

UNIVERSIDADE DE SÃO PAULO ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA MECÂNICA

ALGORITMO PARA RASTREAMENTO DE PACIENTES COM
PROBABILIDADE DE APRESENTAR SÍNDROME EPILÉPTICA

Bruno Velloso Rodrigues
Iohana Lima de Barros

São Paulo
2017

UNIVERSIDADE DE SÃO PAULO ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA MECÂNICA

ALGORITMO PARA RASTREAMENTO DE PACIENTES COM
PROBABILIDADE DE APRESENTAR SÍNDROME EPILÉPTICA

Trabalho de formatura apresentado à Escola
Politécnica da Universidade de São Paulo para
obtenção do título de Graduação em Engenharia

Bruno Velloso Rodrigues
Iohana Lima de Barros

Orientador: Raul Gonzáles Lima

Área de Concentração:
Engenharia Mecânica

São Paulo
2017

Catálogo-na-publicação

Rodrigues, Bruno Velloso

Algoritmo para rastreamento de pacientes com probabilidade de apresentar síndrome epilética / B. V. Rodrigues, I. L. De Barros -- São Paulo, 2017.

p[RB1]. 80

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Mecânica.

1.EPILEPSIA 2.DIAGNÓSTICO POR IMAGEM 3.SINAIS BIOMÉDICOS
I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Mecânica II.t. III.De Barros, Iohama Lima

RESUMO

O presente trabalho pretende apresentar e comparar métodos de leitura de sinais elétricos para exames de Eletroencefalograma no auxílio da identificação de pacientes epiléticos. Embora normalmente seja uma doença de fácil tratamento, o diagnóstico costuma requerer a atenção de um especialista e ser fastidioso. Levando em consideração que se tratam de sinais não lineares e não estacionários, os métodos são divididos em duas partes: a transformação dos sinais e a classificação do paciente em possivelmente saudável ou possivelmente epilético. A partir de duas bases de dados da literatura, é possível comparar o desempenho de cada método e determinar as vantagens e desvantagens de cada um. Foram aplicados diferentes métodos (Transformada de Fourier, Método de Frequências, Método de Entropias e combinações) com classificadores (SVM e Árvore Complexa).

ABSTRACT

This report intends to present and compare methods to classify electrical signs from EEG exams in order to identify epileptic patients. Although the treatment to this syndrome is usually simple, the diagnosis is a tedious and time-consuming process. Taking into account the fact EEG signals are nonlinear and nonstationary, the methods were presented in two steps: transformation of EEG signals and classification of the patients in possibly epileptic or possibly healthy. From two different databases available in the literature, we are able to compare the efficiency of each method and determine their pros and cons. Different methods were used (Frequency method, Entropy Method, Fourier Transformation Method and combinations) with classifiers (SVM and Complex Tree).

SUMÁRIO

LISTA DE FIGURAS	1
LISTA DE TABELAS	2
1. Introdução	1
2. Objetivo	3
3. Revisão do material técnico	4
3.1. A epilepsia como doença	4
3.1.1. Crises Epilépticas e a Epilepsia	4
3.1.2. Crises Focais	6
3.1.3. Crises Generalizadas	7
3.1.4. Diagnóstico	7
3.1.5. Eletroencefalograma	8
3.1.6. Tratamento	9
3.1.7. Prognóstico da epilepsia	9
4. As bases de dados	10
4.1. Bases de dados	10
4.1.1. Database de Bonn	10
4.1.2. Database CHT-MIT	14
4.2. Tratamento de sinais de exames de EEG	16
4.2.1. Método das Entropias	17
4.2.1.1. Entropia Approximate Entropy (AppEn)	18
4.2.2. Método das Frequências	19
4.2.2.1. Introdução	19
4.2.2.2. Método	20
4.2.2.3. Transformada de Hilbert-Huang	21
4.2.3. Método da Transformada de Fourier	18
4.2.4. Machine Learning	20
4.2.4.1. Classificação – SVM	20
4.2.4.2. Classificação - COMPLEX TREE	22
5. METODOLOGIA	25
5.1. Estrutura de diagnóstico	25
5.1.1. Métodos das Frequências	25
5.1.2. Métodos da Transformada de Fourier	26
5.1.3. Método das Entropias	27

5.2.	MATLAB	28
6.	RESULTADOS E DISCUSSÃO	30
6.1.	Resultados Método de Frequências + SVM Quadrático	30
6.2.	Resultados Método da Transformada de Fourier + Árvore complexa	33
6.3.	Resultados Método das Entropias + SVM Quadrático	36
6.4.	Extensão dos métodos para classificação tripla	38
6.4.1.	Resultados Método de Fourier para classificação tripla + Árvore Complexa	39
6.4.2.	Resultados Método das Entropias para classificação tripla + Árvore Complexa	40
6.4.3.	Resultados Método combinado de Fourier e Entropias para classificação tripla + Árvore Complexa	41
7.	ANÁLISE DOS RESULTADOS E CONSIDERAÇÕES FINAIS	42
8.	REFERENCIAS	44

LISTA DE FIGURAS

Figura 1 - Áreas funcionais do cérebro.....	5
Figura 2 - Posicionamento dos eletrodos para EEG externo.	11
Figura 3 - Sinal resultado de EEG para o caso A – Univ de Bonn	12
Figura 4 - Sinal resultado de EEG para o caso B – Univ de Bonn	12
Figura 5 - Sinal resultado de EEG para o caso C – Univ de Bonn	13
Figura 6 - Sinal resultado de EEG para o caso D – Univ de Bonn	13
Figura 7 - Sinal resultado de EEG para o caso D – Univ de Bonn	14
Figura 8 - Segmento da CHB-MIT database por eletrodo.....	16
Figura 9 - Exemplo de árvore	24
Figura 10 - Fluxograma do método das Frequências	26
Figura 11 - Fluxograma do método da Transformada de Fourier	27
Figura 12 - Fluxograma do método das entropias.	28
Figura 13 - Dispersão de pontos com o método de frequências e classificador SVM Quadrático	32
Figura 14 - Matrix de Conflito para método das entropias e classificador SVM Quadrático	32
Figura 15 - Potência Espectral de Densidade para o caso A.....	34
Figura 16 - Potência Espectral de Densidade para o caso E.....	34
Figura 17 – Dispersão de pontos com classificador Árvore Complexa no método de Fourier	35
Figura 18 - Matriz de Conflito de classificador árvore no método de Fourier	36
Figura 19 - Matriz de Conflito de classificador SVM quadrático no método de Entropias.....	37
Figura 20 - Dispersão de pontos com classificador SVM Quadrático no método de Entropias.....	37
Figura 21 - Matriz de Conflito de classificador Árvore Complexa no método de Fourier em valores relativos e absolutos em classificação tripla.....	39
Figura 22 - Dispersão de pontos com classificador Árvore Complexa no método de Fourier em classificação tripla	39
Figura 23 - Matriz de Conflito de classificador Árvore Complexa no método de Entropias em valores relativos e absolutos em classificação tripla	40
Figura 24 - Dispersão de pontos com classificador Árvore Complexa no método de Entropias em classificação tripla	40
Figura 25 - Matriz de Conflito de classificador Árvore Complexa no método de Entropias em valores relativos e absolutos em classificação tripla	41
Figura 26 - Dispersão de pontos com classificador Árvore Complexa para método combinado de Entropias e Fourier em classificação tripla	41

LISTA DE TABELAS

Tabela 1 - Dados de pacientes (Sexo e Idade)	15
Tabela 2 - Faixa de frequências por banda	25
Tabela 3 – Comparação de classificadores para o método de frequências	32

Lista de acrônimos

AEDs, Anti-Epileptic Drugs ou Medicamentos anti-epiléticos
AppEn, Approximate Entropy
DFT, Discrete Fourier Transformation ou Transformada de Fourier Discreta
EEG, Electroencephalogram ou Eletroencefalograma
EMD, Empirical Mode Decomposition ou Método Empírico de Decomposição
FFT, Fourier Transformation ou Transformada de Fourier
HHT, Hilbert–Huang Transformation ou Transformada de Hilbert–Huang
HSA, Hilbert Spectral Analysis ou Análise Espectral de Hilbert
IMF, Intrinsic Mode Functions ou Funções de Modo Intrínseco
MATLAB, MATrix LABoratory ou Laboratório Matrix
OMS, Organização Mundial de Saúde
PSD, Potencial Spectral Density ou Densidade Espectral de Potência
RBF, Radial Basis Function ou Função Radial Base
SVM, Support Vector Machine ou Máquina de Vetor de Suporte
TFI, Time-Frequency Image ou Imagem em Tempo-Frequência
TFR, Time-Frequency Representation ou Representação de Tempo-Frequência
WHO, World Health Organization

1. Introdução

Segundo a Organização Mundial de Saúde, a OMS, hoje no mundo aproximadamente 50 milhões de pessoas sofrem com epilepsia, fazendo dela uma das doenças neurológicas mais comuns globalmente. Perto de 80% das pessoas com epilepsia vivem em países de baixa e média renda. Destes mesmos 80%, 70% não recebe um tratamento adequado para a doença.

A epilepsia, embora possa ser classificada em diversos tipos diferentes, não é uma doença contagiosa. A forma mais comum da síndrome, a epilepsia idiopática, ainda não possui causa identificada e responde por 6 em cada 10 casos da doença.

As síndromes de causa conhecida são chamadas de epilepsia secundária ou sintomática. Entre as diversas causas, podemos citar dano cerebral provocado por lesões pré-natais ou perinatais; anormalidades congênitas ou condições genéticas; ferimento grave na cabeça; infecções no cérebro como meningite, encefalite e neurocisticercose; etc.

Acredita-se que boa parte destas causas de epilepsia secundária expliquem a maior incidência da síndrome em países de menor renda devido às maiores taxas de doenças endêmicas, a maior incidência de acidentes de trânsito, lesões relacionadas ao nascimento e variações na infraestrutura médica, como na disponibilidade de programas preventivos de saúde e cuidados acessíveis.

O tratamento da doença pode ser facilmente acessível a um custo baixo de medicação diária, que pode custar tão pouco quanto 5 dólares ao ano. Estudos recentes em países de média e baixa renda mostram que até 70% de crianças e adultos com epilepsia podem ter suas crises completamente controladas com medicamentos antiepilépticos (AEDs, ou *anti-epileptic drugs*, em inglês). Mais surpreendente, é o fato de que após 2 a 5 anos de tratamento bem-sucedido e sem convulsões registradas, as drogas podem ser

suspendidas em até 70% das crianças e 60% dos adultos sem complicações subsequentes. Em alguns casos, procedimentos cirúrgicos podem ser indicados a pacientes que não reajam ao uso da medicação.

Controversamente, sabe-se que 75% das pessoas com epilepsia em países de média e baixa renda não recebem o tratamento necessário. Além disso, nesses países a disponibilidade de AEDs é baixa, chegando a menos de 50% no setor público de saúde, dificultando o acesso.

É possível diagnosticar e tratar a maior parte das pessoas portadoras da síndrome em um nível primário de atendimento à saúde, sem a necessidade de equipamentos sofisticados. Estudos da OMS mostram que o treinamento de prestadores de cuidados de nível básico para diagnosticar e tratar epilepsia pode reduzir substancialmente a lacuna do tratamento da doença.

A falta de capacitação e preparo de profissionais da saúde passa a ser um dos gargalos e torna-se uma barreira importante ao tratamento de pessoas com epilepsia. Faz-se necessária, portanto, formas de aumentar a capacitação de tais agentes e facilitar o processo de diagnóstico.

2. Objetivo

Esse relatório tem como objetivo mostrar diferentes métodos matemáticos para o tratamento e classificação dos sinais coletados através de exames de eletroencefalograma, ferramenta fundamental para o diagnóstico da epilepsia. A comparação dos métodos permitirá uma posterior análise em termos de eficiência e precisão de cada um.

Tendo como alicerce para esse estudo duas bases dados de domínio público com informações e exames de pacientes tanto saudáveis quanto epiléticos, será aplicado um método reverso para a partir dos exames chegar-se a uma possível indicação da saúde dos indivíduos. Estas duas database, Bonn e CHT-MIT, que serão tratadas de forma mais aprofundada posteriormente, são resultado de uma série de exames de eletroencefalograma de pacientes epiléticos e saudáveis que foram diagnosticados por médicos especialistas.

Para tal, aplicar-se-ão à essas duas bases métodos designados método das Frequências, método das entropias, método da Transformada de Fourier e *Machine Learning*. Em seguida, uma classificação resultante dos pacientes em saudáveis ou epiléticos será apresentada através de classificadores dos sinais tratados, como *Decision Tree* e SVM.

Finalmente, tendo como parâmetro os diagnósticos dados pelos médicos, será possível comparar os resultados obtidos com o estado de saúde real de cada paciente, determinando assim a eficiência de cada combinação: método-classificador.

3. Revisão do material técnico

3.1. A epilepsia como doença

3.1.1. Crises Epilépticas e a Epilepsia

Uma crise epiléptica nada mais é do que sinais ou sintomas transitórios resultados de uma atividade neuronal anormal, síncrona ou excessiva. Podem ser classificadas de duas maneiras: crises focais ou generalizadas, que serão mais detalhadas em seguida.

É importante salientar que crises epiléticas não implicam diretamente em epilepsia. Alguns tipos de ataques cerebrais agudos podem provocar crises que não resultam em tendência crônica e não necessitam de tratamento com antiepiléticos. Crises ditas sintomáticas agudas correspondem a 40% dos casos incidentes de crises não febris.

Epilepsias são grupos de condições em que um distúrbio neurológico subjacente resulta em crises recorrentes não provocadas de forma crônica.

A ocorrência de duas ou mais crises nessas condições configura o diagnóstico de epilepsia. Existem vários tipos, causas e expressões clínicas da epilepsia, mas para alguns casos é possível identificar síndromes epiléticas específicas, para as quais é possível criar um grupo com características clínicas e eletroencefalográficas semelhantes. Para esses casos, é possível determinar etiologias específicas que respondem a tratamentos adequados com implicações prognósticas bem definidas.

Um em cada dez pessoas no mundo irá passar por uma crise pelo menos uma vez na vida, apesar de a maioria estar relacionada a eventos agudos. Por esta mesma razão, a maior incidência está em recém-nascidos e idosos, por conta de causas metabólicas, infecciosas ou encefalopáticas durante o período de neonatal e doenças degenerativas e cerebrovasculares em idosos. Diferença entre países desenvolvidos e em desenvolvimento.

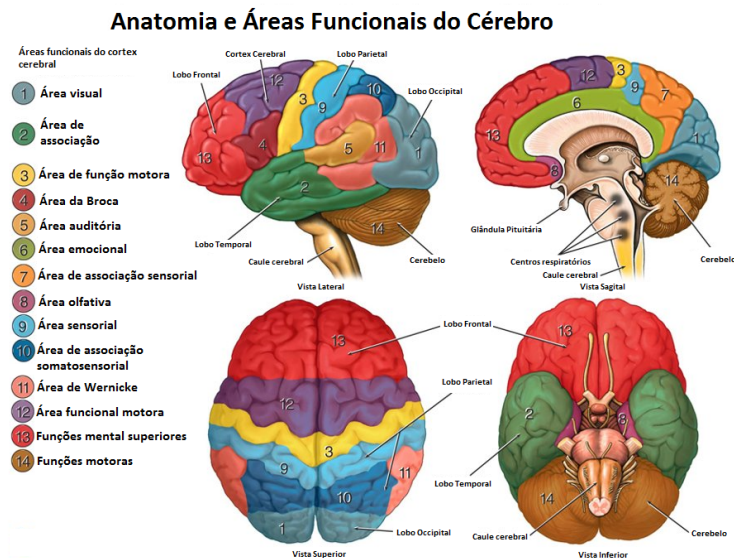
Para as crises sintomáticas agudas, os principais fatores de risco são: traumatismo craniano, AVE, doenças infecciosas, distúrbios tóxicos-metabólicos e abstinência de drogas e álcool.

