



CARACTERIZAÇÃO DE PERÍODOS ANTECIPATÓRIOS DE CRISES EPILEPTICAS

Kleber de Aguiar

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientadores: Felipe Maia Galvão França
Valmir Carneiro Barbosa
César Alexandre Domingues
Teixeira

Rio de Janeiro
Dezembro de 2022

CARACTERIZAÇÃO DE PERÍODOS ANTECIPATÓRIOS DE CRISES
EPILEPTICAS

Kleber de Aguiar

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientadores: Felipe Maia Galvão França
Valmir Carneiro Barbosa
César Alexandre Domingues Teixeira

Aprovada por: Prof. Valmir Carneiro Barbosa
Prof. Felipe Maia Galvão França
Prof. Adilson Elias Xavier
Prof. Fábio Protti
Prof. Lúcia Maria de Assumpção Drummond

RIO DE JANEIRO, RJ – BRASIL
DEZEMBRO DE 2022

de Aguiar, Kleber

Caracterização de Períodos Antecipatórios de Crises Epiléticas/Kleber de Aguiar. – Rio de Janeiro: UFRJ/COPPE, 2022.

XIV, 116 p.: il.; 29,7cm.

Orientadores: Felipe Maia Galvão França

Valmir Carneiro Barbosa

César Alexandre Domingues Teixeira

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2022.

Referências Bibliográficas: p. 112 – 116.

1. Crise Epilética. 2. Pré-crise. 3. Períodos Antecipatórios. 4. Inteligência Computacional. I. Maia Galvão França, Felipe *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*Às minhas filhas, Mariah e
Luiza, à memória do meu
falecido pai, Nicanor Pires de
Aguiar e ao meu irmão Paulo de
Aguiar, portador de epilepsia e
genuína inspiração para esta
pesquisa.*

*“Meu filho, se me ouvires com
atenção, serás instruído; se
submeteres o teu espírito, tu te
tornarás sábio.”
(Bíblia Sagrada:
Eclesiástico 6, 33)*

Agradecimentos

Ao Nosso Deus, Trino e Uno, rendo graças e louvores de gratidão na figura de Suas três pessoas: o Pai Eterno, o Filho Redentor e o Espírito Santo. Sua misericórdia, sabedoria e amor-caridade (agape) são imensuráveis! Sem Ele, após tantas dificuldades e dissabores ao longo dessa jornada acadêmica e de vida, a conclusão deste doutorado não teria se tornado uma doce realidade. Tudo posso Naquele que me fortalece!

À Santíssima Virgem Maria, mãe do Deus e Senhor Jesus Cristo, Nossa Senhora de Fátima e Aparecida, pela incansável intercessão junto a Cristo em meu favor, pelo colo de mãe Cheia de Graça nas “noites traiçoeiras” e nos inúmeros momentos em que eu duvidei de mim mesmo. *Totus Tuus, Mariae!*

À minha família, pelo apoio, suporte emocional constante, orações e eventuais auxílios financeiros durante esse longos anos de doutorado que enfim chegaram a bom termo. Um *muito obrigado* especial a minha mãe Elisa Maria, personificação da providência divina e do carinho mariano em minha existência. Gratidão *ad eternum!*

Aos meus irmãos de Caminhada com Cristo, pelas orações e pensamentos positivos, especialmente aos meus companheiros de “Santa Escravidão por Amor”, um sentimento extra de gratidão. *Apostolado da Servidão Perpétua a Jesus em Maria*, minha tábua de salvação espiritual e secular, Deus vos recompense, amém!

Aos meus orientadores, professores Felipe França, Valmir Barbosa e César Teixeira, pelo apoio, paciência, resiliência, ensinamentos, parceria e amizade. Vocês formaram a pedra angular sob o qual essa minha (nossa) vitória foi erguida. Sentimentos inefáveis de gratidão é o que eu sinto e sempre sentirei por cada um de vocês. Obrigado!

Aos colegas de PESC, mesmo aqueles que não mais fazem parte da minha realidade, o meu muito obrigado pelo tempo em que passamos juntos em maravilhosa convivência acadêmica. E um agradecimento especial aos meus amigos Mestre Daniel Nunes e Doutor Douglas Cardoso. Suas “digitais intelectuais” estão marcadas nas páginas desta tese. Obrigado! Gratidão!

À CAPES, pelo fomento a esta pesquisa, à COPPE/UFRJ, pela bênção de ter sido, por tantos anos, aluno de tão respeitável instituição de ensino e pesquisa e

ao CISUC/DEI (Universidade de Coimbra) pelo suporte acadêmico no ano em que estive além-mar, no período “sanduíche” deste doutorado. Obrigado e até algum dia!

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

CARACTERIZAÇÃO DE PERÍODOS ANTECIPATÓRIOS DE CRISES EPILEPTICAS

Kleber de Aguiar

Dezembro/2022

Orientadores: Felipe Maia Galvão França
Valmir Carneiro Barbosa
César Alexandre Domingues Teixeira

Programa: Engenharia de Sistemas e Computação

Segundo dados da Organização Mundial da Saúde, a epilepsia é considerada um dos transtornos neurológicos de maior incidência global. Os sintomas da epilepsia são caracterizados pela ocorrência de crises epiléticas, que manifestam-se de forma súbita, impactando de maneira negativa aqueles que são acometidos por esse mal. As consequências de um episódio de crise epilética variam de lesões físicas a distúrbios de natureza psicológica, sem contar o grau de incerteza que acomete os pacientes.

A possibilidade de antever o início de uma crise epilética tem o potencial de atenuar esses efeitos negativos, pois essa seria uma forma de anular ou mitigar a característica de a incidência das crises ser repentina. Para que isso seja possível, é essencial a caracterização de períodos nos quais possam ser detetadas alterações na atividade normal do cérebro relacionadas a uma crise iminente.

Nesta tese é apresentada uma metodologia capaz de caracterizar tais períodos, através da análise dos registros encefalográficos do paciente. Contrastando com o que é usual em estudos relacionados, são utilizados os dados encefalográficos após serem modificados apenas por operações matemáticas simples e de cálculo eficiente, além de um teste estatístico. Os resultados obtidos sugerem que tal metodologia pode ser usada como uma ferramenta auxiliar aos sistemas de predição de crises, particularmente no que diz respeito às suas fases de treinamento e aprendizado.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

CHARACTERIZATION OF ANTICIPATORY PERIODS OF EPILEPTIC SEIZURES

Kleber de Aguiar

December/2022

Advisors: Felipe Maia Galvão França
Valmir Carneiro Barbosa
César Alexandre Domingues Teixeira

Department: Systems Engineering and Computer Science

According to data from the World Health Organization, epilepsy is a neurological disorder with one of the highest global incidence rates. Epilepsy symptoms are characterized by the occurrence of epileptic seizures, which happen suddenly and impact those affected by the disease negatively. The consequences of an epileptic seizure range from physical injuries to psychological disorders, not to mention the degree of uncertainty that afflicts epilepsy patients.

Anticipating the onset of an epileptic seizure has the potential to attenuate such negative effects, since it might work toward nullifying or mitigating the suddenness of epileptic events. In order for anticipation to be possible, an essential first step is to characterize those periods when alterations in the normal activity of the brain can be detected that relate to an imminent seizure.

In this thesis a methodology is presented for the characterization of such periods, based on analyses of a patient's encephalographic records. In strong contrast with the common approach in related studies, only simple and computationally efficient modifications are applied to the data, and also a statistical test. The results obtained suggest that this methodology can be used as an auxiliary tool for seizure prediction systems, particularly in regard to the systems' training and learning phases.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiii
Lista de Abreviaturas	xiv
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	2
1.3 Trabalhos Correlatos	2
1.3.1 <i>Optimal Preictal Period in Seizure Prediction</i>	2
1.3.2 <i>Preictal State Identification by Synchronization Changes in Long-term Intracranial EEG Recordings</i>	3
1.4 Estrutura da Tese	3
2 Epilepsia	5
2.1 Conceitos	6
2.1.1 Definição de Epilepsia	6
2.1.2 Etiologia	6
2.2 Crises Epiléticas	7
2.2.1 Crises Focais (Parciais)	7
2.2.2 Crises Generalizadas (desde o início)	7
2.3 Eletroencefalografia	9
3 Métodos e Base de Dados	11
3.1 Banco de Dados EPILEPSIAE	11
3.2 Análise das Componentes Principais	12
3.3 Detecção de Mudança nos Dados	13
3.3.1 CUSUM e o Teste de Page-Hinkley	14
3.3.2 Drift Detection Method	16
3.3.3 ADWIN	17
3.4 Teste Estatístico <i>Two-Sample Kolmogorov-Smirnov</i>	20

4	Metodologia para a Caracterização de Períodos Antecipatórios de Crises Epiléticas	22
4.1	Segmentação dos Dados em Entradas	22
4.2	Representação Vetorial do Conjunto de Entradas em um Espaço Multidimensional	24
4.3	Unidimensionalização dos vetores do espaço multidimensional	25
4.4	Deteção de <i>Concept Drifts</i>	26
4.5	Avaliação dos Períodos Antecipatórios de Crises	28
5	Resultados e Discussões	30
5.1	Base Experimental	30
5.2	Resultados Obtidos	30
5.2.1	Paciente N ^o 3500	32
5.2.2	Paciente N ^o 4000	42
5.2.3	Paciente N ^o 4200	50
5.2.4	Paciente N ^o 4400	55
5.2.5	Paciente N ^o 4500	65
5.2.6	Paciente N ^o 4600	73
5.2.7	Paciente N ^o 5800	83
5.2.8	Paciente N ^o 5900	90
5.2.9	Paciente N ^o 6300	96
5.2.10	Paciente N ^o 6600	102
6	Conclusões	108
6.1	Resumo	108
6.2	Principais Contribuições	109
6.3	Trabalhos Futuros	110
	Referências Bibliográficas	112

Lista de Figuras

2.1	Fragmentos de um EEG	9
3.1	Base de dados Iris antes e depois da PCA	13
3.2	Exemplo de desvio de conceito dos dados	13
3.3	Padrões de <i>concept drift</i>	14
3.4	FDAs e <i>Dmax</i>	21
4.1	Segmento <i>earlyictal E</i> e seus prefixos	28
5.1	Paciente 3500 - Crise do bloco n ^o 41 do EEG	33
5.2	Paciente 3500 - Crise do bloco n ^o 46 do EEG	34
5.3	Paciente 3500 - Crise do bloco n ^o 48 do EEG	36
5.4	Paciente 3500 - Crise do bloco n ^o 50 do EEG	38
5.5	Paciente 3500 - Crise do bloco n ^o 54 do EEG	40
5.6	Paciente 4000 - Crise do bloco n ^o 8 do EEG	43
5.7	Paciente 4000 - Crise do bloco n ^o 12 do EEG	44
5.8	Paciente 4000 - Crise do bloco n ^o 15 do EEG	46
5.9	Paciente 4000 - Crise do bloco n ^o 18 do EEG	48
5.10	Paciente 4000 - Crise do bloco n ^o 22 do EEG	49
5.11	Paciente 4200 - Crise do bloco n ^o 27 do EEG	51
5.12	Paciente 4200 - Crise do bloco n ^o 31 do EEG	52
5.13	Paciente 4200 - Crise do bloco n ^o 37 do EEG	54
5.14	Paciente 4400 - Crise do bloco n ^o 36 do EEG	56
5.15	Paciente 4400 - Crise do bloco n ^o 39 do EEG	58
5.16	Paciente 4400 - Crise do bloco n ^o 41 do EEG	59
5.17	Paciente 4400 - Crise do bloco n ^o 43 do EEG	61
5.18	Paciente 4400 - Crise do bloco n ^o 49 do EEG	63
5.19	Paciente 4500 - Crise do bloco n ^o 45 do EEG	66
5.20	Paciente 4500 - Crise do bloco n ^o 56 do EEG	68
5.21	Paciente 4500 - Crise do bloco n ^o 62 do EEG	70
5.22	Paciente 4500 - Crise do bloco n ^o 71 do EEG	71
5.23	Paciente 4600 - Crise do bloco n ^o 25 do EEG	74

5.24	Paciente 4600 - Crise do bloco n ^o 26 do EEG	76
5.25	Paciente 4600 - Crise do bloco n ^o 30 do EEG	78
5.26	Paciente 4600 - Crise do bloco n ^o 34 do EEG	80
5.27	Paciente 4600 - Crise do bloco n ^o 49 do EEG	81
5.28	Paciente 5800 - Crise do bloco n ^o 0 do EEG	84
5.29	Paciente 5800 - Crise do bloco n ^o 18 do EEG	85
5.30	Paciente 5800 - Crise do bloco n ^o 32 do EEG	87
5.31	Paciente 5800 - Crise do bloco n ^o 42 do EEG	88
5.32	Paciente 5900 - Crise do bloco n ^o 7 do EEG	91
5.33	Paciente 5900 - Crise do bloco n ^o 17 do EEG	93
5.34	Paciente 5900 - Crise do bloco n ^o 18 do EEG	95
5.35	Paciente 6300 - Crise do bloco n ^o 28 do EEG	97
5.36	Paciente 6300 - Crise do bloco n ^o 33 do EEG	99
5.37	Paciente 6300 - Crise do bloco n ^o 39 do EEG	100
5.38	Paciente 6600 - Crise do bloco n ^o 9 do EEG	103
5.39	Paciente 6600 - Crise do bloco n ^o 35 do EEG	105
5.40	Paciente 6600 - Crise do bloco n ^o 39 do EEG	106

Lista de Tabelas

2.1	Tipos de crises generalizadas.	8
4.1	Parâmetros usados na segmentação dos dados.	23
4.2	Valores da variância acumulada ε	25
4.3	Classes dos <i>concepts</i>	27
5.1	Atributos do paciente n ^o 3500.	32
5.2	Deteção de <i>concepts</i> do paciente n ^o 3500.	32
5.3	Atributos do paciente n ^o 4000.	42
5.4	Deteção de <i>concepts</i> do paciente n ^o 4000.	42
5.5	Atributos do paciente n ^o 4200.	50
5.6	Deteção de <i>concepts</i> do paciente n ^o 4200.	51
5.7	Atributos do paciente n ^o 4400.	55
5.8	Deteção de <i>concepts</i> do paciente n ^o 4400.	56
5.9	Atributos do paciente n ^o 4500.	65
5.10	Deteção de <i>concepts</i> do paciente n ^o 4500.	66
5.11	Atributos do paciente n ^o 4600.	73
5.12	Deteção de <i>concepts</i> do paciente n ^o 4600.	73
5.13	Atributos do paciente n ^o 5800.	83
5.14	Deteção de <i>concepts</i> do paciente n ^o 5800.	84
5.15	Atributos do paciente n ^o 5900.	90
5.16	Deteção de <i>concepts</i> do paciente n ^o 5900.	91
5.17	Atributos do paciente n ^o 6300.	96
5.18	Deteção de <i>concepts</i> do paciente n ^o 6300.	97
5.19	Atributos do paciente n ^o 6600.	102
5.20	Deteção de <i>concepts</i> do paciente n ^o 6600.	102

Lista de Abreviaturas

2S-KS Test	<i>Two-Sample Kolmogorov-Smirnov Test</i> , p. 20
ADWIN	<i>ADaptive WINdowing</i> , p. 13
CUSUM	<i>Cumulative Sum</i> , p. 13
DDM	<i>Drift Detection Method</i> , p. 13
EEG	Eletroencefalograma, p. 9
EPILEPSIAE	<i>Evolving Platform for Improving Living Expectation of Patients Suffering from IctAl Events</i> , p. 11
FDA	Função de Distribuição Acumulada, p. 20
ILAE	<i>International League Against Epilepsy</i> , p. 3
PCA	<i>Principal Component Analysis</i> , p. 12
ROC	<i>Receiver operating characteristic curve</i> , p. 3
SVD	<i>Singular Value Decomposition</i> , p. 12
Vídeo-EEG	Videoeletroencefalograma, p. 9

Capítulo 1

Introdução

De acordo com a Organização Mundial da Saúde, a epilepsia é uma doença crônica do cérebro que afeta cerca de 50 milhões de pessoas no mundo inteiro, de diferentes faixas etárias[1]. Pessoas portadoras de epilepsia apresentam uma tendência em ter mais problemas físicos causados pela ocorrência de crises epiléticas (fraturas e outros ferimentos)[2], assim como possuem elevados índices de problemas de ordem psicológica, como a depressão e o transtorno de ansiedade[1]. Além disso, o risco de morte prematura é três vezes maior em pacientes epiléticos do que na população em geral[3]. A detecção de uma crise epilética com um certo tempo de antecedência em relação à sua ocorrência seria uma alternativa aos portadores de epilepsia, possibilitando a estes a tomada de medidas precatórias no sentido de mitigar os inconvenientes causados por tais eventos epiléticos, contribuindo não só para a salvaguarda da integridade física dessas pessoas, como também para ajudá-las a ter uma melhor condição de vida.

