 10 | 0,907 | 0,950 | 0,500 | 0,860 |
| 4 | 30 10 10 | 0,911 | 0,988 | 0,500 | 0,871 |
| 5 | 30 10 10 | 0,953 | 0,962 | 0,500 | 0,925 |

Para o caso apresentado na Tabela 16, o algoritmo de classificação que obteve melhores resultados (0,996) foi o *Random Forest* e foi alcançado com a 2ª aplicação da reamostragem. De referir também que tanto o algoritmo J48, como o Regressão Logística obtiveram resultados bastante acima dos 0,9. Também é possível visualizar que os algoritmos *Random Forest* e J48 são os algoritmos que reagem mais rapidamente à reamostragem, ou seja, logo na 1ª aplicação obtêm resultados acima dos 0,9 mantendo-se assim para todas as aplicações.
Na Tabela 17 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 20 10.

Tabela 17 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 30 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 20 10 | 0,888 | 0,929 | 0,500 | 0,773 |
| 2 | 30 20 10 | 0,919 | **0,997** | 0,500 | 0,927 |
| 3 | 30 20 10 | 0,919 | 0,970 | 0,500 | 0,917 |
| 4 | 30 20 10 | 0,954 | 0,990 | 0,500 | 0,917 |
| 5 | 30 20 10 | 0,953 | 0,987 | 0,500 | 0,957 |

É possível observar na Tabela 17 que o *Random Forest* foi o algoritmo de classificação que obteve melhores resultados (0,997). O melhor resultado foi obtido logo na 2ª aplicação da reamostragem. Tanto aqui, como na tabela anterior também se verificam valores acima dos

0,9 para os algoritmos J48 e Regressão Logística. O algoritmo SVM manteve-se inalterado mesmo após a 5ª reamostragem o que significa que não será o ideal.

Na Tabela 17 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 10 10.

Tabela 18 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 10 10 | 0,684 | 0,942 | 0,500 | 0,671 |
| 2 | 40 10 10 | 0,956 | 0,981 | 0,500 | 0,731 |
| 3 | 40 10 10 | 0,897 | 0,978 | 0,500 | 0,853 |
| 4 | 40 10 10 | 0,937 | **0,992** | 0,500 | 0,892 |
| 5 | 40 10 10 | 0,899 | 0,964 | 0,500 | 0,956 |

É possível observar na Tabela 18 que o algoritmo *Random Forest* voltou a obter o melhor resultado (0,992), este foi alcançado na 4ª aplicação da reamostragem. Também é possível visualizar valores interessantes para os algoritmos J48 e Regressão Logística. O SVM continua a apresentar sempre o mesmo valor.

Na Tabela 19 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 10.

Tabela 19 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 10 | 0,744 | 0,949 | 0,500 | 0,711 |
| 2 | 40 20 10 | 0,899 | 0,979 | 0,500 | 0,888 |
| 3 | 40 20 10 | 0,854 | 0,974 | 0,500 | 0,929 |
| 4 | 40 20 10 | 0,898 | 0,987 | 0,500 | 0,888 |
| 5 | 40 20 10 | 0,937 | 0,963 | 0,500 | **1,000** |

Em relação aos valores da Tabela 19 é possível constatar que o algoritmo Regressão Logística alcançou o máximo, ou seja, 1,000. O *Random Forest* manteve os bons resultados e o J48 também se manteve dentro do que tem apresentado. O SVM continua a apresentar sempre o mesmo valor.

Na Tabela 20 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 20.

Tabela 20 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 20 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 20 | 0,706 | 0,935 | 0,500 | 0,721 |
| 2 | 40 20 20 | 0,882 | 0,981 | 0,500 | 0,910 |
| 3 | 40 20 20 | 0,836 | 0,968 | 0,500 | 0,917 |
| 4 | 40 20 20 | 0,907 | 0,987 | 0,500 | 0,917 |
| 5 | 40 20 20 | 0,929 | 0,982 | 0,500 | **1,000** |

Para os parâmetros apresentados na Tabela 20, foi o algoritmo Regressão Logística que obteve o melhor resultado (1,000), este foi alcançado na 5ª aplicação da reamostragem e atingiu novamente o valor máximo. De salientar que o *Random Forest* ficou muito próximo.
Na Tabela 21 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 10.

Tabela 21 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 30 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 10 | 0,899 | 0,968 | 0,500 | 0,792 |
| 2 | 40 30 10 | 0,899 | 0,975 | 0,500 | 0,903 |
| 3 | 40 30 10 | 0,886 | 0,974 | 0,500 | 0,942 |
| 4 | 40 30 10 | 0,892 | **0,979** | 0,500 | 0,887 |
| 5 | 40 30 10 | 0,892 | **0,979** | 0,500 | 0,892 |

A Tabela 21 mostra que o algoritmo *Random Forest* volta a obter a melhor classificação (0,979 em duas reamostragens). Apenas o algoritmo SVM fica mais distante dos valores ideais.
Na Tabela 22 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 20.

Tabela 22 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 40 30 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 20 | 0,810 | 0,966 | 0,500 | 0,835 |
| 2 | 40 30 20 | 0,923 | 0,979 | 0,500 | 0,909 |
| 3 | 40 30 20 | 0,827 | 0,969 | 0,500 | 0,930 |
| 4 | 40 30 20 | 0,829 | 0,979 | 0,500 | 0,937 |
| 5 | 40 30 20 | 0,929 | 0,987 | 0,500 | **1,000** |

É possível verificar que para os parâmetros da Tabela 22, o algoritmo Regressão Logística volta a obter o melhor resultado, que foi obtida na 5ª aplicação da reamostragem e também atingiu o valor máximo (1,000). Tal como na tabela anterior, apenas o algoritmo SVM fica mais distante dos valores ideais.

Na Tabela 23Tabela 22 apresentam-se os resultados obtidos do Paciente 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 60 20 10.

Tabela 23 – Resultados do Paciente 2 para AUC com reamostragem (parâmetros 60 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 60 20 10 | 0,723 | 0,933 | 0,500 | 0,762 |
| 2 | 60 20 10 | 0,878 | 0,948 | 0,500 | 0,887 |
| 3 | 60 20 10 | 0,894 | 0,970 | 0,500 | 0,977 |
| 4 | 60 20 10 | 0,953 | 0,973 | 0,500 | 0,936 |
| 5 | 60 20 10 | 0,875 | **0,995** | 0,500 | 0,983 |

Finalmente, para os dados da Tabela 23 é possível verificar que o algoritmo *Random Forest* obtém o melhor resultado (0,995). O algoritmo Regressão Logística volta a ter um bom desempenho (0,983), seguido pelo J48 (0,953). Como habitualmente o SVM manteve-se estático e com o resultado bastante inferior aos restantes.

A Figura 17 apresenta um gráfico com os melhores resultados para a medida AUC no Paciente 2, para cada um dos algoritmos testados com reamostragem.

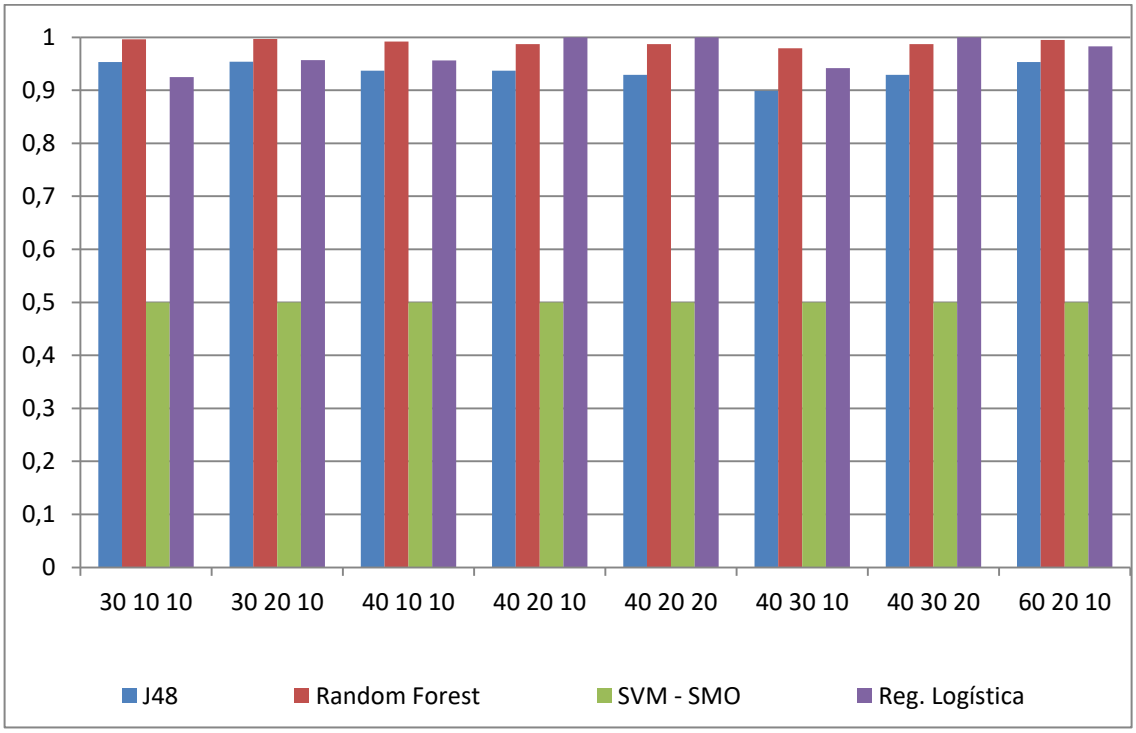


Figura 17 – Gráfico com os melhores resultados para AUC com reamostragem (Paciente 2)

A Figura 17 mostra que no Paciente 2, independentemente dos parâmetros colocados para a extração dos atributos, o algoritmo que obteve as piores classificações foi o SVM. As melhores classificações foram repartidas entre o *Random Forest* e a Regressão Logística, com a diferença em que o primeiro é mais consistente para os diferentes parâmetros utilizados, enquanto o segundo obtém três classificações máximas para os parâmetros 40 20 10, 40 20 20 e 40 30 20.

Na Figura 18 apresentam-se as curvas ROC para a classe preictal (P), correspondentes aos quatro algoritmos utilizados para o caso do Paciente 2. Os parâmetros utilizados para a visualização das curvas ROC na Figura 18 foram 40 20 20 com cinco aplicações da reamostragem. Estes, foram uns dos parâmetros que levaram à obtenção do melhor resultado para o Paciente 2 (Regressão Logística 1,000).

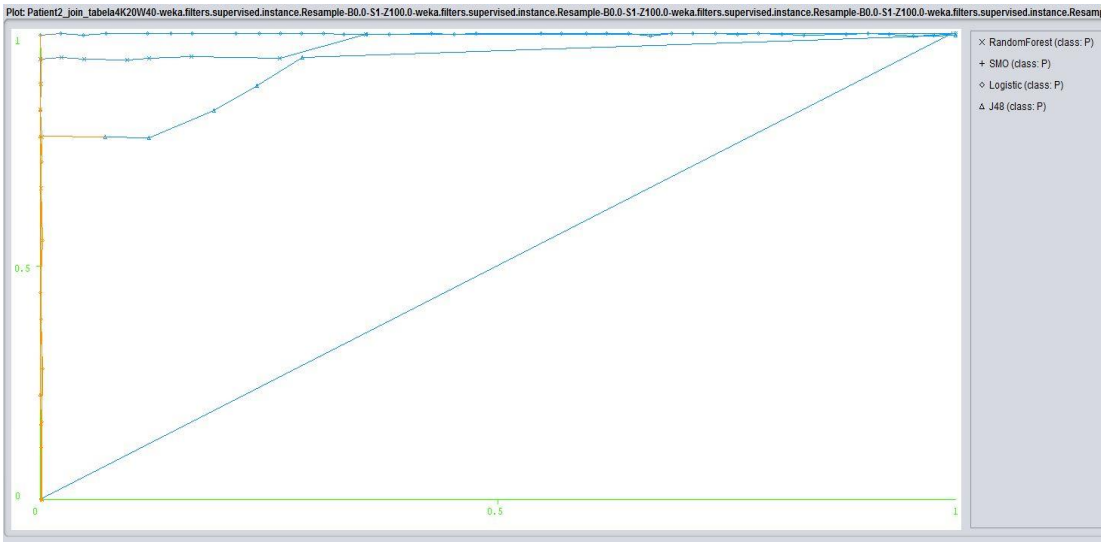


Figura 18 – Curvas ROC para os 4 algoritmos para o Paciente 2 com reamostragem

No gráfico da Figura 18 é possível visualizar que para a classe preictal, o algoritmo Regressão Logística é aquele que apresenta uma melhor curva ROC, pois a sua trajetória é na maioria dos casos superior às restantes linhas. Também é possível verificar que o algoritmo SVM apresenta resultados bastante inferiores aos restantes algoritmos, portanto muito longe dos valores ideais.

### 5.4.3 Cão 1

Todas as medidas apresentadas nesta subsecção são referentes ao Cão 1 para diferentes parâmetros de extração de atributos e várias aplicações da reamostragem.

Na Tabela 24 apresentam-se os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 10 10.

Tabela 24 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 30 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 10 10 | 0,688 | 0,935 | 0,500 | 0,825 |
| 2 | 30 10 10 | 0,707 | 0,952 | 0,500 | 0,734 |
| 3 | 30 10 10 | 0,787 | 0,958 | 0,500 | 0,781 |
| 4 | 30 10 10 | 0,822 | **0,992** | 0,500 | 0,750 |
| 5 | 30 10 10 | 0,866 | 0,950 | 0,500 | 0,727 |

Na Tabela 24 é possível verificar que o algoritmo de classificação que conseguiu melhores resultados (0,992) e de uma forma destacada foi o *Random Forest* com a 4ª aplicação da reamostragem. Também é importante salientar o algoritmo *Random Forest* é o que reage mais rapidamente à reamostragem, ou seja, logo na 1ª aplicação obtém resultados acima dos 0,9 mantendo-se assim para todas as aplicações.

Na Tabela 25 apresentam-se os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 20 10.

Tabela 25 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 30 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 20 10 | 0,731 | 0,882 | 0,500 | 0,793 |
| 2 | 30 20 10 | 0,651 | 0,914 | 0,500 | 0,665 |
| 3 | 30 20 10 | 0,846 | 0,951 | 0,500 | 0,803 |
| 4 | 30 20 10 | 0,817 | **0,994** | 0,500 | 0,822 |
| 5 | 30 20 10 | 0,861 | 0,917 | 0,500 | 0,740 |

É possível observar na Tabela 25 que o *Random Forest* foi o algoritmo de classificação que obteve melhores resultados (0,994). O melhor resultado foi obtido novamente na 4ª aplicação da reamostragem. Tanto aqui, como na tabela anterior também se verifica que os restantes algoritmos ficam longe do *Random Forest*. O algoritmo SVM manteve-se inalterado mesmo após a 5ª reamostragem o que significa que não deverá ser considerado.

Na Tabela 26 apresentam-se os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 10 10.

Tabela 26 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 10 10 | 0,649 | 0,903 | 0,500 | 0,642 |
| 2 | 40 10 10 | 0,567 | 0,902 | 0,500 | 0,588 |
| 3 | 40 10 10 | 0,779 | 0,934 | 0,500 | 0,532 |
| 4 | 40 10 10 | 0,936 | **0,996** | 0,500 | 0,427 |
| 5 | 40 10 10 | 0,852 | 0,903 | 0,500 | 0,396 |

Tal como nas duas tabelas anteriores, para os valores da Tabela 26 também foi o algoritmo *Random Forest* que obteve o melhor resultado (0,996), e também na 4ª aplicação da reamostragem. Apenas o algoritmo J48 se aproxima dos valores do *Random Forest*.

Na Tabela 27 apresentam-se os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 10.

Tabela 27 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 10 | 0,800 | 0,937 | 0,500 | 0,507 |
| 2 | 40 20 10 | 0,759 | 0,937 | 0,500 | 0,600 |
| 3 | 40 20 10 | 0,916 | 0,946 | 0,500 | 0,694 |
| 4 | 40 20 10 | 0,825 | **0,980** | 0,500 | 0,580 |
| 5 | 40 20 10 | 0,900 | 0,961 | 0,500 | 0,634 |

Nos valores da Tabela 27 é possível constatar que o algoritmo *Random Forest* continua a obter o melhor resultado (0,980) e o J48 já se aproximou mais consistentemente dos valores obtidos pelo *Random Forest*. Os restantes continuam a ficar muito longe e sem possibilidades de serem considerados úteis.

Na Tabela 28 apresentam-se os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 20.

Tabela 28 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 20 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 20 | 0,592 | 0,911 | 0,500 | 0,685 |
| 2 | 40 20 20 | 0,690 | 0,937 | 0,500 | 0,658 |
| 3 | 40 20 20 | 0,668 | 0,917 | 0,500 | 0,776 |
| 4 | 40 20 20 | 0,826 | **0,993** | 0,500 | 0,770 |
| 5 | 40 20 20 | 0,865 | 0,946 | 0,500 | 0,837 |

O algoritmo *Random Forest* continua a ser o que obtém o melhor resultado (0,993) e tal como nas tabelas anteriores, também na 4ª aplicação da reamostragem.

Na Tabela 29 apresentam-se os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 10.

Tabela 29 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 30 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 10 | 0,681 | 0,904 | 0,500 | 0,686 |
| 2 | 40 30 10 | 0,802 | 0,938 | 0,500 | 0,619 |
| 3 | 40 30 10 | 0,884 | 0,956 | 0,500 | 0,719 |
| 4 | 40 30 10 | 0,927 | **0,996** | 0,500 | 0,761 |
| 5 | 40 30 10 | 0,853 | 0,948 | 0,500 | 0,813 |

A Tabela 29 mostra que o algoritmo *Random Forest* volta a obter a melhor classificação (0,996). Apenas o algoritmo SVM fica mais distante dos valores ideais e se mantém inalterado.

A Tabela 30 mostra os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 20.

Tabela 30 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 40 30 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 20 | 0,592 | 0,879 | 0,500 | 0,720 |
| 2 | 40 30 20 | 0,718 | 0,937 | 0,500 | 0,735 |
| 3 | 40 30 20 | 0,669 | 0,925 | 0,500 | 0,861 |
| 4 | 40 30 20 | 0,821 | **0,986** | 0,500 | 0,901 |
| 5 | 40 30 20 | 0,859 | 0,941 | 0,500 | 0,898 |

A Tabela 30 mostra que o algoritmo *Random Forest* volta a obter o melhor resultado (0,986) e novamente na 4ª aplicação da reamostragem. Tal como na tabela anterior, apenas o algoritmo SVM fica mais distante dos valores ideais.

A Tabela 30Tabela 31 mostra os resultados obtidos do Cão 1 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 60 20 10.

Tabela 31 – Resultados do Cão 1 para AUC com reamostragem (parâmetros 60 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 60 20 10 | 0,604 | 0,914 | 0,500 | 0,670 |
| 2 | 60 20 10 | 0,829 | 0,973 | 0,500 | 0,805 |
| 3 | 60 20 10 | 0,845 | 0,960 | 0,500 | 0,805 |
| 4 | 60 20 10 | 0,940 | 0,948 | 0,500 | 0,904 |
| 5 | 60 20 10 | 0,890 | **0,992** | 0,500 | 0,891 |

Finalmente, nos dados da Tabela 31 é possível verificar que, como habitualmente, o algoritmo *Random Forest* obtém o melhor resultado (0,992). O algoritmo J48 volta a ter um bom desempenho (0,940), seguido pelo Regressão Logística (0,904). Como habitualmente o SVM manteve-se estático e com o resultado bastante inferior aos restantes.

Apesar de ser na 5ª reamostragem que o melhor resultado é alcançado (*Random Forest* 0,992), a 4ª reamostragem é a que permite os melhores resultados para os algoritmos J48 e Regressão Logística.

Na Figura 15 é possível visualizar um gráfico que apresenta um resumo dos melhores resultados para o Cão 1 de cada um dos algoritmos testados (J48, *Random Forest*, SVM e Regressão Logística) para a medida AUC com recurso às várias reamostragens.