Já para o desenvolvimento da epilepsia, os fatores de risco dependem da faixa etária. Na infância, exceto epilepsias hereditárias, podemos citar traumatismo craniano, infecções cerebrais, retardo mental, paralisia cerebral. Em adultos, os maiores responsáveis são: traumatismo craniano, infecções cerebrais, AVE e a doença de Alzheimer.

Em países em desenvolvimento, com na América Latina, em áreas endêmicas, a neurocisticercose é responsável por 10% dos novos casos de epilepsia.

Apenas 15% dos pacientes epiléticos possuem algum parente portador da síndrome, sendo que três quartos possuem apenas um familiar com epilepsia. Porém, estudos de base populacional mostram que o risco é maior para pacientes com parentes de primeiro grau (pais, irmãos) doentes.

As manifestações clínicas de crises varia consideravelmente em função do tipo de crise (motora, visual, auditória, vertiginosa, somatossensorial, etc.) e das áreas cerebrais afetadas (frontal, parietal, rolândica, córtex, etc.). A identificação precisa dos tipos de crise implicará na medicação necessária para o controle da doença.



1

Figura 1 - Áreas funcionais do cérebro

¹ Imagem traduzida da Texila American University

3.1.2. Crises Focais

Crises focais surgem em redes neuronais de uma região particular de um hemisfério cerebral, produzindo efeitos relacionados àquela área do córtex. Tais crises que implicam em perda de consciência ou de estado de alerta são chamadas de crises discognitivas.

É de grande valia introduzir também o conceito de aura. Aura consiste em sintomas sensoriais, autonômicos ou psíquicos que as pessoas experimentam no início de uma crise clínica. Como o senso comum atribui às crises epiléticas as convulsões e perdas de consciência, muitas vezes auras passam despercebidas por pacientes e médicos.

A identificação da aura é fundamental, pois ela aponta para a área cerebral de início das crises focais. Sendo assim, é possível prosseguir a partir de maiores investigações sobre anormalidades e disfunções cerebrais na área, como tumores ou que pode ser específica para tratamento cirúrgico. Além disso, crises focais tem impacto direto no tratamento e prognóstico.

A descarga neuronal de uma crise focal pode restringir-se à sua região de origem (como uma aura) ou propagar-se por outras regiões cerebrais. Independente da área de origem, crises focais podem atingir estruturas límbicas bilateralmente, causando alteração da consciência. Crises focais unilaterais podem também se propagar envolvendo áreas cerebrais bilaterais causando crise convulsiva.

A evolução de uma crise focal é diretamente refletida nas mudanças registradas em no EEG. Como o resultado de algumas crises podem ser sintomas breves ou não usuais, elas podem acabar passando despercebidas, sendo identificadas somente quando os sintomas estereotipados se apresentam e alterações são identificadas no EEG. O diagnóstico se complica ainda mais quando em cada crise, a descarga neuronal se propaga para regiões diferentes do cérebro, produzindo às mais diversas reações possíveis.

3.1.3. Crises Generalizadas

Afetam ambos os hemisférios cerebrais e tem sintomas mais definidos. As crises convulsivas podem começar nos dois hemisférios ao mesmo tempo ou ser uma decorrência de crises focais. Elas podem começar com um grito, contrações musculares, pausa respiratória, cianose e incontinência urinária seguido de contrações clônicas rítmicas dos 4 membros.

As crises de ausência são o segundo tipo mais comum de crises generalizadas. São lapsos breves e abruptos de perda de consciência. Os pacientes não têm percepção da ocorrência do evento e podem não perceber a passagem do tempo, embora percam a sequência do raciocínio. Isso ocorre sem sintomas de aura, já que perda e retomada de consciência são abruptas. Professores são os primeiros a notarem as crises desse tipo em crianças durante a infância.

3.1.4. Diagnóstico

Diagnóstico diferencial

O diagnóstico básico é dado pela história clínica, que quando cuidadosa é capaz de revelar se os episódios do paciente se tratam de crises epiléticas. Exames de EEG, imagens e estudos laboratoriais podem auxiliar na determinação do tipo de epilepsia, local de origem das crises focais.

O primeiro passo dos médicos é a determinação se os episódios descritos se tratam ou não de crises epiléticas. Existem três fatores principais capazes de distingui-las de outras causas de perda de consciência, como síncope:

- O contexto clínico, como o histórico médico e familiar, além das circunstâncias do ocorrido
- Fatores provocadores ou desencadeadores
- O primeiro sinal ou sintoma como a presença de aura/evidência de crise focal. A evolução após o início, o término e se existem déficits neurológicos após a crise.

Como pacientes costumam ter perda de consciência durante episódio e algumas vezes de perda de memória do ocorrido, a participação de acompanhantes na descrição do caso é fundamental.

Investigações diagnósticas

Uma história detalhada, registros de EEG e ressonância magnética fortalecem o diagnóstico para até 50% dos pacientes. O monitoramento contínuo por vídeo-EEG em uma unidade de internação de epilepsia pode aumentar a sensibilidade do diagnóstico.

Vamos em seguida detalhar apenas o exame de eletroencefalograma que possui alta relevância para nosso estudo.

3.1.5. Eletroencefalograma

O exame de EEG é a principal forma investigativa em todos os pacientes com crises e epilepsia. Ele é fundamental para se determinar a síndrome epiléptica e o determinado tratamento. Para epilepsias focais, o EEG ajuda na identificação das áreas de anormalidade.

O EEG pode determinar com certeza o diagnóstico de epilepsia se for realizado durante uma crise clínica e descargas elétricas esperadas forem identificadas. Ele pode falhar nessa identificação caso o foco da crise seja muito pequeno ou muito profundo, se as descargas. Ele sempre será efetivo durante crises de ausência e crises convulsivas generalizadas.

Entre crises, ele pode ajudar na avaliação da função cerebral geral e do tipo, localização e quantidade de descargas epileptiformes (espículas).

Para 60% das pessoas com epilepsia diagnosticada, o EEG inicial é normal. Esse número pode ser melhorado para pessoas com epilepsia focal quando o número de exames realizados passa para três ou mais, registrando assim anormalidades epileptiformes em 80% dos indivíduos. Em epilepsias generalizadas, as descargas anormais são mais facilmente identificáveis.

Nos casos de avaliação cirúrgica de pacientes epilépticos deve-se registrar o exame durante uma crise ao menos. Para pacientes em estado de coma, é altamente sugerido o monitoramento contínuo com vídeo-EEG.

3.1.6. Tratamento

O tratamento de crises e de epilepsia, como já foi citado anteriormente, está estritamente ligado à identificação do tipo de crise e síndrome epilépticas, mas também à probabilidade de crises recorrentes, incidência, gravidade das consequências físicas e psicossociais e se o benefício do tratamento supera substancialmente os possíveis efeitos colaterais.

Deve-se identificar e diminuir impreterivelmente todos e quaisquer tipos de fatores ambientais fisiológicos ou estilo de vida que aumente a probabilidade de crises, como a privação de sono.

Tratamentos cirúrgicos devem ser muito bem analisados, pois envolve a ressecção ou desconexão da região cerebral que contém o foco da crise. No entanto, muitos pacientes que passam pelo processo acabam tendo reincidência de crises e voltando para o tratamento com medicação.

3.1.7. Prognóstico da epilepsia

Em países em desenvolvimento, onde ainda pacientes epilépticos não recebem o diagnóstico e/ou não realizam o tratamento correto, observou-se que para 30 a 40% obtêm remissões depois de 5 a 10 anos sem tratamento. Em países desenvolvidos, onde a epilepsia é diagnosticada após duas crises não provocadas, a probabilidade de remissão depois de 5 anos é de 60% quando acompanhados durante 10 anos e de 70% quando acompanhados por 20 anos.

A duração da epilepsia ativa é uma das maiores preditores de remissão. Quanto mais as crises permanecerem não controladas depois do diagnóstico menores serão as chances de remissão.

4. As bases de dados

As bases de dados são de grande interesse para estudos deste tipo, pois além de ter uma amostra significativa para estudos estatísticos, é possível comparar o resultado dos métodos aplicados com o já conhecido diagnóstico concebido pelos médicos, que também é um dos campos da database.

4.1. Bases de dados

4.1.1. Database de Bonn

A base de dados de Bonn foi extraída em 2001 para Universidade de Bonn, na Alemanha e contém os resultados discretizados dos exames de EEG de pacientes saudáveis e epiléticos. Trata-se de uma base de dados pública disponibilizada pela própria Universidade de Bonn (Andrzejak RG et al, 2001).

O conjunto completo de dados consiste em cinco conjuntos (denotados A-E) cada um contendo 100 segmentos de EEG de canal único. Estes segmentos foram selecionados e cortados de registros EEG multicanais após inspeção visual de artefatos, como por exemplo, devido à atividade muscular ou movimentos oculares.

Os conjuntos A e B consistem em segmentos extraídos de gravações EEG de superfície que foram realizadas em cinco voluntários saudáveis usando um esquema padronizado de colocação de eletrodos (ver figura 2).

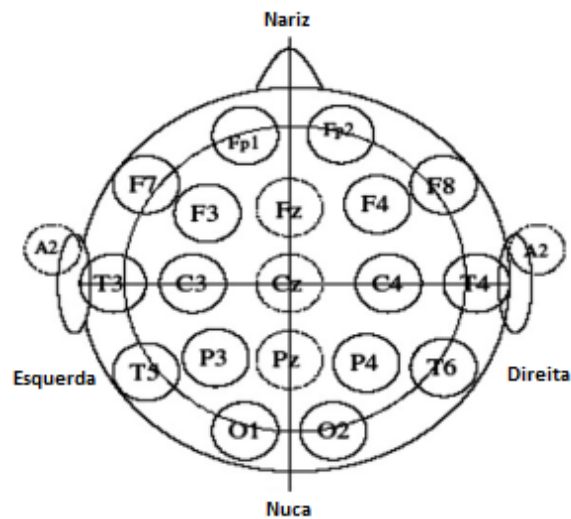


Figura 2 - Posicionamento dos eletrodos para EEG externo.

Os voluntários estavam em estado de vigília com os olhos abertos (A) e os olhos fechados (B), respectivamente. Os conjuntos C-E originados do arquivo EEG de diagnóstico pré-cirúrgico, ou seja, pacientes epiléticos com resistência a medicação que aguardavam cirurgia.

Foram selecionados EEGs de cinco pacientes, todos com controle completo de crises após ressecção de uma das formações do hipocampo, o que foi corretamente diagnosticado como o epileptogênica. Os segmentos no conjunto D foram registrados a partir da zona epileptogênica, e os do conjunto C da formação hipocampal do hemisfério oposto do cérebro.

Enquanto os conjuntos C e D continham apenas atividade medida durante os intervalos livres de convulsões (estado interictal), o conjunto E continha apenas pacientes epiléticos em atividade de convulsão (estado ictal). Aqui os segmentos foram selecionados de todos os locais de registro que exibem atividade ictal (Andrzejak RG et al, 2001).

Todos os sinais EEG foram gravados com o mesmo 128-Canal amplificador,

usando uma referência comum média. Os dados foram digitalizados em 173,61 Hz usando resolução de 12 bits. As configurações do filtro passa banda foram de 0,53-40 Hz (12 dB / oct).