1.1 Motivação

Considerando a natureza súbita da manifestação de uma crise epilética e os inconvenientes e perigos associados à sua incidência, uma tentativa de minimizar os seus efeitos adversos seria o desenvolvimento de sistemas preditores de crises capazes de avisar aos portadores de epilepsia a respeito da sua eminência, eliminando ou ao menos reduzindo a característica de ocorrência súbita de tal evento epilético. A predição de crises epiléticas é um tema recorrente em pesquisas da comunidade científica e uma parte significativa desses trabalhos baseia-se na existência e identificação do chamado período *preictal* (pré-crise), que seria um período no qual alterações nos sinais neuronais sinalizariam que uma crise epilética estaria em vias de ocorrer.

1.2 Objetivos

A presente tese de doutorado tem por objetivo apresentar uma metodologia capaz de realizar a caracterização de períodos antecipatórios de crises epiléticas, tornando factível a caracterização desses períodos e com isso, possibilitar que futuros sistemas de predição de crises epiléticas possam se beneficiar da metodologia proposta neste trabalho, no sentido de obter um aumento de eficiência através do uso das informações a respeito dos períodos antecipatórios na fase de aprendizado/treinamento de sistemas desse tipo. De maneira geral, a metodologia proposta possui cinco partes: tratamento e segmentação dos dados provenientes da eletroencefalografia; representação do conjunto de entradas como vetores no espaço vetorial das principais componentes dos dados; unidimensionalização dos vetores do conjunto de entradas; detecção de mudanças de “conceito” nos dados (*concept drifts detection*); e caracterização dos períodos antecipatórios de crises epiléticas pelo uso de um teste estatístico biamostrais não paramétrico. Em termos de base de dados, esta é formada pelos dados dos registros eletroencefalográficos de dez pacientes do banco de dados do Projeto EPILEPSIAE, que se tratou de um projeto conjunto entre centros de pesquisa de Portugal, França e Alemanha[4] [5] [6]. Ressalte-se que esses dados foram gentilmente disponibilizados de forma gratuita pelo Centro de Informática e Sistemas da Universidade de Coimbra, o centro de pesquisa português integrante do EPILEPSIAE.

1.3 Trabalhos Correlatos

Nesta Seção abordaremos dois trabalhos cujos temas envolvem o uso de sistemas computacionais para a aferição de períodos pré-crise.

1.3.1 *Optimal Preictal Period in Seizure Prediction*

O referido trabalho visa identificar períodos pré-crise para a tarefa de predição de crises epiléticas. Sua motivação foi a de encontrar os períodos pré-crise mais discriminativos, evitando a aferição errônea da duração de tais períodos, o que afetaria negativamente os resultados dos sistemas preditores de crises.

O método apresentado no trabalho em questão baseia-se na utilização de histogramas de distribuição de amplitude em amostras *preictais* (pré-criSES) e *interictais* (períodos entre crises). Um histograma de distribuição de amplitude funciona como a representação de amostras de uma determinada *feature* (característica) dos dados associadas a uma classe, no qual o eixo da *feature* é discretizado em um certo número de pequenos grupos de intervalos, i. e., *bins*, sendo que cada um desses pequenos grupos representa a quantidade acumulada de amplitudes que recaem no intervalo

representado por um *bin*. A ideia central do método consiste na seleção do período pré-crise que possui a menor área comum entre dois histogramas de distribuição de amplitude normalizados.

Os resultados obtidos pelo trabalho mencionado nessa subseção mostraram que o chamado período pré-crise ótimo pode variar de crise para crise. Em relação à base de dados, foram utilizados os registros do EEG intracranial de cinco pacientes portadores de epilepsia que manifestaram apenas crises do tipo focal (complexa ou simples), registros estes oriundos do banco de dados do Projeto EPILEPSIAE[7].

1.3.2 *Preictal State Identification by Synchronization Changes in Long-term Intracranial EEG Recordings*

O trabalho exposto nesta Subseção tem por objetivo investigar as transições nos sinais neuronais que ocorrem previamente ao surgimento de uma crise epilética do lobo temporal mesial, através do desenvolvimento de uma extração automática de padrões representativos da atividade de sincronização *interictal*, fase na qual não há nenhuma crise epilética em andamento.

Basicamente, a identificação do estado pré-crise é composta de três estágios: pré-processamento dos dados no sentido de determinar uma “biblioteca de referência” de padrões característicos de sincronização *interictal*, através do algoritmo de clusterização *K-means* e da aferição de variáveis discriminantes capazes de diferenciar os estágios *preictal* e *interictal*; classificação do padrão de sincronização através do cálculo da distância mínima de Mahalanobis em relação aos padrões de referência, além da detecção de *outliers*; e a avaliação das métricas de sensibilidade e especificidade da detecção por curvas ROC.

Em 70% dos experimentos (36 de um total de 52 crises) foi observada a presença de um estado específico de sincronização do cérebro cerca de horas antes do início de uma crise epilética, estado este caracterizado por alterações nos níveis de sincronização, seja na forma de um aumento ou decréscimo destes. Tal estado de sincronização foi mais frequentemente localizado próximo da zona epileptogênica primária. Foram utilizados como base de dados os registros do EEG intracranial de cinco pacientes portadores de epilepsia.

1.4 Estrutura da Tese

Em relação à organização desta tese, ela encontra-se estruturada da seguinte forma: o Capítulo 2 aborda conceitos a respeito da epilepsia, como a sua definição, etiologia, os tipos de crises epiléticas, de acordo com a classificação adotada pela *International League Against Epilepsy* (ILAE), além dos aspectos básicos da eletroencefalografia;

o Capítulo 3 introduz os métodos que serviram como fundamentos para o desenvolvimento do presente trabalho e a base de dados utilizada; o Capítulo 4 detalha a implementação da metodologia de caracterização de períodos antecipatórios de crises epiléticas; enquanto os resultados obtidos e a análise desses resultados propriamente ditos encontram-se no Capítulo 5; no Capítulo 6 temos o resumo do trabalho realizado, as principais contribuições e os trabalhos a serem desenvolvidos futuramente como forma de aperfeiçoamento da metodologia de caracterização de períodos antecipatórios aqui exposta; por fim, são citadas as referências que serviram de base bibliográfica para esta tese de Doutorado.

Capítulo 2

Epilepsia

A epilepsia foi formalmente definida em 2005 como um distúrbio que acomete o cérebro, caracterizado pela existência de uma predisposição duradoura para gerar crises epiléticas. Essa definição só é aplicável quando da ocorrência de pelo menos duas crises não provocadas em um intervalo maior que 24 horas[8].

Entretanto, posteriormente a ILAE, atendendo as recomendações de uma força-tarefa composta por especialistas, substituiu a a definição prática de epilepsia por circunstâncias especiais que anteriormente não atendiam aos requisitos de quantidade de crises e intervalo de tempo de ocorrência entre crises, assim como passou a classificar a epilepsia como uma **doença** e não mais como um distúrbio, pois este segundo termo acabava por minimizar a natureza grave da epilepsia enquanto problema de saúde[8]. No tocante às circunstâncias especiais, estas seriam:

- No mínimo duas ocorrências de crises epiléticas (ou crises epiléticas reflexas) não provocadas, com intervalo maior do que 24 horas entre elas;
- Ocorrência de uma crise epilética (ou crise epilética reflexa) não provocada e a possibilidade de ocorrência de crises futuras similar ao risco geral de recorrência de crise (probabilidade de pelo menos 60% de ocorrência);
- Diagnóstico de síndrome epilética.

Por crise epilética reflexa, entenda-se como crises epiléticas cuja ocorrência pode ser claramente associada a algum estímulo externo reconhecido, sendo que tais estímulos externos podem ser de natureza sensorial, elétrica ou bioquímica. Crises epiléticas reflexas tem uma prevalência de 4 a 7% entre os pacientes portadores de epilepsia[9][10].

O presente capítulo abordará os principais conceitos a respeito da epilepsia, assim como apresentará uma concisa explicação sobre os tipos de crises epiléticas. Também será apresentado uma breve explanação sobre a eletroencefalografia e a sua

importância no diagnóstico de epilepsia e na identificação da ocorrência de crises epiléticas.

2.1 Conceitos

A seguir, apresentamos conceitos importantes a respeito da epilepsia, a saber, as definições de epilepsia e de síndrome epilética. Além disso, os principais aspectos da etiologia da epilepsia serão abordados.

2.1.1 Definição de Epilepsia

A epilepsia pode ser definida como um transtorno do cérebro caracterizado pela ocorrência transitória de sinais e/ou sintomas devido à presença de atividade neuronal anormal, excessiva ou síncrona, que podem se manifestar em forma de crises ou de comportamento não usual, sendo que em alguns casos ocorre a perda de consciência[11]. Essas descargas elétricas tem como origem a movimentação iônica através da membrana celular[12].

Em relação ao diagnóstico de epilepsia, este pode ser estabelecido clinicamente quando houver recorrência de crises epiléticas não provocadas em pessoa sem fatores desencadeantes agudos ou potencialmente reversíveis. Crises provocadas são aquelas que decorrem de uma causa conhecida, como por exemplo intoxicação aguda, distúrbios metabólicos etc. Em contrapartida, a expressão síndrome epilética é empregada para denotar um conjunto de sinais e sintomas que definem de maneira única uma condição epilética[13].

2.1.2 Etiologia

A epilepsia possui multi fatores etiológicos, de caráter genético, estrutural e metabólico, tais como a epilepsia proveniente de causas genéticas, tumores cerebrais (benignos e malignos), traumas no cérebro, infecções bacterianas e virais, e.g., meningite e encefalite, neurocisticercose, toxoplasmose, esclerose múltipla, acidente vascular encefálico, má formações cavernosas cerebrais e no desenvolvimento do córtex, eventos perinatais adversos, dentre outros[14].

Ademais, é importante citar que a maior parte das doenças que acometem a substância cinzenta do cérebro, algumas patologias da substância cerebral branca (por exemplo, doenças metabólicas) e inúmeras doenças sistêmicas também podem causar crises epiléticas.

2.2 Crises Epiléticas

Apesar da ILAE ter adotado uma nova classificação para os tipos de crises epiléticas[15][16], neste trabalho optou-se pela classificação anterior (1981), por ser esta a classificação adotada na produção da base de dados utilizada nesta pesquisa. Abordaremos a seguir as crises epiléticas parciais, também conhecidas como focais, e as crises generalizadas(desde o início).

2.2.1 Crises Focais (Parciais)

São denominadas assim as crises que afetam um dos dois hemisférios do cérebro. Esse tipo de crise pode ser subdividido em focal simples, focal complexa e secundariamente generalizada[12]. Distingue-se a crise focal complexa da crise focal simples pelo fato de que na focal complexa a consciência é comprometida. Por consciência, entenda-se a capacidade de percepção consciente e de interação com o ambiente de forma adequada. Tanto a crise focal simples quanto a crise focal complexa podem evoluir, afetando não apenas o hemisfério em que se iniciaram, mas também o outro hemisfério cerebral, o que caracteriza uma crise focal secundariamente generalizada.

2.2.2 Crises Generalizadas (desde o início)

Denominam-se dessa forma as crises que atingem, desde o seu início, os dois hemisférios cerebrais. De acordo com a ILAE[15], as crises generalizadas são classificadas em relação aos sintomas que elas manifestam. Nesse ínterim, as crises generalizadas são divididas em dois subtipos: *crises motoras* e *não motoras (crises de ausência)*. Para além dessa classificação inicial, as crises motoras são especificadas como *tônica*, *clônica*, *tônica-clônica*, *mioclônica*, *mioclônica-tônica-clônica*, *mioclônica-atônica*, *atônica* e *espamo epilético*, enquanto que as crises não motoras são subdivididas em *ausência típica*, *ausência atípica*, *ausência mioclônica* e *mioclonia palpebral*, como mostrado na Tabela 2.1 a seguir.

Subtipo	Tipo
Motora	Tônica-clônica
	Tônica
	Clônica
	Mioclônica
	Mioclônica-tônica-clônica
	Mioclônica-atônica
	Atônica
Espamo epilético	
Não motora (ausência)	Ausência típica
	Ausência atípica
	Ausência Mioclônica
	Mioclonia palpebral

Tabela 2.1: Tipos de crises generalizadas.

Crises tônicas se caracterizam pelo enrijecimento muscular, crises clônicas por abalos musculares generalizados[12], por conseguinte, crises tônicas-clônicas apresentam os sintomas dessas duas crises. Em relação às crises mioclônicas, sua duração é breve e sua manifestação ocorre com mais frequência ao despertar. Esse tipo de crise é caracterizado por movimentos bruscos, mais notadamente dos membros superiores, como se a pessoa tivesse sofrido um choque ou tomado um susto. Crises mioclônicas-tônicas-clônicas representam uma junção dos sintomas das crises tônica-clônica e mioclônica. Sobre as crises atônicas, são crises nas quais há perda súbita do tônus muscular[17] e quando associada a essa perda de tônus muscular há sintomas de crises mioclônicas, temos as crises mioclônicas-atônicas. A respeito do espasmo epilético, este é um tipo de crise que se define por sua curta duração (poucos segundos) e por se manifestar como repentinas contrações dos músculos axiais e proximais dos membros, seja em forma de flexão ou extensão[18].

Por outro lado, as crises de ausência típica e atípica se manifestam em breves episódios de comprometimento da consciência (em maior grau na crise de ausência típica) e sintomas motores discretos, tais como automatismos manuais e orais e alteração do tônus muscular. A crise de ausência mioclônica é mais rara que as crises de ausência típica e atípica, tendo em comum com estas o comprometimento da consciência, mas diferindo-se por apresentar espasmos rítmicos constantes de ambos os membros superiores[17]. Por fim, temos a crise de mioclonia palpebral. Esta crise caracteriza-se por breves e repetidos espasmos mioclônicos das pálpebras, i.e., tremores das pálpebras, com a possibilidade de haver movimento dos globos oculares para cima e movimento da cabeça ligeiramente para trás. A duração da mioclonia palebral geralmente é inferior a seis segundos, porém ela pode ocorrer

várias vezes ao dia[19].

2.3 Eletroencefalografia

A eletroencefalografia, comumente chamado de eletroencefalograma (EEG), é o procedimento mais útil no diagnóstico da epilepsia e contribui de maneira significativa para a correta identificação dos tipos de crise, além de ser importante na descoberta da existência de determinadas síndromes epiléticas[20][21]. O EEG é um registro da atividade elétrica cerebral usando eletrodos colocados no escalpo, sendo estes os responsáveis por captar qualquer alteração na atividade neuronal.

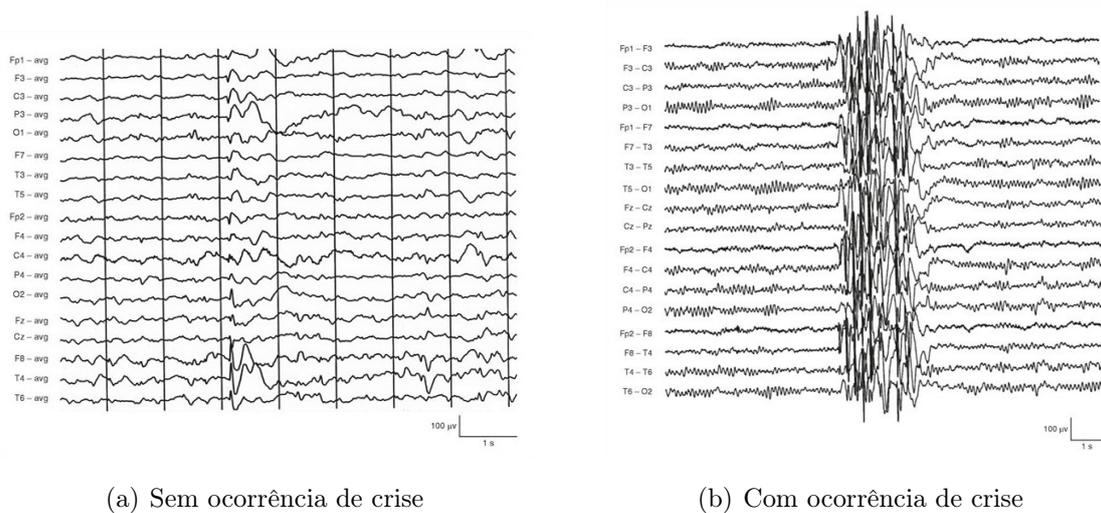


Figura 2.1: Fragmentos de um EEG

[22]

Através da eletroencefalografia torna-se possível obter as respostas para as indagações mais importantes que são feitas durante a investigação diagnóstica dos pacientes suspeitos de sofrerem de epilepsia:

1. O paciente tem epilepsia?
2. Onde se localiza a zona epileptogênica (região neuronal responsável pela geração das crises epiléticas)?
3. O quão boa é a terapia?