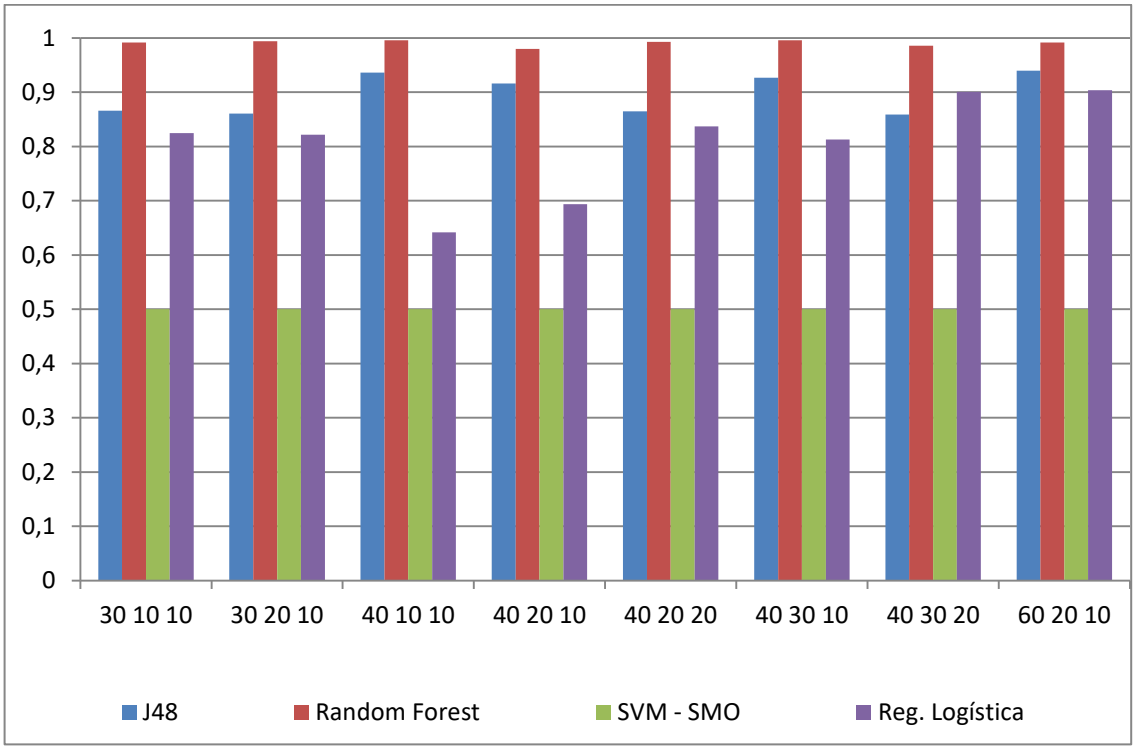


Figura 19 – Gráfico com os melhores resultados para AUC com reamostragem (Cão 1)

É possível verificar que no Cão 1, independentemente dos parâmetros colocados para a extração dos atributos o algoritmo *Random Forest* foi o que claramente obteve os melhores resultados. Por outro lado, o algoritmo que obteve as piores classificações foi o SVM. Os algoritmos J48 e Regressão Logística obtiveram resultados bons com determinados parâmetros.

Na Figura 20Figura 13 apresentam-se as curvas ROC para a classe preictal (P), correspondentes aos quatro algoritmos utilizados para o caso do Cão 1. Os parâmetros utilizados para a visualização das curvas ROC foram 60 20 10 com cinco aplicações da reamostragem. O motivo pelo qual foram estes os parâmetros escolhidos prendeu-se com o facto de terem os resultados mais altos e equilibrados para os algoritmos *Random Forest*, J48 e Regressão Logística.

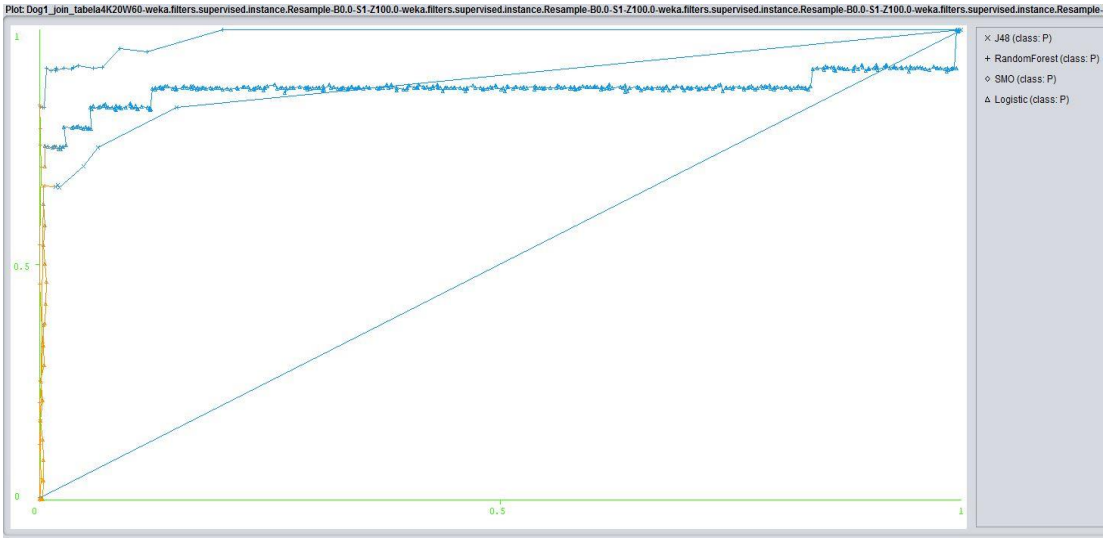


Figura 20 – Curvas ROC para os 4 algoritmos para o Cão 1 com reamostragem

No gráfico da Figura 20 é possível visualizar que para a classe preictal, o algoritmo *Random Forest* é aquele que claramente apresenta uma melhor curva ROC, pois a sua trajetória é superior às restantes linhas.

### 5.4.4 Cão 2

Todas as medidas apresentadas nesta subsecção são referentes ao Cão 2 para diferentes parâmetros de extração de atributos e várias aplicações da reamostragem.

Na Tabela 32 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 10 10.

Tabela 32 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 30 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 10 10 | 0,767 | 0,939 | 0,500 | 0,592 |
| 2 | 30 10 10 | 0,721 | 0,927 | 0,500 | 0,505 |
| 3 | 30 10 10 | 0,897 | 0,985 | 0,500 | 0,659 |
| 4 | 30 10 10 | 0,955 | 0,995 | 0,500 | 0,743 |
| 5 | 30 10 10 | 0,954 | **0,997** | 0,500 | 0,743 |

Na Tabela 32 é possível verificar que o algoritmo de classificação que conseguiu o melhor resultado (0,997) foi o *Random Forest* com a 5ª aplicação da reamostragem. Também é

importante salientar o algoritmo *Random Forest* é o que reage mais rapidamente à reamostragem, ou seja, logo na 1ª aplicação obtém resultados acima dos 0,9 mantendo-se assim para todas as aplicações. O algoritmo SVM manteve-se inalterado mesmo após a 5ª reamostragem o que significa que não deverá ser considerado.

Na Tabela 33 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 30 20 10.

Tabela 33 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 30 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 30 20 10 | 0,753 | 0,955 | 0,500 | 0,567 |
| 2 | 30 20 10 | 0,873 | 0,945 | 0,500 | 0,616 |
| 3 | 30 20 10 | 0,939 | 0,978 | 0,500 | 0,707 |
| 4 | 30 20 10 | 0,949 | **0,999** | 0,500 | 0,737 |
| 5 | 30 20 10 | 0,937 | 0,998 | 0,500 | 0,774 |

A Tabela 33 mostra que o algoritmo *Random Forest* continua a obter o melhor resultado (0,999), este foi obtido na 4ª aplicação da reamostragem. Tanto aqui, como na tabela anterior também se verifica que o algoritmo J48 é o que mais se aproxima. Os restantes algoritmos ficam bastante mais longe.

Na Tabela 34 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 10 10.

Tabela 34 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 10 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 10 10 | 0,652 | 0,922 | 0,500 | 0,602 |
| 2 | 40 10 10 | 0,840 | 0,934 | 0,500 | 0,600 |
| 3 | 40 10 10 | 0,915 | 0,987 | 0,500 | 0,658 |
| 4 | 40 10 10 | 0,928 | 0,992 | 0,500 | 0,706 |
| 5 | 40 10 10 | 0,958 | **0,996** | 0,500 | 0,748 |

Tal como nas duas tabelas anteriores, a Tabela 34 continua a mostrar que algoritmo *Random Forest* foi o que obteve melhores resultados (0,996). Apenas o algoritmo J48 se aproxima dos valores do *Random Forest*.

Na Tabela 35Tabela 33 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 10.

Tabela 35 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 10 | 0,727 | 0,931 | 0,500 | 0,629 |
| 2 | 40 20 10 | 0,821 | 0,932 | 0,500 | 0,683 |
| 3 | 40 20 10 | 0,908 | 0,988 | 0,500 | 0,725 |
| 4 | 40 20 10 | 0,930 | 0,993 | 0,500 | 0,861 |
| 5 | 40 20 10 | 0,974 | **0,999** | 0,500 | 0,877 |

Nos valores da Tabela 35 é possível observar que o algoritmo *Random Forest* continua a obter os melhores resultados (0,999). O algoritmo J48 também obtém valores interessantes. Os restantes continuam muito longe dos valores ideais.

Na Tabela 36 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 20 20.

Tabela 36 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 20 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 20 20 | 0,785 | 0,922 | 0,500 | 0,619 |
| 2 | 40 20 20 | 0,840 | 0,943 | 0,500 | 0,640 |
| 3 | 40 20 20 | 0,925 | 0,971 | 0,500 | 0,720 |
| 4 | 40 20 20 | 0,911 | **0,995** | 0,500 | 0,840 |
| 5 | 40 20 20 | 0,974 | 0,993 | 0,500 | 0,934 |

O algoritmo *Random Forest* continua a ser o que obtém os melhores resultados (0,995). O algoritmo J48 e Regressão Logística já apresentam valores acima dos 0,9 na 5ª reamostragem.

Na Tabela 37 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 10.

Tabela 37 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 30 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 10 | 0,734 | 0,934 | 0,500 | 0,655 |
| 2 | 40 30 10 | 0,762 | 0,952 | 0,500 | 0,720 |
| 3 | 40 30 10 | 0,882 | 0,988 | 0,500 | 0,819 |
| 4 | 40 30 10 | 0,920 | 0,990 | 0,500 | 0,938 |
| 5 | 40 30 10 | 0,950 | **0,999** | 0,500 | 0,967 |

O algoritmo *Random Forest* volta a obter a melhor classificação (0,999) e a ficar muito próximo do valor máximo. Apenas o algoritmo SVM fica mais distante dos valores ideais e se mantém inalterado.

Na Tabela 38 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 40 30 20.

Tabela 38 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 40 30 20)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 40 30 20 | 0,746 | 0,935 | 0,500 | 0,645 |
| 2 | 40 30 20 | 0,818 | 0,949 | 0,500 | 0,664 |
| 3 | 40 30 20 | 0,873 | 0,978 | 0,500 | 0,779 |
| 4 | 40 30 20 | 0,945 | **0,994** | 0,500 | 0,875 |
| 5 | 40 30 20 | 0,962 | 0,990 | 0,500 | 0,951 |

A Tabela 38 mostra que o algoritmo *Random Forest* volta a obter o melhor resultado (0,994). Tal como na tabela anterior, apenas o algoritmo SVM fica mais distante dos valores ideais.

Por fim, na Tabela 39 apresentam-se os resultados obtidos do Cão 2 analisando a medida AUC com cada uma das cinco reamostragens aplicadas e para os quatro algoritmos executados com os parâmetros de extração de atributos 60 20 10.

Tabela 39 – Resultados do Cão 2 para AUC com reamostragem (parâmetros 60 20 10)

| Reamostragem | Parâmetros extração atributos | J48 | Random Forest | SVM - SMO | Reg. Logística |
|---|---|---|---|---|---|
| 1 | 60 20 10 | 0,740 | 0,930 | 0,500 | 0,642 |
| 2 | 60 20 10 | 0,906 | 0,961 | 0,500 | 0,716 |
| 3 | 60 20 10 | 0,898 | 0,963 | 0,500 | 0,784 |
| 4 | 60 20 10 | 0,961 | **0,998** | 0,500 | 0,908 |
| 5 | 60 20 10 | 0,981 | 0,992 | 0,500 | 0,922 |

Finalmente, para os dados da Tabela 39 é possível verificar que, como habitualmente, o algoritmo *Random Forest* obtém o melhor resultado (0,998). O algoritmo J48 volta a ter um bom desempenho (0,981), seguido pelo Regressão Logística (0,922). Como habitualmente o SVM manteve-se estático e com o resultado bastante inferior aos restantes.

Na Figura 21 é possível visualizar um gráfico que apresenta um resumo dos melhores resultados para o Cão 2 de cada um dos algoritmos testados (J48, *Random Forest*, SVM e Regressão Logística) para a medida AUC com recurso às várias reamostragens.

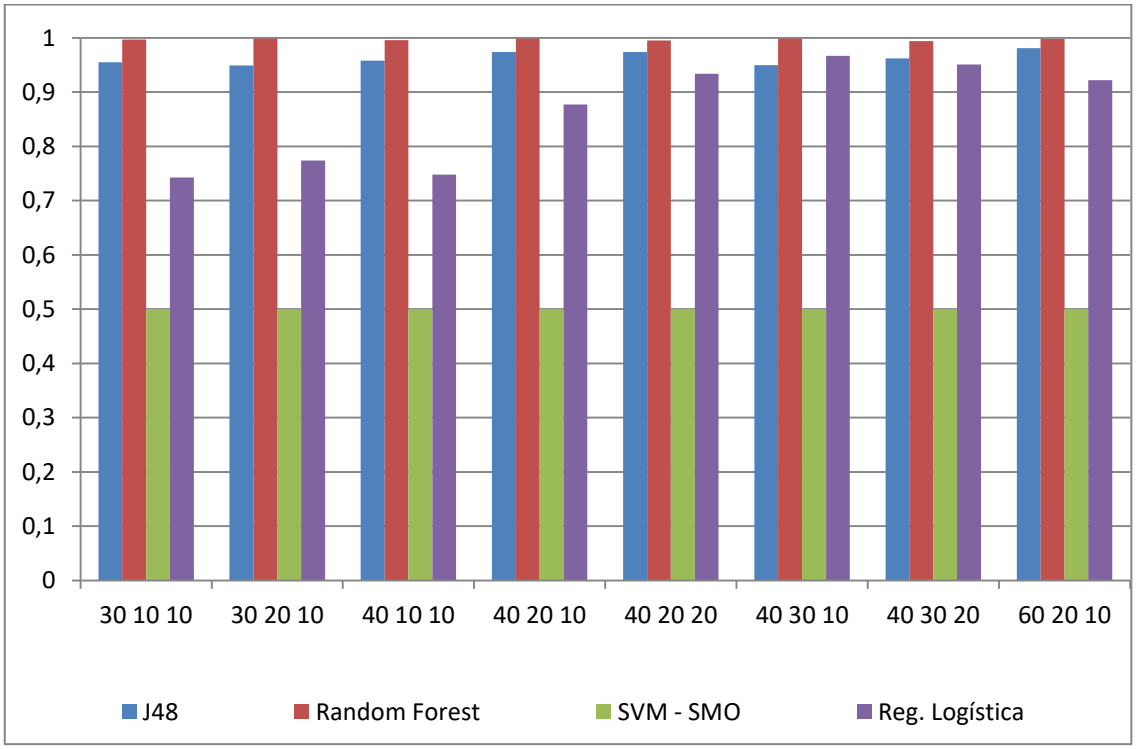


Figura 21 – Gráfico com os melhores resultados para AUC com reamostragem (Cão 2)

É possível verificar que para o Cão 2, o algoritmo *Random Forest* foi o que obteve as melhores classificações em todas as situações. Em situação inversa encontra-se o algoritmo SVM que foi claramente o que obteve as piores classificações. Os algoritmos J48 e Regressão Logística obtiveram bons resultados para determinados parâmetros.

Na Figura 22Figura 13 apresentam-se as curvas ROC para a classe preictal (P), correspondentes aos quatro algoritmos utilizados para o caso do Cão 2. Os parâmetros utilizados para a visualização das curvas ROC foram 40 30 10 com cinco aplicações da reamostragem. O motivo pelo qual foram estes os parâmetros escolhidos prendeu-se com o facto de terem os resultados mais altos e equilibrados para os algoritmos *Random Forest*, J48 e Regressão Logística.

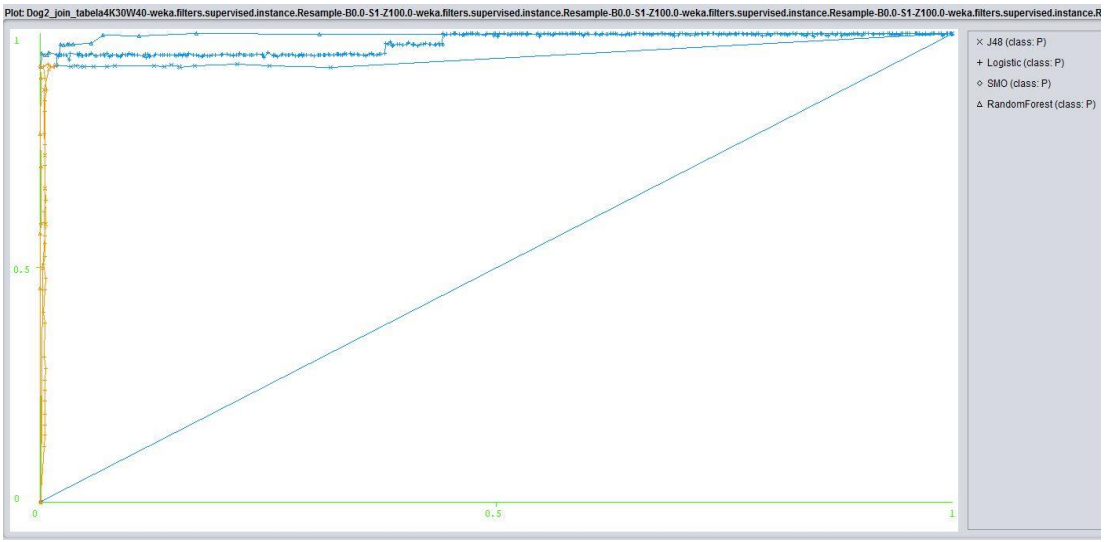


Figura 22 – Curvas ROC para os 4 algoritmos para o Cão 2 com reamostragem

No gráfico da Figura 22 é possível visualizar que para a classe preictal, o algoritmo *Random Forest* é aquele que claramente apresenta uma melhor curva ROC, pois a sua trajetória é superior às restantes linhas. Os resultados obtidos pelos algoritmos J48 e Regressão Logística também são dignos de assinalar. Já o SVM é claramente o algoritmo que apresenta o pior resultado.

## 5.5 Avaliação *Accuracy* e *F-Measure*

Nesta secção serão apresentados os melhores resultados referentes às medidas de avaliação *accuracy* e *f-measure.* Serão expostos os resultados obtidos para os diferentes parâmetros de extração dos atributos para o Paciente 1, Paciente 2, Cão 1 e Cão 2 com e sem o recurso de reamostragem. Também será mostrada a matriz confusão dos melhores resultados obtidos.

### 5.5.1 Paciente 1

Na Tabela 40, apresentam-se os resultados obtidos do Paciente 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo J48. Também estará disponível nas colunas "Reamostragem *Accuracy*" e "Reamostragem *f-measure*" o número de reamostragens realizadas para a obtenção do resultado.

Tabela 40 – Resultados do Paciente 1 para *accuracy* e *f-measure* para J48

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | **0,985** | 5 | **0,985** | 5 |
| 30 20 10 | 0,971 | 5 | 0,971 | 5 |
| 40 10 10 | **0,985** | 5 | **0,985** | 5 |
| 40 20 10 | 0,985 | 5 | 0,985 | 5 |
| 40 20 20 | 0,941 | 2 | 0,940 | 2 |
| 40 30 10 | **0,985** | 5 | **0,985** | 5 |
| 40 30 20 | 0,971 | 5 | 0,971 | 5 |
| 60 20 10 | 0,955 | 4 | 0,956 | 4 |

É possível observar na Tabela 40 que os melhores resultados para a *accuracy* (0,985) e para *f-measure* (0,985) foram obtidos com os parâmetros 30 10 10, 40 10 10 e 40 30 10 e todos eles foram alcançados na 5ª reamostragem.

Na Figura 23 é possível observar a matriz confusão para um dos casos anteriormente descritos.

| | I | P |
|---|---|---|
| I | 49 | 1 |
| P | 0 | 18 |

Figura 23 – Matriz confusão (5ª reamostragem - J48 - parâmetros 30 10 10)

A Figura 23 mostra que a classe interictal (I) tinha 50 registos e que 49 foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que a sua totalidade foi corretamente atribuída.

Na Tabela 41, apresentam-se os resultados obtidos do Paciente 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo *Random Forest*.