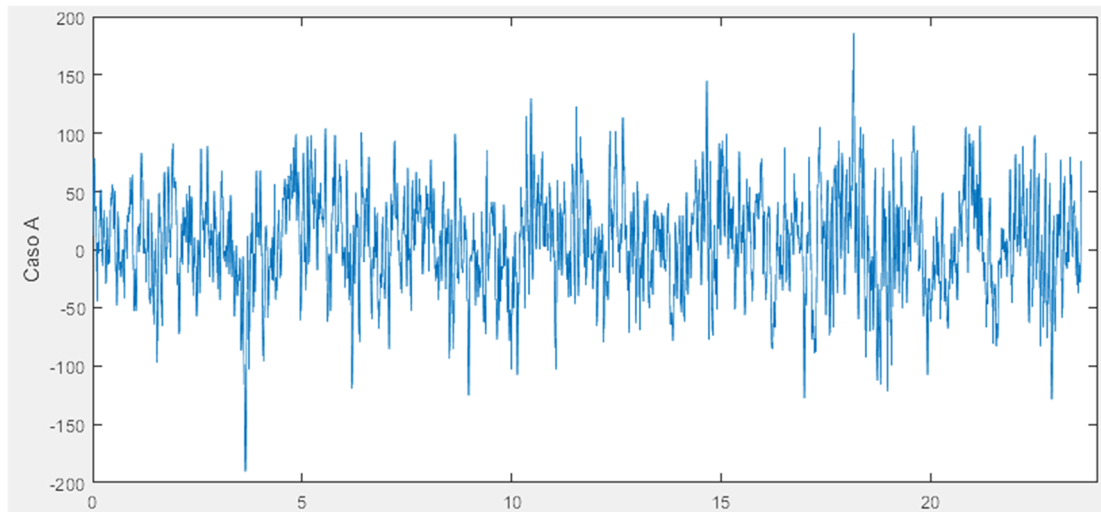


Figura 3 - Sinal resultado de EEG para o caso A – Univ de Bonn

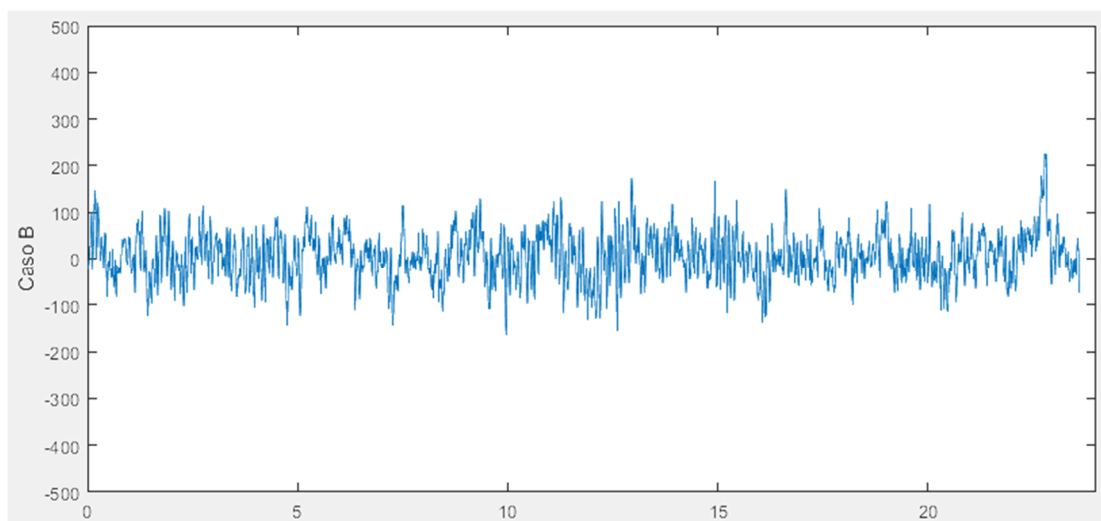


Figura 4 - Sinal resultado de EEG para o caso B – Univ de Bonn

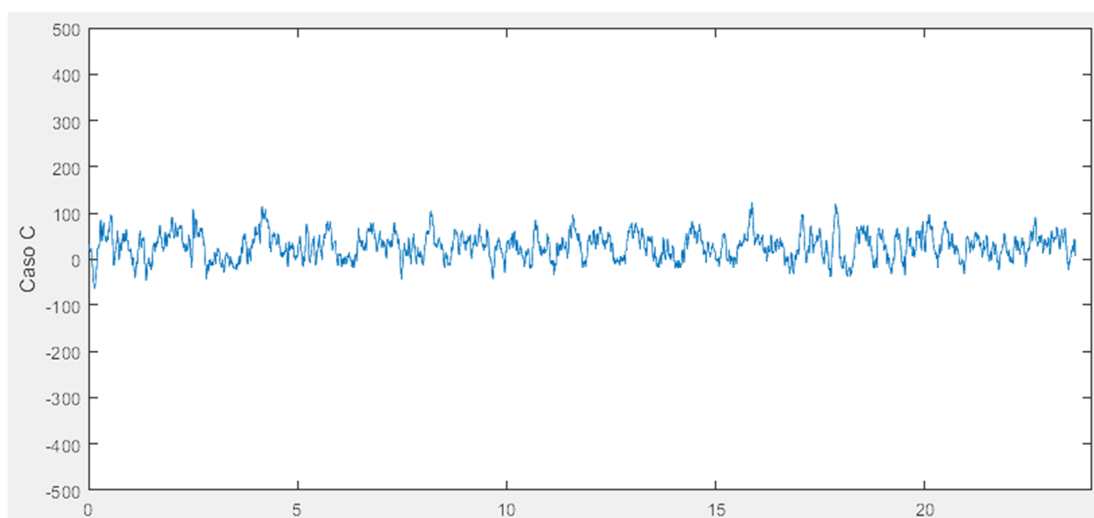


Figura 5 - Sinal resultado de EEG para o caso C – Univ de Bonn

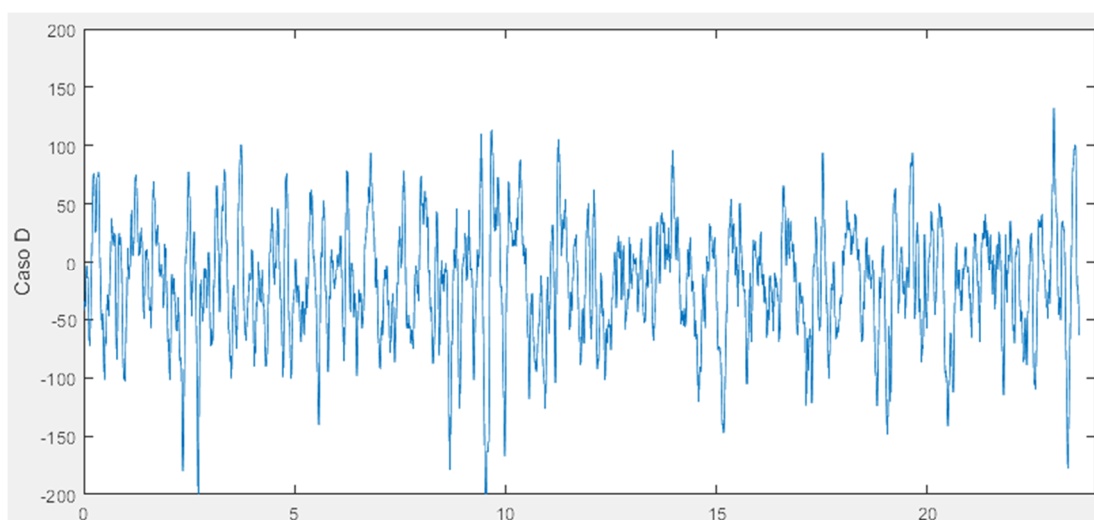


Figura 6 - Sinal resultado de EEG para o caso D – Univ de Bonn

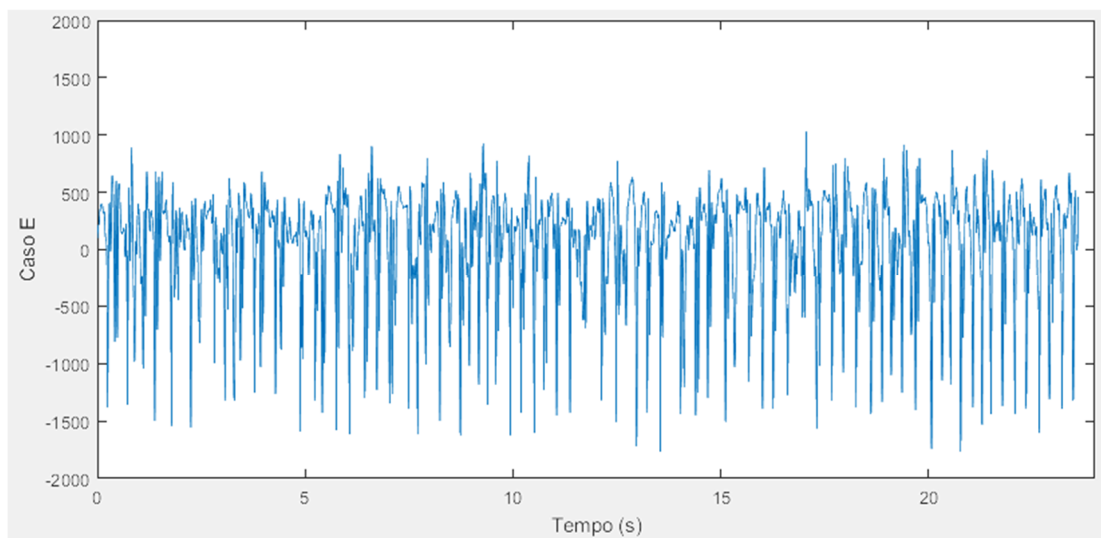


Figura 7 - Sinal resultado de EEG para o caso D – Univ de Bonn

4.1.2. Database CHT-MIT

Este banco de dados, coletado no *Children's Hospital Boston*, consiste de registros de EEG de pacientes pediátricos com convulsões intratáveis.

Os sujeitos foram monitorados por até vários dias após a retirada da medicação anti-convulsões, a fim de caracterizar essas convulsões e avaliar a sua candidatura para a intervenção cirúrgica.

As gravações, agrupadas em 23 casos, foram coletadas de 22 indivíduos (5 do sexo masculino, entre 3 e 22 anos e 17 do sexo feminino, entre 1,5 e 19 anos). Dois casos pertencem ao mesmo indivíduo do sexo feminino com um intervalo de 1,5 anos entre cada medição (Caso chb01 e chb21 respectivamente). Na tabela 1 podemos perceber a relação entre cada caso e o sexo e idade do indivíduo.

Tabela 1 - Dados de pacientes (Sexo e Idade)

Caso	Sexo	Idade	Caso	Sexo	Idade
CHB 01	F	11	CHB 13	F	3
CHB 02	M	11	CHB 14	F	9
CHB 03	F	14	CHB 15	M	16
CHB 04	M	22	CHB 16	F	7
CHB 05	F	7	CHB 17	F	12
CHB 06	F	1.5	CHB 18	F	18
CHB 07	F	14.5	CHB 19	F	19
CHB 08	M	3.5	CHB 20	F	6
CHB 09	F	10	CHB 21	F	13
CHB 10	M	3	CHB 22	F	9
CHB 11	F	12	CHB 23	F	6
CHB 12	F	2			

Cada caso (chb01, chb02, etc.) contém entre 9 e 42 arquivos .edf contínuos de um único objeto. Limitações de hardware resultaram em intervalos entre arquivos .edf numerados consecutivamente, durante os quais os sinais não foram gravados. Na maioria dos casos, as lacunas são de 10 segundos ou menos, mas ocasionalmente há lacunas muito mais longas.

Para proteger a privacidade dos sujeitos, todas as informações de saúde são protegidas, mas as relações de tempo entre os arquivos individuais pertencentes a cada caso foram preservadas. Na maioria dos casos, os arquivos .edf contém exatamente uma hora de sinais de EEG digitalizados, embora aqueles pertencentes ao caso chb10 tenham duas horas de duração, e aqueles pertencentes aos casos chb04, chb06, chb07, chb09 e chb23 têm quatro horas de duração, ocasionalmente, os arquivos em que as convulsões registradas são mais curtas (Shoeb A., 2009).

Todos os sinais foram amostrados a 256 Hz com resolução de 16 bits. A maioria dos arquivos contém 23 sinais EEG (24 ou 26 em alguns casos). O

sistema internacional 10-20 de posições de eletrodos EEG e nomenclatura foi usado para estas gravações (ver figura 8).

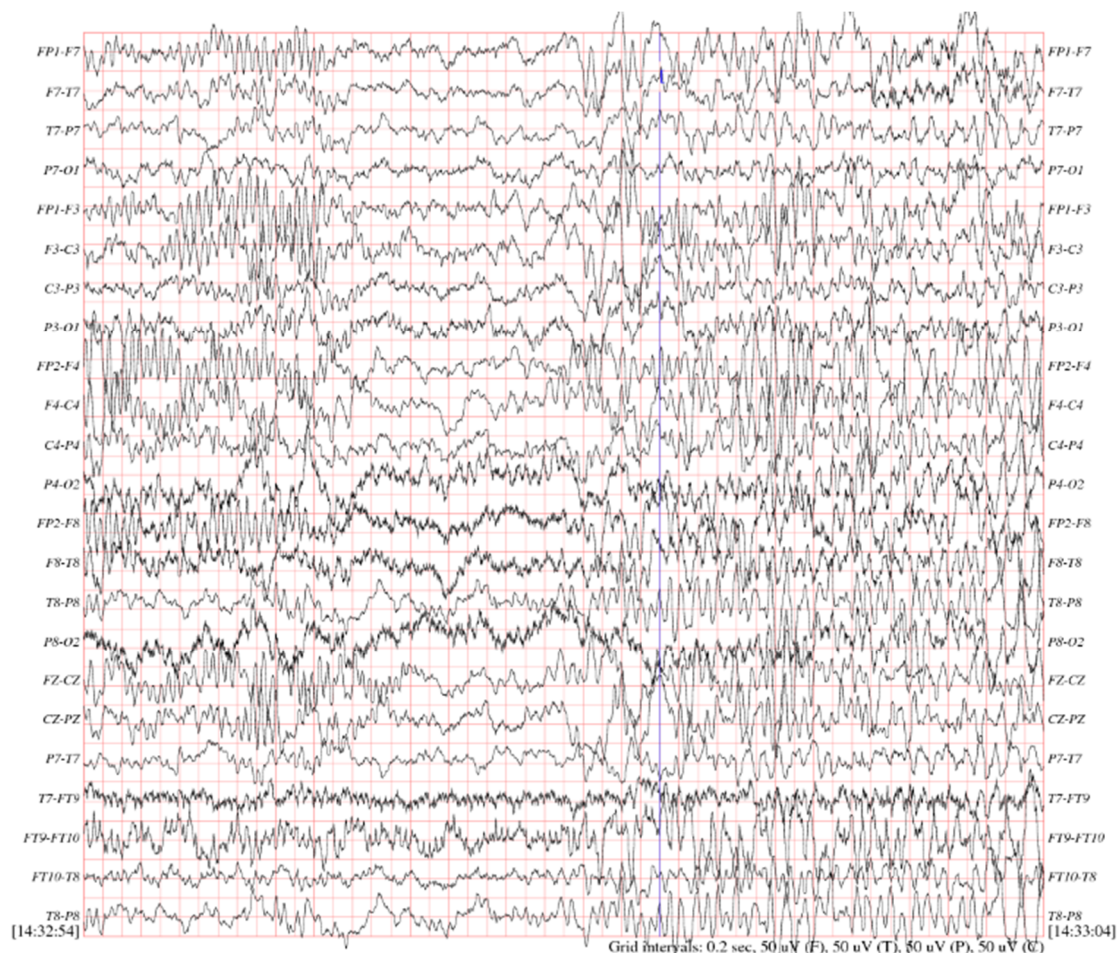


Figura 8 - Segmento da CHB-MIT database por eletrodo

4.2. Tratamento de sinais de exames de EEG

Diversos métodos matemáticos existem na literatura acadêmica para interpretação dos sinais de EEG. Alguns dos métodos foram revisados e testados neste trabalho de formatura e são apresentados a seguir.

4.2.1. Método das Entropias

Existem duas definições relacionadas de entropia: a definição termodinâmica e a definição da mecânica estatística, na qual nós daremos mais atenção nesse trabalho.

Historicamente, a clássica definição termodinâmica desenvolveu-se em primeiro lugar. No ponto de vista termodinâmico clássico, o sistema é composto por um grande número de constituintes (átomos, moléculas) e o estado do sistema é descrito pelas propriedades termodinâmicas médias desses constituintes; os detalhes dos constituintes do sistema não são diretamente considerados, mas seu comportamento é descrito por propriedades macroscopicamente médias como temperatura, pressão, capacidade de calor.

A descrição da mecânica estatística do comportamento de um sistema é necessária uma vez que a definição das propriedades de um sistema que utiliza a termodinâmica clássica se torna um método cada vez mais confiável de prever o estado final de um sistema que está sujeito a algum processo.

Em resumo, para a mecânica estatística, a entropia é um índice não-linear que reflete o grau de desordem de um sistema. Também é uma medida da taxa de informação gerada, que pode ser usada no processamento de sinal para separar o sinal desejado de qualquer ruído (Acharya, U. et al, 2015).

Geralmente, entropia de valor significativo corresponde a um sinal não regular e imprevisível, e uma entropia de baixo valor corresponde a um sinal regular. Como já foi visto no decorrer deste trabalho, o sinal EEG de paciente epiléticos é mais regular, logo de menor entropia que aqueles de pacientes saudáveis.

Na literatura existem diversos tipos de entropias que estão sendo estudadas para o pré diagnóstico da Epilepsia como por exemplo a Approximate Entropy, a Sample Entropy, a Fuzzy Entropy, a Wavelet Entropy e a Permutation Entropy. Neste trabalho vamos nos aprofundar na Entropia Approximate Entropy ou “Entropia Aproximada”.

4.2.1.1. Entropia Approximate Entropy (AppEn)

A AppEn é usada para avaliar a instabilidade da variação de um sinal, ela é capaz de detectar mudanças no comportamento episódico do paciente e compara essas mudanças com amostras de resultado conhecido através do comprimento do padrão (m) do coeficiente de similaridade (r). AppEn é uma medida invariante na escala porque o critério de similaridade é comparável ao conceito estatístico de desvio padrão dos dados.

Logo, uma série de tempo altamente irregular resulta em uma AppEn elevada e uma série de tempo com mais número de padrões semelhantes resulta em um valor baixo AppEn. Matematicamente, o ApEn é calculado por:

$$AppEn = \ln\left(\frac{C_m(r)}{C_{m+1}(r)}\right) \quad (1)$$

Onde $C_m(r)$ é a média padrão do comprimento m e $C_{m+1}(r)$ é a média padrão de comprimento $m+1$. A média padrão é calculada através da contagem de padrões semelhantes de comprimento m e comprimento $m+1$. Neste trabalho, escolhemos $m = 2$ e $r = 0,2$ vezes o desvio padrão do sinal EEG, pois a literatura mostra que esses dois parâmetros, definidos de forma empírica,

resultam em resultados mais precisos na leitura de sinais EEG. A média $C_m(r)$ é calculada por:

$$C_m(r) = (\text{quantidade de } x(j) \text{ onde } d[x(i), x(j)] \leq r) / (N + m - 1) \quad (2)$$

Onde d é a distância entre $x(i)$ e $x(j)$ e N é o tamanho da amostra. Assim, a AppEn depende do comprimento padrão, que é inconsistente e conta a sequência que corresponde a si mesma para evitar a condição $\ln(0)$.

A AppEn possui diversas vantagens; ela pode ser calculada para um período relativamente pequeno de série de dados ruidosos, como o EEG, ela pode ainda potencialmente diferenciar uma variedade de sistemas, como periódicos e múltiplos periódicos, sistemas caóticos e sistemas estocásticos.

As desvantagens da AppEN são; fortemente dependente do comprimento do sinal de entrada. Os sinais especialmente curtos tendem a um valor inferior ao esperado e sinais muito ruidosos dificultam a interpretação significativa desta entropia.

4.2.2. Método das Frequências

Todo o método apresentado a seguir tem como base o artigo publicado pela universidade chinesa de Chongqing, School of Automation, Chongqing University, Fu et al. (2014).

4.2.2.1. Introdução

A detecção de sinais de convulsão em exames de eletroencefalograma é fundamental para o diagnóstico médico de casos de epilepsia. Embora estes

sinais possam ser identificados com clareza, a ocorrência destes costuma ser bastante irregular e imprevisível tornando a leitura de exames de EEG (*electroencephalogram*) um processo um tanto quanto fastidioso e consumidor de esforços.

Uma análise completa de um exame requer a atenção e concentração total de um expert em leitura de sinais ao longo de toda a extensão dos resultados.