Notadamente, se faz necessário ressaltar que importantes avanços na epileptologia foram alcançados graças ao surgimento de uma técnica avançada de EEG, o vídeo-eletroencefalograma(Vídeo-EEG). O Vídeo-EEG permite um diagnóstico mais preciso acerca dos tipos de crises epiléticas, além de ser bastante útil no processo de aferição do número de crises e da localização da área epileptogênica[21].

Por fim, nem o caráter interpretativo dos resultados do EEG pode ser considerada uma desvantagem no tocante ao tratamento e ao diagnóstico de epilepsia, pois dificilmente um especialista na área se equivocaria ao ponto de confundir as alterações epileptiformes com as variantes normais de padrões cerebrais.

Capítulo 3

Métodos e Base de Dados

Visando favorecer a compreensão da metodologia para detecção de períodos antecipatórios de crises epiléticas, tema deste trabalho, neste Capítulo serão abordados os métodos sob os quais a metodologia em questão foi desenvolvida. Além disso, será apresentada a base de dados adotada para a realização dos experimentos pertinentes.

3.1 Banco de Dados EPILEPSIAE

A respeito da base de dados da presente tese, foi utilizado o banco de dados integrante do projeto EPILEPSIAE, um projeto de pesquisa financiado pela União Européia, voltado para o desenvolvimento de algoritmos de predição de crises epiléticas que atuam no monitoramento do EEG visando a identificação de padrões precursores dessas crises[4][5].

O banco de dados EPILEPSIAE disponibiliza gravações de EEG de longa duração de 275 pacientes, assim como uma extensa quantidade de metadados e anotações padronizadas de cada um dos *datasets* que o compõem[5]. Os dados oriundos das gravações foram divididos e armazenados em blocos, sendo disponibilizados à comunidade científica sob a forma de arquivos binários de formato próprio. Os metadados referentes às gravações foram armazenados em uma tabela específica do bando de dados e podem ser acessados pela comunidade através de *webclients* mantidos pelos centros integrantes do projeto[6]. Cada leitura do aparelho de EEG foi nomeada pelo termo *sample* e a taxa de amostragem das gravações variou entre 256, 512 e 1024 *samples*/segundo.

Ainda sobre os dados eletroencefalográficos do EPILEPSIAE, em tese estes seriam gravados de forma contínua durante o período de coleta ao qual os pacientes participantes do projeto foram submetidos. Porém, houve situações em que isto não foi possível e o EEG teve que ser interrompido e reiniciado, seja pela necessidade de adição de eletrodos de superfície, realização de avaliações clínicas ou até mesmo por causas técnicas relativas ao sistema de gravação dos dados do EEG[4].

3.2 Análise das Componentes Principais

A análise das componentes principais (PCA, do inglês *Principal Component Analysis*) é um método estatístico multivariado que tem por finalidade a análise dos dados em busca de redução de dimensionalidade, eliminação de sobreposições e a obtenção de formas mais representativas de dados a partir de combinações lineares das variáveis que compõem os dados originais. Basicamente, a PCA provê uma maneira de representar os dados (conjunto de observações de variáveis possivelmente correlacionadas) como um conjunto de novas variáveis ortogonais (linearmente não correlacionadas) chamadas de componentes principais[23], por meio de uma transformação ortogonal que modifica os dados, transpondo-os para um novo sistema de coordenadas no qual, por qualquer projeção dos dados, a maior variância fica ao longo da primeira coordenada, i.e., primeira componente principal, a segunda maior variância fica ao longo da segunda coordenada e assim sucessivamente. Matematicamente, a PCA depende da decomposição de valores singulares [24].

Em relação ao processo de obtenção das componentes principais, ele se dá da seguinte forma:

- obtenção dos dados (conjunto de vetores de dimensão n);
- cálculo do *vetor médio* do conjunto de vetores (média dos dados);
- subtração do vetor médio de todos os vetores do conjunto de vetores;
- determinação da **matriz de covariância** a partir das subtrações obtidas:
 - resultado da média do produto de cada subtração por ela mesma
 - de dimensão $n \times n$;
- cálculo dos **autovalores** e **autovetores** da matriz de covariância;
- reordenação das colunas da matriz de covariância:
 - colunas ordenadas em ordem decrescente de autovalor
 - primeira coluna corresponde à primeira componente principal; a segunda coluna corresponde à segunda componente principal e assim por diante.

A título de exemplificação, demonstraremos o uso da PCA para a redução de dimensionalidade. Para este fim, aplicaremos a PCA sobre os dados da base de dados pública *Iris*[25]. A base *Iris* é composta por 3 subconjuntos, correspondendo as 3 espécies da flor do gênero *iris* e possui 4 atributos: “*Sepal length*” e “*Sepal width*”(respectivamente, o comprimento e a largura das sépalas¹); “*Petal length*” e

¹Localizadas abaixo das pétalas, geralmente de coloração verde, sua função é garantir a proteção das outras partes da flor no botão floral

“*Petal width*” (comprimento e largura das pétalas, respectivamente). Na figura 3.1.a, temos os dados da base Iris considerando 3 de seus atributos (“*Sepal length*”, “*Petal length*” e “*Petal width*”), enquanto na Figura 3.1.b, temos o resultado da aplicação da PCA, considerando apenas as suas duas primeiras componentes principais.

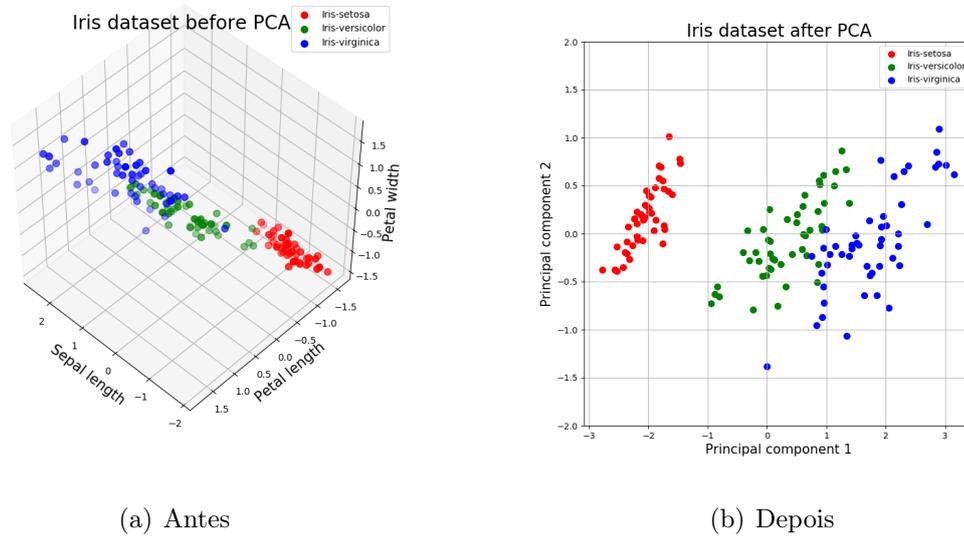


Figura 3.1: Base de dados Iris antes e depois da PCA

3.3 Detecção de Mudança nos Dados

A detecção de mudança nos dados compreende uma vasta gama de métodos, porém nem todos eles aplicáveis em situações de fluxo de dados, como por exemplo, dados oriundos de séries temporais. Trata-se de um campo com uma longa tradição na área de estatística[26]. Uma mudança nos dados é verificada quando ocorre um desvio de conceito dos dados (*concept drift*)[27], ou seja, quando o “*conceito*” sobre quais dados estão sendo coletados muda de tempos em tempos, após um período mínimo de estabilidade, algo comum em ambientes não estacionários, nos quais a distribuição dos dados, estatisticamente falando, pode mudar ao longo do tempo[28].

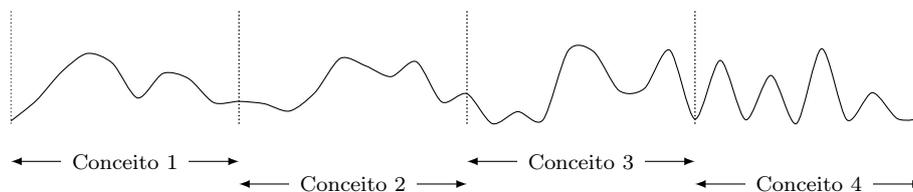


Figura 3.2: Exemplo de desvio de conceito dos dados

Em termos de desvio de conceito dos dados, este pode ocorrer de modo súbito (com os dados passando de um conceito K para um conceito W de forma abrupta),

de maneira incremental (a mudança de um conceito K para um conceito W se dá através da ocorrência de muitos conceitos intermediários entre eles), de forma gradual (entre a mudança definitiva entre conceitos K e W , ocorrem muitos retornos ao conceito anterior ao longo do tempo até que W se estabeleça)[28][29] ou quando há conceitos recorrentes (conceitos vistos no passado reaparecem após um certo tempo)[30]. Uma situação crítica a ser evitada em se tratando de detecção de mudança é considerar que um *outlier* ou um ruído nos dados sejam tomados por *concept drift*, quando na verdade tratam-se de anomalias[29].

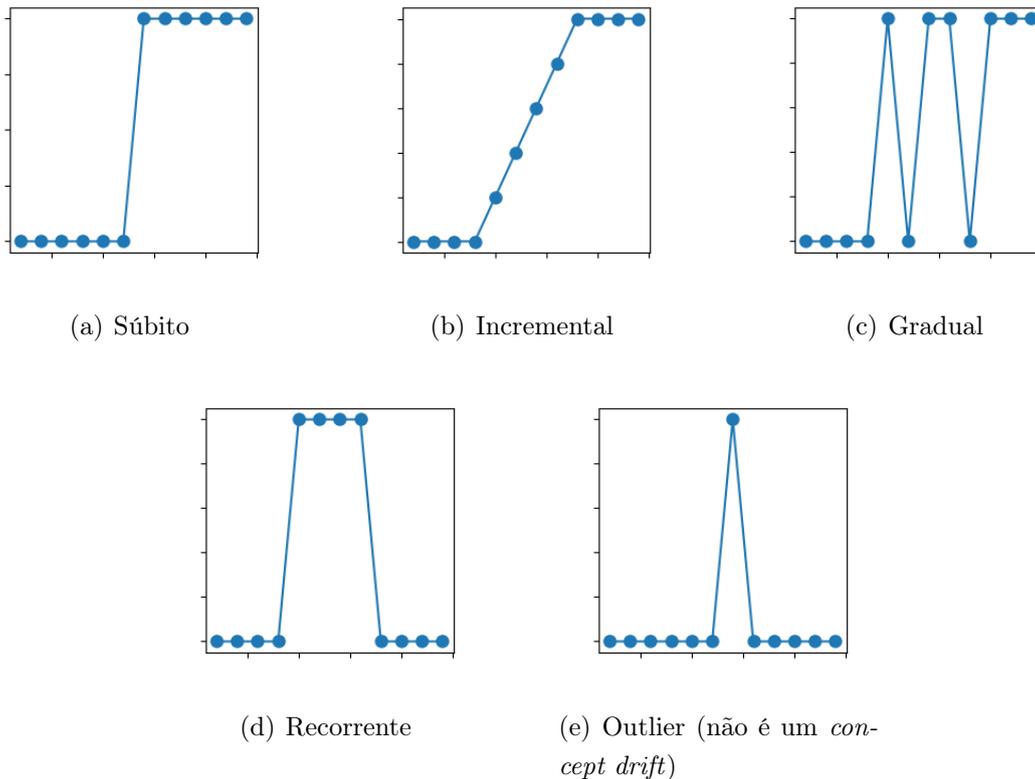


Figura 3.3: Padrões de *concept drift*

Na presente Seção abordaremos métodos de detecção de *concept drifts*. Mais precisamente, apresentaremos alguns dos métodos que são passíveis de serem usados em fluxo de dados, como o *Cumulative Sum*(CUSUM) e sua variante, o teste de *Page-Hinkley*; o *Drift Detection Method*(DDM); e o *ADaptive WINdowing*(ADWIN).

3.3.1 CUSUM e o Teste de Page-Hinkley

O teste de soma cumulativa, CUSUM (do inglês *cumulative sum*), como o próprio nome diz, baseia-se no cálculo de uma soma cumulativa s e funciona da seguinte forma: dada uma sequência de n observações X , define-se $z_i = (X_i - \mu) / \sigma$, onde μ é a esperança (ou valor esperado) de X_i e σ é o desvio padrão. E para cada um dos

elementos X_i de X calcula-se a soma acumulada $s_i = \max(0, s_{i-1} + z_i - k)$, onde k é um parâmetro usado para ajustar a sensibilidade da detecção de *concept drifts*. Quando s_i é maior que um *threshold* h , então um alarme é disparado, sinalizando a detecção da mudança nos dados[31][32]. A seguir, temos uma possível implementação do algoritmo padrão do CUSUM.

Algoritmo 1: CUSUM(implementação padrão)

Entrada: X, k, h

Saída: α (alarme de detecção de *concept drift*)

```

1 início
2    $s_0 = 0$ 
3   para cada  $i \in 1, \dots, n$  faça
4      $\mu(X_i)$ 
5      $\sigma(X_i)$ 
6      $z_i \leftarrow (X_i - \mu)/\sigma$ 
7      $s_i \leftarrow \max(0, s_{i-1} + z_i - k)$ 
8     se  $s_i > h$  então
9       alarme  $\alpha(i)$ 
10       $s_i \leftarrow 0$ 
11    fim
12  fim
13 fim
```

Uma variante do CUSUM como método de detecção de *concept drifts* é o teste de *Page-Hinkley*[28][33], considerado adequado especificamente para a detecção de mudanças nos dados em ambientes de fluxo de dados. O algoritmo do teste de *Page-Hinkley* é dado a seguir, nele há duas variáveis $s_i = s_{i-1} + z_i - k$ e $S_i = \min(s_i, S_{i-1})$. Quando a diferença entre a soma acumulada s_i e o valor mínimo anterior S_{i-1} é superior a um determinado *threshold* h , um alarme α é disparado[34]. Assim como o CUSUM, o teste de *Page-Hinkley* também é sensível à escolha dos valores para h e k , sendo k o nível de tolerância a ser adotado.

Algoritmo 2: CUSUM(teste de *Page-Hinkley*)

Entrada: X, k, h **Saída:** α (alarme de detecção de *concept drift*)

```
1 início
2    $s_0 = 0$ 
3    $S_0 = 0$ 
4   para cada  $i \in 1, \dots, n$  faça
5      $\mu(X_i)$ 
6      $\sigma(X_i)$ 
7      $z_i \leftarrow (X_i - \mu)/\sigma$ 
8      $s_i \leftarrow s_{i-1} + z_i - k$ 
9      $S_i \leftarrow \min(s_i, S_{i-1})$ 
10    se  $s_i - S_i > h$  então
11      alarme  $\alpha(i)$ 
12       $s_i \leftarrow 0$ 
13       $S_i \leftarrow 0$ 
14    fim
15  fim
16 fim
```

3.3.2 Drift Detection Method

O método de detecção de *concept drifts* proposto por Gama et al.[35], DDM, é aplicável no contexto de modelos de predição. DDM utiliza distribuição binomial e tem por objetivo monitorar a quantidade de erros produzidos por um modelo cujo aprendizado foi baseado nos elementos anteriores de uma sequência qualquer. A detecção de mudança no contexto dos dados pelo DDM ocorre quando a taxa de erro de predição aumenta.

Em termos práticos, considerando que o número de erros em uma amostra sequencial s de tamanho n é modelado por uma distribuição binomial, para cada elemento s_i de s , sendo p_i a taxa de erro de um modelo de predição no instante de tempo i , o DDM calcula o desvio padrão st através da equação $st_i = \text{sqr}t(p_i(1 - p_i)/i)$, onde *sqr*t significa o cálculo da raiz quadrada. Os valores mínimos da taxa de erros e do desvio padrão, p_{min} e st_{min} , respectivamente, observados até um instante de tempo i são usados na detecção de possíveis *concept drifts*, como demonstrado no algoritmo a seguir.

Algoritmo 3: DDM

Entrada: *sequência s***Saída:** α (alarme de detecção), γ (aviso de *concept drift* iminente)

```
1 início
2    $p_{mins} = []$ 
3    $st_{mins} = []$ 
4    $ini \leftarrow 1$ 
5   para cada  $i \in 1, \dots, n$  faça
6      $p_i \leftarrow \text{calcula\_taxa\_de\_erro}(s[ini\dots i])$ 
7      $st_i \leftarrow \text{sqr}(p_i(1 - p_i)/i)$ 
8      $p_{mins} \leftarrow p_i$ 
9      $st_{mins} \leftarrow st_i$ 
10     $p_{min} \leftarrow \text{min}(p_{mins})$ 
11     $st_{min} \leftarrow \text{min}(st_{mins})$ 
12    se  $p_i + st_i > p_{min} + 3 * st_{min}$  então
13      alarme  $\alpha(i)$ 
14       $ini \leftarrow i$ 
15       $p_{mins} \leftarrow p_{mins}[i]$ 
16       $st_{mins} \leftarrow st_{mins}[i]$ 
17    fim
18    senão se  $p_i + st_i > p_{min} + 2 * st_{min}$  então
19      aviso  $\gamma(i)$ 
20    fim
21  fim
22 fim
```

O DDM tem como desvantagem a possibilidade de em determinadas ocasiões ser um pouco lento para detetar os *concept drifts*, dado que p_i é aferido levando-se em conta todos os elementos da amostra sequencial desde a última mudança (*drift*) observada, o que pode acarretar a necessidade de um número considerável de elementos da amostra para fazer com que p_i seja significamente maior que p_{min} . Ademais, o número de elementos da amostra armazenados em memória depois da emissão de um alarme pode se tornar grande.