Tabela 41 – Resultados do Paciente 1 para *accuracy* e *f-measure* para *Random Forest*

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | **0,985** | 5 | **0,985** | 5 |
| 30 20 10 | 0,971 | 5 | 0,971 | 5 |
| 40 10 10 | **0,985** | 5 | **0,985** | 5 |
| 40 20 10 | **0,985** | 5 | **0,985** | 5 |
| 40 20 20 | **0,985** | 5 | **0,985** | 5 |
| 40 30 10 | **0,985** | 5 | **0,985** | 5 |
| 40 30 20 | **0,985** | 5 | **0,985** | 5 |
| 60 20 10 | **0,985** | 4 | **0,985** | 4 |

É possível observar na Tabela 41 que os melhores resultados para a *accuracy* (0,985) e para *f-measure* (0,985) foram obtidos com todos os parâmetros, exceto com 30 20 10 e todos eles foram alcançados na 5ª reamostragem, exceto o último que foi alcançado na 4ª reamostragem.

Na Figura 24 é possível observar a matriz confusão para um dos casos anteriormente descritos.

| | I | P |
|---|---|---|
| I | 49 | 1 |
| P | 0 | 18 |

Figura 24 – Matriz confusão (5ª reamostragem - *Random Forest* - parâmetros 30 10 10)

A Figura 24 mostra que a classe interictal (I) tinha 50 registos e que 49 foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que a sua totalidade foi corretamente atribuída.

Na Tabela 42, apresentam-se os resultados obtidos do Paciente 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo SVM - SMO.

Tabela 42 – Resultados do Paciente 1 para *accuracy* e *f-measure* para SVM - SMO

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,868 | 5 | 0,851 | 5 |
| 30 20 10 | 0,868 | 5 | 0,851 | 5 |
| 40 10 10 | 0,868 | 5 | 0,851 | 5 |
| 40 20 10 | 0,868 | 5 | 0,851 | 5 |
| 40 20 20 | 0,853 | 4 | 0,831 | 4 |
| 40 30 10 | 0,853 | 5 | 0,837 | 5 |
| 40 30 20 | **0,882** | 5 | **0,874** | 5 |
| 60 20 10 | 0,868 | 5 | 0,851 | 5 |

É possível observar na Tabela 42, que os melhores resultados para a *accuracy* (0,882) e para *f-measure* (0,874) foram obtidos com os parâmetros 40 30 20 e foi alcançado na 5ª reamostragem.

Na Figura 25 é possível observar a matriz confusão paro o caso anteriormente descrito.

| | I | P |
|---|---|---|
| I | 49 | 1 |
| P | 9 | 9 |

Figura 25 – Matriz confusão (5ª reamostragem - SVM - SMO - parâmetros 40 30 20)

A Figura 25 mostra que a classe interictal (I) tinha 50 registos e que 49 foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que apenas 50% foram corretamente atribuídos.

Na Tabela 43, apresentam-se os resultados obtidos do Paciente 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo Regressão Logística.

Tabela 43 – Resultados do Paciente 1 para *accuracy* e *f-measure* para Regressão Logística

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,971 | 4 | 0,971 | 4 |
| 30 20 10 | 0,971 | 4 | 0,971 | 4 |
| 40 10 10 | **0,985** | 4 | **0,985** | 4 |
| 40 20 10 | 0,956 | 5 | 0,956 | 5 |
| 40 20 20 | **0,985** | 5 | **0,985** | 5 |
| 40 30 10 | **0,985** | 5 | **0,985** | 5 |
| 40 30 20 | 0,971 | 2 | 0,970 | 2 |
| 60 20 10 | **0,985** | 4 | **0,985** | 4 |

É possível observar na Tabela 43, que os melhores resultados para a *accuracy* (0,985) e para *f-measure* (0,985) foram obtidos com os parâmetros 40 10 10, 40 20 20, 40 30 10 e 60 20 10. O primeiro e o último foram obtidos na 4ª reamostragem enquanto os restantes foram obtidos na 5ª reamostragem.

Na Figura 26 é possível observar a matriz confusão paro um dos casos anteriormente descrito.

| | I | P |
|---|---|---|
| I | 50 | 0 |
| P | 1 | 17 |

Figura 26 – Matriz confusão (4ª reamostragem - Regressão Logística - parâmetros 40 10 10)

A Figura 25Figura 26 mostra que a classe interictal (I) tinha 50 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que 17 deles foram corretamente atribuídos.

### 5.5.2 Paciente 2

Na Tabela 44, apresentam-se os resultados obtidos do Paciente 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo J48. Também estará disponível nas colunas "Reamostragem *Accuracy*" e "Reamostragem *f-measure*" o número de reamostragens realizadas para a obtenção do resultado.

Tabela 44 – Resultados do Paciente 2 para *accuracy* e *f-measure* para J48

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,950 | 1 | 0,950 | 1 |
| 30 20 10 | 0,933 | 4 | 0,932 | 4 |
| 40 10 10 | **0,967** | 2 | **0,967** | 2 |
| 40 20 10 | 0,917 | 4 | 0,914 | 4 |
| 40 20 20 | 0,933 | 5 | 0,931 | 5 |
| 40 30 10 | 0,900 | 4 | 0,898 | 4 |
| 40 30 20 | 0,933 | 5 | 0,931 | 5 |
| 60 20 10 | **0,967** | 4 | **0,967** | 4 |

É possível observar na Tabela 44 que os melhores resultados para a *accuracy* (0,967) e para *f-measure* (0,967) foram obtidos com os parâmetros, 40 10 10 e 60 20 10. O primeiro foi alcançado na 2ª reamostragem e o segundo na 4ª reamostragem.

Na Figura 27 é possível observar a matriz confusão para um dos casos anteriormente descritos.

| | I | P |
|---|---|---|
| I | 41 | 1 |
| P | 1 | 17 |

Figura 27 – Matriz confusão (2ª reamostragem - J48 - parâmetros 40 10 10)

A Figura 27 mostra que a classe interictal (I) tinha 42 registos e que 41 foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que 17 deles foram corretamente atribuídos.

Na Tabela 45 apresentam-se os resultados obtidos do Paciente 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo *Random Forest*.

Tabela 45 – Resultados do Paciente 2 para *accuracy* e *f-measure* para *Random Forest*

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | **0,983** | 2 | **0,983** | 2 |
| 30 20 10 | 0,967 | 2 | 0,966 | 2 |
| 40 10 10 | 0,967 | 2 | 0,967 | 2 |
| 40 20 10 | 0,967 | 5 | 0,966 | 5 |
| 40 20 20 | 0,967 | 5 | 0,966 | 5 |
| 40 30 10 | 0,933 | 2 | 0,934 | 2 |
| 40 30 20 | 0,967 | 5 | 0,966 | 5 |
| 60 20 10 | **0,983** | 4 | **0,983** | 4 |

É possível observar na Tabela 45 que os melhores resultados para a *accuracy* (0,983) e para *f-measure* (0,983) foram obtidos com os parâmetros, 30 10 10 e 60 20 10. O primeiro foi alcançado na 2ª reamostragem e o segundo na 4ª reamostragem.

Na Figura 28 é possível observar a matriz confusão para um dos casos anteriormente descritos.

| | I | P |
|---|---|---|
| I | 42 | 0 |
| P | 1 | 17 |

Figura 28 – Matriz confusão (2ª reamostragem – *Random Forest* - parâmetros 30 10 10)

A Figura 28 mostra que a classe interictal (I) tinha 42 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que 17 deles foram corretamente atribuídos.

Na Tabela 46 apresentam-se os resultados obtidos do Paciente 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo SVM - SMO.

Tabela 46 – Resultados do Paciente 2 para *accuracy* e *f-measure* para SVM - SMO

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,700 | 1 | 0,576 | 1 |
| 30 20 10 | **0,700** | 0 | **0,604** | 0 |
| 40 10 10 | 0,700 | 1 | 0,576 | 1 |
| 40 20 10 | 0,700 | 1 | 0,584 | 1 |
| 40 20 20 | 0,700 | 1 | 0,576 | 1 |
| 40 30 10 | 0,700 | 1 | 0,576 | 1 |
| 40 30 20 | 0,700 | 1 | 0,576 | 1 |
| 60 20 10 | 0,700 | 1 | 0,576 | 1 |

É possível observar na Tabela 46 que os melhores resultados para a *accuracy* (0,700) e para *f-measure* (0,604) foram obtidos com os parâmetros, 30 20 10 sem recurso à reamostragem.
Na Figura 29 é possível observar a matriz confusão para o caso anteriormente descrito.

| | I | P |
|---|---|---|
| I | 41 | 1 |
| P | 17 | 1 |

Figura 29 – Matriz confusão (sem reamostragem – SVM - SMO - parâmetros 30 20 10)

A Figura 29 mostra que a classe interictal (I) tinha 42 registos e que 41 foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que apenas 1 deles foi corretamente atribuído.
Na Tabela 47, apresentam-se os resultados obtidos do Paciente 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo Regressão Logística.

Tabela 47 – Resultados do Paciente 2 para *accuracy* e *f-measure* para Regressão Logística

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,967 | 5 | 0,966 | 5 |
| 30 20 10 | 0,983 | 5 | 0,983 | 5 |
| 40 10 10 | 0,967 | 5 | 0,966 | 5 |
| 40 20 10 | **1,000** | 5 | **1,000** | 5 |
| 40 20 20 | 0,983 | 5 | 0,983 | 5 |
| 40 30 10 | 0,950 | 2 | 0,950 | 2 |
| 40 30 20 | 0,983 | 5 | 0,983 | 5 |
| 60 20 10 | 0,983 | 5 | 0,983 | 5 |

É possível observar na Tabela 44Tabela 47 que os melhores resultados para a *accuracy* (1,000) e para *f-measure* (1,000) foram obtidos com os parâmetros, 40 20 10 na 5ª aplicação da reamostragem.

Na Figura 30 é possível observar a matriz confusão para o caso anteriormente descrito.

| | I | P |
|---|---|---|
| I | 42 | 0 |
| P | 0 | 18 |

Figura 30 – Matriz confusão (5ª reamostragem – Regressão Logística - parâmetros 40 20 10)

A Figura 30 mostra que a classe interictal (I) tinha 42 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 18 registos e que também todos eles foram corretamente atribuídos.

### 5.5.3 Cão 1

Na Tabela 48, apresentam-se os resultados obtidos do Cão 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo J48. Também estará disponível nas colunas "Reamostragem *Accuracy*" e "Reamostragem *f-measure*" o número de reamostragens realizadas para a obtenção do resultado.

Tabela 48 – Resultados do Cão 1 para *accuracy* e *f-measure* para J48

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,986 | 5 | 0,986 | 5 |
| 30 20 10 | 0,982 | 5 | 0,982 | 5 |
| 40 10 10 | 0,974 | 4 | 0,974 | 4 |
| 40 20 10 | **0,990** | 5 | **0,990** | 5 |
| 40 20 20 | 0,984 | 5 | 0,984 | 5 |
| 40 30 10 | 0,976 | 5 | 0,977 | 5 |
| 40 30 20 | 0,978 | 5 | 0,978 | 5 |
| 60 20 10 | 0,986 | 5 | 0,986 | 5 |

É possível observar na Tabela 48, que o melhor resultado para a *accuracy* (0,990) e para *f-measure* (0,990) foram obtidos com os parâmetros 40 20 10 na 5ª reamostragem.

Na Figura 31 é possível observar a matriz confusão para o caso anteriormente descrito.

| | I | P |
|---|---|---|
| I | 480 | 0 |
| P | 5 | 19 |

Figura 31 – Matriz confusão (5ª reamostragem - J48 - parâmetros 40 20 10)

A Figura 31 mostra que a classe interictal (I) tinha 480 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 24 registos e que 19 deles foram corretamente atribuídos.

Na Tabela 49 apresentam-se os resultados obtidos do Cão 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo *Random Forest*.

Tabela 49 – Resultados do Cão 1 para *accuracy* e *f-measure* para *Random Forest*

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,990 | 4 | 0,990 | 4 |
| 30 20 10 | **0,992** | 4 | **0,992** | 4 |
| 40 10 10 | 0,990 | 2 | 0,990 | 2 |
| 40 20 10 | **0,992** | 4 | **0,992** | 4 |
| 40 20 20 | **0,992** | 4 | **0,992** | 4 |
| 40 30 10 | 0,990 | 5 | 0,990 | 5 |
| 40 30 20 | **0,992** | 4 | **0,992** | 4 |
| 60 20 10 | 0,990 | 4 | 0,990 | 4 |

É possível observar na Tabela 49, que os melhores resultados para a *accuracy* (0,992) e para *f-measure* (0,992) foram obtidos com os parâmetros 30 20 10, 40 20 10, 40 20 20 e 40 30 20 na 4ª reamostragem.

Na Figura 32 é possível observar a matriz confusão para um dos casos anteriormente descrito.

| | I | P |
|---|---|---|
| I | 480 | 0 |
| P | 4 | 20 |

Figura 32 – Matriz confusão (5ª reamostragem - *Random Forest* - parâmetros 40 20 10)

A Figura 32 mostra que a classe interictal (I) tinha 480 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 24 registos e que 20 deles foram corretamente atribuídos.

Na Tabela 50, apresentam-se os resultados obtidos do Cão 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo SVM - SMO.

Tabela 50 – Resultados do Cão 1 para *accuracy* e *f-measure* para SVM - SMO

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,952 | 0 | 0,929 | 0 |
| 30 20 10 | 0,952 | 0 | 0,929 | 0 |
| 40 10 10 | 0,952 | 0 | 0,929 | 0 |
| 40 20 10 | 0,952 | 0 | 0,929 | 0 |
| 40 20 20 | 0,952 | 0 | 0,929 | 0 |
| 40 30 10 | 0,952 | 0 | 0,929 | 0 |
| 40 30 20 | 0,952 | 0 | 0,929 | 0 |
| 60 20 10 | 0,952 | 0 | 0,929 | 0 |

É possível observar na Tabela 50, que todos os resultados para a *accuracy* (0,952) e para *f-measure* (0,929) são iguais, independentemente dos parâmetros passados. Também a reamostragem não produziu qualquer efeito.

Na Figura 33 é possível observar a matriz confusão para um dos casos anteriormente descrito.

| | I | P |
|---|---|---|
| I | 480 | 0 |
| P | 24 | 0 |

Figura 33 – Matriz confusão (5ª reamostragem - SVM - SMO - parâmetros 40 20 10)

A Figura 32 mostra que a classe interictal (I) tinha 480 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 24 registos e que nenhum deles foi corretamente atribuído.

Na Tabela 51, apresentam-se os resultados obtidos do Cão 1 analisando as medidas *accuracy* e *f-measure* para o algoritmo Regressão Logística.

Tabela 51 – Resultados do Cão 1 para *accuracy* e *f-measure* para Regressão Logística

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,960 | 2 | 0,956 | 2 |
| 30 20 10 | 0,966 | 1 | 0,961 | 1 |
| 40 10 10 | 0,960 | 3 | 0,949 | 3 |
| 40 20 10 | 0,958 | 5 | 0,947 | 5 |
| 40 20 20 | 0,964 | 5 | 0,962 | 5 |
| 40 30 10 | 0,956 | 5 | 0,949 | 5 |
| 40 30 20 | **0,988** | 5 | **0,988** | 5 |
| 60 20 10 | 0,960 | 2 | 0,952 | 2 |

É possível observar na Tabela 51, que o melhor resultado para a *accuracy* (0,988) e para *f-measure* (0,988) foram obtidos com os parâmetros 40 30 20 na 5ª reamostragem.
Na Figura 34 é possível observar a matriz confusão para o caso anteriormente descrito.

| | I | P |
|---|---|---|
| I | 478 | 2 |
| P | 4 | 20 |

Figura 34 – Matriz confusão (5ª reamostragem - Regressão Logística - parâmetros 40 30 20)

A Figura 34 mostra que a classe interictal (I) tinha 480 registos e que 478 foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 24 registos e que 20 deles foram corretamente atribuídos.

### 5.5.4 Cão 2

Na Tabela 52 apresentam-se os resultados obtidos do Cão 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo J48. Também estará disponível nas colunas "Reamostragem *Accuracy*" e "Reamostragem *f-measure*" o número de reamostragens realizadas para a obtenção do resultado.

Tabela 52 – Resultados do Cão 2 para *accuracy* e *f-measure* para J48

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,989 | 5 | 0,989 | 5 |
| 30 20 10 | 0,989 | 5 | 0,989 | 5 |
| 40 10 10 | 0,985 | 5 | 0,985 | 5 |
| 40 20 10 | **0,994** | 5 | **0,994** | 5 |
| 40 20 20 | 0,991 | 5 | 0,991 | 5 |
| 40 30 10 | 0,987 | 5 | 0,987 | 5 |
| 40 30 20 | 0,989 | 5 | 0,989 | 5 |
| 60 20 10 | 0,987 | 5 | 0,987 | 5 |

É possível observar na Tabela 52Tabela 48, que o melhor resultado para a *accuracy* (0,994) e para *f-measure* (0,994) foram obtidos com os parâmetros 40 20 10 na 5ª reamostragem.
Na Figura 35 é possível observar a matriz confusão para o caso anteriormente descrito.

| | I | P |
|---|---|---|
| I | 499 | 1 |
| P | 2 | 40 |

Figura 35 – Matriz confusão (5ª reamostragem - J48 - parâmetros 40 20 10)

A Figura 35 mostra que a classe interictal (I) tinha 500 registos e que apenas um foi incorretamente atribuído, mostra também que a classe preictal (P) tinha 42 registos e que 40 deles foram corretamente atribuídos.
Na Tabela 53 apresentam-se os resultados obtidos do Cão 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo *Random Forest*.

Tabela 53 – Resultados do Cão 2 para *accuracy* e *f-measure* para *Random Forest*

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,994 | 4 | 0,994 | 4 |
| 30 20 10 | 0,992 | 4 | 0,992 | 4 |
| 40 10 10 | 0,994 | 4 | 0,994 | 4 |
| 40 20 10 | 0,994 | 5 | 0,994 | 5 |
| 40 20 20 | 0,994 | 5 | 0,994 | 5 |
| 40 30 10 | 0,994 | 5 | 0,994 | 5 |
| 40 30 20 | 0,994 | 5 | 0,994 | 5 |
| 60 20 10 | 0,994 | 5 | 0,994 | 5 |

É possível observar na Tabela 53, que os melhores resultados para a *accuracy* (0,994) e para *f-measure* (0,994) foram obtidos com tosos parâmetros exceto com 30 20 10. Todos foram alcançados na 4ª ou 5ª reamostragem

Na Figura 32Figura 36 é possível observar a matriz confusão para um dos casos anteriormente descrito.

| | I | P |
|---|---|---|
| I | 500 | 0 |
| P | 3 | 39 |

Figura 36 – Matriz confusão (5ª reamostragem - *Random Forest* - parâmetros 40 20 10)

A Figura 32Figura 36 mostra que a classe interictal (I) tinha 500 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 42 registos e que 39 deles foram corretamente atribuídos.

Na Tabela 54 apresentam-se os resultados obtidos do Cão 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo SVM - SMO.

Tabela 54 – Resultados do Cão 2 para *accuracy* e *f-measure* para SVM - SMO

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,923 | 0 | 0,885 | 0 |
| 30 20 10 | 0,923 | 0 | 0,885 | 0 |
| 40 10 10 | 0,923 | 0 | 0,885 | 0 |
| 40 20 10 | 0,923 | 0 | 0,885 | 0 |
| 40 20 20 | 0,923 | 0 | 0,885 | 0 |
| 40 30 10 | 0,923 | 0 | 0,885 | 0 |
| 40 30 20 | 0,923 | 0 | 0,885 | 0 |
| 60 20 10 | 0,923 | 0 | 0,885 | 0 |

É possível observar na Tabela 54, que todos os resultados para a *accuracy* (0,923) e para *f-measure* (0,885) são iguais, independentemente dos parâmetros passados. Também a reamostragem não produziu qualquer efeito.

Na Figura 37 é possível observar a matriz confusão para um dos casos anteriormente descrito.

| | I | P |
|---|---|---|
| I | 500 | 0 |
| P | 42 | 0 |

Figura 37 – Matriz confusão (5ª reamostragem - SVM - SMO - parâmetros 40 20 10)

A Figura 32Figura 37 mostra que a classe interictal (I) tinha 500 registos e que todos foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 42 registos e que nenhum deles foi corretamente atribuído.

Na Tabela 55 apresentam-se os resultados obtidos do Cão 2 analisando as medidas *accuracy* e *f-measure* para o algoritmo Regressão Logística.