Embora diversas iniciativas tenham surgido no meio, os sinais resultantes de um exame de EEG é de uma complexidade alta dado seu padrão não linear e não estacionário. Tais padrões fazem de análises como os métodos baseados na transformada de Fourier terem certas limitações, visto que estas assumem por definição o estudo de um sinal estacionário.

A leitura e estudo de sinais a partir dos resultados de tais exames continua sendo, no entanto, de vital importância para o diagnóstico. Métodos de tratamento de sinais em frequência auxiliam na identificação de padrões comuns em sinais de pacientes epiléticos.

4.2.2.2. Método

Com uma abordagem direcionada ao estudo de tempo-frequência e segmentação de padrões, o método propõe o tratamento de sinais através da “Transformada de Hilbert–Huang” (HHT). Esta ferramenta permite observação do sinal tanto em informações relacionadas ao domínio da frequência quanto ao domínio do tempo. Baseado em um método empírico de decomposição (empirical mode decomposition, EMD), o conjunto de dados é decomposto em número finito de funções de modo intrínseco (ou em inglês, intrinsic mode

functions - IMFs).

4.2.2.3. Transformada de Hilbert-Huang

A transformada de Hilbert-Huang é um método de análise de dados baseado em dados empíricos. Comparado com o método de análise de frequência de tempo tradicional, como STFT, HHT consegue analisar consistentemente sinais não lineares e não estacionários. A técnica HHT para análise de dados consiste em dois componentes: EMD e análise espectral de Hilbert (Hilbert Spectral Analysis, ou HSA).

O princípio do EMD é decompor um sinal não linear e não estacionário em um conjunto de IMFs que tem banda limitada. Cada IMF satisfaz duas condições básicas:

- O número de pontos extremos Ne e o número de raízes Nz devem ser distanciados de no máximo 1 unidade:

$$(Nz - 1) \leq Ne \leq (Nz + 1) \quad (3)$$

- Em qualquer tempo t_i , o valor de média local do envelope que é definido pela média do máximo $f_{max}(t)$ e o mínimo $f_{min}(t)$ do envelope é zero.

$$\frac{[f_{max}(t_i) + f_{min}(t_i)]}{2} = 0, t_i \in [t_a, t_b] \quad (4)$$

As propriedades do IMF permitem a definição de frequência e instantâneas de

uma forma inequívoca, de modo que a transformação de Hilbert pode ser aplicada a cada modo intrínseco.

O algoritmo EMD para um sinal $x(t)$ pode ser implementado da seguinte maneira:

- Determinar os extremos máximos e mínimos do sinal $x(t)$
- Gerar envelopes superiores e inferiores, $x_{up}(t)$ e $x_{low}(t)$, conectando todos os máximos e mínimos locais com uma interpolação spline cúbica
- Calcular a média point-by-point como

$$m(t) = \frac{[x_{up}(t) + x_{low}(t)]}{2} \quad (5)$$

- Subtrair a média local do sinal:

$$d(t) = x(t) - m(t) \quad (6)$$

- Decidir se $d(t)$ é um IMF ou não verificando as duas condições básicas descritas anteriormente. Caso os critérios das condições tenham sido satisfeitos, então $imf_i = d(t)$. Caso contrário, $x(t) = d(t)$, e o processo deve ser reiterado
- Sendo $r(t) = r(t) - imf_i$, repetir as etapas anteriores até que $r(t)$ seja um resíduo monotônico, caso contrário $x(t) = r(t)$ e o processo deve recomeçar

No final deste processo, o sinal $x(t)$ pode ser expresso da seguinte forma:

$$x(t) = \sum_{i=1}^N c_i(t) + r_N(t) \quad (7)$$

Onde N é o número de modos intrínsecos, $c_i(t)$ é o i -ésimo FMI, em $r_N(t)$ é o resíduo final que pode ser interpretado como o componente DC do sinal.

A segunda da parte do método envolve a análise espectral de Hilbert. Com o objetivo de computar instantaneamente frequências e amplitudes e descrever o sinal localmente, a transformada de Hilbert é aplicada para cada IMF obtido pelo EMD. Para qualquer sinal $x(t)$ de classe L^p , sua transformada de Hilbert $y(t)$ é:

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \quad (8)$$

Onde P é o valor principal de Cauchy da integral única. Podemos construir a função analítica:

$$z(t) = x(t) + jy(t) = a(t)e^{j\phi(t)} \quad (9)$$

A amplitude do pré-envelope $a(t)$ e a fase instantânea $\phi(t)$ são definidas como:

$$a(t) = \sqrt{x(t)^2 + y(t)^2} \quad (10)$$

$$\phi(t) = \arctan \frac{y(t)}{x(t)} \quad (11)$$

A frequência instantânea pode então ser escrita como uma derivada no tempo da fase:

$$w(t) = \frac{d\phi(t)}{dt} \quad (12)$$

Ou

$$f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} \quad (13)$$

Uma representação de tempo-frequência (time-frequency representation, ou TFR) é uma perspectiva do sinal representado ao longo do tempo e da frequência e é muito útil para a análise de sinais não estacionários. As TFRs são frequentemente representadas pela amplitude ou densidade de energia ao longo do tempo e da frequência.

A distribuição de densidade de energia denominada espectro de Hilbert-Huang do sinal original pode ser obtida por HHT. As funções de amplitude e frequência são expressas como funções do tempo que podem ser construídas como $H(f, t)$. Ela exhibe as contribuições de energia relativa para uma frequência específica em um momento específico. O espectro de Hilbert-Huang é definido como:

$$H(f, t) = \text{Re} \sum_{i=1}^n a_i(t) \exp(j \int f_i(t) dt) \quad (14)$$

Onde a_i e f_i são amplitude e frequência instantâneas do i -ésimo FMI, respectivamente. A distribuição tempo-frequência do sinal EEG pode ser considerada como um TFI, e assim o método de processamento de imagem pode ser usado para lidar com o problema de classificação de ataques epiléticos.

O próximo passo é a segmentação da imagem de tempo-frequência para localizar estruturas significantes e extrair diferentes componentes do sinal. O objetivo da segmentação TFI dos sinais EEG é obter regiões que correspondam às bandas de frequência dos ritmos. O TFI pode ser dividido em 5 sub-imagens correspondentes a tais bandas. O TFI é convertido então em uma imagem de escala cinza de 8-bit para obter-se um histograma de

escala cinza neste método. A classificação dos principais ritmos do EEG baseada nos intervalos de suas frequências:

Tabela 2 - Faixa de frequências por banda

Banda	Frequência
Delta	0–4 Hz
Teta	4–8 Hz
Alfa	8–12 Hz
Beta	12–30 Hz
Gama	30–50 Hz

O histograma é uma representação gráfica da distribuição de dados. O histograma de uma imagem representa a distribuição dos pixels na imagem sobre a escala. O histograma da imagem pode nos dar uma visão intuitiva e informações detalhadas sobre a classificação de sinais EEG. Considerando uma imagem I , na qual a intensidade em pixel com coordenadas (x, y) pode ser representada como $I(x, y)$. Para o histograma h , o h_i indica que a intensidade i aparece h_i vezes na imagem. Então o h_i pode ser definido como:

$$h_i = \sum_x \sum_y^i I(x, y) \quad (15)$$

A forma no histograma da sub-imagem fornece-nos informações detalhadas para detectar ataques epiléticos a partir dos sinais EEG. Quatro funções estatísticas, incluindo média, variância, assimetria e curtose foram retiradas do histograma de sub-imagens de 8 bits em tons de cinza de sinais EEG. O uso dessas características é interessante pelo fato de que a distribuição do histograma é muitas vezes caracterizada pelo seu nível de dispersão, assimetria e concentração em torno da média. Esses recursos são definidos da seguinte forma:

$$\mu = \frac{1}{L} \sum_{i=0}^{L-1} h_i \quad (16)$$

$$\alpha = \frac{1}{\sigma^3} \sum_{i=0}^{L-1} (h_i - \mu)^3 \quad (17)$$

$$\sigma^2 = \frac{1}{L-1} \sum_{i=0}^{L-1} (h_i - \mu)^2 \quad (18)$$

$$\beta = \frac{1}{\sigma^4} \sum_{i=0}^{L-1} (h_i - \mu)^4 \quad (19)$$

Onde σ é o desvio padrão, h_i é a intensidade do i -ésimo pixel no histograma e L é o número de níveis de intensidade na sub-imagem de escala de cinza.

A média do histograma pode refletir os valores médios das imagens em escala de cinza, enquanto a variância mede quanto os valores estão espalhados.

Assimetria pode refletir o grau de assimetria do histograma. Se o histograma for simétrico, então a assimetria é zero. Se a cauda da mão esquerda for mais longa, a assimetria será negativa e se a cauda da mão direita for mais longa, a assimetria será positiva.

Curtose é uma medida para o grau de planicidade na distribuição variável. Alta curtose tende a ter um pico distinto perto da média, declínio bastante rápido, e têm caudas pesadas, enquanto curtose baixa tende a ter um topo plano perto do meio do que um pico afiado. A curva normal padrão do histograma tem uma curtose de zero.

4.2.3. Método da Transformada de Fourier

Se a informação disponível sobre o sinal consiste nas amostras $\{x(n)\}_{n=1}^N$, a

periodograma de estimação espectral é dada por:

$$\hat{P}_{PER}(f) = \frac{1}{N} |\sum_{n=1}^N x(n) \exp(-j2\pi fn)|^2 \quad (20)$$

Onde $\hat{P}_{PER}(f)$ é o periodograma de estimação espectral, f é a frequência do sinal EEG e N é o número total de amostras. No método de Welch, os sinais são divididos em segmentos sobrepostos, cada segmento de dados é 'windowed', periodogramas são calculados e então a média de periodogramas pode ser calculada. $\{x_l(n)\}, l = 1, \dots, S$ são os segmentos da database e cada segmento tem comprimento igual a M . Então a sobreposição é frequentemente adotada como 50%. O Espectro Estimado de Welch é dado por:

$$\hat{P}_w(f) = \frac{1}{S} \sum_{l=1}^S \hat{P}_l(f) \quad (21)$$

$$\hat{P}_l(f) = \frac{1}{M} \frac{1}{P} |\sum_{n=1}^M v(n) x_l(n) \exp(-j2\pi fn)|^2 \quad (22)$$

Onde $\hat{P}_w(f)$ é o periodograma estimado do l -ésimo segmento e $v(n)$ é a 'data-window', P é a média total de $v(n)$ e P é dado por:

$$P = \frac{1}{M} \sum_{n=1}^M |v(n)|^2 \quad (23)$$

$\hat{P}_w(f)$ é o Welch PSD estimado, M é o comprimento de cada segmento de sinal e S é o número de segmentos.

Logo, a avaliação do Welch PSD estimado na frequência das amostras é dada pela Transformada de Fourier Discreta (DFT):

$$X(k) = \sum_{n=1}^N x(n) \exp\left(-j \frac{2\pi}{N} nk\right), \quad k = 0, \dots, N-1 \quad (24)$$

Onde $X(k)$ é o coeficiente da DFT, N é o tamanho da database $x(n)$ (Polat K et Günes S, 2007).

4.2.4. Machine Learning

4.2.4.1. Classificação – SVM

SVM é uma técnica de aprendizado de máquina, fundamentada nos princípios da Minimização do Risco Estrutural. Esta técnica busca minimizar o erro com relação ao conjunto de treinamento (risco empírico), assim como o erro com relação ao conjunto de teste, isto é, conjunto de amostras não empregadas no treinamento do classificador (risco na generalização). O objetivo de SVM consiste em obter um equilíbrio entre esses erros, minimizando o excesso de ajustes com respeito às amostras de treinamento (overfitting) e aumentando conseqüentemente a capacidade de generalização (VAPNIK, 1999). O problema denominado de overfitting consiste em o classificador memorizar os padrões de treinamento, gravando suas peculiaridades e ruídos, ao invés de extrair as características gerais que permitirão a generalização ou reconhecimento de padrões não utilizados no treinamento do classificador (SMOLA et al., 2000).

A máquina de vetor de suporte é baseada na teoria de dimensão em aprendizado estatístico de Vapnik-Chervonenkis e resolve o problema de uma amostra pequena. A SVM maximiza a margem a partir da determinação de um hiperplano separador para identificar diferentes classes de dados. Para um problema de duas classes, considere um dado conjunto teste $\{x_k, y_k\}_{k=1}^N$ com dados de entrada $x_k \in R^n$, e dados de saída $y_k \in R$ com rótulos de

classe $y_k \in \{-1, 1\}$. O seguinte classificador é considerado:

$$y(x) = \text{sign}[w^T x + b] \quad (25)$$

Para garantir que falsas classificações sejam toleradas, o hiperplano ótimo de separação deve garantir a condição:

$$y_k[w^T x + b] \geq 1 - \xi_k, \quad k = 1, \dots, N \quad (26)$$

Onde $\xi_k > 0$ são variáveis de folga.

O método SVM é feito dentro de um contexto de teoria de otimização convexa. Para obter o hiperplano ótimo, precisamos resolver o seguinte problema inicial em w e ξ_k :

$$\min J_p(w, \xi) = \frac{1}{2} w^T w + c \sum_{k=1}^N \xi_k \quad (27)$$

Onde c é uma constante real positiva que permite a ponderação na exclusão de amostras mal classificadas.

O problema no espaço de peso primário é um problema de otimização restringido, em seguida formulando o Lagrangeano, para então implementar as condições para a solução ótima e finalmente resolver o problema no espaço dual. O classificador SVM não linear pode assumir a forma:

$$y(x) = \text{sign}\left[\sum_{k=1}^N \alpha_k y_k K(x, x_k) + b\right] \quad (28)$$

Onde α_k são multiplicadores de Lagrange, $K(x, x_k)$ é uma função de núcleo,

que nos permite trabalhar em espaços de característica dimensionais enorme sem a necessidade de ter que fazer cálculos explícitos neste espaço. O núcleo RBF é usado neste método, pode ser formulado como:

$$K(x, x_i) = \exp\left(\frac{-\|x-x_i\|^2}{2\sigma^2}\right) \quad (29)$$

Onde SIGMA controla a largura do núcleo RFB.

A partir dos resultados das características dos sinais de EEG obtidos inicialmente, aplicados na SVM pode-se obter um resultado de classificação.

4.2.4.2. Classificação - COMPLEX TREE

A classificação por árvores (*decision tree*, em inglês) apesar de ser um método simples é amplamente utilizado na literatura em problemas de classificação. Podemos imaginar este método como uma série de perguntas que ajudam a indicar se um determinado conjunto de características faz parte de um grupo A, B, C, etc.

Vamos tomar o caso de um experimento que se proponha a identificar se um livro de autor desconhecido foi escrito por algum pensador famoso. A partir de uma série de perguntas que já se conhecem as respostas para as obras de tal pensador, vamos usar o método para tentar decidir se o livro pode pertencer ou não ao autor especificado.

Poderíamos nos perguntar se a problemática do livro desconhecido é tema recorrente deste autor. Ou então, se a maneira de apresentar os argumentos é parecida. Ou ainda, se os elementos utilizados são contemporâneos ao autor. Através das respostas a todas essas perguntas, iríamos refinar nossa

hipótese inicial e afirmar com certa certeza se ela deve ser mantida ou descartada.

Certamente, quanto maior for a base de dados de livros que possuímos do autor e quanto maior o número de perguntas que soubermos a resposta, mais preciso sério nosso julgamento sobre o livro desconhecido.

Não necessariamente precisaremos realizar todo nosso portfólio de perguntas para chegar a uma conclusão se nossa classificação está correta, mas ao mesmo nem sempre elas serão suficientes para nos dar 100% de certeza sobre nossas hipóteses.

Chamaremos de nó-raiz (*root node*) a primeira pergunta que faremos (ou seja, o primeiro critério que gostaríamos de testar). Esta pergunta pode nos levar por dois caminhos: a negativa ou a confirmação.