3.3.3 ADWIN

O ADWIN é um detetor de *concept drifts* que utiliza uma janela deslizante w de tamanho variável para a detecção de mudanças nos dados. O algoritmo do ADWIN

analisa w para tentar identificar duas subjanelas w_1 e w_2 de w que possuam médias diferentes entre si e quando isso ocorre, conclui-se que os valores esperados de w_1 e w_2 são díspares, o que significa que w_1 , a parte mais “antiga” de w , possui dados oriundos de uma distribuição diferente de w_2 , a parte mais “recente”. Logo, w_1 e w_2 representam dois “conceitos” de dados distintos[36].

Quando uma detecção ocorre, w diminui em tamanho, pois os dados que correspondem ao “conceito” mais antigo, a subjanela w_1 , são deletados. Por outro lado, w tende a crescer se nenhuma detecção acontecer, porém o tamanho máximo que w pode atingir é “estatisticamente consistente com a hipótese de que não houve nenhuma mudança na média dos dados dentro da janela”[37]. Por consequência, no ADWIN a detecção de *concept drifts* está intrinsecamente relacionada com a redução do tamanho de w , ou seja, toda vez que w diminui, pode-se afirmar com segurança que uma mudança no “conceito” dos dados ocorreu. Além disso, a média atual em w , que pode ser obtida em qualquer instante de tempo, funciona como uma estimativa confiável da média corrente do próprio fluxo de dados ou série temporal que está sendo analisada.

Sobre o algoritmo do ADWIN, este tem por parâmetros de entrada a sequência $s(s = s_1, s_2, \dots, s_t, s_{t+1} \dots s_n)$ de dados, o nível de confiança γ , o tamanho b da sequência a ser analisada a cada interação do algoritmo e o “passo” k para o deslocamento da janela deslizante. O algoritmo é parametrizado por um teste T , que é responsável pela tarefa de comparar as médias de w_1 e w_2 e definir se os dados das subjanelas são, com confiança igual a $1 - \gamma$, provenientes da mesma distribuição. A analogia usada é de que a fração de tamanho b de s será armazenada em um *histograma exponencial* composto por b compartimentos e que $b - 1$ testes T serão executados sobre esse histograma.

Em relação ao acesso aos dados da sequência s , o valor de s_t só é possível de ser obtido no respectivo instante de tempo t . Um detalhe a ser destacado a respeito do algoritmo é o seu consumo de memória, que está atrelado ao fato do histograma exponencial (janela deslizante) do ADWIN, como dito anteriormente, não possuir tamanho fixo, diferente do que ocorre com os histogramas exponenciais convencionais. Tal fato acarreta em um consumo de memória na ordem de $O(\log W)$. A seguir, o algoritmo é descrito detalhadamente.

Algoritmo 4: ADWIN

Entrada: s, γ, b, k **Saída:** δ (*conceito detetado*)

```
1 início
2    $W \leftarrow s[...b-1]$ 
3   para cada  $i \in (b...n)$  faça
4      $w_i \leftarrow W + s[i]$ 
5      $\delta \leftarrow False$ 
6     para cada  $j \in (1...tamanho(w_i) - 1, k)$  faça
7        $w_1 \leftarrow w_i[...j]$ 
8        $w_2 \leftarrow w_i[j+1...]$ 
9        $R \leftarrow T(w_1, w_2, \gamma)$ 
10      se  $R ==$  “mudança” então
11         $W \leftarrow w_2$ 
12         $\delta \leftarrow True$ 
13      fim
14    fim
15    se  $\delta \neq False$  então
16      retorna  $\delta$ 
17    fim
18  senão
19     $W \leftarrow w_i$ 
20  fim
21 fim
22 fim
```

Analisando o algoritmo, nota-se que a escolha do valor do parâmetro k determina o número de testes T que serão executados em cada janela w_i da sequência s a ser analisada pelo ADWIN. Se por um lado não é obrigatório que k seja igual a 1, valores de k maiores que 1 podem causar um atraso na detecção de mudança nos dados na ordem de k instantes de tempo. Além disso, para evitar que o consumo de memória cresça indiscriminadamente no caso de nenhuma detecção de mudança nos dados acontecer, pode-se fixar um valor máximo para o tamanho do histograma exponencial do algoritmo.

3.4 Teste Estatístico *Two-Sample Kolmogorov-Smirnov* (2S-KS Test)

O teste estatístico biamostrado Kolmogorov-Smirnov (2S-KS Test) é um teste de caráter não paramétrico usado para verificar se duas amostras são provenientes de uma mesma distribuição[38]. Para isso, o teste obtém a função de distribuição acumulada (FDA)[39] de duas amostras x de tamanho m e y de tamanho n , compara as FDAs de ambas e encontra o ponto que representa a distância máxima $Dmax$ entre as FDAs, sendo essa distância máxima definida como uma medida de dessemelhança (ou semelhança) entre as duas amostras testadas[40]. O 2S-KS Test avalia duas hipóteses, a nula H_0 e a alternativa H_1 , sendo estas descritas por:

$$\begin{aligned} H_0 &: \text{ as amostras são provenientes de uma mesma distribuição} \\ H_1 &: \text{ as amostras são provenientes de distribuições distintas} \end{aligned} \quad (3.1)$$

Se $Dmax$ for maior que um valor crítico Δ_α estabelecido, então a hipótese nula é rejeitada e a hipótese alternativa não é rejeitada com nível de certeza $1 - \alpha$. Usualmente, defini-se $\alpha = 0.05$, o que acarreta em um valor crítico de 95%. As equações que expressam $Dmax$ e Δ_α são as seguintes:

$$Dmax_{m,n} = \max(x, y) |FDA(x) - FDA(y)| \quad (3.2)$$

$$\Delta_{0,05} = 1.36 \sqrt{\frac{1}{m} + \frac{1}{n}} \quad (3.3)$$

, sendo m e n , como dito anteriormente, o tamanho das amostras x e y .

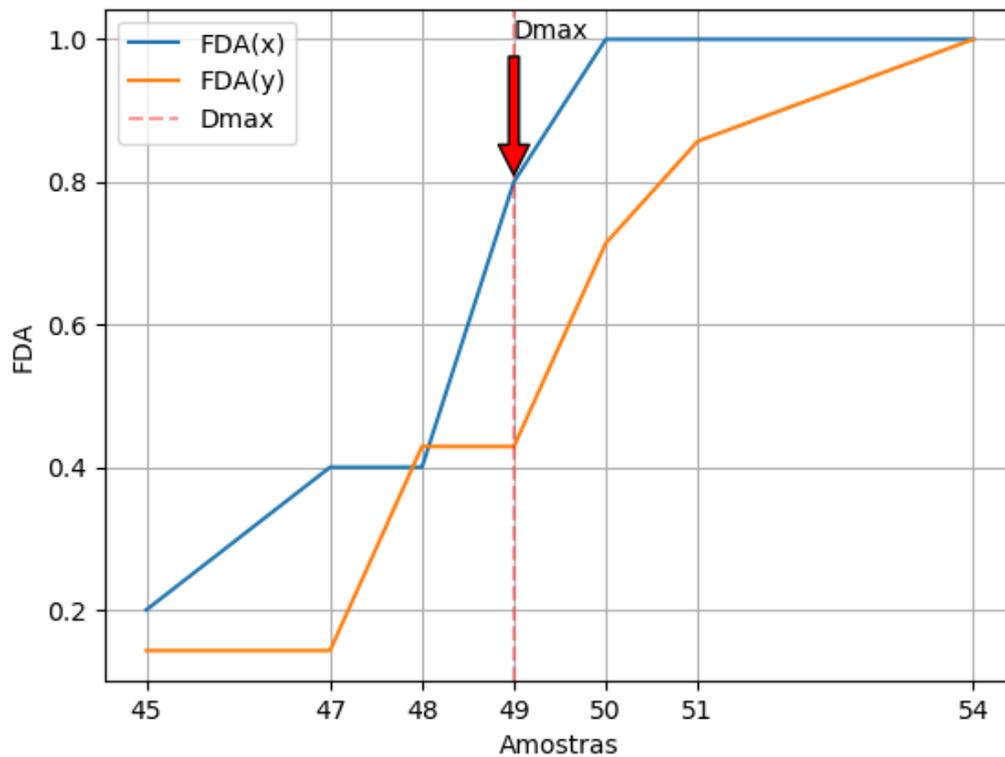


Figura 3.4: FDAs e D_{max}

Se por um lado o 2S-KS Test tem a vantagem de não estabelecer nenhum pressuposto sobre a distribuição das amostras analisadas, ele tem como limitação o fato de que as amostras tem que ser unidimensionais, dado que para uma amostra n dimensional seria possível a obtenção de $1 - n$ FDAs diferentes.

Capítulo 4

Metodologia para a Caracterização de Períodos Antecipatórios de Crises Epiléticas

Neste capítulo abordaremos a metodologia desenvolvida na presente tese de doutorado, que tem por objetivo a caracterização de períodos antecipatórios de crises epiléticas baseada na análise dos dados obtidos através do eletroencefalograma em sua forma “crua”, ou seja, sem submetê-los a nenhuma técnica usual de processamento de sinais. Por “períodos antecipatórios”, entenda-se frações ou segmentos do EEG correspondentes ao estágio *interictal* (período entre crises epiléticas, i.e., períodos em que o paciente de epilepsia não desenvolveu crise alguma) que apresentem alguma similaridade com a parte do EEG relativa ao estágio *ictal* (período no qual uma crise está em andamento). A metodologia de caracterização de períodos antecipatórios de crises epiléticas em questão dividi-se em cinco fases:

- segmentação dos dados em entradas;
- representação vetorial do conjunto de entradas em um espaço multidimensional;
- unidimensionalização dos vetores do espaço;
- detecção multidimensional de *concept drifts*;
- avaliação dos períodos antecipatórios de crises.

4.1 Segmentação dos Dados em Entradas

Os dados eletroencefalográficos que formaram a base de dados para esta pesquisa são provenientes do EEG dos pacientes partícipes do projeto EPILEPSIAE, cujas coletas

de dados tiveram lugar em Coimbra, Portugal. Os dados obtidos com o EEG de cada paciente foram subdivididos em blocos e cada um desses blocos foi armazenado em arquivo binário, cujo formato foi especificado pelo próprio EPILEPSIAE. Ou seja, para cada paciente há um conjunto de arquivos que representa o total dos dados eletroencefalográficos obtidos. Os dados entre blocos são em regra contínuos, porém há casos em que existe um intervalo entre um bloco e seu subsequente devido a alguma situação técnica, como por exemplo a necessidade de troca ou inclusão de eletrodos durante o exame EEG.

No tocante ao uso dos dados na metodologia de caracterização de períodos antecipatórios de crises epiléticas, estes foram segmentados em entradas de dados de acordo com a taxa de amostragem do EEG (1024 amostras por segundo), com a quantidade de eletrodos usados no EEG (29 ou 27, dependendo do paciente) e as seguintes premissas: uma janela deslizante de tamanho relativo a 1(um) segundo de coleta de dados foi definida para a segmentação dos dados em **entradas**; o *passo* para a movimentação da janela foi definido em 0.5 (meio) segundo, garantindo a sobreposição durante o deslocamento da janela deslizante sobre os dados na criação das entradas propriamente ditas. Na Tabela 4.1 a seguir apresentamos um resumo dos parâmetros mencionados.

Parâmetro	Valor
taxa de amostragem (sr)	1024 amostras/segundo
número de eletrodos (n)	29 ou 27
tamanho da janela (j)	1 segundo
passo da janela (st)	0.5 segundo

Tabela 4.1: Parâmetros usados na segmentação dos dados.

Para fins de criação das entradas a serem utilizadas na caracterização de períodos antecipatórios de crises epiléticas, foram desconsiderados os eletrodos auxiliares, como os de captação de batimentos cardíacos e movimentos dos olhos usados na coleta de dados do projeto EPILEPSIAE, levando-se em conta apenas os eletrodos usados para captação dos dados eletroencefalográficos propriamente ditos. Assim, inicialmente cada entrada criada foi representada sob forma de uma matriz M de dimensão $1024 \times n$, sendo n o número de eletrodos usado no EEG (específico por paciente, 27 ou 29 eletrodos) e 1024 a quantidade de amostras referentes a 1 segundo de coleta de dados, de acordo com a taxa de amostragem adotada (leituras obtidas pelo aparelho de EEG por segundo).

$$M_{1024 \times n} = \begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,n} \\ m_{2,1} & m_{2,2} & \dots & m_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ m_{1023,1} & m_{1023,2} & \dots & m_{1023,n} \\ m_{1024,1} & m_{1024,2} & \dots & m_{1024,n} \end{bmatrix}$$

Seguindo com o processo de segmentação dos dados, aplicou-se um fator de compressão ψ em cada entrada do conjunto de entradas inicialmente obtido, de modo que as entradas pudessem ser representadas como vetores de dimensão $1 \times n$. O fator ψ corresponde ao cálculo da média aritmética dos valores de cada eletrodo obtidos durante 1 segundo (taxa de amostragem), representados pelas colunas da matriz M . Assim, cada atributo de um vetor de entrada v_i corresponde à média dos valores do eletrodo i coletados a cada leitura do EEG.

$$\psi(M_{1024 \times n}) \left(\begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,n} \\ m_{2,1} & m_{2,2} & \dots & m_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ m_{1023,1} & m_{1023,2} & \dots & m_{1023,n} \\ m_{1024,1} & m_{1024,2} & \dots & m_{1024,n} \end{bmatrix} \right) \rightarrow v_{1 \times n} = [m_1, m_2, \dots, m_n]$$

4.2 Representação Vetorial do Conjunto de Entradas em um Espaço Multidimensional

Para fins de análise, os vetores do conjunto de entradas obtidos na segmentação dos dados foram visualizados como pontos em um espaço multidimensional (R^{27} ou R^{29} , a depender do paciente), no intuito de aferir algum tipo de relação entre eles, mais especificamente entre vetores referentes a um mesmo estágio, *interictal* ou *ictal* do EEG. No entanto, para evitar que a análise fosse prejudicada pela *maldição da dimensionalidade*[41], foi aplicada a análise das componentes principais (PCA) dos dados, de modo a reduzir a dimensionalidade do espaço amostral de R^n ($n = 27$ ou 29) para R^p , sendo p o número de componentes principais a serem mantidos na nova representação vetorial das entradas, os vetores \hat{v} de dimensão $1 \times p$.

$$\text{PCA}(v_{1 \times n}) \left([m_1, m_2, \dots, m_n] \right) \rightarrow \hat{v}_{1 \times p} = [c_1, c_2, \dots, c_p]$$

A respeito do número p ($p < n$), de componentes principais, este dependeria da variância acumulada ε desejada. Neste trabalho optou-se por não definir um valor ε único, mas por cinco valores distintos, garantindo assim uma escolha posterior de acordo com o modelo desejado de grau de restrição ou generalidade, caso a caso. Detalhamos os valores de ε adotados na Tabela 4.2 a seguir.

Modelo	Variância (ε)
A	90%
B	80%
C	70%
D	60%
E	50%

Tabela 4.2: Valores da variância acumulada ε .

4.3 Unidimensionalização dos vetores do espaço multidimensional

Nessa terceira etapa, o conjunto de entradas de dados encontra-se representado por um conjunto de vetores \hat{v}_i de dimensão $1 \times p$: $\hat{V} = \{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_p\}$. Nessa Seção buscaremos a unidimensionalização desses vetores, visando a adequação dos dados para a próxima etapa da metodologia, a detecção de *concept drifts*. Para isso, analisaremos a relação entre esses vetores \hat{v}_i s no espaço amostral R^p provenientes da aplicação da PCA sobre os dados, mais precisamente, a relação de distância entre vetores \hat{v}_{i-1} e \hat{v}_i subsequentes. Para o cálculo da distância entre vetores, utilizou-se a distância euclidiana[42].