Tabela 55 – Resultados do Cão 2 para *accuracy* e *f-measure* para Regressão Logística

| Parâmetros extração atributos | *Accuracy* | Reamostragem *Accuracy* | *f-measure* | Reamostragem *f-measure* |
|---|---|---|---|---|
| 30 10 10 | 0,923 | 0 | 0,885 | 0 |
| 30 20 10 | 0,923 | 0 | 0,887 | 5 |
| 40 10 10 | 0,923 | 0 | 0,885 | 0 |
| 40 20 10 | 0,928 | 5 | 0,921 | 5 |
| 40 20 20 | 0,954 | 5 | 0,951 | 5 |
| 40 30 10 | 0,980 | 5 | 0,980 | 5 |
| 40 30 20 | **0,985** | 5 | **0,985** | 5 |
| 60 20 10 | 0,961 | 5 | 0,961 | 5 |

É possível observar na Tabela 55, que o melhor resultado para a *accuracy* (0,985) e para *f-measure* (0,985) foram obtidos com os parâmetros 40 30 20 na 5ª reamostragem.

Na Figura 34Figura 38 é possível observar a matriz confusão para o caso anteriormente descrito.

| | I | P |
|---|---|---|
| I | 495 | 5 |
| P | 3 | 39 |

Figura 38 – Matriz confusão (5ª reamostragem - Regressão Logística - parâmetros 40 30 20)

A Figura 34Figura 38 mostra que a classe interictal (I) tinha 500 registos e que 495 foram corretamente atribuídos, mostra também que a classe preictal (P) tinha 42 registos e que 39 deles foram corretamente atribuídos.

## 5.6 Discussão da metodologia de avaliação

Nesta dissertação foi dada uma maior importância à medida AUC (visível através do gráfico da Curva ROC), no entanto, as medidas *accuracy* e *f-measure* também foram analisadas.

Esta escolha foi feita tendo em consideração que as classes utilizadas são bastante desbalanceadas, logo, uma alta taxa de *accuracy* ou *f-measure* poderá não significar um resultado útil. É possível visualizar um exemplo prático na Tabela 50, onde é apresentado um resultado para a *accuracy* de 0,952 e para a *f-measure* de 0,929. Este resultado poderia levar a crer que os resultados são bons, mas como é possível verificar através da Figura 33, onde está

representada a matriz confusão referente às classificações referidas anteriormente, o algoritmo SVM-SMO acertou em todas as classes interictal (480), mas falhou em todas as classes preictal (24). Significa que, para este *dataset*, com este algoritmo, não seria possível prever nenhuma convulsão. Este mesmo caso, quando analisado com as curvas ROC (Figura 11) mostra claramente que os resultados ficam muito longe de serem considerados satisfatórios.

## 5.7 Teste estatístico

Para suportar as afirmações relativas à comparação dos modelos de classificação, usamos o *t-test* para amostras emparelhadas.

Como já foi referido, utilizamos o *Weka* para realizar as nossas experiências. Assim, para efetuar a análise dos resultados, usamos o *Paired t-test* disponibilizado pelo Experimenter, uma ferramenta de análise de experiências do *Weka*.

Nas figuras seguintes, apresentamos os resultados obtidos para o *t-test* quando comparamos o desempenho do algoritmo *Random Forest* com os algoritmos J48, SVM e Regressão Logística, pois foi o *Random Forest* que apresentou os melhores resultados. Nos testes seguintes pretendemos determinar se os resultados são estatisticamente significativos, para um nível de significância 0.05.

Na Figura 39, apresenta-se o resultado obtido para o Paciente 1, nas condições correspondentes à 1ª linha da Tabela 4 (situação em que observámos um melhor desempenho do modelo, para a medida de avaliação AUC).

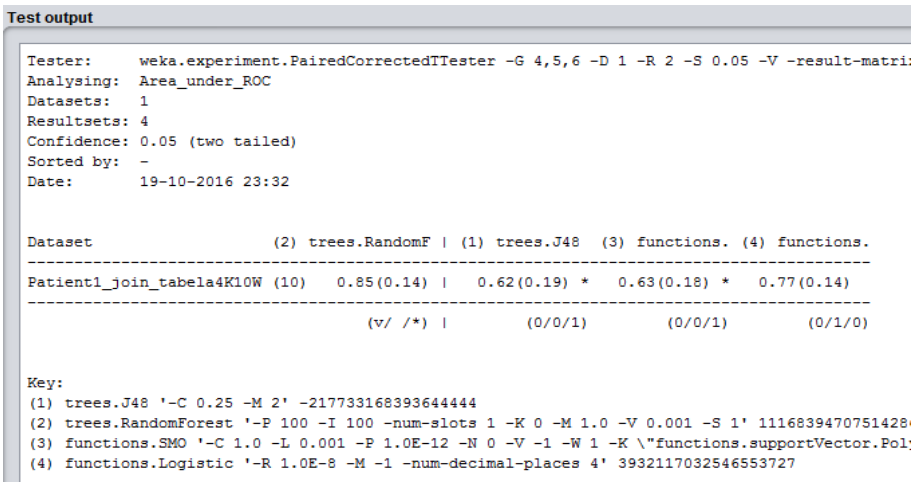


Figura 39 – Resultado do *t-test* – Paciente 1

A Figura 39 apresenta uma tabela com os quatros algoritmos e o respetivo resultado. Na primeira linha da tabela, surge o nome de cada algoritmo. Na segunda, os valores obtidos para a medida de avaliação escolhida (AUC) e, entre parênteses, o desvio padrão (opcional). Na última linha surge o resultado da análise efetuada para os diferentes algoritmos. A notação **v** ou ***** indica que um determinado resultado é estatisticamente melhor (v) ou pior (*) que o modelo de comparação (*baseline scheme*), neste caso com o algoritmo *Random Forest*, para o nível de significância especificado (neste caso, 0.05). Para este nível de significância, os resultados obtidos pelo J48 e SMO, são considerados piores que os resultados obtidos pelo *Random Forest*.

No final de cada coluna da tabela, surge uma contagem (xx/ yy/ zz), do número de vezes que o modelo foi melhor do que o modelo (xx), (yy), ou (zz). Neste exemplo, podemos observar que o J48 foi uma vez melhor do que a Regressão Logística; o SVM foi uma vez melhor do que a Regressão Logística; a Regressão Logística foi uma vez melhor do que o J48.

Na Figura 40, apresenta-se o resultado obtido para o Paciente 2, nas condições correspondentes à 2ª linha da Tabela 5 (situação em que observámos um melhor desempenho do modelo, para a medida de avaliação AUC).

**Test output**

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -V -result-matrix "w
Analysing:  Area_under_ROC
Datasets:   1
Resultsets: 4
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       22-10-2016 2:07


Dataset                   (2) trees.RandomF | (1) trees.J48  (3) functions. (4) functions.
-----------------------------------------------------------------------------------------
Patient2_join_tabela4K20W (10)   0.76(0.25) |   0.68(0.24)     0.64(0.41)     0.54(0.17)
-----------------------------------------------------------------------------------------
                                 (v/ /*) |        (0/1/0)        (0/1/0)        (0/1/0)


Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644444
(2) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698
(3) functions.Logistic '-R 1.0E-8 -M -1 -num-decimal-places 4' 3932117032546553727
(4) functions.SMO '-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"functions.supportVector.PolyKer
```

Figura 40 - Resultado do *t-test* – Paciente 2

Neste caso, os resultados não são estatisticamente conclusivos, para o nível de significância 0.05 (não surge a notação **v** ou *****). Apesar de o *Random Forest* ter apresentado melhor

desempenho, para este nível de significância, os resultados não são estatisticamente significativos.

Na Figura 41 apresenta-se o resultado obtido para o Cão 1, nas condições correspondentes à 4ª linha da Tabela 6 (situação em que observámos um melhor desempenho do modelo, para a medida de avaliação AUC).

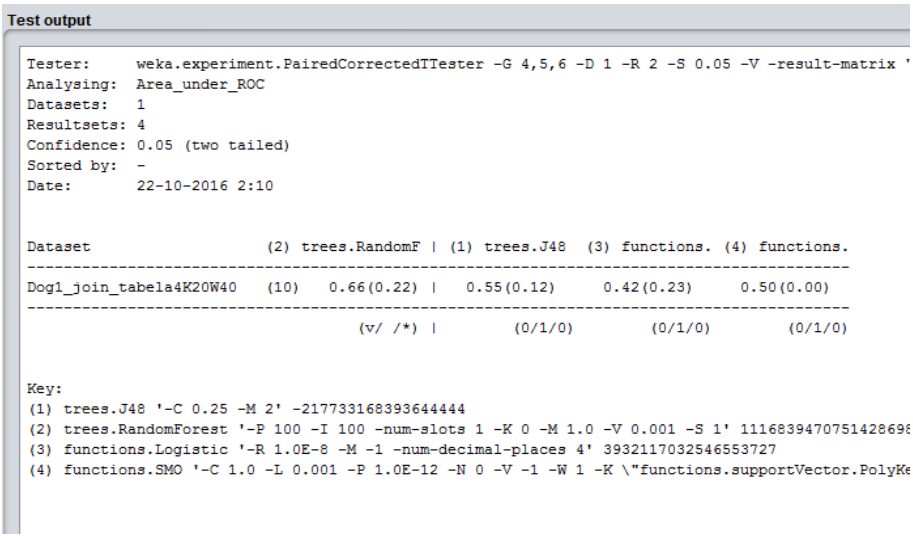


Figura 41 - Resultado do *t-test* – Cão 1

Neste caso, tal como no caso anterior, apesar do *Random Forest* apresentar o melhor desempenho durante as nossas experiências, os resultados não são estatisticamente conclusivos, para o nível de significância 0.05.

Na Figura 42, apresenta-se o resultado obtido para o Cão 2, nas condições correspondentes à 5ª linha da Tabela 7 (situação em que observámos um melhor desempenho do modelo, para a medida de avaliação AUC).

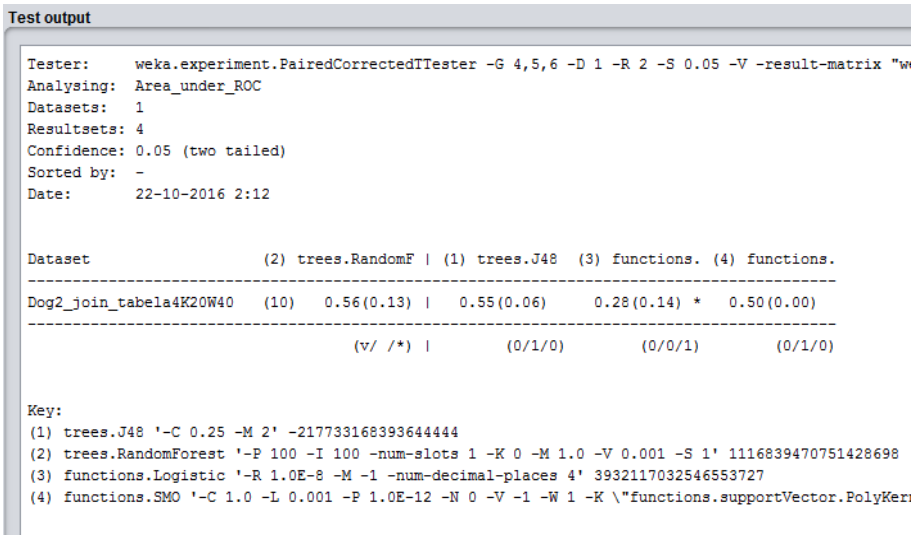


Figura 42 - Resultado do *t-test* – Cão 2

No caso do Cão2, para o nível de significância 0.05, os resultados não são estatisticamente conclusivos para os algoritmos J48 e para o SVM. No entanto, para este nível de significância, a Regressão Logistica apresenta um pior desempenho quando comparada com o algoritmo *Random Forest*.

# 6 Conclusão

O principal objetivo desta dissertação é contribuir para o desenvolvimento de uma metodologia que permita efetuar a previsão da ocorrência de convulsões epiléticas e como consequência ajudar a melhorar a vida de todas as pessoas afetadas por esta doença.

Neste trabalho, apresenta-se uma metodologia para tratamento de dados obtidos a partir de eletroencefalogramas. O objetivo foi o de diferenciar as classes existentes (preictal e interictal) aplicando e avaliando vários algoritmos de classificação depois de aplicadas técnicas de extração de atributos.

Os resultados obtidos pelos algoritmos de classificação mostram que o *Random Forest* obtém os resultados mais consistentes ao longo dos diferentes processamentos. Foram efetuadas variadas análises aos resultados obtidos, nomeadamente com as medidas AUC, *accuracy* e *f-measure.* Foi dada uma maior importância à medida AUC e ao gráfico da Curva ROC onde é possível ver toda a evolução da curva e comparar com os restantes algoritmos de classificação. Globalmente, o algoritmo que obteve a melhor curva e por consequência melhor desempenho foi o *Random Forest*.

Os melhores resultados obtidos, para os quatro casos de estudo, podem ser visualizados através das curvas de ROC na Figura 16, Figura 18, Figura 20 e Figura 22, para o Paciente 1, Paciente 2, Cão 1 e Cão 2 respetivamente. Em todos estes casos os resultados obtidos foram bastante satisfatórios e também é possível identificar o algoritmo *Random Forest* como sendo o que obtém os melhores resultados. O algoritmo Regressão Logística também mostra resultados bastante interessantes, sendo que, no caso do Paciente 2 foi o que obteve a melhor classificação.

Assim, é possível acreditar que esta metodologia poderá ser um suporte útil no desenvolvimento de sistemas de previsão de convulsões em EEG.

## 6.1 Trabalho futuro

Numa perspetiva de continuação deste trabalho, são de considerar as seguintes tarefas:

- Testar técnicas referidas na literatura para lidar com dados desbalanceados;
- Testar o modelo sugerido em novos *datasets* de diferentes proveniências, a fim de consolidar os resultados obtidos;
- Desenvolvimento de um sistema de notificações para os pacientes, ativado pelas previsões do modelo proposto.

# Referências


(Aarabi and He, 2012) Aarabi, A., He, B., 2012. A rule-based seizure prediction method for focal neocortical epilepsy. Clin. Neurophysiol. 123, 1111–1122.

(Acharya et al., 2012) Acharya, U.R., Molinari, F., Sree, S.V., Chattopadhyay, S., Ng, K.-H., Suri, J.S., 2012. Automated diagnosis of epileptic EEG using entropies. Biomed. Signal Process. Control 7, 401–408.

(Adeli et al., 2003) Adeli, H., Zhou, Z., Dadmehr, N., 2003. Analysis of EEG records in an epileptic patient using wavelet transform. J. Neurosci. Methods 123, 69–87.

(Alam and Bhuiyan, 2013) Alam, S., Bhuiyan, M.I.H., 2013. Detection of seizure and epilepsy using higher order statistics in the EMD domain. Biomed. Health Inform. IEEE J. Of 17, 312–318.

(Alam and Bhuiyan, 2011) Alam, S.S., Bhuiyan, M., 2011. Detection of epileptic seizures using chaotic and statistical features in the EMD domain. Presented at the India Conference (INDICON), 2011 Annual IEEE, IEEE, pp. 1–4.

(Almeida et al., 2002) Almeida, A.T. de, Gomes, C.F.S., Gomes, L., 2002. Tomada de Decisão Gerencial–Enfoque Multicritério. São Paulo Atlas.

(Alotaiby et al., 2014) Alotaiby, T.N., Alshebeili, S.A., Alshawi, T., Ahmad, I., El-Samie, F.E.A., 2014. EEG seizure detection and prediction algorithms: a survey. EURASIP J. Adv. Signal Process. 2014, 1–21.

(American Epilepsy Society, 2016) American Epilepsy Society, 2016. What is epilepsy? | FACTS AND FIGURES | American Epilepsy Society [WWW Document]. Am. Epilepsy Soc. URL https://www.aesnet.org/for_patients/facts_figures#Two (accessed 1.17.16).

(Bana e Costa et al., 2011) Bana e Costa, C.A., Corte, J., Vansnick, J., 2011. MACBETH (measuring attractiveness by a categorical based evaluation technique). Wiley Encycl. Oper. Res. Manag. Sci.

(Barnes et al., 2009) Barnes, C., Blake, H., Pinder, D., 2009. Creating and Delivering Your Value Proposition: Managing Customer Experience for Profit. Kogan Page Publishers.

(Bedeeuzzaman et al., 2014) Bedeeuzzaman, M., Fathima, T., Khan, Y.U., Farooq, O., 2014. Seizure prediction using statistical dispersion measures of intracranial EEG. Biomed. Signal Process. Control 10, 338–341.

(Boas and de Lima, 2010). Boas, V., de Lima, C., 2010. Modelo multicritérios de apoio à decisão aplicado ao uso múltiplo de reservatórios: estudo da barragem do ribeirão João Leite.

(Bray et al., 2015) Bray, S., Caggiani, L., Ottomanelli, M., 2015. Measuring Transport Systems Efficiency Under Uncertainty by Fuzzy Sets Theory Based Data Envelopment Analysis: Theoretical and Practical Comparison with Traditional DEA Model. Transp. Res. Procedia 5, 186–200.

(Breiman, 2001) Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

(Carney et al., 2011) Carney, P.R., Myers, S., Geyer, J.D., 2011. Seizure prediction: methods. Epilepsy Behav. 22, S94–S101.

(Carvalho et al., 2014) Carvalho, V.R., Mendes, E.M., Moraes, M.F., Braga, A. de P., 2014. Classificação de Crises Epilépticas por decomposição de modos de sinais de EEG 1–8.

(Cass, 2015) Cass, S., 2015. The 2015 Top Ten Programming Languages [WWW Document]. URL http://spectrum.ieee.org/computing/software/the-2015-top-

ten-programming-languages (accessed 2.16.16).

(Castro and Azevedo, 2010) Castro, N., Azevedo, P.J., 2010. Multiresolution Motif Discovery in Time Series. Presented at the SDM, SIAM, pp. 665–676.

(Chiang et al., 2011) Chiang, C.-Y., Chang, N.-F., Chen, T.-C., Chen, H.-H., Chen, L.-G., 2011. Seizure prediction based on classification of EEG synchronization patterns with on-line retraining and post-processing scheme. Presented at the Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, IEEE, pp. 7564–7569.

(Chisci et al., 2010) Chisci, L., Mavino, A., Perferi, G., Sciandrone, M., Anile, C., Colicchio, G., Fuggetta, F., 2010. Real-time epileptic seizure prediction using AR models and support vector machines. Biomed. Eng. IEEE Trans. On 57, 1124–1132.

(Classification Methods, 2016) Classification Methods [WWW Document], 2016. URL http://www.d.umn.edu/~padhy005/Chapter5.html (accessed 10.9.16).

(Comon, 1994) Comon, P., 1994. Independent component analysis, a new concept? Signal Process. 36, 287–314.

(Dalton et al., 2012) Dalton, A., Patel, S., Chowdhury, A.R., Welsh, M., Pang, T., Schachter, S.C., OLaighin, G., Bonato, P., 2012. Development of a body sensor network to detect motor patterns of epileptic seizures. Biomed. Eng. IEEE Trans. On 59, 3204–3211.

(Data | Kaggle, 2016) Data | Kaggle [WWW Document], 2016. URL https://www.kaggle.com/c/seizure-prediction/data (accessed 10.20.16).

(de Moor and Weigand, 2004) de Moor, A., Weigand, H., 2004. Business Negotiation Support: Theory and Practice. Int. Negot. 9, 31–57.

(Eftekhar et al., 2008) Eftekhar, A., Vohra, F., Toumazou, C., Drakakis, E.M., Parker, K., 2008. Hilbert-Huang transform: preliminary studies in epilepsy and cardiac arrhythmias. Presented at the Biomedical Circuits and Systems Conference, 2008. BioCAS 2008. IEEE, IEEE, pp. 373–376.

(Evaluation | Kaggle, 2016) Evaluation | Kaggle [WWW Document], 2016. URL https://www.kaggle.com/c/seizure-prediction/details/evaluation (accessed 10.16.16).

(Fisher et al., 2000) Fisher, R.S., Vickrey, B.G., Gibson, P., Hermann, B., Penovich, P., Scherer, A., Walker, S., 2000. The impact of epilepsy from the patient's perspective I. Descriptions and subjective perceptions. Epilepsy Res. 41, 39–51. doi:10.1016/S0920-1211(00)00126-1

(Frei, 2013) Frei, M.G., 2013. Seizure detection. Scholarpedia 8, 5780.

(Gajic et al., 2014). Gajic, D., Djurovic, Z., Di Gennaro, S., Gustafsson, F., 2014. Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. Biomed. Eng. Appl. Basis Commun. 26, 1450021.

(Gama et al., 2012) Gama, J., Carvalho, A., Oliveira, M., Faceli, K., Lorena, A., 2012. Extração de Conhecimento de Dados, Data Mining. JC Gama Extração Conhecimento Dados Data Min. 101.

(Ghosh-Dastidar et al., 2008) Ghosh-Dastidar, S., Adeli, H., Dadmehr, N., 2008. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. Biomed. Eng. IEEE Trans. On 55, 512–518.

(Gomes and Batista, 2015a), Gomes, E.F., Batista, F., 2015a. Classifying Urban Sounds using Time Series Motifs. Adv. Sci. Technol. Lett. 97, 52–57.

Gomes and Batista, 2015b) Gomes, E.F., Batista, F., 2015b. Using Multiresolution Time Series Motifs to Classify Urban Sounds. Int. J. Softw. Eng. Its Appl. 9, 189–196.

(Gomes et al., 2014) Gomes, E.F., Jorge, A.M., Azevedo, P.J., 2014. Classifying heart sounds

using SAX motifs, random forests and text mining techniques. Presented at the Proceedings of the 18th International Database Engineering & Applications Symposium, ACM, pp. 334–337.

(Gomes et al., 2013) Gomes, E.F., Jorge, A.M., Azevedo, P.J., 2013. Classifying heart sounds using multiresolution time series motifs: an exploratory study. Presented at the Proceedings of the International C* Conference on Computer Science and Software Engineering, ACM, pp. 23–30.

(Gomide et al., 1995) Gomide, F., Gudwin, R.R., Tanscheit, R., 1995. Conceitos fundamentais da teoria de conjuntos fuzzy, lógica fuzzy e aplicações. Presented at the Proc. 6 th IFSA Congress-Tutorials, pp. 1–38.

(Hearst et al., 1998) Hearst, M.A., Dumais, S.T., Osman, E., Platt, J., Scholkopf, B., 1998. Support vector machines. Intell. Syst. Their Appl. IEEE 13, 18–28.

(Howbert et al., 2014) Howbert, J.J., Patterson, E.E., Stead, S.M., Brinkmann, B., Vasoli, V., Crepeau, D., Vite, C.H., Sturges, B., Ruedebusch, V., Mavoori, J., 2014. Forecasting seizures in dogs with naturally occurring epilepsy. PloS One 9, e81920.

(Hung et al., 2010) Hung, S.-H., Chao, C.-F., Wang, S.-K., Lin, B.-S., Lin, C.-T., 2010. VLSI implementation for epileptic seizure prediction system based on wavelet and chaos theory. Presented at the TENCON 2010-2010 IEEE Region 10 Conference, IEEE, pp. 364–368.

(Java Essentials, 2016) Java Essentials [WWW Document], 2016. URL http://www.oracle.com/technetwork/java/compile-136656.html#platform (accessed 2.16.16).

(Kaggle.com, 2016) Kaggle.com [WWW Document], 2016. . Am. Epilepsy Soc. Seizure Predict. Chall. Kaggle. URL https://www.kaggle.com/c/seizure-prediction (accessed 2.18.16).

(Khamis et al., 2013) Khamis, H., Mohamed, A., Simpson, S., 2013. Frequency–moment signatures: a method for automated seizure detection from scalp EEG. Clin. Neurophysiol. 124, 2317–2327.

(Khan et al., 2012) Khan, Y.U., Rafiuddin, N., Farooq, O., 2012. Automated seizure detection in scalp EEG using multiple wavelet scales. Presented at the Signal Processing, Computing and Control (ISPCC), 2012 IEEE International Conference on, IEEE, pp. 1–5.

(Lee, 2014) Lee, P., 2014. Resampling Methods Improve the Predictive Power of Modeling in Class-Imbalanced Datasets. Int. J. Environ. Res. Public. Health 11, 9776–9789. doi:10.3390/ijerph110909776

(Liou and Tzeng, 2010) Li, S., Zhou, W., Yuan, Q., Liu, Y., 2013. Seizure prediction using spike rate of intracranial EEG. Neural Syst. Rehabil. Eng. IEEE Trans. On 21, 880–886.

(Li et al., 2013) Liou, J.J.H., Tzeng, G.-H., 2010. A Dominance-based Rough Set Approach to customer behavior in the airline market. Inf. Sci. 180, 2230–2238.

(Liu, 2007) Liu, B., 2007. Web data mining: exploring hyperlinks, contents, and usage data. Springer Science & Business Media.

(Liu et al., 2012) Liu, Y., Zhou, W., Yuan, Q., Chen, S., 2012. Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG. Neural Syst. Rehabil. Eng. IEEE Trans. On 20, 749–755.

(Miri and Nasrabadi, 2011) Miri, S.M.R., Nasrabadi, A.M., 2011. A new seizure prediction method based on return map. Presented at the Biomedical Engineering (ICBME), 2011 18th Iranian Conference of, IEEE, pp. 244–248.

(Moghim and Corne, 2011) Moghim, N., Corne, D., 2011. Evaluating bio-inspired approaches for advance prediction of epileptic seizures. Presented at the Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on, IEEE, pp. 540–545.

(Mormann, 2008) Mormann, F., 2008. Seizure prediction. Scholarpedia 3, 5770.

(Mormann et al., 2007) Mormann, F., Andrzejak, R.G., Elger, C.E., Lehnertz, K., 2007. Seizure prediction: the long and winding road. Brain 130, 314–333.

(Morte, 2013) Morte, R.C.G., 2013. Modelo de Apoio à Decisão Multicritério para a Avaliação de Desempenho de Motoristas numa Empresa Portuguesa de Transporte Rodoviário.

(Mota and Almeida, 2007) Mota, C.M. de M., Almeida, A.T. de, 2007. Método multicritério ELECTRE IV-H para priorização de atividades em projetos. Pesqui. Oper. 27, 247–269.

(NetBeans, 2016) NetBeans [WWW Document], 2016. URL https://netbeans.org/index_pt_PT.html (accessed 2.17.16).

(Nunes de Oliveira and Rosado, 2004). Nunes de Oliveira, S., Rosado, P., 2004. Electroencefalograma Interictal. Sensibilidade e Especificidade no Diagnóstico de Epilepsia. Acta Médica Port. 465–470.

(Panda et al., 2010) Panda, R., Khobragade, P., Jambhule, P., Jengthe, S., Pal, P., Gandhi, T., 2010. Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction. Presented at the Systems in Medicine and Biology (ICSMB), 2010 International Conference on, IEEE, pp. 405–408.