Um nó-interno (*internal node*) é uma pergunta subsequente ao nó-raiz ou a outro nó-interno, que por sua vez nos levará a outras duas possíveis conclusões. Um terminal ou folha (*terminal* ou *leaf*) é um fim de galho (*branch*), ou caminho, desta nossa árvore.

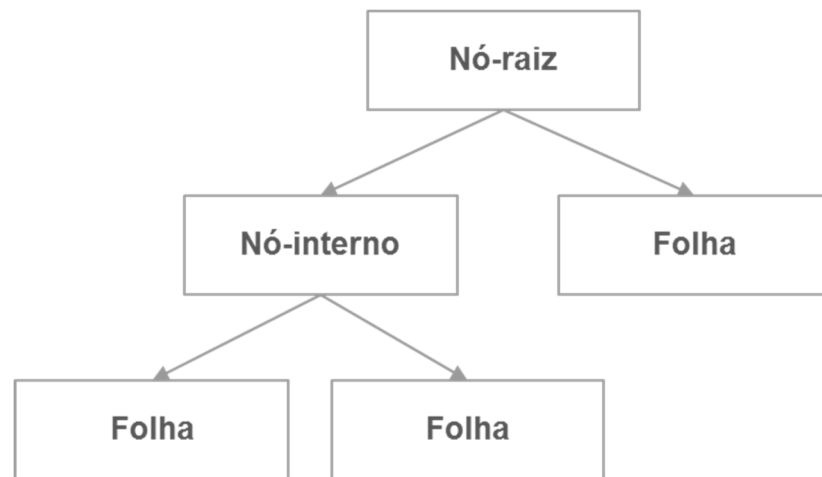


Figura 9 - Exemplo de árvore

Em princípio, poderíamos pensar que existem diversas árvores que poderiam ser montadas através de diversos arranjos das perguntas. As árvores mais interessantes seriam aquelas nos levariam a folhas mais rapidamente, pois são nelas que encontramos as respostas que queremos. Encontrar uma árvore ótima não é uma tarefa simples quando percebemos que o formato de resposta binária das árvores nos resulta em um tamanho exponencial de pesquisa.

Não entraremos no detalhe da comparação entre diferentes métodos de árvores de decisão, pois este não é o objetivo deste trabalho. Através do software MatLab, foi utilizado um classificador chamado de complex tree, ou árvore complexa.

5. METODOLOGIA

5.1. Estrutura de diagnóstico

5.1.1. Métodos das Frequências

Em conjunto a esta metodologia, o uso de máquina de vetor de suporte (support vector machine, ou SVM), uma técnica de classificação baseada em *machine learning*, foi empregada. SVM mostrou bons resultados na identificação ataques epilépticos, pois foi desenvolvida como um classificador binário, possuindo assim vantagens em classificações binárias. Ela constrói um hiperplano separador ótimo no espaço apresentado e faz com que a máquina aprenda e ofereça um ponto ótimo como resposta.

Para este tipo de análise, a representação de tempo-frequência (TFR) baseada em HHT foi considerada como tempo-frequência Imagem (TFI). A segmentação do TFI foi implementada com base nas bandas de frequência dos ritmos de sinais de EEG.

São extraídas também funções estatísticas, incluindo média, variância, assimetria e curtose da intensidade de pixels no histograma da imagem segmentada em escala de cinza em tempo-frequência (time-frequency image, ou TFI).

A partir de ferramentas estatísticas, um teste de hipóteses com menores valores de p indica que as quatro formas de onda (teta, alfa, beta e gama) com suas características de média, variância e assimetria são altamente determinantes. As características ótimas das ondas teta, alfa, beta e gama são alimentadas no SVM com núcleo de função radial base (radial basis

function, RBF) para a classificação de sinais de EEG de convulsões ou não. Os experimentos são conduzidos em dez trilhas independentes para testar o desempenho do método proposto.

A curva de precisão de classificação e características de operação do receptor (ROC) foi utilizada para avaliar a precisão do método. Os resultados mostram que esta pode atingir uma média de 99,125% e proporcionar uma precisão de classificação melhor do que algumas abordagens estudadas previamente.

O diagrama a seguir é apresentado no artigo do método e retrata a metodologia passo-a-passo para o chamado método das frequências:

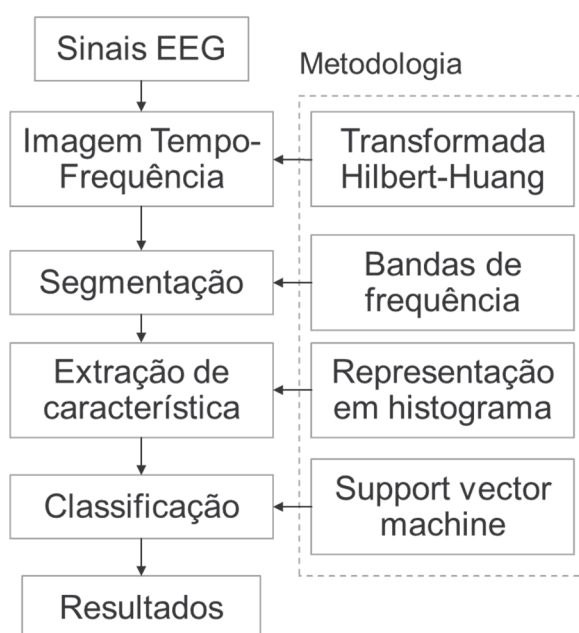


Figura 10 - Fluxograma do método das Frequências

5.1.2. Métodos da Transformada de Fourier

Este método proposto consiste de dois estágios: método baseado em FFT

Welch e classificador de SVM. Quando se aplica o método de análise espectral de Welch aos sinais EEG obtém-se 129 características a partir do método Welch baseado em FFT foram aplicadas ao classificador de SVM para a detecção de crises epiléticas. O fluxograma do método proposto é apresentado na figura 11.

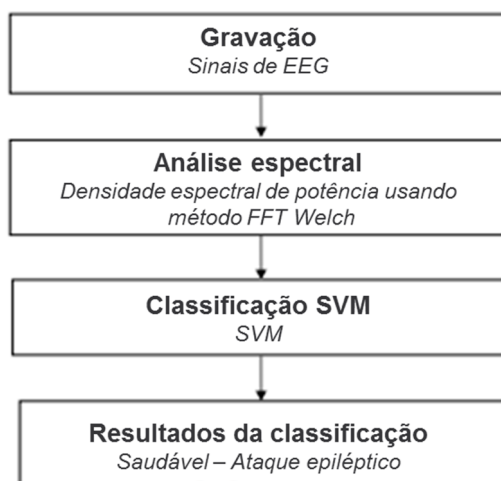


Figura 11 - Fluxograma do método da Transformada de Fourier

A forma como os resultados dos métodos de tratamento de sinal se relacionam com a classificação para separar pacientes doentes de saudáveis é dada na seção resultados deste relatório.

5.1.3. Método das Entropias

Para utilização da Entropia AppEn nesse trabalho os sinais EEG Primeiro, são divididos em dois conjuntos (50% / 50%), um conjunto de "treinamento" e um conjunto de validação. Este primeiro conjunto foi usado para direcionar os parâmetros do modelo e o segundo conjunto foi usado para aplicar o modelo e avaliar seu desempenho. Ou seja, a AppEn é calculada para a totalidade da base, 50% das entropias é usada para calibrar o classificador, no caso o

Complex Tree, já descrito anteriormente, e em seguida o restante da base é tratado pelo classificador que atribui o “diagnóstico” a base, resultado esse que é confrontado com os resultados pré-estabelecidos para Universidade de Bonn e assim a eficiência do método pode ser definida.

O fluxograma do método proposto é dado na Figura 12.

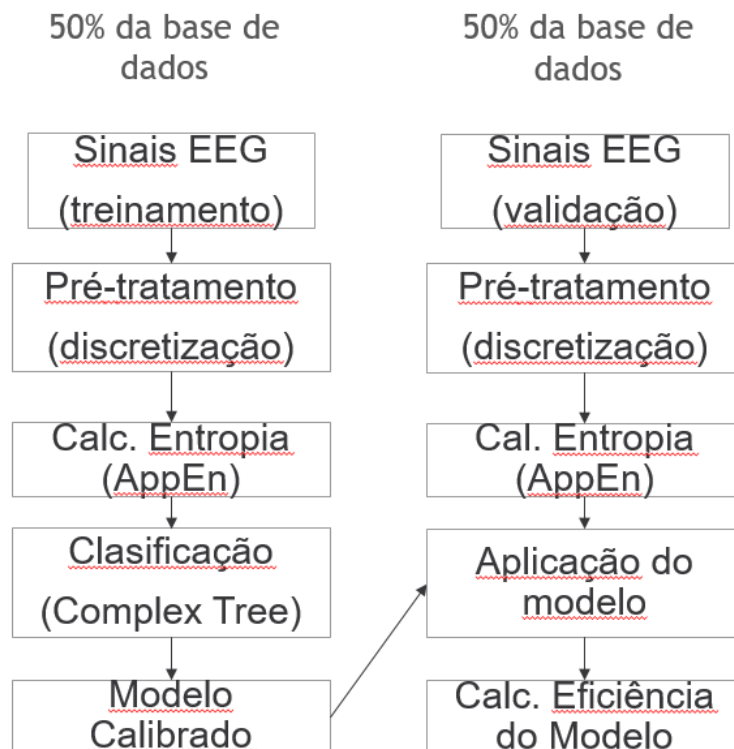


Figura 12 - Fluxograma do método das entropias.

5.2. MATLAB

Todos os métodos e classificadores aqui apresentados foram implementados no software MATLAB. O MATLAB, abreviação para MATrix LABoratory, é um software com linguagem simples e direta, que realiza cálculos de engenharia, e por isso possui uma ampla biblioteca de funções matemáticas pré-definidas.

Todos os códigos utilizados para demonstrar os resultados desse trabalho podem ser encontrados no apêndice deste documento.

6. RESULTADOS E DISCUSSÃO

6.1. Resultados Método de Frequências + SVM Quadrático

A aplicação do Método de Frequências em um primeiro trabalho realizado anteriormente na *Ecole Centrale de Marseille*, França, (Fontenelle et al., 2015), permitiu seu desenvolvimento através do software MatLab e o teste de sua eficiência.

O estudo foi baseado em exames de EEG discretizados, caracterizados por um conjunto de valores e com uma frequência de amostragem. O trabalho é baseado em dados da Universidade de Bonn em que os exames são apresentados em formato .txt, e a frequência de amostragem é 173,61 Hz. Cada arquivo tem 4097 números inteiros.

Seguida a primeira leitura da base de dados, a decomposição é aplicada em forma modal intrínseca. Para isso, foi utilizado um módulo disponível em *MATLAB Central*. Para cada uma das funções, foram aplicadas transformadas de Hilbert para calcular os valores de frequência e intensidade em cada instante.

Para discretizar os valores de frequência e intensidade, de modo a criar a TRI desejada, com a intensidade como uma função de tempo e frequência na forma de uma imagem. Tempo e frequência foram indexados em uma matriz e a intensidade das cores em um elemento (i, j). As frequências são apresentadas em 10 bits (1024 valores).

Estamos interessados em valores que não são zero, então a matriz é dividida em bandas de frequência: delta (0-4 Hz), teta (08/04 Hz), alfa (8-12 Hz), beta

(12 a 30 Hz), gama (30-50 Hz). O programa resgata os valores diferentes de zero para cada banda e os coloca em um vetor. Em seguida, é feita a amostragem em intensidades de 8 bits (256), com uma função *round_vector* criada pelo projeto.

Para cada vetor, parâmetros estatísticos foram calculados: média, variância, assimetria, e curtose usando a função *fund_stats*. O método é então replicado para cada paciente e série, gravando os resultados em um arquivo *.mat* (um para cada paciente). Com esses dados, foi aplicada a função *Classification Learner*, que para ser usada teve-se de embaralhar aleatoriamente os dados de 100 séries, com o seguinte código na tabela de Mnew.

Estes dados são tabelas de 100x5 para cada banda de frequência e cada paciente. Cada linha corresponde a uma série e as colunas de 1 a 4 aos parâmetros estatísticos calculados. A quinta coluna é a que contém a informação sobre o estado de saúde do paciente, -1 não é doente ou se 1 doente.

Finalmente, lança-se o método de classificação de aprendizagem automática SVM para a tabela Mnew através da função *Classification Learner*. As colunas de 1 a 4 são utilizadas como preditores, e coluna 5 como uma resposta.

Para o conjunto destas 500 séries, temos como resultado uma comparação entre os diversos classificadores:

Tabela 3 - Comparação de classificadores para o método de frequências

Classificador	Eficácia do método
Árvore complexa	97.0%
SVM linear	97.4%
SVM quadrático	98.2%

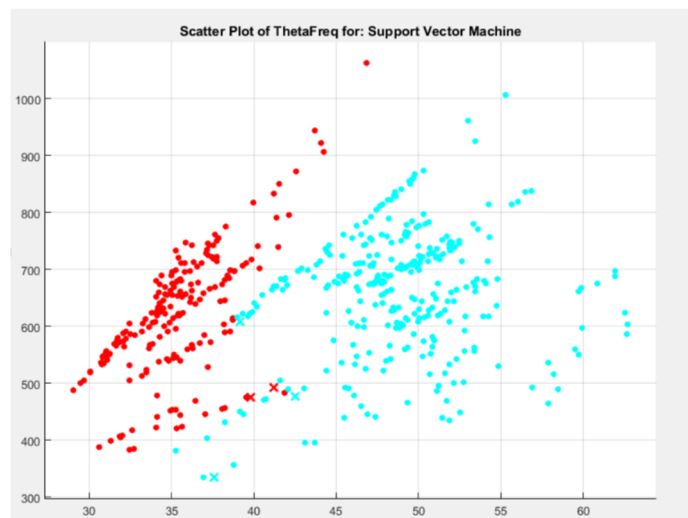


Figura 13 - Dispersão de pontos com o método de frequências e classificador SVM Quadrático

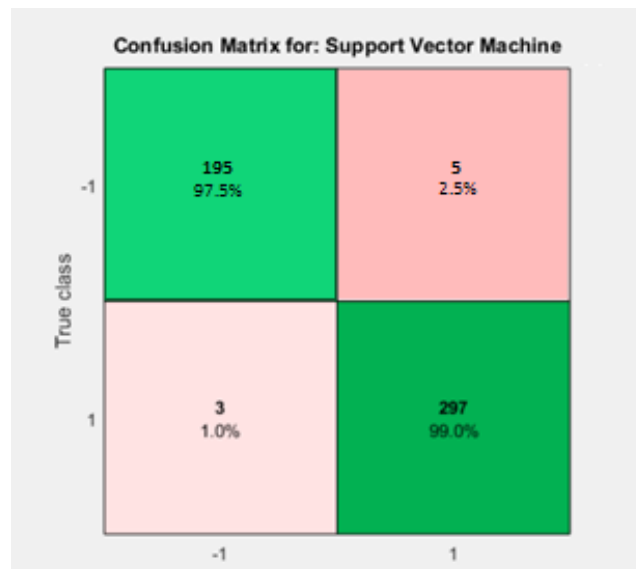


Figura 14 - Matrix de Conflito para método das entropias e classificador SVM Quadrático

6.2.Resultados Método da Transformada de Fourier + Árvore complexa

Em um primeiro momento apenas a Database de Bonn foi usada para testar esta metodologia.

Foram utilizados como base os casos A ao E, saudáveis de olhos abertos e epiléticos em estado convulsivo. Esta abordagem foi possível apenas pois a database de Bonn oferece um conjunto de sinais pré-tratado e discretizado como foi visto no item 4.1.1. Cada caso é apressado em formato .txt com 100 segmentos e a frequência de amostragem é 173,61 Hz. Cada arquivo tem 4097 números inteiros que representam os picos de sinal do EEG.

Aplicando-se o método da Transformada de Fourier para cada caso, ou seja, 409700 pontos para o caso A e 409700 pontos para o caso E, com 100 segmentos de comprimento 4097 podemos calcular a PSD para cada caso, com 129 pontos em cada. O cálculo teórico da PSD foi abordado anteriormente neste relatório na Revisão do Material Teórico e o código de implementação do método em MATLAB se encontra no apêndice deste documento. Nas figuras 15 e 16 podemos ver o PSD para os casos A e E.