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (4.1)$$

Algoritmo 5: Unidimensionalização dos vetores de entrada

Entrada: \hat{V} **Saída:** D

```
1 início
2    $D = \{\}$ 
3   para cada  $i \in 2, \dots, p$  faça
4      $distance \leftarrow$  euclidiana ( $\hat{V}_{i-1}, \hat{V}_i$ )
5      $D \{i\} \leftarrow distance$ 
6   fim
7   retorna  $D$ 
8 fim
```

Após a unidimensionalização dos vetores de entrada \hat{v}_i , obtivemos a transformação de \hat{V} em um “sinal” unidimensional D com $z - 1$ elementos ϕ (as distâncias obtidas entre pares $\langle \hat{v}_{i-1}, \hat{v}_i \rangle$), onde z é a quantidade de vetores no conjunto de entradas \hat{V} antes da unidimensionalização. Ao final dessa etapa da metodologia, observa-se como a representação de uma entrada de dados sofreu transformações, passando de uma matriz ($1024 \times n$) a um valor escalar ϕ , como resumimos a seguir.

$$M_{1024 \times n} = \begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,n} \\ m_{2,1} & m_{2,2} & \dots & m_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ m_{1023,1} & m_{1023,2} & \dots & m_{1023,n} \\ m_{1024,1} & m_{1024,2} & \dots & m_{1024,n} \end{bmatrix} \rightarrow v^T_{n \times 1} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} \rightarrow \hat{v}^T_{p \times 1} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} \rightarrow \phi$$

4.4 Deteção de *Concept Drifts*

Dado que o conjunto de entradas está nessa altura representado como um sinal de distância D unidimensional, no qual cada elemento d_i representa um instante de tempo i e o intervalo temporal entre instantes consecutivos i e $i + 1$ é de 0.5 segundo, sobre este sinal D é aplicado o detetor de mudança de conceitos (*concept drifts*) ADWIN com nível de significância $\delta = 0.002$ (o valor de δ é usado para determinar se duas subjanelas correspondem a uma mesmo conceito, o que nesse caso significaria que não houve mudança no conceito dos dados). O sinal D é percorrido de maneira incremental, com passo 1, e ao final da interação é obtido o conjunto de *concepts* \mathbb{C} correspondente.

Relembrando que o sinal D corresponde aos dados do EEG após estes terem

sido segmentados e transformados pelo uso do fator de compressão ψ e da análise das componentes principais (PCA), portanto cada *concept* c_j de \mathbb{C} trata-se de um subsinal de D , cujos elementos representam algum intervalo temporal do EEG. Por conseguinte, cada elemento de c_j refere-se a um dos dois estágios do EEG: *interictal* ou *ictal*. Assim, os *concepts* do conjunto \mathbb{C} podem ser compostos de:

- elementos *interictais*;
- elementos *ictais*;
- uma sequência de elementos *interictais* e uma sequência de elementos *ictais*;
- uma sequência de elementos *ictais* e uma sequência de elementos *interictais*;
- uma sequência de elementos *interictais*, uma sequência de elementos *ictais* e outra sequência de elementos *interictais*.

Considerando a natureza dos elementos que formam cada conceito, a maior parte dos c_j de \mathbb{C} será composta de elementos correspondentes ao estágio *interictal* do EEG e apenas alguns c_j s possuirão elementos do estágio *ictal*, pois os estágios *interictais* são consideravelmente maiores em termos de duração do que os estágios *ictais*. Isto posto, cada *concept* c_j é então rotulado de acordo com a natureza de seus elementos, como demonstrado na Tabela 4.3 seguinte.

Classe do <i>concept</i>	Elementos
Interictal	Interictais
Ictal	Ictais
Earlyictal-Ictal	Interictais Ictais
Ictal-Postictal	Ictais Interictais
Earlyictal-Ictal-Postictal	Interictais Ictais Interictais

Tabela 4.3: Classes dos *concepts*

Dentre as cinco classes de *concepts* possíveis, as classes de interesse são aquelas em que há elementos *interictais* que precedem elementos *ictais* em um mesmo *concept*, pois isso significaria que esse segmento do estágio *interictal* tem alguma similaridade com o estágio *ictal* que o sucede, o que torna tal segmento *interictal* o foco da última fase da metodologia para caracterização de períodos antecipatórios de crises epiléticas. Esses segmentos *interictais* são chamados nesta pesquisa

de períodos *earlyictal*. Portanto, as referidas classes de interesse são os *concepts* **Earlyictal-Ictal** e **Earlyictal-Ictal-Postictal**.

4.5 Avaliação dos Períodos Antecipatórios de Crises

Enfim abordaremos o último passo para que de fato seja(m) caracterizado(s) o(s) período(s) antecipatório(s) de uma crise epilética, passo este que consiste em realizar uma análise estatística sobre a parte *earlyictal* dos *concepts* de interesse, os *concepts* *Earlyictal-Ictal* e *Earlyictal-ictal-Postictal*. Para efetuar a análise estatística utilizou-se o teste biamostrai Kolmogorov-Smirnov (2S-KS Test) na comparação do segmento *earlyictal* E com prefixos dele mesmo, afim de identificar quais prefixos seriam estatisticamente similares ao segmento E como um todo e quais não o seriam. O tamanho do prefixo inicial foi de 10 elementos, por ser esse o tamanho mínimo exigido para que um sinal possa ser avaliado pelo 2S-KS Test. O exemplo seguinte mostra um segmento *earlyictal* E de tamanho k e o seus prefixos.

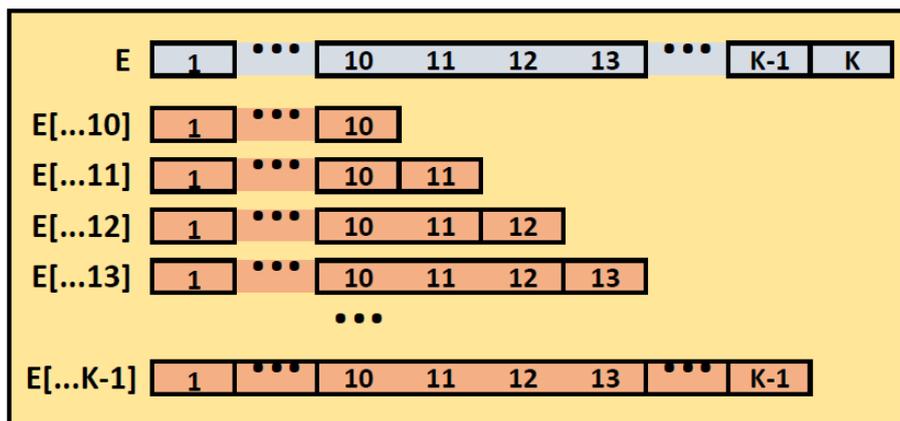


Figura 4.1: Segmento *earlyictal* E e seus prefixos

Algoritmo 6: 2S - KS Test - Avaliação de períodos antecipatórios de crise

Entrada: *earlyictal concept* E

Saída: *períodos antecipatórios* α

```
1 início
2    $\alpha = [ ]$ 
3   para cada  $i \in 10, \dots, tamanho(E)$  faça
4      $ho \leftarrow \text{KS Test}(E, E[...i])$ 
5     se  $ho ==$  “não rejeitada” então
6        $\alpha \leftarrow E[...i]$ 
7     fim
8   fim
9   retorna  $\alpha$ 
10 fim
```

No algoritmo para a avaliação de períodos antecipatórios de crise, todos os prefixos de um *earlyictal* E nos quais o 2S-KS Test não rejeitou a hipótese nula ho (*ambos os sinais comparados são provenientes de uma mesma distribuição*) são classificados como períodos antecipatórios, pois tais prefixos α isoladamente refletem as características do segmento *earlyictal* E como um todo e por isso podem detetá-lo se forem usados para esse fim em um sistema de predição de crises dotado de algum algoritmo ou técnica de classificação de padrões. Por fim, descartando os prefixos nos quais a hipótese nula ho é rejeitada, eliminam-se os possíveis ruídos do segmento *earlyictal* dos *concepts* de interesse.

Capítulo 5

Resultados e Discussões

Nesta Seção apresentamos os resultados obtidos nos experimentos realizados para esta tese de Doutorado. Os resultados são apresentados de maneira individualizada, assim como as respectivas análises propriamente ditas.

5.1 Base Experimental

A base experimental utilizada é composta por dados de dez pacientes do banco de dados EPILEPSIAE, cujos dados eletroencefalográficos foram coletados em Coimbra, Portugal. As crises epiléticas desenvolvidas por esses pacientes durante a coleta de dados foram crises focais (simples, complexas e secundariamente generalizadas). Não há registros de crises generalizadas desde o princípio, seja nos pacientes que compõem a base experimental desta pesquisa ou nos demais pacientes do projeto EPILEPSIAE.

Os registros eletroencefalográficos dos pacientes foram gravados em conjuntos de blocos de dados. Tais blocos de dados provenientes das gravações do EEG são compostos de *samples* que refletem os valores calculados a partir das leituras obtidas dos eletrodos do EEG em um determinado período de gravação (taxa de amostragem). Esses dados não foram armazenados no banco de dados relacional do projeto EPILEPSIAE, mas em arquivos binários, sendo um arquivo para cada bloco de dados EEG, em formato estabelecido pelo próprio projeto. Esses arquivos são do tipo *m-file*, ou seja, arquivos de comandos do MatLab/Octave.

5.2 Resultados Obtidos

Em relação aos períodos antecipatórios, apresentamos para cada crise epilética o resultado obtido com o uso do 2S-KS Teste sobre o respectivo *concept de interesse*, i.e., conceito que contém elementos do tipo *earlyictal*, ou seja, elementos *interictais*

que antecedem os elementos do estágio *ictal* em um mesmo conceito. Destaca-se o fato de que os instantes de tempo são apresentados como números negativos, pois foi tomado por base o início das crises epiléticas. Tomando esses inícios como o instante de tempo 0, o instante de tempo imediatamente anterior é o instante -0.5 segundo, o seu predecessor, o instante -1.0 segundo e assim sucessivamente, de acordo com o “passo” de 0.5 segundo utilizado na formação dos *inputs*, i.e., dados de entrada, utilizados nessa metodologia. Por questão de ética médica, os pacientes estão identificados pelos seus números de registro no projeto EPILEPSIAE, mantendo assim o sigilo a respeito da identidade real dos mesmos.

Ainda sobre os experimentos realizados nesta pesquisa, destacamos que para cada uma das crises dos pacientes foram utilizados cinco modelos de teste, referentes as cinco opções de variância acumulada da PCA usadas na fase de representação vetorial do conjunto de entradas em um espaço R^n sobre os dados. Apresentamos essas informações sobre o número das componentes principais e a variância acumulada utilizados, assim como a quantidade de *concepts* identificados em cada modelo de teste. Por fim, informações sobre os pacientes e os dados utilizados nos experimentos desenvolvidos também são disponibilizados.

Sobre a análise estatística impetrada na avaliação dos períodos antecipatórios propriamente ditos, o *threshold* para fins de rejeição da hipótese nula foi 0.05 em termos de *p-value* (nos gráficos representado por uma linha vermelha pontilhada). Ademais, a base de comparação utilizada no 2S-KS Teste foi o período *earlyictal* completo e ele foi comparado pelo teste a todos os seus prefixos. A seguir, apresentamos os atributos de cada paciente, os resultados obtidos em termos de períodos antecipatórios das suas crises epiléticas analisadas e uma visão geral acerca dos *concepts* detetados.

5.2.1 Paciente N^o 3500

Atributo	Valor
Idade à época do EEG	34 anos
Número de eletrodos (EEG)	29
Data da coleta de dados (EEG)	16/11/2009
Duração da coleta de dados (EEG)	93 horas
Número de blocos (EEG)	61 blocos
Primeira crise epilética	12 anos
Sexo	Masculino
Número de crises analisadas	5
Tipos de crises analisadas	Focais complexas
Padrões de ondas	Alfas rítmicas
	Tetas rítmicas

Tabela 5.1: Atributos do paciente n^o 3500.

Em termos de modelos de experimentos referente aos dados do paciente n^o 3500, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	N ^o de componentes	N ^o de <i>concepts</i>
50%	58.19%	3	206
60%	64.15%	4	208
70%	73.28%	6	221
80%	82.14%	9	231
90%	90.61%	11	233

Tabela 5.2: Detecção de *concepts* do paciente n^o 3500.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 3500, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

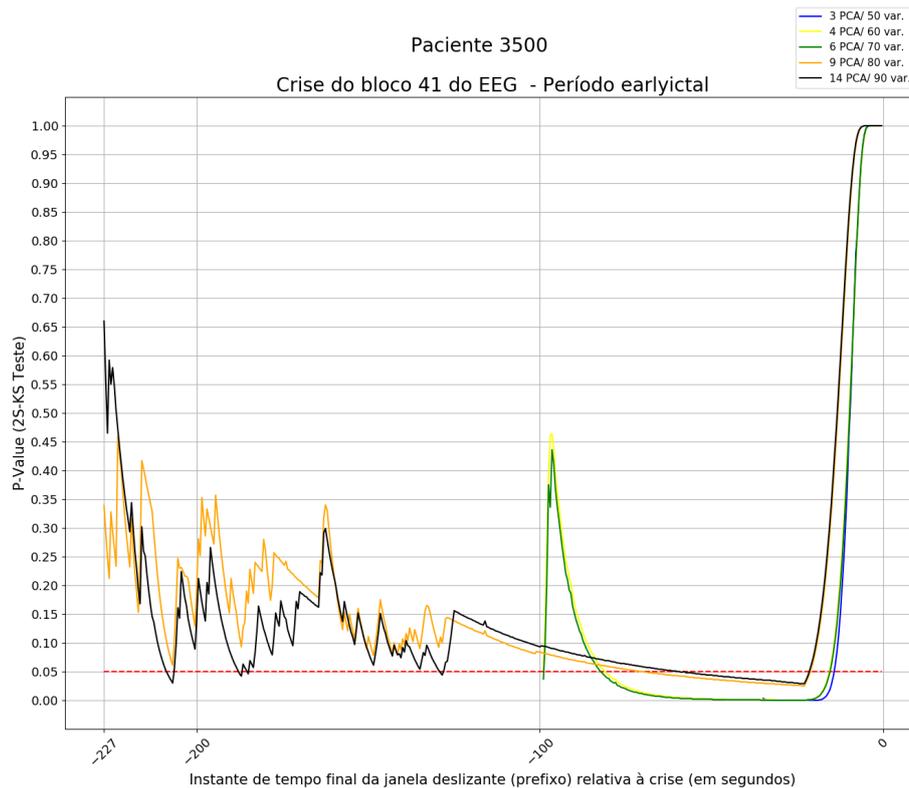


Figura 5.1: Paciente 3500 - Crise do bloco nº 41 do EEG

Os resultados obtidos no tocante aos períodos antecipatórios de crise apresentam três comportamentos distintos, de acordo com o modelo de teste (variância acumulada/PKA) observado. No modelo de teste com 50% de variância na PKA temos apenas um único período antecipatório de crise epilética, que possui uma curta duração, correspondendo ao intervalo -14.5 a -0.5 . Por outro lado, os modelos com 60% e 70% apresentam resultados muito parecidos, com 2 períodos antecipatórios cada um, sendo estes:

- modelo 60% de variância:
 1. -99 a -81.5
 2. -15.5 a -0.5
- modelo 70% de variância:
 1. -98.5 a -82.5
 2. -15.5 a -0.5

Por fim, temos os modelos com 80% e 90% de variância na PKA que indicam a existência de períodos antecipatórios até o instante de tempo -227 . O modelo com 80% apresenta apenas dois períodos antecipatórios: -227 a -70 & -21 a -0.5 .