(Park et al., 2011) Park, Y., Luo, L., Parhi, K.K., Netoff, T., 2011. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines: Seizure Prediction with Spectral Power of EEG. Epilepsia 52, 1761–1770. doi:10.1111/j.1528-1167.2011.03138.x

(Peng et al., 2002) Peng, C.-Y.J., Lee, K.L., Ingersoll, G.M., 2002. An Introduction to Logistic Regression Analysis and Reporting. J. Educ. Res. 96, 3–14. doi:10.1080/00220670209598786

(Platt, 1998) Platt, J., 1998. Sequential minimal optimization: A fast algorithm for training support vector machines.

(Qi et al., 2012) Qi, Y., Wang, Y., Zheng, X., Zhang, J., Zhu, J., Guo, J., 2012. Efficient epileptic seizure detection by a combined IMF-VoE feature. Presented at the Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, IEEE, pp. 5170–5173.

(Rana et al., 2012) Rana, P., Lipor, J., Lee, H., Van Drongelen, W., Kohrman, M.H., Van Veen, B., 2012. Seizure detection using the phase-slope index and multichannel ECoG. Biomed. Eng. IEEE Trans. On 59, 1125–1134.

(Random Forests, 2016). Random Forests [WWW Document], 2016. URL https://www.stat.berkeley.edu/~breiman/RandomForests/ (accessed 2.19.16).

(Runarsson and Sigurdsson, 2005) Runarsson, T.P., Sigurdsson, S., 2005. On-line detection of patient specific neonatal seizures using support vector machines and half-wave attribute histograms. Presented at the null, IEEE, pp. 673–677.

(Sartini et al., 2004) Sartini, B.A., Garbugio, G., Bortolossi, H.J., Santos, P.A., Barreto, L.S., 2004. Uma introdução à teoria dos jogos. II Bien. SBM–Universidade Fed. Bahia.

(Schelter et al., 2011) Schelter, B., Feldwisch-Drentrup, H., Ihle, M., Schulze-Bonhage, A., Timmer, J., 2011. Seizure prediction in epilepsy: From circadian concepts via probabilistic forecasting to statistical evaluation. IEEE, pp. 1624–1627. doi:10.1109/IEMBS.2011.6090469

(Shahid et al., 2013) Shahid, A., Kamel, N., Malik, A.S., Jatoi, M.A., 2013. Epileptic Seizure Detection using the singular values of EEG signals. Presented at the Complex Medical Engineering (CME), 2013 ICME International Conference on, IEEE, pp. 652–655.

(Shahidi Zandi et al., 2013) Shahidi Zandi, A., Tafreshi, R., Javidan, M., Dumont, G.A., 2013. Predicting epileptic seizures in scalp EEG based on a variational Bayesian Gaussian mixture model of zero-crossing intervals. Biomed. Eng.

IEEE Trans. On 60, 1401–1413.

(Shieh and Keogh, 2008) Shieh, J., Keogh, E., 2008. i SAX: indexing and mining terabyte sized time series. Presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 623–631.

(Shieh et al., 2006) Shieh, J.-M., Lou, D.-C., Chang, M.-C., 2006. A semi-blind digital watermarking scheme based on singular value decomposition. Comput. Stand. Interfaces 28, 428–440.

(Skålén et al., 2015) Skålén, P., Gummerus, J., Koskull, C., Magnusson, P., 2015. Exploring value propositions and service innovation: a service-dominant logic study. J. Acad. Mark. Sci. 43, 137–158. doi:10.1007/s11747-013-0365-2

(Soleimani-B et al., 2012) Soleimani-B, H., Lucas, C., Araabi, B.N., Schwabe, L., 2012. Adaptive prediction of epileptic seizures from intracranial recordings. Biomed. Signal Process. Control 7, 456–464.

(Stacey and Litt, 2008) Stacey, W.C., Litt, B., 2008. Technology insight: neuroengineering and epilepsy—designing devices for seizure control. Nat. Clin. Pract. Neurol. 4, 190–201.

(Subasi and Ercelebi, 2005) Subasi, A., Ercelebi, E., 2005. Classification of EEG signals using neural network and logistic regression. Comput. Methods Programs Biomed. 78, 87–99.

(Tafreshi et al., 2008) Tafreshi, A.K., Nasrabadi, A.M., Omidvarnia, A.H., 2008. Epileptic seizure detection using empirical mode decomposition. Presented at the Signal Processing and Information Technology, 2008. ISSPIT 2008. IEEE International Symposium on, IEEE, pp. 238–242.

(Vanrumste et al., 2002) Vanrumste, B., Jones, R.D., Bones, P.H.J., 2002. Detection of focal epileptiform activity in the EEG: an SVD and dipole model approach. Presented at the Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint, IEEE, pp. 2031–2032.

(Wang et al., 2010) Wang, S., Chaovalitwongse, W.A., Wong, S., 2010. A novel reinforcement learning framework for online adaptive seizure prediction. Presented at the Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on, IEEE, pp. 499–504.

(Witten et al., 2011) Witten, I.H., Frank, E., Hall, M.A., 2011. Data mining: practical machine learning tools and techniques, 3rd ed. ed, Morgan Kaufmann series in data management systems. Morgan Kaufmann, Burlington, MA.

(Wong, 2015) Wong, T.-T., 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit. 48, 2839–2846. doi:10.1016/j.patcog.2015.03.009

(Woodall, 2003) Woodall, T., 2003. Conceptualising “value for the customer”: an attributional, structural and dispositional analysis. Acad. Mark. Sci. Rev. 12, 1–42.

(Xie and Krishnan, 2011) Xie, S., Krishnan, S., 2011. Signal decomposition by multi-scale PCA and its applications to long-term EEG signal classification. Presented at the Complex Medical Engineering (CME), 2011 IEEE/ICME International Conference on, IEEE, pp. 532–537.

(Yoo et al., 2013) Yoo, J., Yan, L., El-Damak, D., Altaf, M.A.B., Shoeb, A.H., Chandrakasan, A.P., 2013. An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor. Solid-State Circuits IEEE J. Of 48, 214–228.

(Zandi et al., 2010) Zandi, A.S., Tafreshi, R., Javidan, M., Dumont, G.A., 2010. Predicting temporal lobe epileptic seizures based on zero-crossing interval

analysis in scalp EEG. Presented at the Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, IEEE, pp. 5537–5540.

(Zhou et al., 2013) Zhou, W., Liu, Y., Yuan, Q., Li, X., 2013. Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG. Biomed. Eng. IEEE Trans. On 60, 3375–3381.

# Anexos

Tabela com resultados ROC Area (Parâmetros 30 10 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 30 10 2 2 10 | 0 | 0,450 | 0,572 | 0,500 | 0,613 |
| Cão 1 | 30 10 2 2 10 | 1 | 0,688 | 0,935 | 0,500 | 0,825 |
| Cão 1 | 30 10 2 2 10 | 2 | 0,707 | 0,952 | 0,500 | 0,734 |
| Cão 1 | 30 10 2 2 10 | 3 | 0,787 | 0,958 | 0,500 | 0,781 |
| Cão 1 | 30 10 2 2 10 | 4 | 0,822 | 0,992 | 0,500 | 0,750 |
| Cão 1 | 30 10 2 2 10 | 5 | 0,866 | 0,950 | 0,500 | 0,727 |
| Cão 2 | 30 10 2 2 10 | 0 | 0,530 | 0,673 | 0,500 | 0,416 |
| Cão 2 | 30 10 2 2 10 | 1 | 0,767 | 0,939 | 0,500 | 0,592 |
| Cão 2 | 30 10 2 2 10 | 2 | 0,721 | 0,927 | 0,500 | 0,505 |
| Cão 2 | 30 10 2 2 10 | 3 | 0,897 | 0,985 | 0,500 | 0,659 |
| Cão 2 | 30 10 2 2 10 | 4 | 0,955 | 0,995 | 0,500 | 0,743 |
| Cão 2 | 30 10 2 2 10 | 5 | 0,954 | 0,997 | 0,500 | 0,743 |
| Paciente 1 | 30 10 2 2 10 | 0 | 0,650 | 0,847 | 0,639 | 0,737 |
| Paciente 1 | 30 10 2 2 10 | 1 | 0,929 | 0,967 | 0,667 | 0,701 |
| Paciente 1 | 30 10 2 2 10 | 2 | 0,877 | 0,978 | 0,667 | 0,950 |
| Paciente 1 | 30 10 2 2 10 | 3 | 0,866 | 0,964 | 0,722 | 0,942 |
| Paciente 1 | 30 10 2 2 10 | 4 | 0,939 | 0,966 | 0,722 | 0,980 |
| Paciente 1 | 30 10 2 2 10 | 5 | 0,981 | 0,996 | 0,750 | 0,949 |
| Paciente 2 | 30 10 2 2 10 | 0 | 0,646 | 0,757 | 0,488 | 0,630 |
| Paciente 2 | 30 10 2 2 10 | 1 | 0,936 | 0,941 | 0,500 | 0,759 |
| Paciente 2 | 30 10 2 2 10 | 2 | 0,948 | 0,996 | 0,500 | 0,859 |
| Paciente 2 | 30 10 2 2 10 | 3 | 0,907 | 0,950 | 0,500 | 0,860 |
| Paciente 2 | 30 10 2 2 10 | 4 | 0,911 | 0,988 | 0,500 | 0,871 |
| Paciente 2 | 30 10 2 2 10 | 5 | 0,953 | 0,962 | 0,500 | 0,925 |

Tabela com resultados ROC Area (Parâmetros 30 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 30 20 2 2 10 | 0 | 0,450 | 0,613 | 0,500 | 0,653 |
| Cão 1 | 30 20 2 2 10 | 1 | 0,731 | 0,882 | 0,500 | 0,793 |
| Cão 1 | 30 20 2 2 10 | 2 | 0,651 | 0,914 | 0,500 | 0,665 |
| Cão 1 | 30 20 2 2 10 | 3 | 0,846 | 0,951 | 0,500 | 0,803 |
| Cão 1 | 30 20 2 2 10 | 4 | 0,817 | 0,994 | 0,500 | 0,822 |
| Cão 1 | 30 20 2 2 10 | 5 | 0,861 | 0,917 | 0,500 | 0,740 |
| Cão 2 | 30 20 2 2 10 | 0 | 0,511 | 0,529 | 0,500 | 0,386 |
| Cão 2 | 30 20 2 2 10 | 1 | 0,753 | 0,955 | 0,500 | 0,567 |
| Cão 2 | 30 20 2 2 10 | 2 | 0,873 | 0,945 | 0,500 | 0,616 |
| Cão 2 | 30 20 2 2 10 | 3 | 0,939 | 0,978 | 0,500 | 0,707 |
| Cão 2 | 30 20 2 2 10 | 4 | 0,949 | 0,999 | 0,500 | 0,737 |
| Cão 2 | 30 20 2 2 10 | 5 | 0,937 | 0,998 | 0,500 | 0,774 |
| Paciente 1 | 30 20 2 2 10 | 0 | 0,772 | 0,846 | 0,639 | 0,596 |
| Paciente 1 | 30 20 2 2 10 | 1 | 0,881 | 0,975 | 0,667 | 0,775 |
| Paciente 1 | 30 20 2 2 10 | 2 | 0,886 | 0,972 | 0,667 | 0,907 |
| Paciente 1 | 30 20 2 2 10 | 3 | 0,787 | 0,917 | 0,722 | 0,907 |
| Paciente 1 | 30 20 2 2 10 | 4 | 0,939 | 0,986 | 0,722 | 0,950 |
| Paciente 1 | 30 20 2 2 10 | 5 | 0,964 | 0,997 | 0,750 | 0,935 |
| Paciente 2 | 30 20 2 2 10 | 0 | 0,646 | 0,789 | 0,516 | 0,572 |
| Paciente 2 | 30 20 2 2 10 | 1 | 0,888 | 0,929 | 0,500 | 0,773 |
| Paciente 2 | 30 20 2 2 10 | 2 | 0,919 | 0,997 | 0,500 | 0,927 |
| Paciente 2 | 30 20 2 2 10 | 3 | 0,919 | 0,970 | 0,500 | 0,917 |
| Paciente 2 | 30 20 2 2 10 | 4 | 0,954 | 0,990 | 0,500 | 0,917 |
| Paciente 2 | 30 20 2 2 10 | 5 | 0,953 | 0,987 | 0,500 | 0,957 |

Tabela com resultados ROC Area (Parâmetros 40 10 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 10 2 2 10 | 0 | 0,455 | 0,535 | 0,500 | 0,373 |
| Cão 1 | 40 10 2 2 10 | 1 | 0,649 | 0,903 | 0,500 | 0,642 |
| Cão 1 | 40 10 2 2 10 | 2 | 0,567 | 0,902 | 0,500 | 0,588 |
| Cão 1 | 40 10 2 2 10 | 3 | 0,779 | 0,934 | 0,500 | 0,532 |
| Cão 1 | 40 10 2 2 10 | 4 | 0,936 | 0,996 | 0,500 | 0,427 |
| Cão 1 | 40 10 2 2 10 | 5 | 0,852 | 0,903 | 0,500 | 0,396 |
| Cão 2 | 40 10 2 2 10 | 0 | 0,532 | 0,703 | 0,500 | 0,311 |
| Cão 2 | 40 10 2 2 10 | 1 | 0,652 | 0,922 | 0,500 | 0,602 |
| Cão 2 | 40 10 2 2 10 | 2 | 0,840 | 0,934 | 0,500 | 0,600 |
| Cão 2 | 40 10 2 2 10 | 3 | 0,915 | 0,987 | 0,500 | 0,658 |
| Cão 2 | 40 10 2 2 10 | 4 | 0,928 | 0,992 | 0,500 | 0,706 |
| Cão 2 | 40 10 2 2 10 | 5 | 0,958 | 0,996 | 0,500 | 0,748 |
| Paciente 1 | 40 10 2 2 10 | 0 | 0,728 | 0,769 | 0,601 | 0,648 |
| Paciente 1 | 40 10 2 2 10 | 1 | 0,708 | 0,917 | 0,657 | 0,662 |
| Paciente 1 | 40 10 2 2 10 | 2 | 0,780 | 0,889 | 0,667 | 0,814 |
| Paciente 1 | 40 10 2 2 10 | 3 | 0,844 | 0,867 | 0,722 | 0,771 |
| Paciente 1 | 40 10 2 2 10 | 4 | 0,919 | 0,954 | 0,722 | 0,958 |
| Paciente 1 | 40 10 2 2 10 | 5 | 0,990 | 0,998 | 0,750 | 0,985 |
| Paciente 2 | 40 10 2 2 10 | 0 | 0,479 | 0,719 | 0,476 | 0,389 |
| Paciente 2 | 40 10 2 2 10 | 1 | 0,684 | 0,942 | 0,500 | 0,671 |
| Paciente 2 | 40 10 2 2 10 | 2 | 0,956 | 0,981 | 0,500 | 0,731 |
| Paciente 2 | 40 10 2 2 10 | 3 | 0,897 | 0,978 | 0,500 | 0,853 |
| Paciente 2 | 40 10 2 2 10 | 4 | 0,937 | 0,992 | 0,500 | 0,892 |
| Paciente 2 | 40 10 2 2 10 | 5 | 0,899 | 0,964 | 0,500 | 0,956 |

Tabela com resultados ROC Area (Parâmetros 40 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 20 2 2 10 | 0 | 0,531 | 0,631 | 0,500 | 0,396 |
| Cão 1 | 40 20 2 2 10 | 1 | 0,800 | 0,937 | 0,500 | 0,507 |
| Cão 1 | 40 20 2 2 10 | 2 | 0,759 | 0,937 | 0,500 | 0,600 |
| Cão 1 | 40 20 2 2 10 | 3 | 0,916 | 0,946 | 0,500 | 0,694 |
| Cão 1 | 40 20 2 2 10 | 4 | 0,825 | 0,980 | 0,500 | 0,580 |
| Cão 1 | 40 20 2 2 10 | 5 | 0,900 | 0,961 | 0,500 | 0,634 |
| Cão 2 | 40 20 2 2 10 | 0 | 0,549 | 0,659 | 0,500 | 0,236 |
| Cão 2 | 40 20 2 2 10 | 1 | 0,727 | 0,931 | 0,500 | 0,629 |
| Cão 2 | 40 20 2 2 10 | 2 | 0,821 | 0,932 | 0,500 | 0,683 |
| Cão 2 | 40 20 2 2 10 | 3 | 0,908 | 0,988 | 0,500 | 0,725 |
| Cão 2 | 40 20 2 2 10 | 4 | 0,930 | 0,993 | 0,500 | 0,861 |
| Cão 2 | 40 20 2 2 10 | 5 | 0,974 | 0,999 | 0,500 | 0,877 |
| Paciente 1 | 40 20 2 2 10 | 0 | 0,572 | 0,756 | 0,601 | 0,551 |
| Paciente 1 | 40 20 2 2 10 | 1 | 0,871 | 0,946 | 0,657 | 0,749 |
| Paciente 1 | 40 20 2 2 10 | 2 | 0,867 | 0,934 | 0,667 | 0,931 |
| Paciente 1 | 40 20 2 2 10 | 3 | 0,979 | 0,902 | 0,722 | 0,825 |
| Paciente 1 | 40 20 2 2 10 | 4 | 0,912 | 0,953 | 0,722 | 0,902 |
| Paciente 1 | 40 20 2 2 10 | 5 | 0,983 | 0,996 | 0,750 | 0,934 |
| Paciente 2 | 40 20 2 2 10 | 0 | 0,569 | 0,667 | 0,492 | 0,344 |
| Paciente 2 | 40 20 2 2 10 | 1 | 0,744 | 0,949 | 0,500 | 0,711 |
| Paciente 2 | 40 20 2 2 10 | 2 | 0,899 | 0,979 | 0,500 | 0,888 |
| Paciente 2 | 40 20 2 2 10 | 3 | 0,854 | 0,974 | 0,500 | 0,929 |
| Paciente 2 | 40 20 2 2 10 | 4 | 0,898 | 0,987 | 0,500 | 0,888 |
| Paciente 2 | 40 20 2 2 10 | 5 | 0,937 | 0,963 | 0,500 | 1,000 |

Tabela com resultados ROC Area (Parâmetros 40 20 2 2 20)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 20 2 2 20 | 0 | 0,461 | 0,495 | 0,500 | 0,507 |
| Cão 1 | 40 20 2 2 20 | 1 | 0,592 | 0,911 | 0,500 | 0,685 |
| Cão 1 | 40 20 2 2 20 | 2 | 0,690 | 0,937 | 0,500 | 0,658 |
| Cão 1 | 40 20 2 2 20 | 3 | 0,668 | 0,917 | 0,500 | 0,776 |
| Cão 1 | 40 20 2 2 20 | 4 | 0,826 | 0,993 | 0,500 | 0,770 |
| Cão 1 | 40 20 2 2 20 | 5 | 0,865 | 0,946 | 0,500 | 0,837 |
| Cão 2 | 40 20 2 2 20 | 0 | 0,565 | 0,562 | 0,500 | 0,287 |
| Cão 2 | 40 20 2 2 20 | 1 | 0,785 | 0,922 | 0,500 | 0,619 |
| Cão 2 | 40 20 2 2 20 | 2 | 0,840 | 0,943 | 0,500 | 0,640 |
| Cão 2 | 40 20 2 2 20 | 3 | 0,925 | 0,971 | 0,500 | 0,720 |
| Cão 2 | 40 20 2 2 20 | 4 | 0,911 | 0,995 | 0,500 | 0,840 |
| Cão 2 | 40 20 2 2 20 | 5 | 0,974 | 0,993 | 0,500 | 0,934 |
| Paciente 1 | 40 20 2 2 20 | 0 | 0,517 | 0,764 | 0,601 | 0,593 |
| Paciente 1 | 40 20 2 2 20 | 1 | 0,845 | 0,961 | 0,657 | 0,705 |
| Paciente 1 | 40 20 2 2 20 | 2 | 0,907 | 0,978 | 0,667 | 0,909 |
| Paciente 1 | 40 20 2 2 20 | 3 | 0,741 | 0,904 | 0,694 | 0,878 |
| Paciente 1 | 40 20 2 2 20 | 4 | 0,903 | 0,958 | 0,722 | 0,910 |
| Paciente 1 | 40 20 2 2 20 | 5 | 0,962 | 0,995 | 0,722 | 0,997 |
| Paciente 2 | 40 20 2 2 20 | 0 | 0,501 | 0,620 | 0,476 | 0,352 |
| Paciente 2 | 40 20 2 2 20 | 1 | 0,706 | 0,935 | 0,500 | 0,721 |
| Paciente 2 | 40 20 2 2 20 | 2 | 0,882 | 0,981 | 0,500 | 0,910 |
| Paciente 2 | 40 20 2 2 20 | 3 | 0,836 | 0,968 | 0,500 | 0,917 |
| Paciente 2 | 40 20 2 2 20 | 4 | 0,907 | 0,987 | 0,500 | 0,917 |
| Paciente 2 | 40 20 2 2 20 | 5 | 0,929 | 0,982 | 0,500 | 1,000 |

Tabela com resultados ROC Area (Parâmetros 40 30 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 30 2 2 10 | 0 | 0,452 | 0,555 | 0,500 | 0,270 |
| Cão 1 | 40 30 2 2 10 | 1 | 0,681 | 0,904 | 0,500 | 0,686 |
| Cão 1 | 40 30 2 2 10 | 2 | 0,802 | 0,938 | 0,500 | 0,619 |
| Cão 1 | 40 30 2 2 10 | 3 | 0,884 | 0,956 | 0,500 | 0,719 |
| Cão 1 | 40 30 2 2 10 | 4 | 0,927 | 0,996 | 0,500 | 0,761 |
| Cão 1 | 40 30 2 2 10 | 5 | 0,853 | 0,948 | 0,500 | 0,813 |
| Cão 2 | 40 30 2 2 10 | 0 | 0,491 | 0,654 | 0,500 | 0,171 |
| Cão 2 | 40 30 2 2 10 | 1 | 0,734 | 0,934 | 0,500 | 0,655 |
| Cão 2 | 40 30 2 2 10 | 2 | 0,762 | 0,952 | 0,500 | 0,720 |
| Cão 2 | 40 30 2 2 10 | 3 | 0,882 | 0,988 | 0,500 | 0,819 |
| Cão 2 | 40 30 2 2 10 | 4 | 0,920 | 0,990 | 0,500 | 0,938 |
| Cão 2 | 40 30 2 2 10 | 5 | 0,950 | 0,999 | 0,500 | 0,967 |
| Paciente 1 | 40 30 2 2 10 | 0 | 0,801 | 0,784 | 0,583 | 0,678 |
| Paciente 1 | 40 30 2 2 10 | 1 | 0,920 | 0,933 | 0,629 | 0,762 |
| Paciente 1 | 40 30 2 2 10 | 2 | 0,869 | 0,948 | 0,667 | 0,840 |