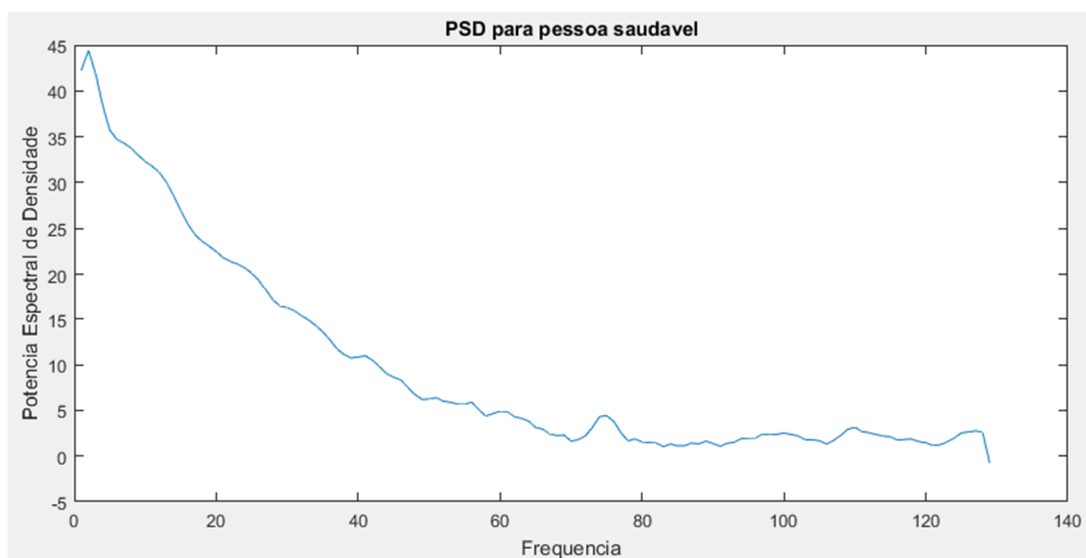


Figura 15 - Potência Espectral de Densidade para o caso A

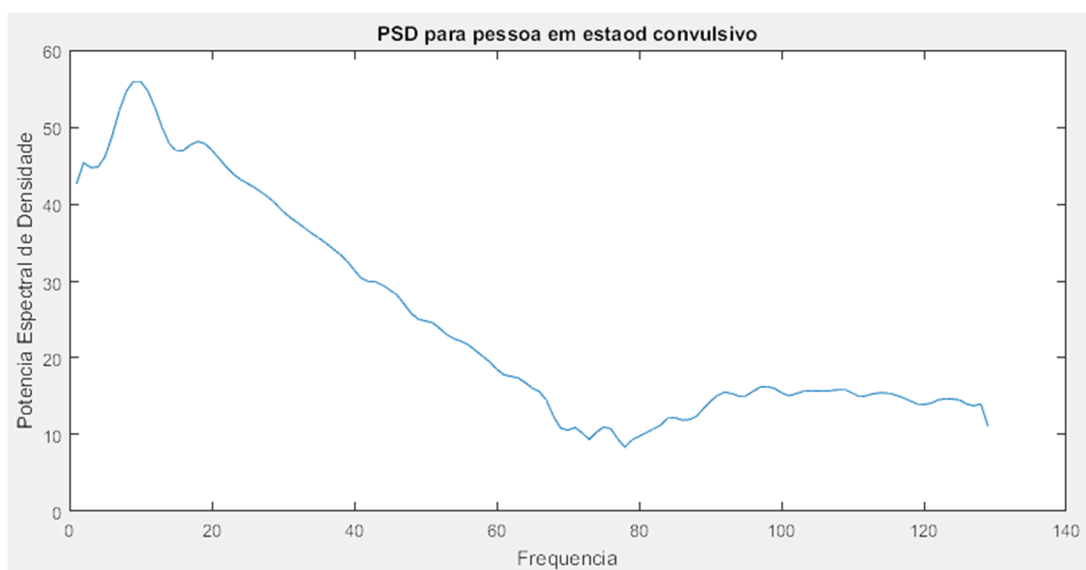


Figura 16 - Potência Espectral de Densidade para o caso E

Uma vez que PSD foi calculado, ou seja, que os dados de cada caso foram tratados é preciso classificá-los em doentes e saudáveis e depois comparar obtidos com o resultado estabelecido pelo médico (A é saudável e E é epilético).

Assim como o método da transformada de Fourier, o classificador SVM também foi codado em MATLAB, o código está no apêndice e a explicação teórica se encontra no item Revisão do material teórico. A Figura 17 mostra o resultado da classificação de árvore complexa para os dois casos.

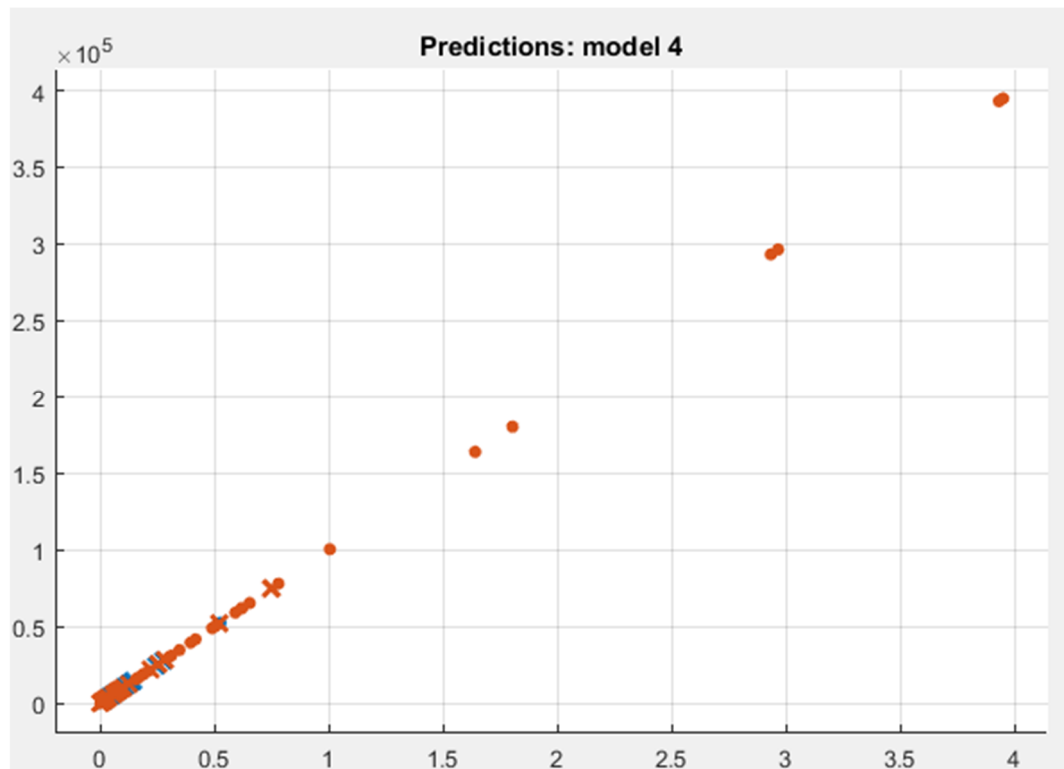


Figura 17 – Dispersão de pontos com classificador Arvore Complexa no método de Fourier

Desde já podemos perceber que o método não resultará em uma eficiência ótima pois os pontos não se encontram perfeitamente agrupados. Aplicando-se a matriz de conflito ao classificador em árvore, temos:

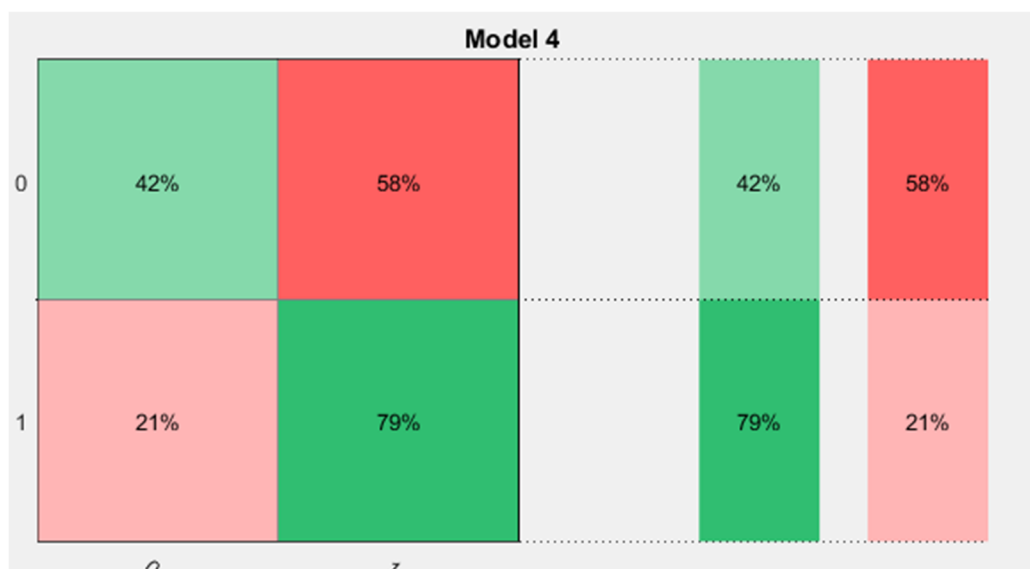


Figura 18 - Matriz de Conflito de classificador árvore no método de Fourier

Através da Figura 18 podemos perceber que o método da transformada de Fourier + SVM quadrático oferece uma taxa de 42% de acertos no diagnóstico de pacientes doentes e de 79% de pacientes saudáveis.

6.3. Resultados Método das Entropias + SVM Quadrático

Como nos outros métodos, classificamos primeiramente como doentes os indivíduos em que sinal resultante do exame de EEG tenha sido interictal ou ictal.

Combinando o método de entropias com um classificador do tipo SVM quadrático, chegamos a uma eficiência de 78% para a combinação. Comparado ao método de Fourier com classificador de árvore complexa, temos uma eficiência mais satisfatória, mas longe do resultado para o método das frequências com classificador SVM quadrático.

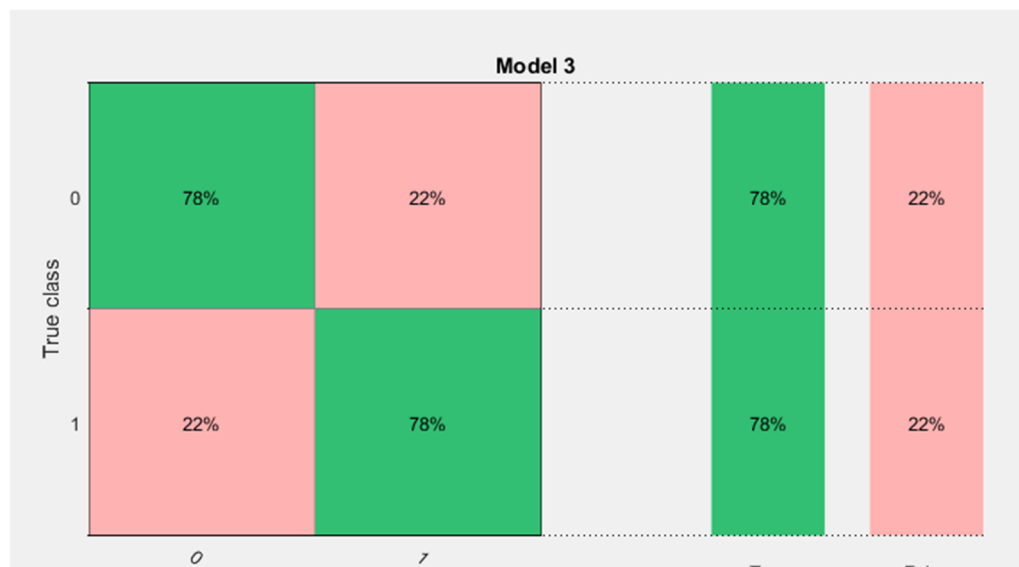


Figura 19 - Matriz de Conflito de classificador SVM quadrático no método de Entropias

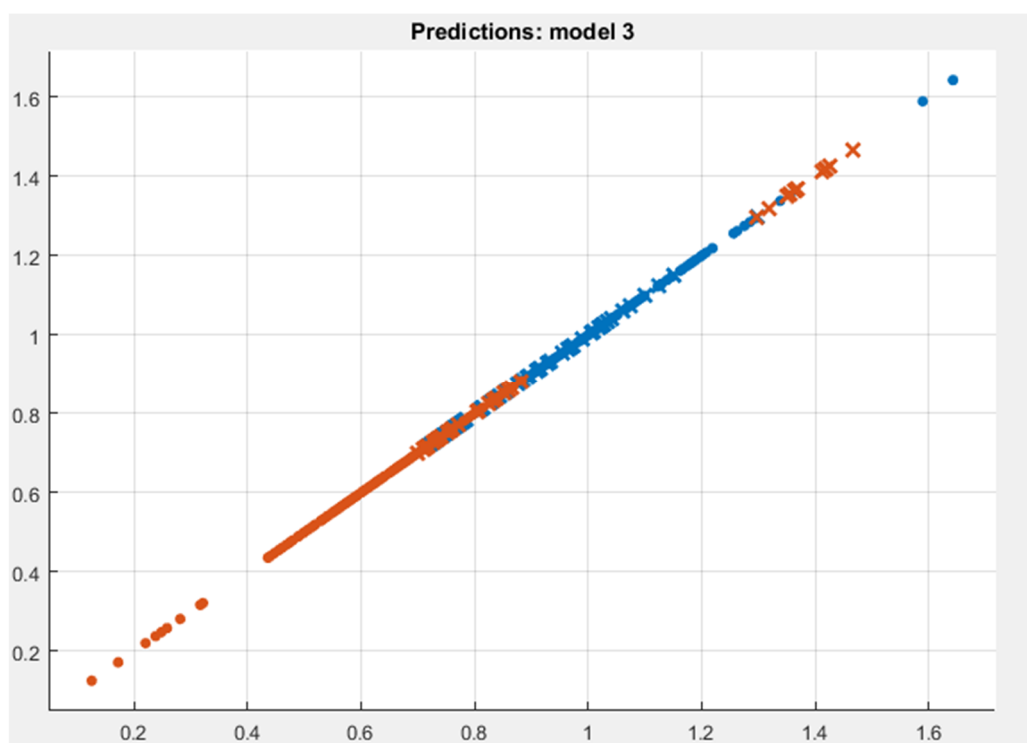


Figura 20 - Dispersão de pontos com classificador SVM Quadrático no método de Entropias

6.4. Extensão dos métodos para classificação tripla

Levando os métodos de classificação a um nível superior, podemos decidir classificar os pacientes não só entre saudáveis e doentes, mas em três categorias diferentes: normal, interictal (entre crises), ictal (em crise); definidos respectivamente por 0, 1, 2.

Embora o problema passe a ser mais complexo, ele é muito útil quando a crise detectada durante o exame não é a imagem clássica de uma síndrome epilética, a convulsão, mas sim as crises que podem passar despercebidas.

Para isso, iremos utilizar os métodos das entropias e de Fourier, pois apesar de terem apresentado eficiência menor quando comparados ao método das frequências, são os únicos capazes de rodar tal classificação tripla, mais complexa, em um tempo razoável.

Como esperado, as eficiências dos métodos diminuiram por conta da complexidade da nova análise. Para mitigar essa deficiência encontrada, será apresentado em seguida um método combinando os métodos de Fourier e Entropias.

A metodologia para a resolução desta classificação segue as mesmas teorias para os métodos e classificadores apresentados na seção Metodologia deste relatório. Novamente, os classificadores são escolhidos de forma a maximizar a eficiência atingida por cada método.