Já o modelo com 90% apresenta um número maior de períodos antecipatórios por identificar alguns *outliers* entre o intervalo $[-227, -0.5]$, como pode ser observado a seguir:

- modelo 90% de variância:

1. -227 a -209
2. -206.5 a -188
3. -186.5 a -188.5
4. -184.5 a -129.5
5. -128 a -59.5
6. -58.5
7. -21.5 a -0.5

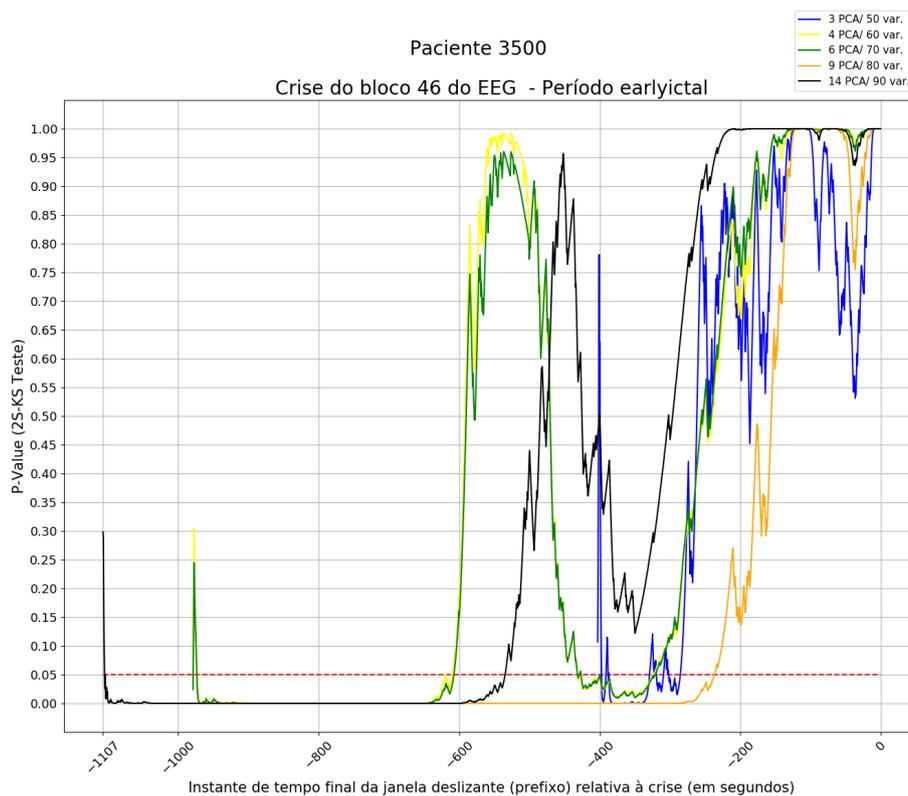


Figura 5.2: Paciente 3500 - Crise do bloco nº 46 do EEG

Os modelos de teste (variância acumulada/PCA) com 50%, 80% e 90% apresentam resultados bastante variados, tanto em número de períodos antecipatórios, quanto na localização destes ao longo do EEG, como demonstrado abaixo:

- modelo 50% de variância:

1. -403 a -398
2. -391 a -388.5
3. -387.5
4. -329 a -320
5. -309 a -302.5
6. -285 a -0.5

- modelo 80% de variância:

1. -237 a -0.5

- modelo 90% de variância:

1. -1107 a -1105
2. -534.5 a -0.5

Em relação aos modelos de teste com 60% e 70%, estes possuem algumas características em comum. Ambos indicam a existência de 3 períodos antecipatórios de crises, sendo estas bem similares, como podemos notar a seguir:

- modelo 60% de variância:

1. -99 a -81.5
2. -15.5 a -0.5

- modelo 70% de variância:

1. -98.5 a -82.5
2. -15.5 a -0.5

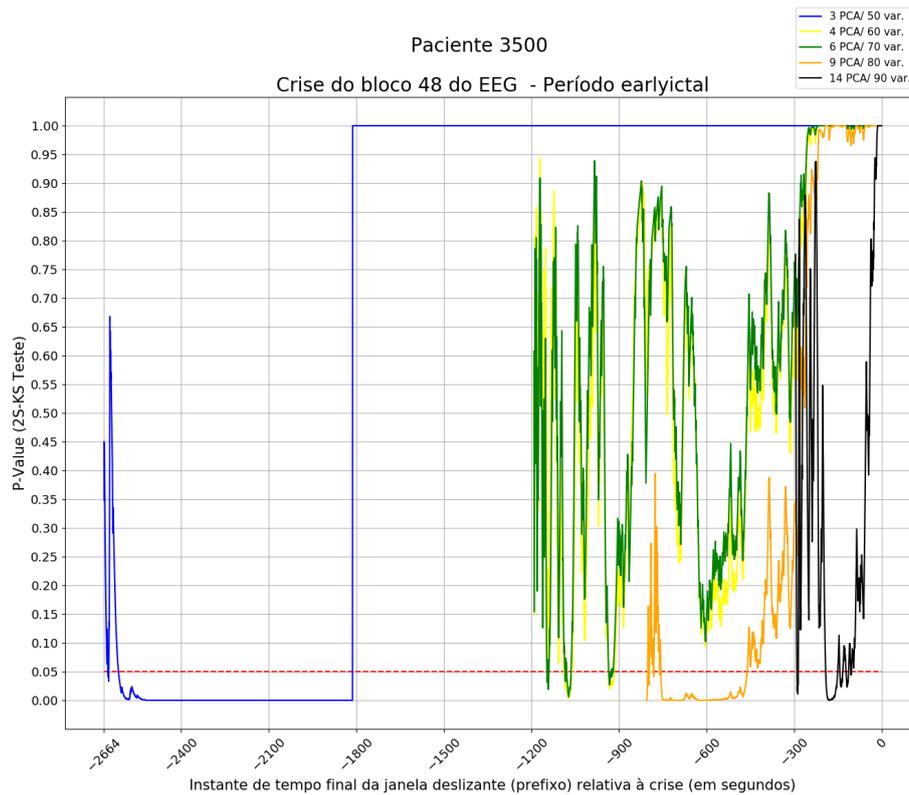


Figura 5.3: Paciente 3500 - Crise do bloco nº 48 do EEG

Nessa crise, a extensão do intervalo de tempo no qual são encontrados períodos antecipatórios de crise varia de maneira significativa a depender do modelo de crise utilizado. Porém, destaca-se o fato de que os modelos com 60% e 50% de variância acumulada na PCA apresentarem similaridades (o período antecipatório mais distante da crise inicia a 1191 segundos da crise; a maior parte dos demais períodos tem início e final semelhantes).

O modelo com 50% de variância identifica períodos antecipatórios mais distantes do início da crise, enquanto que os modelos com 80% e, principalmente, com 90% apresentam períodos antecipatórios mais próximos. Vale ressaltar que esta crise ocorreu próxima à crise anterior (bloco nº 46) e esse fato pode ter alguma correlação com a identificação de períodos antecipatórios com grande antecedência ao início do período ictal desta referida crise. Os resultados obtidos foram os seguintes:

- modelo 50% de variância:
 1. -2633.5 a -2650.5
 2. -2649.5 a -2648.5
 3. -2647 a -2614
 4. -1812 a -0.5

- modelo 60% de variância:

1. -1191 a -1089
2. -1088
3. -1087
4. -1060.5 a -936
5. -916 a -0.5

- modelo 70% de variância:

1. -1191 a -1147
2. -1144.5 a -1143.5
3. -1139.5 a -1085.5
4. -1082.5 a -1079
5. -1063.5 a -935
6. -925 a -924.5
7. -922
8. -921 a -920.5
9. -919.5 a -0.5

- modelo 80% de variância:

1. -801.5 a -798.5
2. -797.5
3. -795.5 a -782.5
4. -781.5
5. -780.5 a -761
6. -760 a -756.5
7. -460.5 a -0.5

- modelo 90% de variância:

1. -295 a -290.5
2. -285.5 a -192
3. -150.5 a -143.5
4. -132.5 a -122.5

5. -113 a -106
6. -104.5
7. -103.5 a -0.5

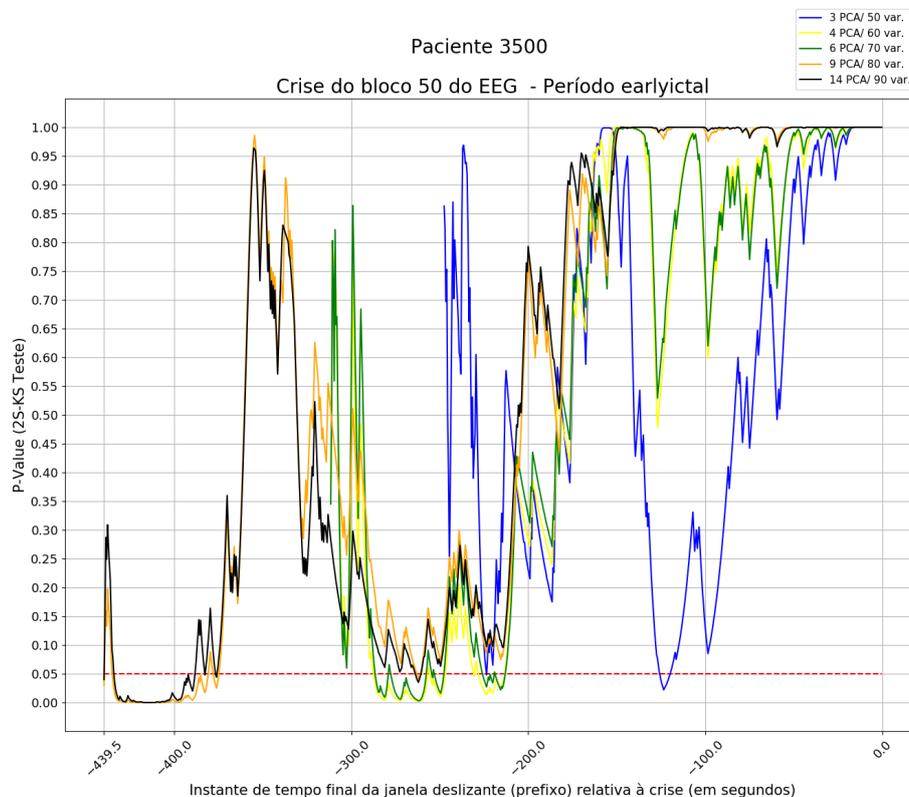


Figura 5.4: Paciente 3500 - Crise do bloco n^o 50 do EEG

Em relação a esta crise, percebe-se um comportamento mais linear de todos os modelos de teste, nos quais os n períodos antecipatórios podem ser vistos como um único intervalo contendo alguns *outliers*, que seriam os pontos faltantes entre um período e outro. Em suma, as distâncias temporais entre os períodos antecipatórios correspondem a um curto espaço de tempo.

Ademais, há similaridades entre os resultados dos modelos com 60% e 70% de variância acumulada na PCA, assim como entre os resultados dos modelos com 80% e 90% de variância. Além disso, desconsiderando os pequenos “gaps”, nota-se que os períodos antecipatórios do modelo com 50% de variância está contido nos resultados obtidos com os modelos com 60% e 70%. De maneira análoga, o mesmo ocorre com os períodos antecipatórios dos modelos com 60% e 70% e os resultados dos modelos com 80% e 90% de variância acumulada na PCA. A seguir, os períodos antecipatórios do modelo de teste com 50% de variância:

- modelo 50% de variância:

1. -247.5 a -224
2. -223 a -126
3. -119.5 a -0.5

Em seguida, apresentamos os períodos antecipatórios obtidos com os modelos com 60% e 70% de variância acumulada (PCA). Além da similaridade entre ambos, percebe-se que o intervalo abrangido pelo modelo com 50%, $[-247.5, -0.5]$ está de fato contido nos resultados dos modelos com 60% e 70%:

- modelo 60% de variância:

1. -311.5 a -287.5
2. -256.5 a -256
3. -246.5 a -231
4. -230 a -228
5. -213 a -0.5

- modelo 70% de variância:

1. -311.5 a -286.5
2. -279 a -278
3. -257.5 a -254.5
4. -253.5
5. -247.5 a -226
6. -219
7. -212.5 a -0.5

Por fim, os períodos antecipatórios obtidos com os modelos com 80% e 90% de variância acumulada (PCA), cujos períodos antecipatórios abrangem o intervalo obtido pelos modelos com 60% e 70%, $[-311.5, -0.5]$. Nota-se também a semelhança entre os resultados dos modelos com 80% e 90%, cujos períodos antecipatórios estão listados a seguir:

- modelo 80% de variância:

1. -439 a -435
2. -380.5 a -378

3. -374.5 a -263

4. -260.5 a -0.5

• modelo 90% de variância:

1. -439 a -434

2. -388 a -383

3. -382 a -377

4. -375.5 a -264

5. -260.5 a -0.5

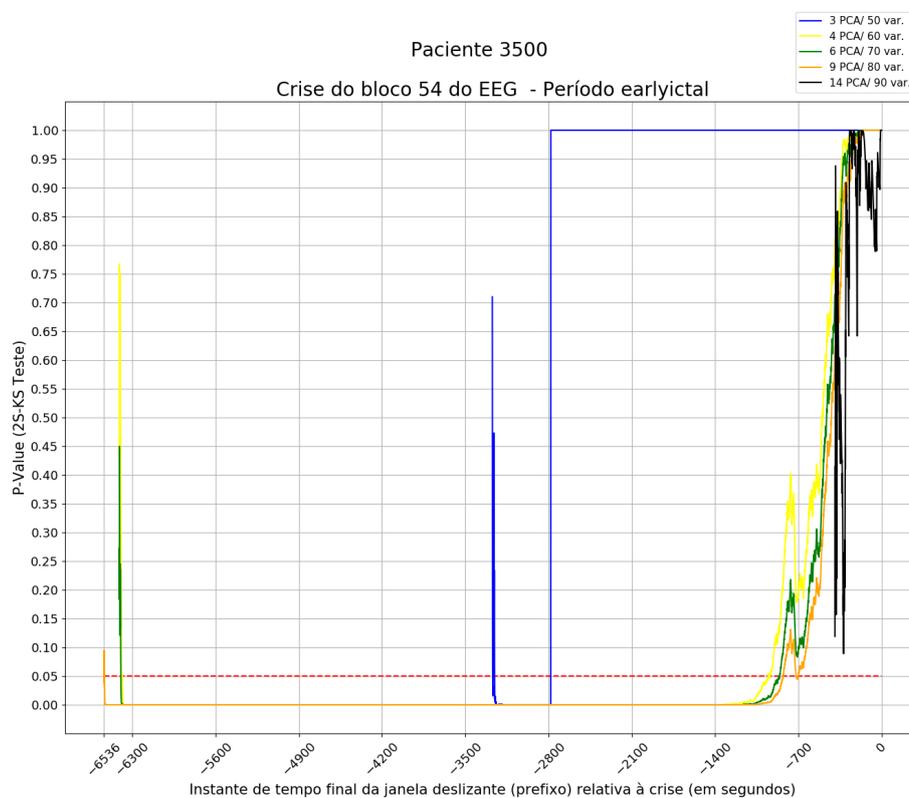


Figura 5.5: Paciente 3500 - Crise do bloco nº 54 do EEG

Nessa crise, os modelos de testes apresentaram resultados com quase nenhuma correlação entre si, seja na identificação dos períodos antecipatórios ou no intervalo de tempo nos quais tais períodos estão contidos. Exceto pelo fato de que os modelos com variância de 60%, 70% e 80% conseguiram aferir curtos períodos antecipatórios bem distantes do início da crise e dos períodos antecipatórios anteriores a estes.

Ainda cabe destacar que diferente do que ocorreu com as crises anteriores, nesta o modelo com 90% de variância identificou apenas um período antecipatório, sendo este de curta duração e bem próximo ao início desta crise propriamente dita. A

seguir, temos os períodos antecipatórios de crise epilética obtidos com todos os modelos de teste:

- modelo 50% de variância:

1. -3272.5 a -3269
2. -3266 a -3254
3. -2780.5 a -0.5

- modelo 60% de variância:

1. -6408 a -6391
2. -6384.5
3. -954.5 a -947.5
4. -941
5. -938 a -0.5

- modelo 70% de variância:

1. -6408.5 a -6395.5
2. -863 a -862.5
3. -857 a -0.5

- modelo 80% de variância:

1. -6535 a -6534.5
2. -6533
3. -828 a -719
4. -711.5
5. -701 a -0.5

- modelo 90% de variância:

1. -392.5 a -0.5

Análise

Os modelos de teste com menor e maior variância (50% e 90%) ora caracterizaram períodos antecipatórios em geral de curta duração, ora períodos de longa duração, variando de comportamento de acordo com as crises analisadas deste paciente. Tal

comportamento foi parcialmente observado também no modelo com 80% de variância. Os modelos com 60% e 70% de variância foram mais constantes em termos de caracterização dos períodos por eles identificados, apresentando períodos antecipatórios que geralmente ficavam em um meio termo: não eram os mais longos, nem eram os mais curtos no tocante à duração/localização temporal nos sinais analisados. Se fosse necessário optar por algum modelo de teste para caracterização de períodos antecipatórios de crises futuras, os melhores candidatos seriam os modelos com 60% e 70% de variância acumulada (PCA).

5.2.2 Paciente N^o 4000

Atributo	Valor
Idade à época do EEG	55 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	04/03/2010
Duração da coleta de dados (EEG)	67 horas
Número de blocos (EEG)	41 blocos
Primeira crise epilética	35 anos
Sexo	Masculino
Número de crises analisadas	6
Tipos de crises analisadas	Focais complexas
Padrões de ondas	Tetas rítmicas

Tabela 5.3: Atributos do paciente n^o 4000.

Em termos de modelos de experimentos referente aos dados do paciente n^o 4000, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	N ^o de componentes	N ^o de <i>concepts</i>
50%	70.18%	2	71
60%	70.18%	2	71
70%	70.18%	2	71
80%	81.15%	6	99
90%	90.64%	12	108

Tabela 5.4: Detecção de *concepts* do paciente n^o 4000.