| Paciente 1 | 40 30 2 2 10 | 3 | 0,941 | 0,888 | 0,702 | 0,875 |
| Paciente 1 | 40 30 2 2 10 | 4 | 0,938 | 0,958 | 0,712 | 0,938 |
| Paciente 1 | 40 30 2 2 10 | 5 | 0,981 | 0,986 | 0,740 | 0,984 |
| Paciente 2 | 40 30 2 2 10 | 0 | 0,553 | 0,673 | 0,476 | 0,473 |
| Paciente 2 | 40 30 2 2 10 | 1 | 0,899 | 0,968 | 0,500 | 0,792 |
| Paciente 2 | 40 30 2 2 10 | 2 | 0,899 | 0,975 | 0,500 | 0,903 |
| Paciente 2 | 40 30 2 2 10 | 3 | 0,886 | 0,974 | 0,500 | 0,942 |
| Paciente 2 | 40 30 2 2 10 | 4 | 0,892 | 0,979 | 0,500 | 0,887 |
| Paciente 2 | 40 30 2 2 10 | 5 | 0,892 | 0,979 | 0,500 | 0,892 |

Tabela com resultados ROC Area (Parâmetros 40 30 2 2 20)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 30 2 2 20 | 0 | 0,465 | 0,522 | 0,500 | 0,445 |
| Cão 1 | 40 30 2 2 20 | 1 | 0,592 | 0,879 | 0,500 | 0,720 |
| Cão 1 | 40 30 2 2 20 | 2 | 0,718 | 0,937 | 0,500 | 0,735 |
| Cão 1 | 40 30 2 2 20 | 3 | 0,669 | 0,925 | 0,500 | 0,861 |
| Cão 1 | 40 30 2 2 20 | 4 | 0,821 | 0,986 | 0,500 | 0,901 |
| Cão 1 | 40 30 2 2 20 | 5 | 0,859 | 0,941 | 0,500 | 0,898 |
| Cão 2 | 40 30 2 2 20 | 0 | 0,563 | 0,588 | 0,500 | 0,207 |
| Cão 2 | 40 30 2 2 20 | 1 | 0,746 | 0,935 | 0,500 | 0,645 |
| Cão 2 | 40 30 2 2 20 | 2 | 0,818 | 0,949 | 0,500 | 0,664 |
| Cão 2 | 40 30 2 2 20 | 3 | 0,873 | 0,978 | 0,500 | 0,779 |
| Cão 2 | 40 30 2 2 20 | 4 | 0,945 | 0,994 | 0,500 | 0,875 |
| Cão 2 | 40 30 2 2 20 | 5 | 0,962 | 0,990 | 0,500 | 0,951 |
| Paciente 1 | 40 30 2 2 20 | 0 | 0,466 | 0,736 | 0,546 | 0,550 |
| Paciente 1 | 40 30 2 2 20 | 1 | 0,859 | 0,952 | 0,629 | 0,739 |
| Paciente 1 | 40 30 2 2 20 | 2 | 0,907 | 0,986 | 0,667 | 0,960 |
| Paciente 1 | 40 30 2 2 20 | 3 | 0,823 | 0,902 | 0,674 | 0,876 |
| Paciente 1 | 40 30 2 2 20 | 4 | 0,949 | 0,963 | 0,712 | 0,883 |
| Paciente 1 | 40 30 2 2 20 | 5 | 0,973 | 0,993 | 0,796 | 0,932 |
| Paciente 2 | 40 30 2 2 20 | 0 | 0,629 | 0,651 | 0,504 | 0,457 |
| Paciente 2 | 40 30 2 2 20 | 1 | 0,810 | 0,966 | 0,500 | 0,835 |
| Paciente 2 | 40 30 2 2 20 | 2 | 0,923 | 0,979 | 0,500 | 0,909 |
| Paciente 2 | 40 30 2 2 20 | 3 | 0,827 | 0,969 | 0,500 | 0,930 |
| Paciente 2 | 40 30 2 2 20 | 4 | 0,829 | 0,979 | 0,500 | 0,937 |
| Paciente 2 | 40 30 2 2 20 | 5 | 0,929 | 0,987 | 0,500 | 1,000 |

Tabela com resultados ROC Area (Parâmetros 60 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 60 20 2 2 10 | 0 | 0,459 | 0,566 | 0,500 | 0,512 |
| Cão 1 | 60 20 2 2 10 | 1 | 0,604 | 0,914 | 0,500 | 0,670 |
| Cão 1 | 60 20 2 2 10 | 2 | 0,829 | 0,973 | 0,500 | 0,805 |
| Cão 1 | 60 20 2 2 10 | 3 | 0,845 | 0,960 | 0,500 | 0,805 |
| Cão 1 | 60 20 2 2 10 | 4 | 0,940 | 0,948 | 0,500 | 0,904 |
| Cão 1 | 60 20 2 2 10 | 5 | 0,890 | 0,992 | 0,500 | 0,891 |
| Cão 2 | 60 20 2 2 10 | 0 | 0,523 | 0,669 | 0,500 | 0,509 |
| Cão 2 | 60 20 2 2 10 | 1 | 0,740 | 0,930 | 0,500 | 0,642 |
| Cão 2 | 60 20 2 2 10 | 2 | 0,906 | 0,961 | 0,500 | 0,716 |
| Cão 2 | 60 20 2 2 10 | 3 | 0,898 | 0,963 | 0,500 | 0,784 |
| Cão 2 | 60 20 2 2 10 | 4 | 0,961 | 0,998 | 0,500 | 0,908 |
| Cão 2 | 60 20 2 2 10 | 5 | 0,981 | 0,992 | 0,500 | 0,922 |
| Paciente 1 | 60 20 2 2 10 | 0 | 0,613 | 0,819 | 0,611 | 0,643 |
| Paciente 1 | 60 20 2 2 10 | 1 | 0,863 | 0,961 | 0,667 | 0,838 |
| Paciente 1 | 60 20 2 2 10 | 2 | 0,861 | 0,946 | 0,667 | 0,949 |
| Paciente 1 | 60 20 2 2 10 | 3 | 0,902 | 0,906 | 0,722 | 0,964 |
| Paciente 1 | 60 20 2 2 10 | 4 | 0,907 | 0,963 | 0,722 | 0,992 |
| Paciente 1 | 60 20 2 2 10 | 5 | 0,978 | 1,000 | 0,750 | 0,982 |
| Paciente 2 | 60 20 2 2 10 | 0 | 0,414 | 0,505 | 0,476 | 0,468 |
| Paciente 2 | 60 20 2 2 10 | 1 | 0,723 | 0,933 | 0,500 | 0,762 |
| Paciente 2 | 60 20 2 2 10 | 2 | 0,878 | 0,948 | 0,500 | 0,887 |
| Paciente 2 | 60 20 2 2 10 | 3 | 0,894 | 0,970 | 0,500 | 0,977 |
| Paciente 2 | 60 20 2 2 10 | 4 | 0,953 | 0,973 | 0,500 | 0,936 |
| Paciente 2 | 60 20 2 2 10 | 5 | 0,875 | 0,995 | 0,500 | 0,983 |

Tabela com resultados *Accuracy* (Parâmetros 30 10 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 30 10 2 2 10 | 0 | 0,952 | 0,952 | 0,952 | 0,952 |
| Cão 1 | 30 10 2 2 10 | 1 | 0,956 | 0,984 | 0,952 | 0,960 |
| Cão 1 | 30 10 2 2 10 | 2 | 0,964 | 0,986 | 0,952 | 0,956 |
| Cão 1 | 30 10 2 2 10 | 3 | 0,968 | 0,986 | 0,952 | 0,948 |
| Cão 1 | 30 10 2 2 10 | 4 | 0,972 | 0,990 | 0,952 | 0,948 |
| Cão 1 | 30 10 2 2 10 | 5 | 0,986 | 0,986 | 0,952 | 0,952 |
| Cão 2 | 30 10 2 2 10 | 0 | 0,921 | 0,921 | 0,923 | 0,923 |
| Cão 2 | 30 10 2 2 10 | 1 | 0,946 | 0,967 | 0,923 | 0,923 |
| Cão 2 | 30 10 2 2 10 | 2 | 0,937 | 0,980 | 0,923 | 0,923 |
| Cão 2 | 30 10 2 2 10 | 3 | 0,982 | 0,989 | 0,923 | 0,923 |
| Cão 2 | 30 10 2 2 10 | 4 | 0,980 | 0,994 | 0,923 | 0,919 |
| Cão 2 | 30 10 2 2 10 | 5 | 0,989 | 0,994 | 0,923 | 0,919 |
| Paciente 1 | 30 10 2 2 10 | 0 | 0,691 | 0,779 | 0,809 | 0,721 |
| Paciente 1 | 30 10 2 2 10 | 1 | 0,882 | 0,897 | 0,824 | 0,750 |
| Paciente 1 | 30 10 2 2 10 | 2 | 0,868 | 0,912 | 0,824 | 0,882 |
| Paciente 1 | 30 10 2 2 10 | 3 | 0,868 | 0,912 | 0,853 | 0,926 |
| Paciente 1 | 30 10 2 2 10 | 4 | 0,926 | 0,971 | 0,853 | 0,971 |
| Paciente 1 | 30 10 2 2 10 | 5 | 0,985 | 0,985 | 0,868 | 0,956 |
| Paciente 2 | 30 10 2 2 10 | 0 | 0,817 | 0,800 | 0,683 | 0,650 |
| Paciente 2 | 30 10 2 2 10 | 1 | 0,950 | 0,933 | 0,700 | 0,767 |
| Paciente 2 | 30 10 2 2 10 | 2 | 0,900 | 0,983 | 0,700 | 0,900 |
| Paciente 2 | 30 10 2 2 10 | 3 | 0,900 | 0,933 | 0,700 | 0,917 |
| Paciente 2 | 30 10 2 2 10 | 4 | 0,900 | 0,950 | 0,700 | 0,933 |
| Paciente 2 | 30 10 2 2 10 | 5 | 0,917 | 0,950 | 0,700 | 0,967 |

Tabela com resultados *Accuracy* (Parâmetros 30 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 30 20 2 2 10 | 0 | 0,952 | 0,952 | 0,952 | 0,940 |
| Cão 1 | 30 20 2 2 10 | 1 | 0,960 | 0,986 | 0,952 | 0,966 |
| Cão 1 | 30 20 2 2 10 | 2 | 0,960 | 0,988 | 0,952 | 0,964 |
| Cão 1 | 30 20 2 2 10 | 3 | 0,982 | 0,990 | 0,952 | 0,960 |
| Cão 1 | 30 20 2 2 10 | 4 | 0,980 | 0,992 | 0,952 | 0,950 |
| Cão 1 | 30 20 2 2 10 | 5 | 0,982 | 0,992 | 0,952 | 0,948 |
| Cão 2 | 30 20 2 2 10 | 0 | 0,889 | 0,856 | 0,923 | 0,923 |
| Cão 2 | 30 20 2 2 10 | 1 | 0,945 | 0,970 | 0,923 | 0,923 |
| Cão 2 | 30 20 2 2 10 | 2 | 0,948 | 0,980 | 0,923 | 0,923 |
| Cão 2 | 30 20 2 2 10 | 3 | 0,970 | 0,989 | 0,923 | 0,915 |
| Cão 2 | 30 20 2 2 10 | 4 | 0,974 | 0,993 | 0,923 | 0,913 |
| Cão 2 | 30 20 2 2 10 | 5 | 0,989 | 0,994 | 0,923 | 0,913 |
| Paciente 1 | 30 20 2 2 10 | 0 | 0,765 | 0,794 | 0,809 | 0,618 |
| Paciente 1 | 30 20 2 2 10 | 1 | 0,853 | 0,912 | 0,824 | 0,750 |
| Paciente 1 | 30 20 2 2 10 | 2 | 0,882 | 0,941 | 0,824 | 0,956 |
| Paciente 1 | 30 20 2 2 10 | 3 | 0,882 | 0,912 | 0,853 | 0,926 |
| Paciente 1 | 30 20 2 2 10 | 4 | 0,926 | 0,956 | 0,853 | 0,971 |
| Paciente 1 | 30 20 2 2 10 | 5 | 0,971 | 0,971 | 0,868 | 0,956 |
| Paciente 2 | 30 20 2 2 10 | 0 | 0,817 | 0,833 | 0,700 | 0,683 |
| Paciente 2 | 30 20 2 2 10 | 1 | 0,917 | 0,917 | 0,700 | 0,850 |
| Paciente 2 | 30 20 2 2 10 | 2 | 0,883 | 0,967 | 0,700 | 0,917 |
| Paciente 2 | 30 20 2 2 10 | 3 | 0,917 | 0,950 | 0,700 | 0,950 |
| Paciente 2 | 30 20 2 2 10 | 4 | 0,933 | 0,967 | 0,700 | 0,933 |
| Paciente 2 | 30 20 2 2 10 | 5 | 0,917 | 0,950 | 0,700 | 0,983 |

Tabela com resultados *Accuracy* (Parâmetros 40 10 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 10 2 2 10 | 0 | 0,952 | 0,948 | 0,952 | 0,948 |
| Cão 1 | 40 10 2 2 10 | 1 | 0,958 | 0,984 | 0,952 | 0,950 |
| Cão 1 | 40 10 2 2 10 | 2 | 0,952 | 0,986 | 0,952 | 0,956 |
| Cão 1 | 40 10 2 2 10 | 3 | 0,958 | 0,990 | 0,952 | 0,960 |
| Cão 1 | 40 10 2 2 10 | 4 | 0,974 | 0,988 | 0,952 | 0,950 |
| Cão 1 | 40 10 2 2 10 | 5 | 0,970 | 0,986 | 0,952 | 0,952 |
| Cão 2 | 40 10 2 2 10 | 0 | 0,910 | 0,921 | 0,923 | 0,923 |
| Cão 2 | 40 10 2 2 10 | 1 | 0,923 | 0,967 | 0,923 | 0,923 |
| Cão 2 | 40 10 2 2 10 | 2 | 0,954 | 0,976 | 0,923 | 0,923 |
| Cão 2 | 40 10 2 2 10 | 3 | 0,969 | 0,989 | 0,923 | 0,923 |
| Cão 2 | 40 10 2 2 10 | 4 | 0,969 | 0,993 | 0,923 | 0,923 |
| Cão 2 | 40 10 2 2 10 | 5 | 0,985 | 0,994 | 0,923 | 0,923 |
| Paciente 1 | 40 10 2 2 10 | 0 | 0,765 | 0,779 | 0,779 | 0,721 |
| Paciente 1 | 40 10 2 2 10 | 1 | 0,765 | 0,882 | 0,809 | 0,809 |
| Paciente 1 | 40 10 2 2 10 | 2 | 0,868 | 0,897 | 0,824 | 0,868 |
| Paciente 1 | 40 10 2 2 10 | 3 | 0,912 | 0,912 | 0,853 | 0,882 |
| Paciente 1 | 40 10 2 2 10 | 4 | 0,971 | 0,941 | 0,853 | 0,985 |
| Paciente 1 | 40 10 2 2 10 | 5 | 0,985 | 0,985 | 0,868 | 0,985 |
| Paciente 2 | 40 10 2 2 10 | 0 | 0,683 | 0,733 | 0,667 | 0,533 |
| Paciente 2 | 40 10 2 2 10 | 1 | 0,767 | 0,917 | 0,700 | 0,650 |
| Paciente 2 | 40 10 2 2 10 | 2 | 0,967 | 0,967 | 0,700 | 0,767 |
| Paciente 2 | 40 10 2 2 10 | 3 | 0,900 | 0,933 | 0,700 | 0,817 |
| Paciente 2 | 40 10 2 2 10 | 4 | 0,933 | 0,967 | 0,700 | 0,950 |
| Paciente 2 | 40 10 2 2 10 | 5 | 0,917 | 0,967 | 0,700 | 0,967 |

Tabela com resultados *Accuracy* (Parâmetros 40 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 20 2 2 10 | 0 | 0,946 | 0,952 | 0,952 | 0,946 |
| Cão 1 | 40 20 2 2 10 | 1 | 0,964 | 0,986 | 0,952 | 0,944 |
| Cão 1 | 40 20 2 2 10 | 2 | 0,968 | 0,988 | 0,952 | 0,950 |
| Cão 1 | 40 20 2 2 10 | 3 | 0,974 | 0,990 | 0,952 | 0,954 |
| Cão 1 | 40 20 2 2 10 | 4 | 0,984 | 0,992 | 0,952 | 0,948 |
| Cão 1 | 40 20 2 2 10 | 5 | 0,990 | 0,992 | 0,952 | 0,958 |
| Cão 2 | 40 20 2 2 10 | 0 | 0,895 | 0,923 | 0,923 | 0,923 |
| Cão 2 | 40 20 2 2 10 | 1 | 0,924 | 0,970 | 0,923 | 0,923 |
| Cão 2 | 40 20 2 2 10 | 2 | 0,956 | 0,978 | 0,923 | 0,910 |
| Cão 2 | 40 20 2 2 10 | 3 | 0,969 | 0,989 | 0,923 | 0,917 |
| Cão 2 | 40 20 2 2 10 | 4 | 0,972 | 0,993 | 0,923 | 0,906 |
| Cão 2 | 40 20 2 2 10 | 5 | 0,994 | 0,994 | 0,923 | 0,928 |
| Paciente 1 | 40 20 2 2 10 | 0 | 0,779 | 0,765 | 0,779 | 0,603 |
| Paciente 1 | 40 20 2 2 10 | 1 | 0,868 | 0,912 | 0,809 | 0,824 |
| Paciente 1 | 40 20 2 2 10 | 2 | 0,912 | 0,912 | 0,824 | 0,941 |
| Paciente 1 | 40 20 2 2 10 | 3 | 0,941 | 0,956 | 0,853 | 0,912 |
| Paciente 1 | 40 20 2 2 10 | 4 | 0,971 | 0,941 | 0,853 | 0,956 |
| Paciente 1 | 40 20 2 2 10 | 5 | 0,985 | 0,985 | 0,868 | 0,956 |
| Paciente 2 | 40 20 2 2 10 | 0 | 0,733 | 0,800 | 0,667 | 0,500 |
| Paciente 2 | 40 20 2 2 10 | 1 | 0,833 | 0,900 | 0,700 | 0,817 |
| Paciente 2 | 40 20 2 2 10 | 2 | 0,867 | 0,933 | 0,700 | 0,950 |
| Paciente 2 | 40 20 2 2 10 | 3 | 0,850 | 0,933 | 0,700 | 0,933 |
| Paciente 2 | 40 20 2 2 10 | 4 | 0,917 | 0,933 | 0,700 | 0,950 |
| Paciente 2 | 40 20 2 2 10 | 5 | 0,917 | 0,967 | 0,700 | 1,000 |

Tabela com resultados *Accuracy* (Parâmetros 40 20 2 2 20)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 20 2 2 20 | 0 | 0,950 | 0,952 | 0,952 | 0,950 |
| Cão 1 | 40 20 2 2 20 | 1 | 0,962 | 0,986 | 0,952 | 0,946 |
| Cão 1 | 40 20 2 2 20 | 2 | 0,960 | 0,988 | 0,952 | 0,946 |
| Cão 1 | 40 20 2 2 20 | 3 | 0,958 | 0,990 | 0,952 | 0,954 |
| Cão 1 | 40 20 2 2 20 | 4 | 0,972 | 0,992 | 0,952 | 0,952 |
| Cão 1 | 40 20 2 2 20 | 5 | 0,984 | 0,990 | 0,952 | 0,964 |
| Cão 2 | 40 20 2 2 20 | 0 | 0,904 | 0,923 | 0,923 | 0,923 |
| Cão 2 | 40 20 2 2 20 | 1 | 0,921 | 0,970 | 0,923 | 0,919 |
| Cão 2 | 40 20 2 2 20 | 2 | 0,958 | 0,978 | 0,923 | 0,930 |
| Cão 2 | 40 20 2 2 20 | 3 | 0,969 | 0,989 | 0,923 | 0,924 |
| Cão 2 | 40 20 2 2 20 | 4 | 0,976 | 0,993 | 0,923 | 0,911 |
| Cão 2 | 40 20 2 2 20 | 5 | 0,991 | 0,994 | 0,923 | 0,954 |
| Paciente 1 | 40 20 2 2 20 | 0 | 0,676 | 0,721 | 0,779 | 0,691 |
| Paciente 1 | 40 20 2 2 20 | 1 | 0,853 | 0,926 | 0,809 | 0,838 |
| Paciente 1 | 40 20 2 2 20 | 2 | 0,941 | 0,926 | 0,824 | 0,926 |
| Paciente 1 | 40 20 2 2 20 | 3 | 0,853 | 0,956 | 0,838 | 0,926 |
| Paciente 1 | 40 20 2 2 20 | 4 | 0,926 | 0,956 | 0,853 | 0,956 |
| Paciente 1 | 40 20 2 2 20 | 5 | 0,926 | 0,985 | 0,853 | 0,985 |
| Paciente 2 | 40 20 2 2 20 | 0 | 0,667 | 0,750 | 0,667 | 0,550 |
| Paciente 2 | 40 20 2 2 20 | 1 | 0,783 | 0,917 | 0,700 | 0,817 |
| Paciente 2 | 40 20 2 2 20 | 2 | 0,850 | 0,950 | 0,700 | 0,950 |
| Paciente 2 | 40 20 2 2 20 | 3 | 0,833 | 0,917 | 0,700 | 0,917 |
| Paciente 2 | 40 20 2 2 20 | 4 | 0,900 | 0,950 | 0,700 | 0,950 |
| Paciente 2 | 40 20 2 2 20 | 5 | 0,933 | 0,967 | 0,700 | 0,983 |

Tabela com resultados *Accuracy* (Parâmetros 40 30 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 30 2 2 10 | 0 | 0,950 | 0,952 | 0,952 | 0,940 |
| Cão 1 | 40 30 2 2 10 | 1 | 0,964 | 0,986 | 0,952 | 0,948 |
| Cão 1 | 40 30 2 2 10 | 2 | 0,960 | 0,988 | 0,952 | 0,946 |
| Cão 1 | 40 30 2 2 10 | 3 | 0,968 | 0,986 | 0,952 | 0,954 |
| Cão 1 | 40 30 2 2 10 | 4 | 0,974 | 0,988 | 0,952 | 0,950 |
| Cão 1 | 40 30 2 2 10 | 5 | 0,976 | 0,990 | 0,952 | 0,956 |
| Cão 2 | 40 30 2 2 10 | 0 | 0,908 | 0,923 | 0,923 | 0,923 |
| Cão 2 | 40 30 2 2 10 | 1 | 0,935 | 0,970 | 0,923 | 0,924 |
| Cão 2 | 40 30 2 2 10 | 2 | 0,939 | 0,978 | 0,923 | 0,913 |
| Cão 2 | 40 30 2 2 10 | 3 | 0,965 | 0,989 | 0,923 | 0,911 |
| Cão 2 | 40 30 2 2 10 | 4 | 0,972 | 0,993 | 0,923 | 0,943 |
| Cão 2 | 40 30 2 2 10 | 5 | 0,987 | 0,994 | 0,923 | 0,980 |
| Paciente 1 | 40 30 2 2 10 | 0 | 0,882 | 0,765 | 0,779 | 0,676 |
| Paciente 1 | 40 30 2 2 10 | 1 | 0,897 | 0,897 | 0,794 | 0,838 |
| Paciente 1 | 40 30 2 2 10 | 2 | 0,912 | 0,941 | 0,824 | 0,912 |
| Paciente 1 | 40 30 2 2 10 | 3 | 0,912 | 0,956 | 0,824 | 0,926 |
| Paciente 1 | 40 30 2 2 10 | 4 | 0,941 | 0,956 | 0,838 | 0,956 |
| Paciente 1 | 40 30 2 2 10 | 5 | 0,985 | 0,985 | 0,853 | 0,985 |
| Paciente 2 | 40 30 2 2 10 | 0 | 0,717 | 0,767 | 0,667 | 0,583 |
| Paciente 2 | 40 30 2 2 10 | 1 | 0,883 | 0,933 | 0,700 | 0,850 |
| Paciente 2 | 40 30 2 2 10 | 2 | 0,883 | 0,933 | 0,700 | 0,950 |
| Paciente 2 | 40 30 2 2 10 | 3 | 0,850 | 0,933 | 0,700 | 0,917 |
| Paciente 2 | 40 30 2 2 10 | 4 | 0,900 | 0,933 | 0,700 | 0,950 |
| Paciente 2 | 40 30 2 2 10 | 5 | 0,900 | 0,933 | 0,700 | 0,950 |

Tabela com resultados *Accuracy* (Parâmetros 40 30 2 2 20)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 30 2 2 20 | 0 | 0,950 | 0,952 | 0,952 | 0,946 |
| Cão 1 | 40 30 2 2 20 | 1 | 0,962 | 0,986 | 0,952 | 0,940 |
| Cão 1 | 40 30 2 2 20 | 2 | 0,958 | 0,988 | 0,952 | 0,942 |
| Cão 1 | 40 30 2 2 20 | 3 | 0,956 | 0,990 | 0,952 | 0,956 |
| Cão 1 | 40 30 2 2 20 | 4 | 0,966 | 0,992 | 0,952 | 0,980 |
| Cão 1 | 40 30 2 2 20 | 5 | 0,978 | 0,988 | 0,952 | 0,988 |
| Cão 2 | 40 30 2 2 20 | 0 | 0,871 | 0,921 | 0,923 | 0,923 |
| Cão 2 | 40 30 2 2 20 | 1 | 0,921 | 0,970 | 0,923 | 0,926 |
| Cão 2 | 40 30 2 2 20 | 2 | 0,943 | 0,978 | 0,923 | 0,928 |
| Cão 2 | 40 30 2 2 20 | 3 | 0,965 | 0,989 | 0,923 | 0,917 |
| Cão 2 | 40 30 2 2 20 | 4 | 0,983 | 0,993 | 0,923 | 0,935 |
| Cão 2 | 40 30 2 2 20 | 5 | 0,989 | 0,994 | 0,923 | 0,985 |
| Paciente 1 | 40 30 2 2 20 | 0 | 0,647 | 0,735 | 0,750 | 0,647 |
| Paciente 1 | 40 30 2 2 20 | 1 | 0,897 | 0,912 | 0,794 | 0,853 |
| Paciente 1 | 40 30 2 2 20 | 2 | 0,941 | 0,926 | 0,824 | 0,971 |
| Paciente 1 | 40 30 2 2 20 | 3 | 0,897 | 0,956 | 0,809 | 0,941 |
| Paciente 1 | 40 30 2 2 20 | 4 | 0,956 | 0,956 | 0,838 | 0,941 |
| Paciente 1 | 40 30 2 2 20 | 5 | 0,971 | 0,985 | 0,882 | 0,956 |
| Paciente 2 | 40 30 2 2 20 | 0 | 0,733 | 0,767 | 0,683 | 0,517 |
| Paciente 2 | 40 30 2 2 20 | 1 | 0,850 | 0,917 | 0,700 | 0,850 |