6.4.1. Resultados Método de Fourier para classificação tripla + Árvore Complexa

Os resultados para este método nos mostram que ele é eficiente, em classificar quem são os indivíduos saudáveis, mas se mostra deficiente em classificar para os indivíduos doentes quais estão entre crises e quais estão em crise.

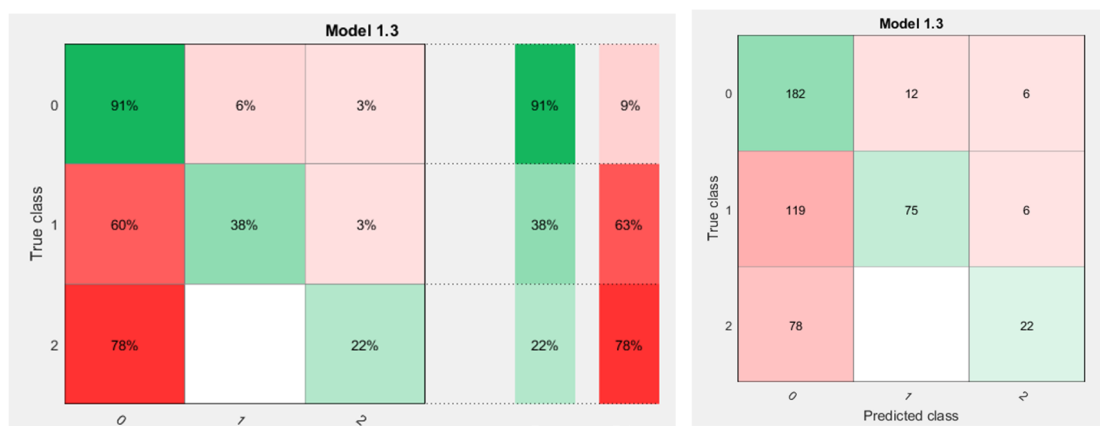


Figura 21 - Matriz de Conflito de classificador Árvore Complexa no método de Fourier em valores relativos e absolutos em classificação tripla

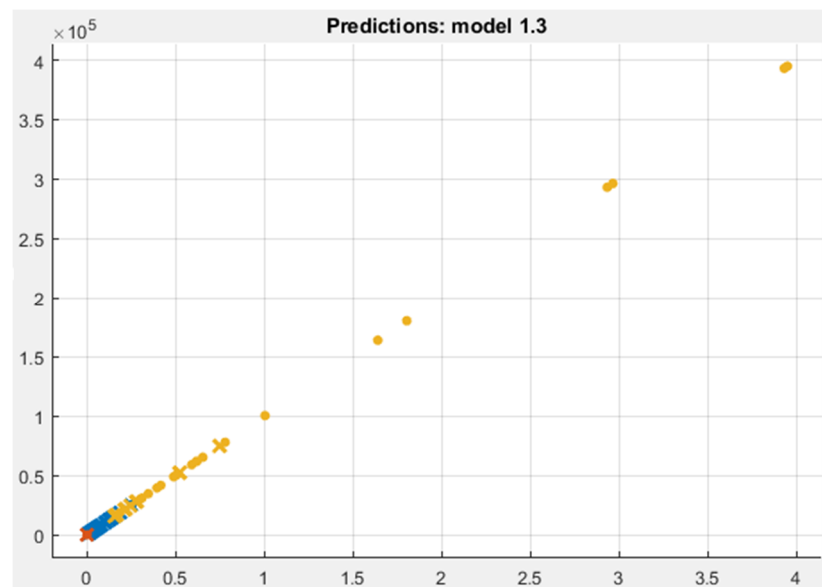


Figura 22 - Dispersão de pontos com classificador Árvore Complexa no método de Fourier em classificação tripla

6.4.2. Resultados Método das Entropias para classificação tripla + Árvore Complexa

Ao contrário do método de Fourier, o método das entropias apresenta um melhor resultado quanto à classificação de pacientes doentes entre o período entre crises e em crise, mas uma eficiência menor em detectar quais pacientes são saudáveis.



Figura 23 - Matriz de Conflito de classificador Árvore Complexa no método de Entropias em valores relativos e absolutos em classificação tripla

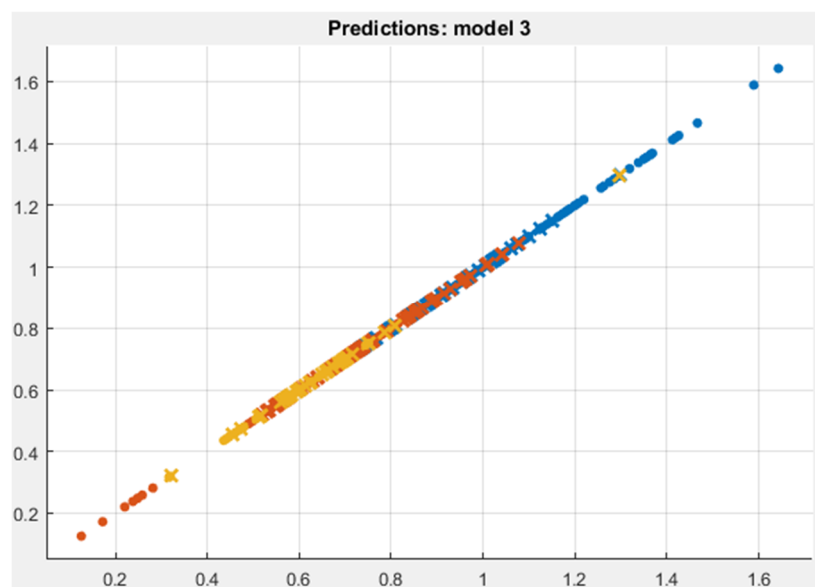


Figura 24 - Dispersão de pontos com classificador Árvore Complexa no método de Entropias em classificação tripla

6.4.3. Resultados Método combinado de Fourier e Entropias para classificação tripla + Árvore Complexa

A combinação dos métodos nos traz um resultado superior quando comparado aos métodos individuais, atingindo uma eficiência de 72% em média. Foi possível atingir um alto nível de precisão na classificação de pessoas saudáveis e um resultado mais satisfatório para a classificação dos pacientes doentes em entre crises ou em crise.



Figura 25 - Matriz de Confusão de classificador Árvore Complexa no método de Entropias em valores relativos e absolutos em classificação tripla

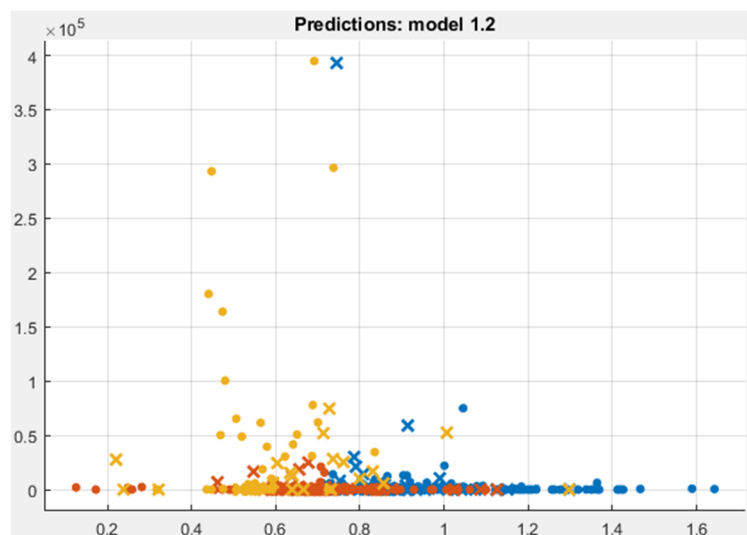


Figura 26 - Dispersão de pontos com classificador Árvore Complexa para método combinado de Entropias e Fourier em classificação tripla

7. ANÁLISE DOS RESULTADOS E CONSIDERAÇÕES FINAIS

Como exposto anteriormente, as análises dos resultados para cada método foram divididas em duas aplicações. Na primeira procuramos identificar os pacientes doentes dos saudáveis. E em um segundo momento procuramos diferenciar pacientes saudáveis, epiléticos em estado ictal e epiléticos em estado interictal.

Na primeira aplicação podemos perceber um resultado bastante expressivo do método das frequências + SVM se comparado aos demais métodos com uma eficiência global de 98%. Isso se deve ao fato do método se apoiar em bandas de frequência, como pudemos ver ao longo desse trabalho. As bandas de frequência conseguiram separar de forma muito precisa o comportamento de pacientes epiléticos, com frequências globais mais baixas, de pacientes saudáveis que pertencem as bandas de frequências mais altas. No entanto, tal eficiência tem um custo de processamento bastante elevado que infelizmente não tivemos a oportunidade de discutir neste trabalho.

Ambos os métodos das entropias e o método da transformada de Fourier tiveram resultados interessantes para a primeira aplicação com eficiência global de 78% e 63% respectivamente. Apesar de ambos os métodos apresentarem uma eficiência de diagnóstico menos interessante que o Método das Frequências eles podem ser construídos com funções já existentes na biblioteca interno do MATLAB o que diminui drasticamente o tempo de cálculo em relação ao método das frequências.

Em um segundo momento os métodos das Entropias e o Método das Frequência foram testados para a aplicação 2, ou seja, a diferenciação de pacientes doentes, epiléticos ictais e epiléticos interictais. Infelizmente o

método das frequências não pode ser utilizado nesta segunda etapa devido a limitações de hardware, uma vez que a complexidade do cálculo se tornava muito importante em relação ao processador utilizado no projeto.

Como esperado ambos os métodos tiveram uma redução na sua eficiência global pois o nível de segmentação do classificador por aumentado. Logo a eficiência global da aplicação 2 foi de 68% para o método das entropias e 54% para o método da Transformada de Fourier.

No entanto, o método de Fourier se revelou mais eficiente para a detecção de pacientes saudáveis com porcentagem de 91% de acerto. Já o método das entropias se revelou mais eficiente que o método de Fourier na detecção dos epiléticos ictais e epiléticos interictais com eficiência de 74 e 40% respectivamente, assim ambos os métodos foram combinados para aumentar a eficiência global da aplicação 2. A combinação aumentou a eficiência global a 72%. Uma melhoria importante em relação aos dois métodos em separado sem o prejuízo ao tempo de processamento.

Finalmente, podemos concluir que todos os métodos geraram resultados satisfatórios em menor ou maior grau e fizeram desta breve pesquisa muito satisfatória para seus integrantes.

Como próximos passos podemos dizer que se faz necessário uma otimização do método das frequências para que se possa testá-lo na aplicação 2. Ainda se faz necessária uma investigação mais aprofundada da base CHB-MIT para discretizá-la e assim poder testar a eficiência dos métodos com uma outra base.

8. REFERENCIAS

Ali Shoeb. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. Tese de PHD, Massachusetts Institute of Technology - USA, 2009.

Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001) Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E, 64, 061907. Artigo, Massachusetts Institute of Technology - USA, 2001.

Fontenelle M, Rosemberg A, Camilli C, Ulian E, Barros IL, Tassoni L. Application des méthodes de l'entropie et des fréquences pour le diagnostique de l'épilepsie. Trabalho de graduação (Eng. Generalista), Ecole Centrale Marseille - França, 2015.

Fu K., Qu J., Chai Y., Dong Y. Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM. Artigo, School of Automation, Chongqing University, Chongqing – China 2014.

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000 (June 13).

GOLDMAN L., AUSIELLO D. Cecil: Medicina. 23^a ed. Rio de Janeiro: Elsevier, 2009. Vol I e II.

GUYTON, A.C.; HALL, J.E. Tratado de Fisiologia Médica. 11^a ed. Rio de Janeiro, Elsevier Ed., 2006.

ROWLAND, L. P. MERRIT – Tratado de Neurologia. 11.ed. Rio de Janeiro: Guanabara Koogan, 2007

Megiddo I, Colson A, Chisholm D, Dua T, Nandi A, and Laxminarayan R (2016). Health and economic benefits of public financing of epilepsy treatment in India: An agent-based simulation model. *Epilepsia Official Journal of the International League Against Epilepsy* doi: 10.1111/epi.13294, [WHO; <http://www.who.int/mediacentre/factsheets/fs999/en/>]

Pang-Ning T., Steinbach M., Kumar V. (2006): Introduction to data mining - Michigan State University, University of Minnesota.

Polat K, Günes S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Artigo, Selcuk University - Turquia*, 2006.

SMOLA, A.J.; BARTLETT P.L.; SCHOLKOPF B.; SCHUURMANS D.; *Advances in large margin classifiers*. Massachusetts Institute of Technology. London, England: Ed. MIT Press, 2000.

U. Rajendra Acharya, H. Fujita, Vidya K. Sudarshan, Shreya Bhat, Joel E.W. Koh. Application of entropies for automated diagnosis of epilepsy using EEG signals: A review.

VAPNIK, V.N.; *The Nature of Statistical Learning Theory*. USA: Springer, 2nd ed., 1999.

Códigos MATLAB

Impressão dos EEGs da Universidade de Bonn

```
Clear

addpath('C:\Users\Iohana\Desktop\TCC\Z');
addpath('C:\Users\Iohana\Desktop\TCC\O');
addpath('C:\Users\Iohana\Desktop\TCC\F');
addpath('C:\Users\Iohana\Desktop\TCC\N');
addpath('C:\Users\Iohana\Desktop\TCC\S');

name = cell(1,100);
data_group = {'Z','O','F','N','S'};

for k=1:5,
    prefix = data_group{k};

    signal = [];

    for i=1:100,

        t = num2str(i);
        if (length(t) == 1),
            t = strcat(prefix,'0','0',t,'.txt');
        end
        if (length(t) == 2),
            t = strcat(prefix,'0',t,'.txt');
        end
        if (length(t) == 3),
            t = strcat(prefix,t,'.txt');
        end
        name{i} = t;

        name{i}
        basedonnees = fopen(name{i});

        y = fscanf(basedonnees,'%f',[4097 1]);
        signal = [signal;y];

        fclose(basedonnees);

    end
end
f= 173.61;
s = 1/f;
fim = length(signal)*s;
```

```
tempo = [0:s:fim];  
tempo(end)=[];  
plot(tempo,signal)  
    hold on  
    ylabel('Caso E')  
    xlabel('Tempo (s)')  
    axis([0, 24, -2000, 2000])  
    hold off
```


Transformada de Fourier pelo método de Welth dos EEGs da Universidade de Bonn

```
clear

addpath('C:\Users\Iohana\Desktop\TCC\Z');
addpath('C:\Users\Iohana\Desktop\TCC\O');
addpath('C:\Users\Iohana\Desktop\TCC\F');
addpath('C:\Users\Iohana\Desktop\TCC\N');
addpath('C:\Users\Iohana\Desktop\TCC\S');

name = cell(1,100);
data_group = {'F','S'};

for k=2:2,
    prefix = data_group{k};

    signal = [];

    for i=1:100,

        t = num2str(i);
        if (length(t) == 1),
            t = strcat(prefix,'0','0',t,'.txt');
        end
        if (length(t) == 2),
            t = strcat(prefix,'0',t,'.txt');
        end
        if (length(t) == 3),
            t = strcat(prefix,t,'.txt');
        end
        name{i} = t;
        basedonnees = fopen(name{i});

        y = fscanf(basedonnees,'%f',[4097 1]);
        signal = [signal;y];

        fclose(basedonnees);

    end

    segmentLength = 100;
    noverlap = [];
    pxx = pwelch(y,segmentLength,noverlap);

end
```

```
f= 173.61;  
s = 1/f;  
fim = length(y)*s;  
tempo = [0:s:fim];  
tempo(end)=[];  
  
plot(10*log10(pxx))
```