Nesse paciente os modelos de teste com limiar de 50%, 60% e 70% em termos de variância acumulada (PCA) atingiram esses valores com o mesmo número de componentes principais (2). Portanto, esses modelos citados apresentam sempre os mesmos resultados em cada uma das crises epiléticas analisadas deste paciente. Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente nº 4000, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

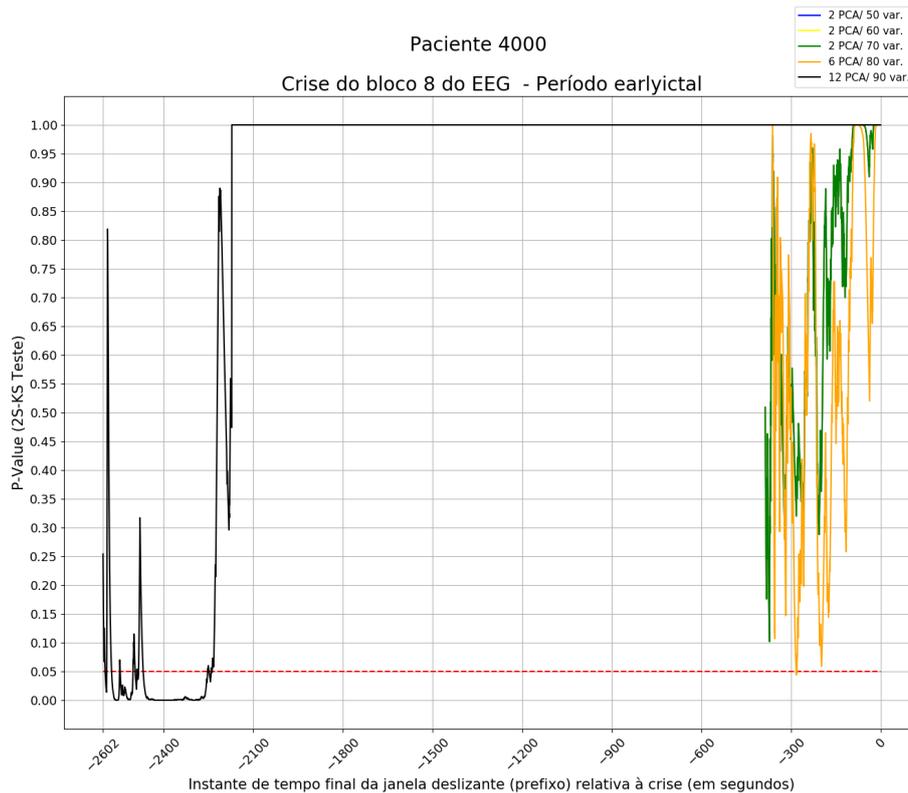


Figura 5.6: Paciente 4000 - Crise do bloco nº 8 do EEG

Os modelos de teste com variância de 50%, 60% e 70% identificaram nessa crise um único período antecipatório no intervalo entre -386.5 e -0.5 segundos. Já o modelo com 80% apresenta três períodos antecipatórios, porém como a distância temporal entre eles é ínfima, pode-se considerar que o período antecipatório de fato corresponde ao intervalo de tempo entre -362.5 e -0.5 segundos, relativamente similar ao período antecipatório obtido com os modelos 50%, 60% e 70%.

Diferente dos modelos anteriores, o modelo de teste com 90% de variância identificou períodos antecipatórios que cobriram um intervalo de tempo maior, até o instante de tempo de -2602.5 segundos anteriores ao início da crise. A relação de períodos antecipatórios dos modelos com 80% e 90% são expressas a seguir:

- modelo 80% de variância:

1. -362.5 a -283
2. -281.5 a -281
3. -280 a -0.5

- modelo 90% de variância:

1. -2602.5 a -2574
2. -2547 a -2545.5
3. -2502 a -2495
4. 2488.5
5. -2483.5 a -2468.5
6. -2252 a -2248
7. -2240.5 a -2239
8. -2237.5 a -0.5

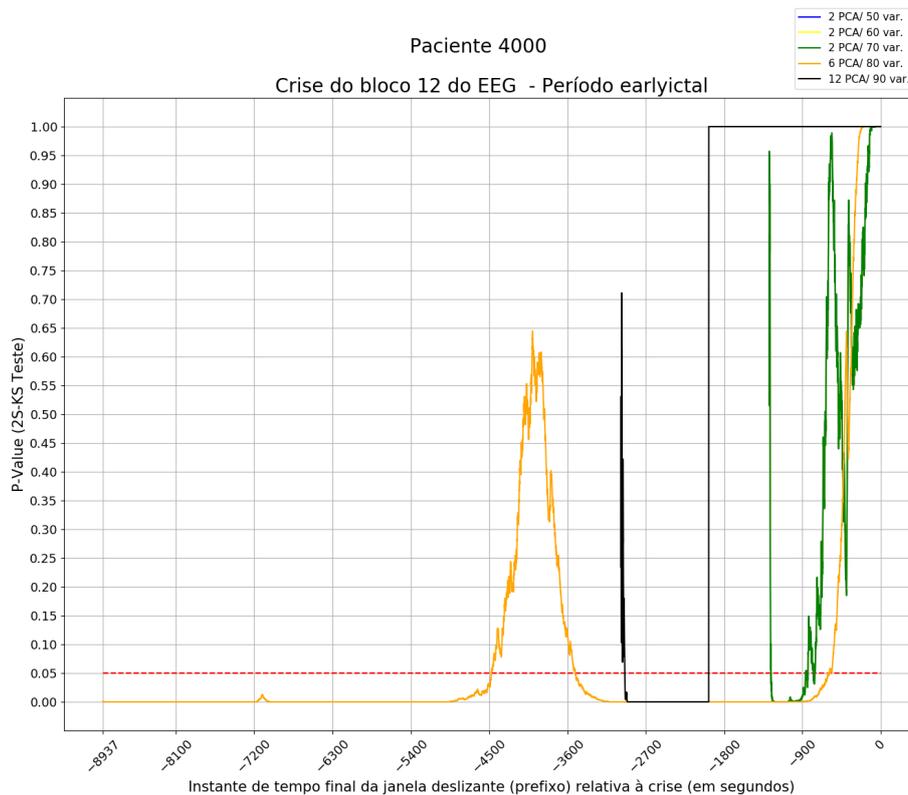


Figura 5.7: Paciente 4000 - Crise do bloco nº 12 do EEG

Nessa crise, os modelos de teste com variância de 50%, 60% e 70% indicaram períodos antecipatórios até o instante de tempo de -1281.5 segundos, enquanto que

os demais modelos identificaram períodos antecipatórios ainda mais distantes. Entre os modelos com 80% e 90% não se percebe alguma correlação entre os resultados.

Embora o modelo com 80% tenha usado 6 componentes principais e o modelo com 90% tenha usado o dobro disso (12), foi o primeiro que apresentou um maior número de períodos antecipatórios e cobriu um intervalo maior de tempo considerando todos os períodos identificados. Os resultados obtidos com os modelos nessa crise são detalhados a seguir:

- modelos com 50%, 60% e 70% de variância:

1. -1281 a -1260.5
2. -1258.5
3. -861 a -860.5
4. -859.5 a -858.5
5. -850.5 a -849.5
6. -848.5 a -848
7. -835 a -774
8. -756 a -0.5

- modelo 80% de variância:

1. -4474 a -3508.5
2. -603 a -601.5
3. -591.5 a -0.5

- modelo 90% de variância:

1. -2985 a -2940
2. -1975 a -0.5

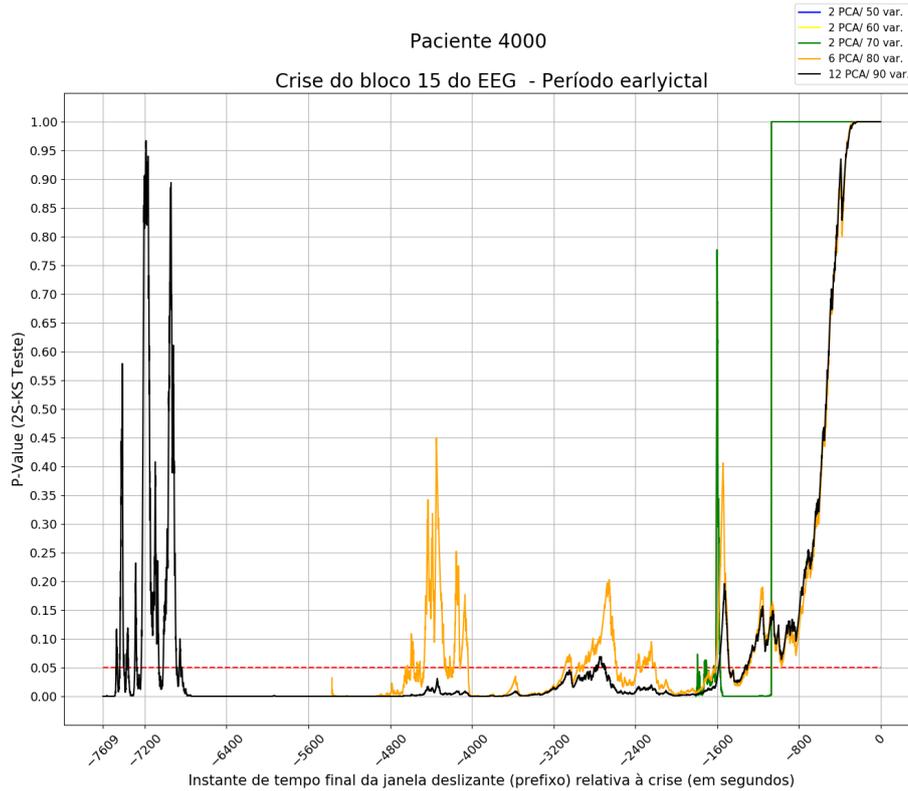


Figura 5.8: Paciente 4000 - Crise do bloco n^o 15 do EEG

Os modelos de teste com variância de 50%, 60% e 70% indicaram sete períodos antecipatórios até o instante de tempo de -1792.5 segundos, que correspondem a quatro intervalos de tempos: $[-1069, -0.5]$, $[-1632, -1570]$, $[-1722, -1703]$ e $[-1792.5, -1791]$. Por outro lado, os modelos com 80% e 90% apresentaram comportamentos diferentes dos demais, com um grande número de períodos antecipatórios. A seguir, os resultados obtidos com os modelos com variância de 50%, 60% e 70% são demonstrados:

- modelos com 50%, 60% e 70% de variância:

1. -1792.5 a -1791
2. -1722 a -1720
3. -1718 a -1715.5
4. -1706 a -1703
5. -1632 a -1630.5
6. -1611.5 a -1570
7. -1069 a -0.5

Em relação aos modelos com 80% e 90% apresentam 31 e 21 períodos antecipatórios, respectivamente. Em ambos os modelos, muitos dos períodos aferidos

são separados por curtos intervalos de tempo. Assim, pode-se considerar que esses períodos formam grupos de períodos antecipatórios, como listados a seguir:

- modelo 80% de variância:

1. $[-4650, -4567]$
2. $[-4536.5, -4531]$
3. $[-4512, -4506]$
4. $[-4468.5, -4249]$
5. $[-4218, -4206]$
6. $[-4188.5, -4036]$
7. $[-3090, -3026.5]$
8. $[-2964, -2583.5]$
9. $[-2387.5, -2203]$
10. $[-1616.5, -1481]$
11. $[-1294.5, -0.5]$

- modelo 90% de variância:

1. $[-7480.5, -7469]$
2. $[-7445, -7404]$
3. $[-7380.5, -7360]$
4. $[-7299, -7278]$
5. $[-7235, -7062.5]$
6. $[-7020.5, -6840.5]$
7. $[-2795, -2695]$
8. $[-1584.5, -1482]$
9. $[-1327, -0.5]$

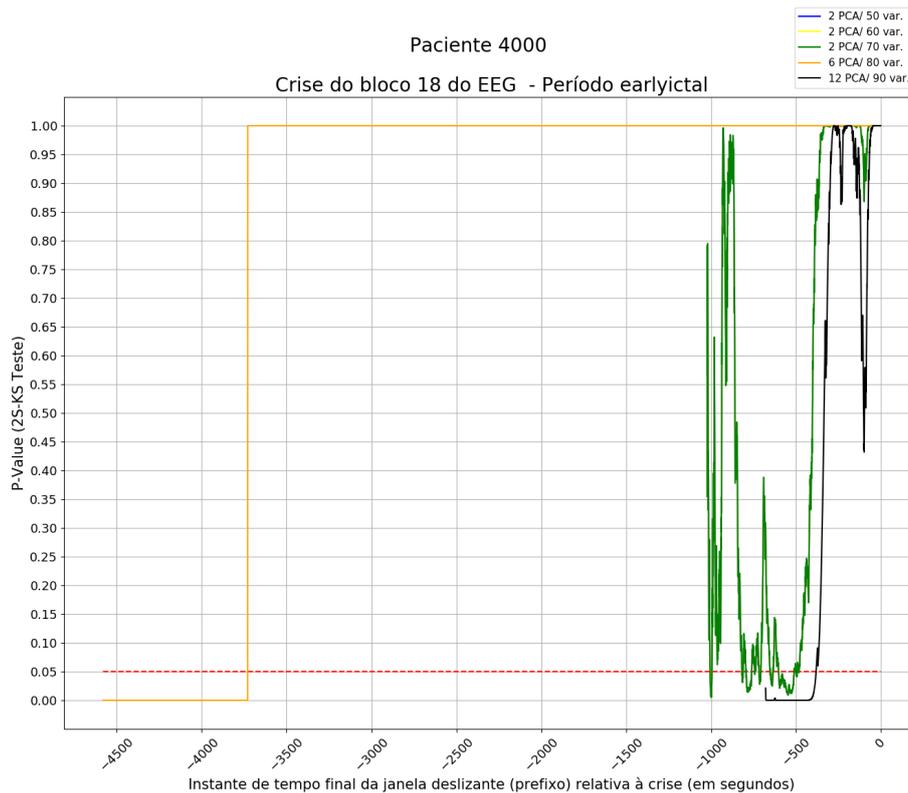


Figura 5.9: Paciente 4000 - Crise do bloco n^o 18 do EEG

Nessa crise, os modelos de teste com variância de 50%, 60% e 70% indicaram quinze períodos antecipatórios que correspondem a três intervalos de tempos. A seguir, os grupos de períodos antecipatórios obtidos com os modelos com variância de 50%, 60% e 70%:

- modelos com 50%, 60% e 70% de variância:

1. $[-1021.5, -793.5]$
2. $[-759, -651]$
3. $[-634, -601.5]$
4. $[-510, -0.5]$

Já os modelos com 80% e 90% apresentaram apenas um único período antecipatório cada um, com valores bem distintos, como pode ser visto abaixo:

- modelo 80% de variância:

1. -3729 a -0.5

- modelo 90% de variância:

1. -380 a -0.5

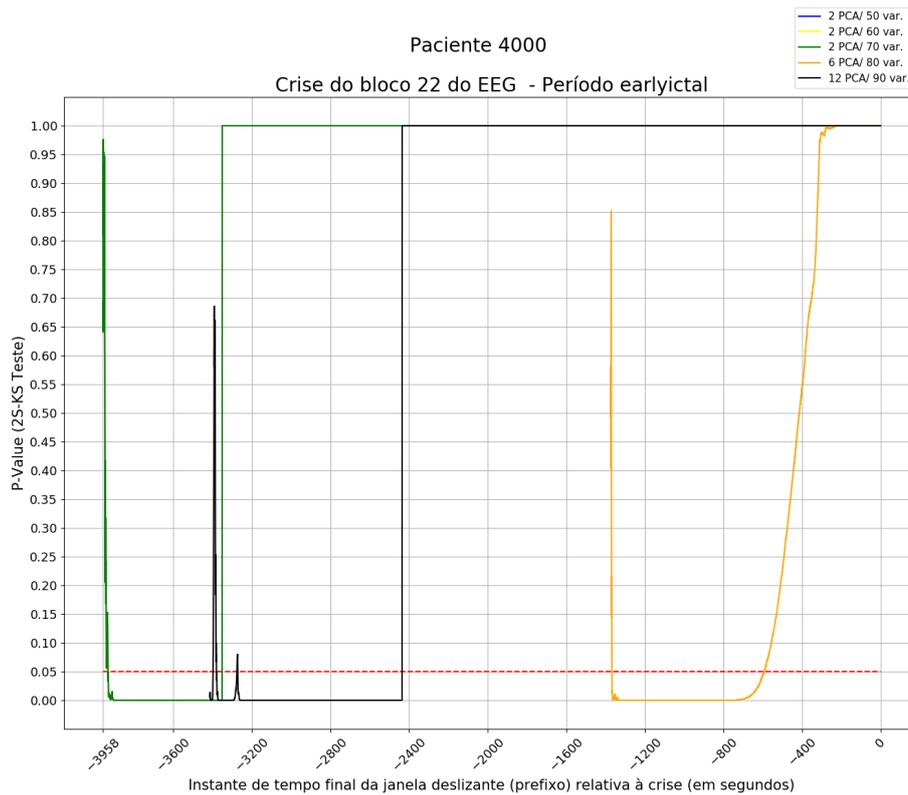


Figura 5.10: Paciente 4000 - Crise do bloco nº 22 do EEG

Os modelos de teste com variância de 50%, 60% e 70% apresentam apenas dois períodos antecipatórios, sendo eles o período de -3958 a -3933.5 segundos e o período de -3351.5 a -0.5 segundos. Coincidentemente, o modelo com 80% de variância, este também identificou apenas dois períodos antecipatórios, a saber: -1374 a -1368 & -593 a -0.5 .