| Paciente 2 | 40 30 2 2 20 | 2 | 0,883 | 0,950 | 0,700 | 0,950 |
| Paciente 2 | 40 30 2 2 20 | 3 | 0,833 | 0,917 | 0,700 | 0,917 |
| Paciente 2 | 40 30 2 2 20 | 4 | 0,883 | 0,950 | 0,700 | 0,950 |
| Paciente 2 | 40 30 2 2 20 | 5 | 0,933 | 0,967 | 0,700 | 0,983 |

Tabela com resultados *Accuracy* (Parâmetros 60 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 60 20 2 2 10 | 0 | 0,952 | 0,952 | 0,952 | 0,952 |
| Cão 1 | 60 20 2 2 10 | 1 | 0,956 | 0,984 | 0,952 | 0,960 |
| Cão 1 | 60 20 2 2 10 | 2 | 0,964 | 0,986 | 0,952 | 0,956 |
| Cão 1 | 60 20 2 2 10 | 3 | 0,968 | 0,986 | 0,952 | 0,948 |
| Cão 1 | 60 20 2 2 10 | 4 | 0,972 | 0,990 | 0,952 | 0,948 |
| Cão 1 | 60 20 2 2 10 | 5 | 0,986 | 0,986 | 0,952 | 0,952 |
| Cão 2 | 60 20 2 2 10 | 0 | 0,911 | 0,924 | 0,923 | 0,923 |
| Cão 2 | 60 20 2 2 10 | 1 | 0,924 | 0,972 | 0,923 | 0,926 |
| Cão 2 | 60 20 2 2 10 | 2 | 0,969 | 0,978 | 0,923 | 0,923 |
| Cão 2 | 60 20 2 2 10 | 3 | 0,967 | 0,989 | 0,923 | 0,913 |
| Cão 2 | 60 20 2 2 10 | 4 | 0,985 | 0,993 | 0,923 | 0,915 |
| Cão 2 | 60 20 2 2 10 | 5 | 0,987 | 0,994 | 0,923 | 0,961 |
| Paciente 1 | 60 20 2 2 10 | 0 | 0,809 | 0,779 | 0,794 | 0,662 |
| Paciente 1 | 60 20 2 2 10 | 1 | 0,882 | 0,926 | 0,824 | 0,868 |
| Paciente 1 | 60 20 2 2 10 | 2 | 0,882 | 0,912 | 0,824 | 0,912 |
| Paciente 1 | 60 20 2 2 10 | 3 | 0,941 | 0,956 | 0,853 | 0,941 |
| Paciente 1 | 60 20 2 2 10 | 4 | 0,956 | 0,985 | 0,853 | 0,985 |
| Paciente 1 | 60 20 2 2 10 | 5 | 0,956 | 0,985 | 0,868 | 0,971 |
| Paciente 2 | 60 20 2 2 10 | 0 | 0,633 | 0,633 | 0,667 | 0,567 |
| Paciente 2 | 60 20 2 2 10 | 1 | 0,833 | 0,883 | 0,700 | 0,833 |
| Paciente 2 | 60 20 2 2 10 | 2 | 0,900 | 0,917 | 0,700 | 0,917 |
| Paciente 2 | 60 20 2 2 10 | 3 | 0,850 | 0,933 | 0,700 | 0,967 |
| Paciente 2 | 60 20 2 2 10 | 4 | 0,967 | 0,983 | 0,700 | 0,950 |
| Paciente 2 | 60 20 2 2 10 | 5 | 0,900 | 0,967 | 0,700 | 0,983 |

Tabela com resultados *F-measure* (Parâmetros 30 10 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 30 10 2 2 10 | 0 | 0,929 | 0,929 | 0,929 | 0,936 |
| Cão 1 | 30 10 2 2 10 | 1 | 0,946 | 0,983 | 0,929 | 0,952 |
| Cão 1 | 30 10 2 2 10 | 2 | 0,961 | 0,985 | 0,929 | 0,944 |
| Cão 1 | 30 10 2 2 10 | 3 | 0,965 | 0,986 | 0,929 | 0,940 |
| Cão 1 | 30 10 2 2 10 | 4 | 0,972 | 0,990 | 0,929 | 0,933 |
| Cão 1 | 30 10 2 2 10 | 5 | 0,986 | 0,986 | 0,929 | 0,936 |
| Cão 2 | 30 10 2 2 10 | 0 | 0,921 | 0,888 | 0,885 | 0,885 |
| Cão 2 | 30 10 2 2 10 | 1 | 0,942 | 0,964 | 0,885 | 0,885 |
| Cão 2 | 30 10 2 2 10 | 2 | 0,929 | 0,978 | 0,885 | 0,885 |
| Cão 2 | 30 10 2 2 10 | 3 | 0,981 | 0,989 | 0,885 | 0,885 |
| Cão 2 | 30 10 2 2 10 | 4 | 0,979 | 0,994 | 0,885 | 0,883 |
| Cão 2 | 30 10 2 2 10 | 5 | 0,989 | 0,994 | 0,885 | 0,883 |
| Paciente 1 | 30 10 2 2 10 | 0 | 0,674 | 0,773 | 0,766 | 0,712 |
| Paciente 1 | 30 10 2 2 10 | 1 | 0,886 | 0,896 | 0,789 | 0,736 |
| Paciente 1 | 30 10 2 2 10 | 2 | 0,869 | 0,912 | 0,789 | 0,877 |
| Paciente 1 | 30 10 2 2 10 | 3 | 0,869 | 0,912 | 0,831 | 0,926 |
| Paciente 1 | 30 10 2 2 10 | 4 | 0,926 | 0,971 | 0,831 | 0,971 |
| Paciente 1 | 30 10 2 2 10 | 5 | 0,985 | 0,985 | 0,851 | 0,956 |
| Paciente 2 | 30 10 2 2 10 | 0 | 0,802 | 0,787 | 0,568 | 0,632 |
| Paciente 2 | 30 10 2 2 10 | 1 | 0,950 | 0,933 | 0,576 | 0,773 |
| Paciente 2 | 30 10 2 2 10 | 2 | 0,898 | 0,983 | 0,576 | 0,898 |
| Paciente 2 | 30 10 2 2 10 | 3 | 0,900 | 0,933 | 0,576 | 0,917 |
| Paciente 2 | 30 10 2 2 10 | 4 | 0,898 | 0,950 | 0,576 | 0,933 |
| Paciente 2 | 30 10 2 2 10 | 5 | 0,916 | 0,950 | 0,576 | 0,966 |

Tabela com resultados *F-measure* (Parâmetros 30 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 30 20 2 2 10 | 0 | 0,929 | 0,929 | 0,929 | 0,923 |
| Cão 1 | 30 20 2 2 10 | 1 | 0,951 | 0,985 | 0,929 | 0,961 |
| Cão 1 | 30 20 2 2 10 | 2 | 0,954 | 0,987 | 0,929 | 0,959 |
| Cão 1 | 30 20 2 2 10 | 3 | 0,980 | 0,990 | 0,929 | 0,957 |
| Cão 1 | 30 20 2 2 10 | 4 | 0,978 | 0,992 | 0,929 | 0,940 |
| Cão 1 | 30 20 2 2 10 | 5 | 0,982 | 0,992 | 0,929 | 0,933 |
| Cão 2 | 30 20 2 2 10 | 0 | 0,877 | 0,860 | 0,885 | 0,885 |
| Cão 2 | 30 20 2 2 10 | 1 | 0,940 | 0,968 | 0,885 | 0,885 |
| Cão 2 | 30 20 2 2 10 | 2 | 0,949 | 0,978 | 0,885 | 0,885 |
| Cão 2 | 30 20 2 2 10 | 3 | 0,970 | 0,989 | 0,885 | 0,882 |
| Cão 2 | 30 20 2 2 10 | 4 | 0,974 | 0,992 | 0,885 | 0,881 |
| Cão 2 | 30 20 2 2 10 | 5 | 0,989 | 0,994 | 0,885 | 0,887 |
| Paciente 1 | 30 20 2 2 10 | 0 | 0,765 | 0,779 | 0,766 | 0,638 |
| Paciente 1 | 30 20 2 2 10 | 1 | 0,847 | 0,910 | 0,789 | 0,756 |
| Paciente 1 | 30 20 2 2 10 | 2 | 0,882 | 0,941 | 0,789 | 0,955 |
| Paciente 1 | 30 20 2 2 10 | 3 | 0,882 | 0,912 | 0,831 | 0,924 |
| Paciente 1 | 30 20 2 2 10 | 4 | 0,926 | 0,956 | 0,831 | 0,971 |
| Paciente 1 | 30 20 2 2 10 | 5 | 0,971 | 0,971 | 0,851 | 0,956 |
| Paciente 2 | 30 20 2 2 10 | 0 | 0,802 | 0,817 | 0,604 | 0,681 |
| Paciente 2 | 30 20 2 2 10 | 1 | 0,917 | 0,916 | 0,576 | 0,849 |
| Paciente 2 | 30 20 2 2 10 | 2 | 0,884 | 0,966 | 0,576 | 0,918 |
| Paciente 2 | 30 20 2 2 10 | 3 | 0,916 | 0,950 | 0,576 | 0,950 |
| Paciente 2 | 30 20 2 2 10 | 4 | 0,932 | 0,966 | 0,576 | 0,933 |
| Paciente 2 | 30 20 2 2 10 | 5 | 0,916 | 0,950 | 0,576 | 0,983 |

Tabela com resultados *F-measure* (Parâmetros 40 10 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 10 2 2 10 | 0 | 0,929 | 0,927 | 0,929 | 0,927 |
| Cão 1 | 40 10 2 2 10 | 1 | 0,954 | 0,983 | 0,929 | 0,935 |
| Cão 1 | 40 10 2 2 10 | 2 | 0,936 | 0,985 | 0,929 | 0,941 |
| Cão 1 | 40 10 2 2 10 | 3 | 0,952 | 0,990 | 0,929 | 0,949 |
| Cão 1 | 40 10 2 2 10 | 4 | 0,974 | 0,988 | 0,929 | 0,928 |
| Cão 1 | 40 10 2 2 10 | 5 | 0,971 | 0,986 | 0,929 | 0,929 |
| Cão 2 | 40 10 2 2 10 | 0 | 0,885 | 0,884 | 0,885 | 0,885 |
| Cão 2 | 40 10 2 2 10 | 1 | 0,910 | 0,964 | 0,885 | 0,885 |
| Cão 2 | 40 10 2 2 10 | 2 | 0,953 | 0,974 | 0,885 | 0,885 |
| Cão 2 | 40 10 2 2 10 | 3 | 0,968 | 0,989 | 0,885 | 0,885 |
| Cão 2 | 40 10 2 2 10 | 4 | 0,968 | 0,992 | 0,885 | 0,885 |
| Cão 2 | 40 10 2 2 10 | 5 | 0,985 | 0,994 | 0,885 | 0,885 |
| Paciente 1 | 40 10 2 2 10 | 0 | 0,765 | 0,760 | 0,730 | 0,705 |
| Paciente 1 | 40 10 2 2 10 | 1 | 0,748 | 0,880 | 0,776 | 0,792 |
| Paciente 1 | 40 10 2 2 10 | 2 | 0,856 | 0,896 | 0,789 | 0,856 |
| Paciente 1 | 40 10 2 2 10 | 3 | 0,908 | 0,912 | 0,831 | 0,880 |
| Paciente 1 | 40 10 2 2 10 | 4 | 0,970 | 0,941 | 0,831 | 0,985 |
| Paciente 1 | 40 10 2 2 10 | 5 | 0,985 | 0,985 | 0,851 | 0,985 |
| Paciente 2 | 40 10 2 2 10 | 0 | 0,615 | 0,707 | 0,560 | 0,487 |
| Paciente 2 | 40 10 2 2 10 | 1 | 0,723 | 0,914 | 0,576 | 0,647 |
| Paciente 2 | 40 10 2 2 10 | 2 | 0,967 | 0,967 | 0,576 | 0,763 |
| Paciente 2 | 40 10 2 2 10 | 3 | 0,898 | 0,933 | 0,576 | 0,815 |
| Paciente 2 | 40 10 2 2 10 | 4 | 0,932 | 0,967 | 0,576 | 0,950 |
| Paciente 2 | 40 10 2 2 10 | 5 | 0,912 | 0,966 | 0,576 | 0,966 |

Tabela com resultados *F-measure* (Parâmetros 40 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 20 2 2 10 | 0 | 0,929 | 0,929 | 0,929 | 0,926 |
| Cão 1 | 40 20 2 2 10 | 1 | 0,960 | 0,985 | 0,929 | 0,928 |
| Cão 1 | 40 20 2 2 10 | 2 | 0,966 | 0,987 | 0,929 | 0,937 |
| Cão 1 | 40 20 2 2 10 | 3 | 0,973 | 0,990 | 0,929 | 0,942 |
| Cão 1 | 40 20 2 2 10 | 4 | 0,983 | 0,992 | 0,929 | 0,931 |
| Cão 1 | 40 20 2 2 10 | 5 | 0,990 | 0,992 | 0,929 | 0,947 |
| Cão 2 | 40 20 2 2 10 | 0 | 0,880 | 0,885 | 0,885 | 0,885 |
| Cão 2 | 40 20 2 2 10 | 1 | 0,922 | 0,967 | 0,885 | 0,885 |
| Cão 2 | 40 20 2 2 10 | 2 | 0,954 | 0,976 | 0,885 | 0,879 |
| Cão 2 | 40 20 2 2 10 | 3 | 0,966 | 0,989 | 0,885 | 0,883 |
| Cão 2 | 40 20 2 2 10 | 4 | 0,972 | 0,992 | 0,885 | 0,887 |
| Cão 2 | 40 20 2 2 10 | 5 | 0,994 | 0,994 | 0,885 | 0,921 |
| Paciente 1 | 40 20 2 2 10 | 0 | 0,742 | 0,740 | 0,730 | 0,617 |
| Paciente 1 | 40 20 2 2 10 | 1 | 0,869 | 0,910 | 0,776 | 0,820 |
| Paciente 1 | 40 20 2 2 10 | 2 | 0,910 | 0,910 | 0,789 | 0,941 |
| Paciente 1 | 40 20 2 2 10 | 3 | 0,940 | 0,955 | 0,831 | 0,910 |
| Paciente 1 | 40 20 2 2 10 | 4 | 0,970 | 0,941 | 0,831 | 0,955 |
| Paciente 1 | 40 20 2 2 10 | 5 | 0,985 | 0,985 | 0,851 | 0,956 |
| Paciente 2 | 40 20 2 2 10 | 0 | 0,696 | 0,780 | 0,584 | 0,491 |
| Paciente 2 | 40 20 2 2 10 | 1 | 0,817 | 0,898 | 0,576 | 0,815 |
| Paciente 2 | 40 20 2 2 10 | 2 | 0,864 | 0,934 | 0,576 | 0,950 |
| Paciente 2 | 40 20 2 2 10 | 3 | 0,846 | 0,933 | 0,576 | 0,934 |
| Paciente 2 | 40 20 2 2 10 | 4 | 0,914 | 0,933 | 0,576 | 0,950 |
| Paciente 2 | 40 20 2 2 10 | 5 | 0,912 | 0,966 | 0,576 | 1,000 |

Tabela com resultados *F-measure* (Parâmetros 40 20 2 2 20)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 20 2 2 20 | 0 | 0,928 | 0,929 | 0,929 | 0,932 |
| Cão 1 | 40 20 2 2 20 | 1 | 0,956 | 0,985 | 0,929 | 0,937 |
| Cão 1 | 40 20 2 2 20 | 2 | 0,954 | 0,987 | 0,929 | 0,937 |
| Cão 1 | 40 20 2 2 20 | 3 | 0,952 | 0,990 | 0,929 | 0,948 |
| Cão 1 | 40 20 2 2 20 | 4 | 0,971 | 0,992 | 0,929 | 0,939 |
| Cão 1 | 40 20 2 2 20 | 5 | 0,984 | 0,990 | 0,929 | 0,962 |
| Cão 2 | 40 20 2 2 20 | 0 | 0,884 | 0,885 | 0,885 | 0,885 |
| Cão 2 | 40 20 2 2 20 | 1 | 0,916 | 0,967 | 0,885 | 0,883 |
| Cão 2 | 40 20 2 2 20 | 2 | 0,958 | 0,976 | 0,885 | 0,907 |
| Cão 2 | 40 20 2 2 20 | 3 | 0,968 | 0,989 | 0,885 | 0,896 |
| Cão 2 | 40 20 2 2 20 | 4 | 0,976 | 0,992 | 0,885 | 0,897 |
| Cão 2 | 40 20 2 2 20 | 5 | 0,991 | 0,994 | 0,885 | 0,951 |
| Paciente 1 | 40 20 2 2 20 | 0 | 0,653 | 0,696 | 0,730 | 0,682 |
| Paciente 1 | 40 20 2 2 20 | 1 | 0,853 | 0,924 | 0,776 | 0,837 |
| Paciente 1 | 40 20 2 2 20 | 2 | 0,940 | 0,924 | 0,789 | 0,926 |
| Paciente 1 | 40 20 2 2 20 | 3 | 0,853 | 0,955 | 0,811 | 0,924 |
| Paciente 1 | 40 20 2 2 20 | 4 | 0,924 | 0,955 | 0,831 | 0,955 |
| Paciente 1 | 40 20 2 2 20 | 5 | 0,927 | 0,985 | 0,831 | 0,985 |
| Paciente 2 | 40 20 2 2 20 | 0 | 0,634 | 0,720 | 0,560 | 0,537 |
| Paciente 2 | 40 20 2 2 20 | 1 | 0,772 | 0,914 | 0,576 | 0,818 |
| Paciente 2 | 40 20 2 2 20 | 2 | 0,846 | 0,950 | 0,576 | 0,950 |
| Paciente 2 | 40 20 2 2 20 | 3 | 0,830 | 0,917 | 0,576 | 0,918 |
| Paciente 2 | 40 20 2 2 20 | 4 | 0,898 | 0,950 | 0,576 | 0,950 |
| Paciente 2 | 40 20 2 2 20 | 5 | 0,931 | 0,966 | 0,576 | 0,983 |

Tabela com resultados *F-measure* (Parâmetros 40 30 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 30 2 2 10 | 0 | 0,928 | 0,929 | 0,929 | 0,923 |
| Cão 1 | 40 30 2 2 10 | 1 | 0,960 | 0,985 | 0,929 | 0,936 |
| Cão 1 | 40 30 2 2 10 | 2 | 0,952 | 0,987 | 0,929 | 0,935 |
| Cão 1 | 40 30 2 2 10 | 3 | 0,966 | 0,986 | 0,929 | 0,951 |
| Cão 1 | 40 30 2 2 10 | 4 | 0,975 | 0,988 | 0,929 | 0,945 |
| Cão 1 | 40 30 2 2 10 | 5 | 0,977 | 0,990 | 0,929 | 0,949 |
| Cão 2 | 40 30 2 2 10 | 0 | 0,878 | 0,885 | 0,885 | 0,885 |
| Cão 2 | 40 30 2 2 10 | 1 | 0,929 | 0,967 | 0,885 | 0,896 |
| Cão 2 | 40 30 2 2 10 | 2 | 0,937 | 0,976 | 0,885 | 0,892 |
| Cão 2 | 40 30 2 2 10 | 3 | 0,962 | 0,989 | 0,885 | 0,897 |
| Cão 2 | 40 30 2 2 10 | 4 | 0,972 | 0,992 | 0,885 | 0,940 |
| Cão 2 | 40 30 2 2 10 | 5 | 0,987 | 0,994 | 0,885 | 0,980 |
| Paciente 1 | 40 30 2 2 10 | 0 | 0,877 | 0,740 | 0,715 | 0,682 |
| Paciente 1 | 40 30 2 2 10 | 1 | 0,898 | 0,896 | 0,754 | 0,837 |
| Paciente 1 | 40 30 2 2 10 | 2 | 0,910 | 0,939 | 0,789 | 0,910 |
| Paciente 1 | 40 30 2 2 10 | 3 | 0,910 | 0,955 | 0,805 | 0,926 |
| Paciente 1 | 40 30 2 2 10 | 4 | 0,940 | 0,955 | 0,818 | 0,955 |
| Paciente 1 | 40 30 2 2 10 | 5 | 0,985 | 0,985 | 0,837 | 0,985 |
| Paciente 2 | 40 30 2 2 10 | 0 | 0,683 | 0,743 | 0,560 | 0,580 |
| Paciente 2 | 40 30 2 2 10 | 1 | 0,882 | 0,932 | 0,576 | 0,851 |
| Paciente 2 | 40 30 2 2 10 | 2 | 0,882 | 0,934 | 0,576 | 0,950 |
| Paciente 2 | 40 30 2 2 10 | 3 | 0,846 | 0,933 | 0,576 | 0,918 |
| Paciente 2 | 40 30 2 2 10 | 4 | 0,898 | 0,933 | 0,576 | 0,950 |
| Paciente 2 | 40 30 2 2 10 | 5 | 0,898 | 0,933 | 0,576 | 0,950 |

Tabela com resultados *F-measure* (Parâmetros 40 30 2 2 20)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 40 30 2 2 20 | 0 | 0,928 | 0,929 | 0,929 | 0,926 |
| Cão 1 | 40 30 2 2 20 | 1 | 0,956 | 0,985 | 0,929 | 0,933 |
| Cão 1 | 40 30 2 2 20 | 2 | 0,951 | 0,987 | 0,929 | 0,934 |
| Cão 1 | 40 30 2 2 20 | 3 | 0,949 | 0,990 | 0,929 | 0,956 |
| Cão 1 | 40 30 2 2 20 | 4 | 0,965 | 0,992 | 0,929 | 0,980 |
| Cão 1 | 40 30 2 2 20 | 5 | 0,978 | 0,988 | 0,929 | 0,988 |
| Cão 2 | 40 30 2 2 20 | 0 | 0,863 | 0,884 | 0,885 | 0,885 |
| Cão 2 | 40 30 2 2 20 | 1 | 0,917 | 0,967 | 0,885 | 0,897 |
| Cão 2 | 40 30 2 2 20 | 2 | 0,943 | 0,976 | 0,885 | 0,914 |
| Cão 2 | 40 30 2 2 20 | 3 | 0,964 | 0,989 | 0,885 | 0,902 |
| Cão 2 | 40 30 2 2 20 | 4 | 0,983 | 0,992 | 0,885 | 0,931 |
| Cão 2 | 40 30 2 2 20 | 5 | 0,989 | 0,994 | 0,885 | 0,985 |
| Paciente 1 | 40 30 2 2 20 | 0 | 0,610 | 0,716 | 0,677 | 0,653 |
| Paciente 1 | 40 30 2 2 20 | 1 | 0,896 | 0,910 | 0,754 | 0,853 |
| Paciente 1 | 40 30 2 2 20 | 2 | 0,940 | 0,924 | 0,789 | 0,970 |
| Paciente 1 | 40 30 2 2 20 | 3 | 0,896 | 0,955 | 0,785 | 0,939 |
| Paciente 1 | 40 30 2 2 20 | 4 | 0,955 | 0,955 | 0,818 | 0,940 |
| Paciente 1 | 40 30 2 2 20 | 5 | 0,971 | 0,985 | 0,874 | 0,956 |
| Paciente 2 | 40 30 2 2 20 | 0 | 0,716 | 0,734 | 0,594 | 0,536 |
| Paciente 2 | 40 30 2 2 20 | 1 | 0,846 | 0,916 | 0,576 | 0,851 |
| Paciente 2 | 40 30 2 2 20 | 2 | 0,880 | 0,950 | 0,576 | 0,950 |
| Paciente 2 | 40 30 2 2 20 | 3 | 0,830 | 0,917 | 0,576 | 0,918 |
| Paciente 2 | 40 30 2 2 20 | 4 | 0,882 | 0,950 | 0,576 | 0,950 |
| Paciente 2 | 40 30 2 2 20 | 5 | 0,931 | 0,966 | 0,576 | 0,983 |

Tabela com resultados *F-measure* (Parâmetros 60 20 2 2 10)

| Dataset | Parâmetros MrMotif | Resample | J48 | Random Forest | SVM - SMO | Logistic |
|---|---|---|---|---|---|---|
| Cão 1 | 60 20 2 2 10 | 0 | 0,929 | 0,929 | 0,929 | 0,936 |
| Cão 1 | 60 20 2 2 10 | 1 | 0,946 | 0,983 | 0,929 | 0,952 |
| Cão 1 | 60 20 2 2 10 | 2 | 0,961 | 0,985 | 0,929 | 0,944 |
| Cão 1 | 60 20 2 2 10 | 3 | 0,965 | 0,986 | 0,929 | 0,940 |
| Cão 1 | 60 20 2 2 10 | 4 | 0,972 | 0,990 | 0,929 | 0,933 |
| Cão 1 | 60 20 2 2 10 | 5 | 0,986 | 0,986 | 0,929 | 0,936 |
| Cão 2 | 60 20 2 2 10 | 0 | 0,883 | 0,890 | 0,885 | 0,885 |
| Cão 2 | 60 20 2 2 10 | 1 | 0,922 | 0,970 | 0,885 | 0,900 |
| Cão 2 | 60 20 2 2 10 | 2 | 0,967 | 0,976 | 0,885 | 0,892 |
| Cão 2 | 60 20 2 2 10 | 3 | 0,966 | 0,989 | 0,885 | 0,884 |
| Cão 2 | 60 20 2 2 10 | 4 | 0,985 | 0,992 | 0,885 | 0,907 |
| Cão 2 | 60 20 2 2 10 | 5 | 0,987 | 0,994 | 0,885 | 0,961 |
| Paciente 1 | 60 20 2 2 10 | 0 | 0,776 | 0,760 | 0,741 | 0,677 |
| Paciente 1 | 60 20 2 2 10 | 1 | 0,880 | 0,924 | 0,789 | 0,869 |
| Paciente 1 | 60 20 2 2 10 | 2 | 0,884 | 0,912 | 0,789 | 0,913 |
| Paciente 1 | 60 20 2 2 10 | 3 | 0,940 | 0,955 | 0,831 | 0,940 |
| Paciente 1 | 60 20 2 2 10 | 4 | 0,955 | 0,985 | 0,831 | 0,985 |
| Paciente 1 | 60 20 2 2 10 | 5 | 0,955 | 0,985 | 0,851 | 0,971 |
| Paciente 2 | 60 20 2 2 10 | 0 | 0,543 | 0,597 | 0,560 | 0,573 |
| Paciente 2 | 60 20 2 2 10 | 1 | 0,810 | 0,882 | 0,576 | 0,836 |
| Paciente 2 | 60 20 2 2 10 | 2 | 0,900 | 0,917 | 0,576 | 0,917 |
| Paciente 2 | 60 20 2 2 10 | 3 | 0,846 | 0,933 | 0,576 | 0,967 |
| Paciente 2 | 60 20 2 2 10 | 4 | 0,967 | 0,983 | 0,576 | 0,950 |
| Paciente 2 | 60 20 2 2 10 | 5 | 0,896 | 0,966 | 0,576 | 0,983 |