Métodos das Entropias - AppEn

```
function [ApEn] = ApEn3(series,dim,r)

control = ~isempty(series);
assert(control,'The user must introduce a time series (first input).');
control = ~isempty(dim);
assert(control,'The user must introduce a embbeding dimension (second
input).');
control = ~isempty(r);
assert(control,'The user must introduce a tolerand (r) (third input).');

N = length(series);
result = zeros(1,2);
r = r*std(series);

for j = 1:2
    m = dim+j-1; % 'm' is the embbeding dimension used each iteration
    % Pre-definition of the varialbes for computational efficiency:
    phi = zeros(1,N-m+1);
    patterns = zeros(m,N-m+1);

    % First, we compose the patterns
    % The columns of the matrix 'patterns' will be the (N-m+1) patterns of
    'm' length:
    if m == 1 % If the embedding dimension is 1, each sample is a pattern:
        patterns = series;
    else % Otherwise, we build the patterns of length 'm':
        for i = 1:m
            patterns(i,:) = series(i:N-m+i);
        end
    end

    % Second, we compute the number of patterns whose distance is less than
    the tolerance.
    % This loop goes over the columns of matrix 'patterns':
    for i = 1:N-m+1
        % 'temp' is an auxiliar matrix whose elements are the maximum
        % absolut difference between the current pattern and the rest:
        if m == 1
            temp = abs(patterns - repmat(patterns(:,i),1,N-m+1));
        else
            temp = max(abs(patterns - repmat(patterns(:,i),1,N-m+1)));
        end
        % We determine which elements of 'temp' are smaller than the
        tolerance:
        bool = any((temp < r),1);
```

```
        % We get the relative frequency of the current pattern:
        phi(i) = sum(bool)/(N-m+1);
    end

    % Finally, we average the natural logarithm of all relative
    frequencies:
    result(j) = mean(log(phi));
end

ApEn = result(1)-result(2);

end
```

Métodos das Entropias – manipulação da base – extração das entropias

```
Clear

addpath('C:\Users\Iohana\Desktop\TCC\Z');
addpath('C:\Users\Iohana\Desktop\TCC\O');
addpath('C:\Users\Iohana\Desktop\TCC\F');
addpath('C:\Users\Iohana\Desktop\TCC\N');
addpath('C:\Users\Iohana\Desktop\TCC\S');
addpath('C:\Users\Iohana\Desktop\TCC\_entropy_functions');

typeEntropie = 'AppEn';

name = cell(1,100);
data_group = {'F','O','N','S','Z'};

for k=1:length(data_group),
    prefix = data_group{k};
    entropie = zeros(1,100);

    for i=1:100,
        t = num2str(i);
        if (length(t) == 1),
            t = strcat(prefix,'0','0',t,'.txt');
        end
        if (length(t) == 2),
            t = strcat(prefix,'0',t,'.txt');
        end
        if (length(t) == 3),
            t = strcat(prefix,t,'.txt');
        end
        name{i} = t;
        name{i}
        basedonnees = fopen(name{i});
        signal = fscanf(basedonnees,'%f',[4097 1]);
        fclose(basedonnees);

        entropie(1,i) = ApEn3(signal,2,0.2);

    end

Folder='_entropies';
if ~exist(Folder,'dir')
```

```
        [PathStr,FolderName]=fileparts(Folder);  
        mkdir(FolderName);  
    end  
  
    fichier =  
    fopen(strcat('_entropies','/', 'entropie',typeEntropie,'_',prefix,'.txt'),'w');  
        fprintf(fichier,'%f\r\n',entropie);  
        fclose all;  
end
```

Métodos das Entropias – manipulação da base – Treinamento do modelo

```
Clear

id = {'Z','O','N','F','S'};
idnum = [1 2 3 4 5];
typeEntropie = 'PermEn';
repertoire = strcat('C:\Users\Iohana\Desktop\TCC\_entropies');
addpath(repertoire);
fichierSortie = strcat('_entropies','\'','entropie','_',typeEntropie,'.txt');

[m n] = size(id);

data_numeroGroupe = zeros(n*100,1);
data_maladie = zeros(n*100,1);
data_entropie = zeros(n*100,1);
data_ictal = zeros(n*100,1);
data_output = zeros(n*100,4);

for i=1:n,
    chemin = strcat('_entropies\'','entropie',typeEntropie,'_',id{i},'.txt');
    fichier = fopen(chemin);
    entropie = fscanf(fichier,'%f',[100 1]);
    fclose(fichier);
    for j=(100*(i-1)+1):(100*i),
        if(j~=i*100),
            k = mod(j,100);  %k vai de 1-100
        else
            k=100;
        end

        data_entropie(j)= entropie(k);

        if idnum(i)==1
            data_ictal(j) = 0;
            data_numeroGroupe(j) = 1;
            data_maladie(j) = 0;
        end
        if idnum(i)==2
            data_ictal(j) = 0;
            data_numeroGroupe(j) = 2;
            data_maladie(j) = 0;
        end
    end
end
```

```

        if idnum(i)==3
            data_ictal(j) = 1;
            data_numeroGroupe(j) = 3;
            data_maladie(j) = 1;
        end
        if idnum(i)==4
            data_ictal(j) = 1;
            data_numeroGroupe(j) = 4;
            data_maladie(j) = 1;
        end
        if idnum(i)==5
            data_ictal(j) = 2;
            data_numeroGroupe(j) = 5;
            data_maladie(j) = 1;
        end

        data_output(j,1) = data_numeroGroupe(j);
        data_output(j,2) = data_maladie(j);
        data_output(j,3) = data_ictal(j);
        data_output(j,4) = entropie(k);

    end
end
if exist(fichierSortie, 'file'),
    display('Fichier déjà existant, variable data_output crée');
else
    for z=1:100*n,
        fprintf(fopen(fichierSortie,'a'), '%1.0f ', data_numeroGroupe(z)); %'a'
    avoids overwrite
        fprintf(fopen(fichierSortie,'a'), '%1.0f ', data_maladie(z));
        fprintf(fopen(fichierSortie,'a'), '%1.0f ', data_ictal(z));
        fprintf(fopen(fichierSortie,'a'), '%f \r\n', data_entropie(z));
        fclose all;
    end
end
end

```


Métodos das frequências – geração da base

```
clear;

for flux=1:5

%open file

if flux == 1
name_begin='Z';
name_end='.txt';
end

if flux == 2
name_begin='O';
name_end='.txt';
end

if flux == 3
name_begin='N';
name_end='.TXT';
end

if flux == 4
name_begin='F';
name_end='.txt';
end

if flux == 5
name_begin='S';
name_end='.txt';
end

v_delta = zeros(100,5);
v_theta = zeros(100,5);
v_alpha = zeros(100,5);
v_beta = zeros(100,5);
v_gamma = zeros(100,5);

for hoa=1:100

str = sprintf('%03d', hoa);
str = sprintf('%s', str);
fileName = strcat(name_begin, str, name_end);
fileID = fopen(fileName, 'r');
formatSpec = '%f';
```

```

x = fscanf(fileID, formatSpec);

f_zero = 173.61;
Ts = 1/f_zero;

addpath('C:\Users\Iohana\Desktop\TCC\Hilbert Huang transform module');
addpath('C:\Users\Iohana\Desktop\TCC\Z');
addpath('C:\Users\Iohana\Desktop\TCC\O');
addpath('C:\Users\Iohana\Desktop\TCC\F');
addpath('C:\Users\Iohana\Desktop\TCC\N');
addpath('C:\Users\Iohana\Desktop\TCC\S');

imf = emd(x);

for k = 1:length(imf)
    th = angle(hilbert(imf{k}));
    d{k} = gradient(th)/Ts/(2*pi);
    abso{k} = abs(hilbert(imf{k}));
end

N = length(x);
c = linspace(0, (N-1)*Ts, N);

cor_final(1024,4097)=0;

for k = 1:length(imf)
    d_normal = [d{k}];
    f_max(k)= max(d_normal);
    abso_normal = [abso{k}];
    absao(k)=max(abso_normal);
end

freq_max = max(f_max);
lim_abs = max(absao);

for k = 1:length(imf)
    cor = zeros(4097,1024);
    d_normal = [d{k}];
    abso_normal = [abso{k}];
    for i= 1:N
        if d_normal(i)>0
            index = round(1024*d_normal(i)/freq_max);
            if index == 0
                index = 1;
            end
            %definit le groupe (temps, frequence, intensite)

```

```

        cor(i,index) = abso_normal(i);
    end
end
cor = cor*1024/lim_abs;
c_t=transpose(cor);
cor_final = cor_final + c_t;
end

%separation des bandes de frequence
ind_delta = round(4*1024/freq_max);
ind_theta = round(8*1024/freq_max);
ind_alpha = round(12*1024/freq_max);
ind_beta = round(30*1024/freq_max);
ind_gamma = round(50*1024/freq_max);

cor_delta = cor_final(1:ind_delta,1:4097);
cor_theta = cor_final(ind_delta+1:ind_theta,1:4097);
cor_alpha = cor_final(ind_theta+1:ind_alpha,1:4097);
cor_beta = cor_final(ind_alpha+1:ind_beta,1:4097);
cor_gamma = cor_final(ind_beta+1:ind_gamma,1:4097);

ens_delta = cor_delta(cor_delta>0);
ens_theta = cor_theta(cor_theta>0);
ens_alpha = cor_alpha(cor_alpha>0);
ens_beta = cor_beta(cor_beta>0);
ens_gamma = cor_gamma(cor_gamma>0);

ens_delta = round_vector(ens_delta, 256);
ens_theta = round_vector(ens_theta, 256);
ens_alpha = round_vector(ens_alpha, 256);
ens_beta = round_vector(ens_beta, 256);
ens_gamma = round_vector(ens_gamma, 256);

v_delta(hoa, 1:4)= fund_stats(ens_delta);
v_theta(hoa, 1:4)= fund_stats(ens_theta);
v_alpha(hoa, 1:4)= fund_stats(ens_alpha);
v_beta(hoa, 1:4)= fund_stats(ens_beta);
v_gamma(hoa, 1:4)= fund_stats(ens_gamma);

end

if flux == 1

v_delta(1:100, 5)= -1;

```

```

v_theta(1:100, 5)= -1;
v_alpha(1:100, 5)= -1;
v_beta(1:100, 5)= -1;
v_gamma(1:100, 5)= -1;

v_delta_A = v_delta;
v_theta_A = v_theta;
v_alpha_A = v_alpha;
v_beta_A = v_beta;
v_gamma_A = v_gamma;

save('A.mat','v_gamma_A', 'v_theta_A', 'v_delta_A', 'v_alpha_A', 'v_beta_A');
end

if flux == 2
v_delta(1:100, 5)= -1;
v_theta(1:100, 5)= -1;
v_alpha(1:100, 5)= -1;
v_beta(1:100, 5)= -1;
v_gamma(1:100, 5)= -1;

v_delta_B = v_delta;
v_theta_B = v_theta;
v_alpha_B = v_alpha;
v_beta_B = v_beta;
v_gamma_B = v_gamma;
save('B.mat','v_gamma_B', 'v_theta_B', 'v_delta_B', 'v_alpha_B', 'v_beta_B');
end

if flux == 3
v_delta(1:100, 5)= 1;
v_theta(1:100, 5)= 1;
v_alpha(1:100, 5)= 1;
v_beta(1:100, 5)= 1;
v_gamma(1:100, 5)= 1;

v_delta_C = v_delta;
v_theta_C = v_theta;
v_alpha_C = v_alpha;
v_beta_C = v_beta;
v_gamma_C = v_gamma;
save('C.mat','v_gamma_C', 'v_theta_C', 'v_delta_C', 'v_alpha_C', 'v_beta_C');
end

if flux == 4
v_delta(1:100, 5)= 1;
v_theta(1:100, 5)= 1;
v_alpha(1:100, 5)= 1;
v_beta(1:100, 5)= 1;
v_gamma(1:100, 5)= 1;

```

```

v_delta_D = v_delta;
v_theta_D = v_theta;
v_alpha_D = v_alpha;
v_beta_D = v_beta;
v_gamma_D = v_gamma;
save('D.mat','v_gamma_D', 'v_theta_D', 'v_delta_D', 'v_alpha_D', 'v_beta_D');
end

if flux == 5
v_delta(1:100, 5)= 1;
v_theta(1:100, 5)= 1;
v_alpha(1:100, 5)= 1;
v_beta(1:100, 5)= 1;
v_gamma(1:100, 5)= 1;

v_delta_E = v_delta;
v_theta_E = v_theta;
v_alpha_E = v_alpha;
v_beta_E = v_beta;
v_gamma_E = v_gamma;
save('E.mat','v_gamma_E', 'v_theta_E', 'v_delta_E', 'v_alpha_E', 'v_beta_E');
end

end

```

Métodos das frequências – Funções Auxiliares

round_vector.m

```
function v = round_vector(a,lim);
    if ~isvector(a)
        error('Input must be a vector')
    end
    max_a=max(a);
    for i=1:size(a)
        v(i) = round(a(i)*lim/max_a);
    end
end
```

fund_stats.m

```
function s = fund_stats(a)
%extracts mean, variance, skewness and kurtosis of a vector/array

    %calculate mean
    mean = 0;
    L = length(a)
    for i=1:L
        mean = mean + a(i);
    end
    mean = mean/L;

    %calculate variance
    variance = 0;
    for i=1:L
        variance = variance + (a(i)-mean)^2;
    end
    variance = variance/(L-1);

    %calculate skewness and kurtosis
    skewness = 0;
    kurtosis = 0;
    for i=1:L
        skewness = skewness + (a(i)-mean)^3;
        kurtosis = kurtosis + (a(i)-mean)^4;
    end
    skewness = skewness/variance^(3/2);
    kurtosis = kurtosis/variance^2;

    s = [mean, variance, skewness, kurtosis];

end
```

emd.m

```
function imf = emd(x)
% Empirical Mode Decomposition (Hilbert-Huang Transform)
% imf = emd(x)
% Func : findpeaks

x = transpose(x(:));
imf = [];
while ~ismonotonic(x)
    x1 = x;
    sd = Inf;
    while (sd > 0.1) | ~isimf(x1)
        s1 = getspline(x1);
        s2 = -getspline(-x1);
        x2 = x1-(s1+s2)/2;

        sd = sum((x1-x2).^2)/sum(x1.^2);
        x1 = x2;
    end

    imf{end+1} = x1;
    x = x-x1;
end
imf{end+1} = x;

% FUNCTIONS

function u = ismonotonic(x)

u1 = length(findpeaks(x))*length(findpeaks(-x));
if u1 > 0, u = 0;
else, u = 1; end

function u = isimf(x)

N = length(x);
u1 = sum(x(1:N-1).*x(2:N) < 0);
u2 = length(findpeaks(x))+length(findpeaks(-x));
if abs(u1-u2) > 1, u = 0;
else, u = 1; end

function s = getspline(x)

N = length(x);
p = findpeaks(x);
s = spline([0 p N+1],[0 x(p) 0],1:N);
```

teste_class_learner.m

```
clear;
addpath('C:\Users\Caio Camilli\Desktop\Faculdade\Centrale\3eme semestre\Projet
S7\Database\Tabelas de dados processados');
load A.mat;
load B.mat;
load C.mat;
load D.mat;
load E.mat;

teste=readEntropy('entropiasZONFS_AppEn.txt', {'Z', 'O', 'N', 'F', 'S'});
teste2=readEntropy('entropiasZONFS_SampEn.txt', {'Z', 'O', 'N', 'F', 'S'});

v_test_1 = v_theta_A(1:100,:);
v_test_2 = v_theta_B(1:100,:);
v_test_3 = v_theta_C(1:100,:);
v_test_4 = v_theta_D(1:100,:);
v_test_5 = v_theta_E(1:100,:);

v_test_1(1:100,6) = teste(1:100,3);
v_test_2(1:100,6) = teste(101:200,3);
v_test_3(1:100,6) = teste(201:300,3);
v_test_4(1:100,6) = teste(301:400,3);
v_test_5(1:100,6) = teste(401:500,3);

v_test_1(1:100,7) = teste2(1:100,3);
v_test_2(1:100,7) = teste2(101:200,3);
v_test_3(1:100,7) = teste2(201:300,3);
v_test_4(1:100,7) = teste2(301:400,3);
v_test_5(1:100,7) = teste2(401:500,3);

mnew = [v_test_1; v_test_2; v_test_3; v_test_4; v_test_5];

k = randperm(size(mnew,1));
ThetaFreq = mnew(k(1:100),:);
```