Por outro lado, o modelo de teste com 90% de variância apresenta em seu resultado quatro períodos antecipatórios, sendo que três destes períodos mostram-se separados por curtos intervalos de tempo, como listado a seguir:

- modelo 90% de variância:
 1. -3398 a -3381
 2. -3379.5
 3. -3276.5 a -3273
 4. -2435.5 a -0.5

Análise

Os modelos de teste com 80% e 90% de variância apresentaram períodos antecipatórios em algumas crises mais curtos, em outras mais longos se comparados aos

períodos caracterizados pelos demais modelos, na maioria das crises. Já os modelos com 50%, 60% e 70%, exceto na última crise analisada, apresentaram períodos antecipatórios de duração mais curta. No caso desse paciente, na PCA, somente com as duas primeiras componentes principais já foi obtido uma variância acumulada que atendeu aos limiares de 50%, 60% e 70%, o que fez com que os modelos de testes relativos a esses limiares se tornassem um mesmo modelo, fato este observável nos gráficos dos resultados. Em caso de crises futuras, um desses modelos seria uma opção confiável para ser usado como modelo de teste.

5.2.3 Paciente N^o 4200

Atributo	Valor
Idade à época do EEG	34 anos
Número de eletrodos (EEG)	29
Data da coleta de dados (EEG)	31/10/2009
Duração da coleta de dados (EEG)	115 horas
Número de blocos (EEG)	68 blocos
Primeira crise epilética	14 anos
Sexo	Feminino
Número de crises analisadas	3
Tipos de crises analisadas	Focais complexas
	Não classificadas
Padrões de ondas	Tetas rítmicas
	Picos repetitivos

Tabela 5.5: Atributos do paciente n^o 4200.

Em termos de modelos de experimentos referente aos dados do paciente n^o 4200, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	Nº de componentes	Nº de <i>concepts</i>
50%	54.32%	3	187
60%	62.36%	4	195
70%	72.99%	6	200
80%	80.35%	9	202
90%	90.79%	16	205

Tabela 5.6: Detecção de *concepts* do paciente n^o 4200.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 4200, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

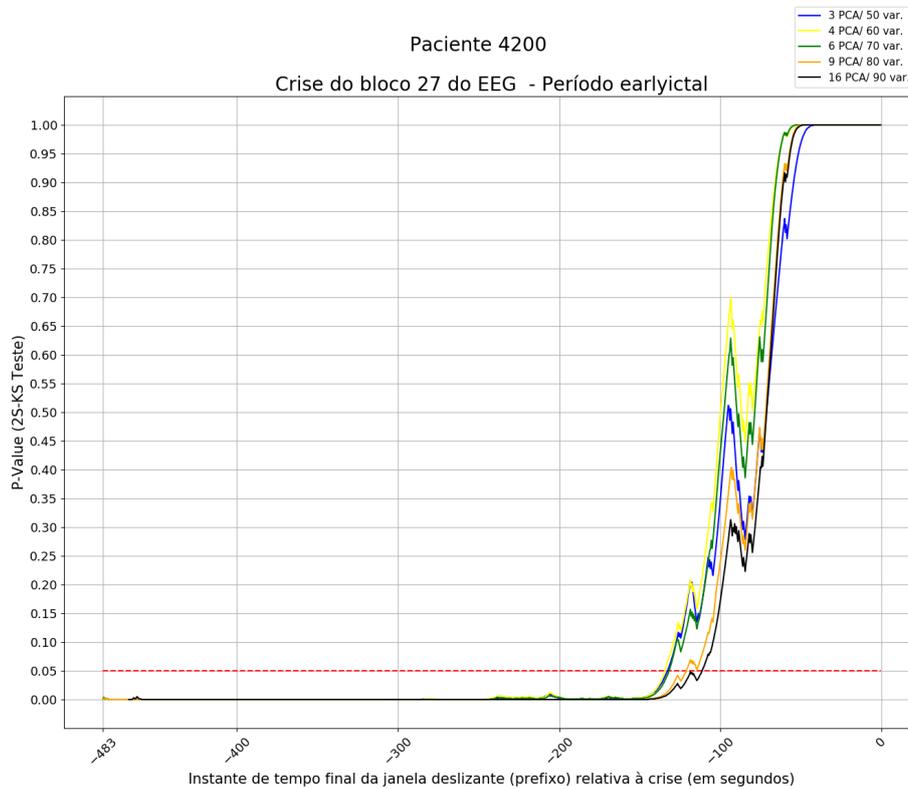


Figura 5.11: Paciente 4200 - Crise do bloco n^o 27 do EEG

Nessa crise, todos os cinco modelos de testes (50%, 60%, 70%, 80% e 90% de variância acumulada na PCA) apresentaram resultados muito semelhantes, pois em cada um deles houve a identificação de um único período antecipatório, com valores de início e término praticamente iguais. A seguir, listamos os resultados obtidos com todos os modelos de teste:

- modelo 50% de variância:
 1. -132.5 a -0.5
- modelo 60% de variância:
 1. -134 a -0.5
- modelo 70% de variância:
 1. -131.5 a -0.5
- modelo 80% de variância:
 1. -121 a -0.5
- modelo 90% de variância:
 1. -111 a -0.5

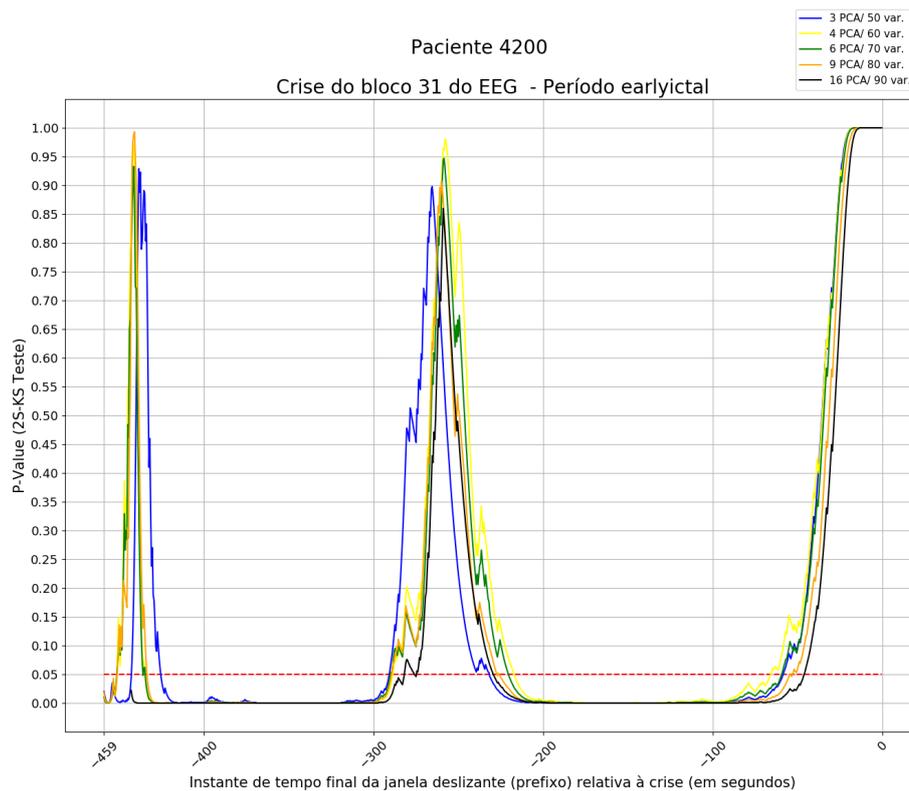


Figura 5.12: Paciente 4200 - Crise do bloco nº 31 do EEG

De maneira semelhante ao ocorrido na crise anterior (crise do bloco nº 27), os modelos de teste apresentaram resultados similares na avaliação dos períodos antecipatórios da crise do bloco nº 31, principalmente os quatro primeiros modelos

(50%, 60%, 70% e 80% de variância). Os resultados propriamente ditos foram os seguintes:

- modelo 50% de variância:

1. -442.5 a -426
2. -290.5 a -232
3. -58.5 a -0.5

- modelo 60% de variância:

1. -451.5 a -435
2. -288.5 a -217.5
3. -65.5 a -0.5

- modelo 70% de variância:

1. -451 a -436.5
2. -435.5 a -435
3. -290 a -220.5
4. -59 a -0.5

- modelo 80% de variância:

1. -451 a -434
2. -289 a -226
3. -54.5
4. -53 a -0.5

- modelo 90% de variância:

1. -281.5 a -276
2. -274.5 a -229.5
3. -46 a -0.5

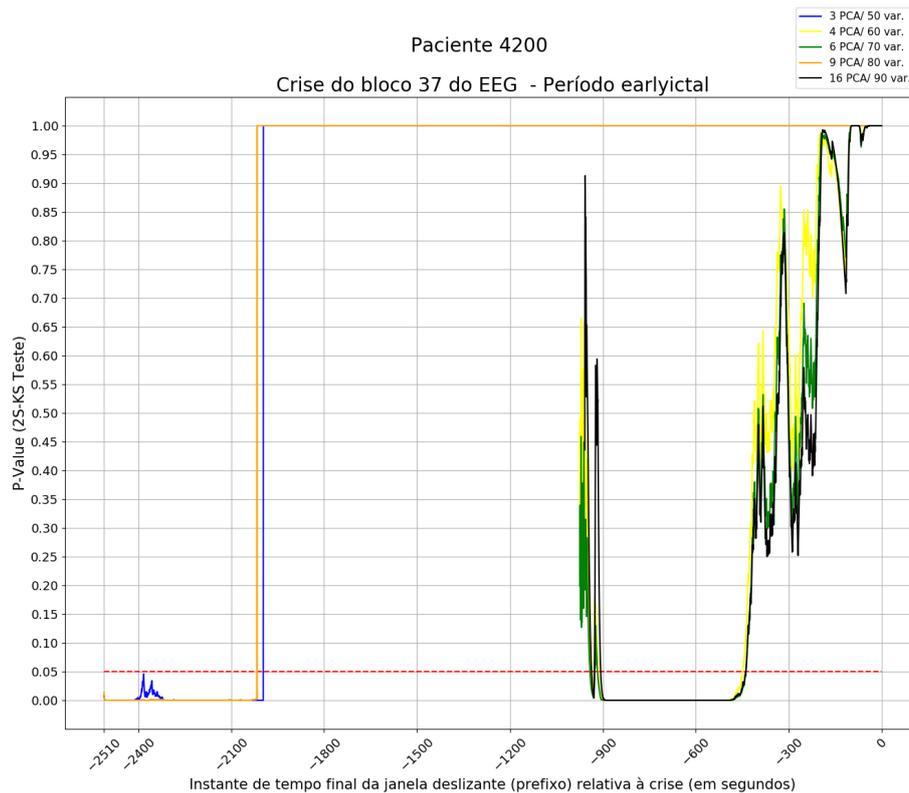


Figura 5.13: Paciente 4200 - Crise do bloco nº 37 do EEG

Nessa crise, pode-se separar os modelos de teste em dois grupos, no tocante à similaridade dos resultados: um composto pelos modelos com 50% e 80% de variância; e o outro composto pelos modelos com 60%, 70% e 90% de variância. O primeiro grupo (50% e 80%) cobriu um intervalo consideravelmente maior do que o segundo grupo (60%, 70% e 80%), como pode-se notar a seguir:

- modelo 50% de variância:
 1. -1996 a -0.5
- modelo 60% de variância:
 1. -974.5 a -937.5
 2. -928 a -915.5
 3. -449 a -0.5
- modelo 70% de variância:
 1. -974.5 a -942
 2. -926.5 a -917.5
 3. -438.5 a -0.5

- modelo 80% de variância:
 1. -2017
 2. -2016 a -0.5
- modelo 90% de variância:
 1. -958.5 a -937
 2. -927 a -907.5
 3. -438 a -0.5

Análise

Nos experimentos realizados, notou-se que os modelos de teste não apresentaram diferenças significativas nas crises analisadas, principalmente nas duas primeiras crises, exceto o modelo com 90% de variância acumulada. Desse modo, caso fosse necessário indicar um modelo para ser usado na caracterização de futuras crises, à exceção do modelo com 90% de variância, qualquer um dos modelos de teste seria uma escolha adequada.

5.2.4 Paciente N^o 4400

Atributo	Valor
Idade à época do EEG	25 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	27/03/2010
Duração da coleta de dados (EEG)	107 horas
Número de blocos (EEG)	69 blocos
Primeira crise epilética	16 anos
Sexo	Masculino
Número de crises analisadas	5
Tipos de crises analisadas	Focais complexas
	Ondas agudas
Padrões de ondas	Tetas rítmicas
	Deltas rítmicas

Tabela 5.7: Atributos do paciente n^o 4400.

Em termos de modelos de experimentos referente aos dados do paciente n^o 4400, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	N ^o de componentes	N ^o de <i>concepts</i>
50%	51.07%	2	175
60%	63.11%	3	196
70%	74.82%	5	214
80%	82.83%	7	225
90%	90.70%	11	221

Tabela 5.8: Detecção de *concepts* do paciente n^o 4400.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 4400, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

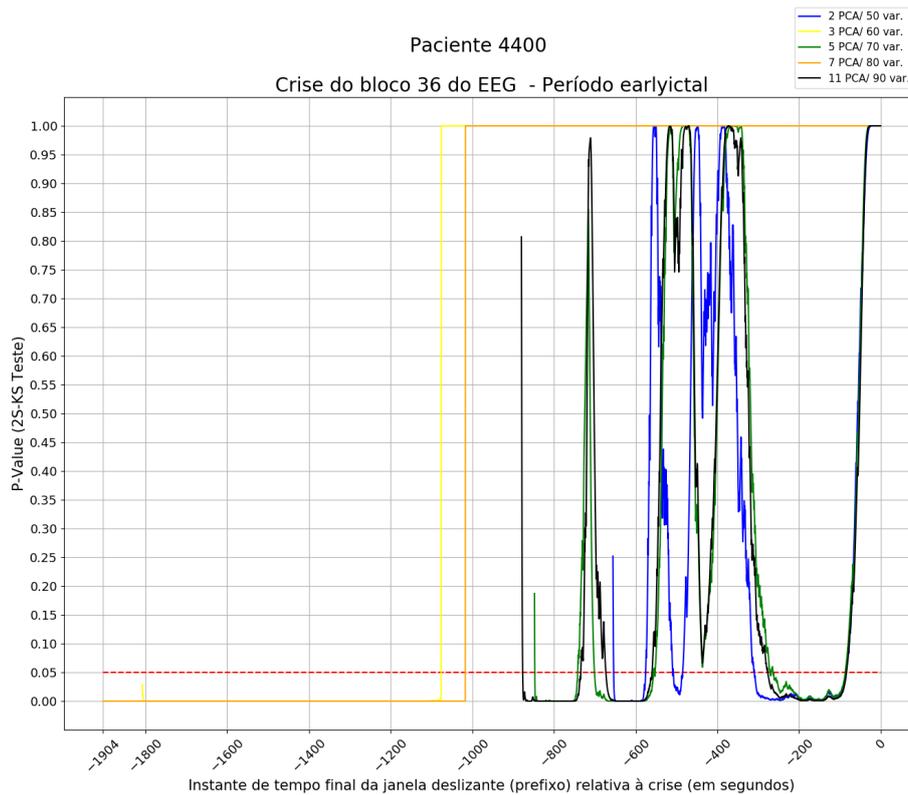


Figura 5.14: Paciente 4400 - Crise do bloco n^o 36 do EEG

Nessa crise, os modelos de testes com 60% e 80% identificaram um único período antecipatório, sendo que em ambos a duração dos períodos aferidos são semelhantes. Por sua vez, os modelos com 70% e 90% de variância apresentaram resultados semelhantes, enquanto que o modelo com 50% de variância foi o único a apresentar

resultados sem nenhuma correlação com os resultados de nenhum dos outros modelos. A seguir, os períodos antecipatórios de crise obtidos com os cinco modelos de teste:

- modelo 50% de variância:

1. -655.5 a -654
2. -575 a -509.5
3. -508.5 a -508
4. -484.5 a -309.5
5. -85 a -0.5

- modelo 60% de variância:

1. -1075.5 a -0.5

- modelo 70% de variância:

1. -847.5 a -847
2. -741.5 a -701
3. -556 a -554
4. -552 a -271
5. -267.5 a -264
6. -86.5 a -0.5

- modelo 80% de variância:

1. -1016.5 a -0.5

- modelo 90% de variância:

1. -879.5 a -876.5
2. -734 a -673
3. -559.5 a -276.5
4. -83.5 a -0.5

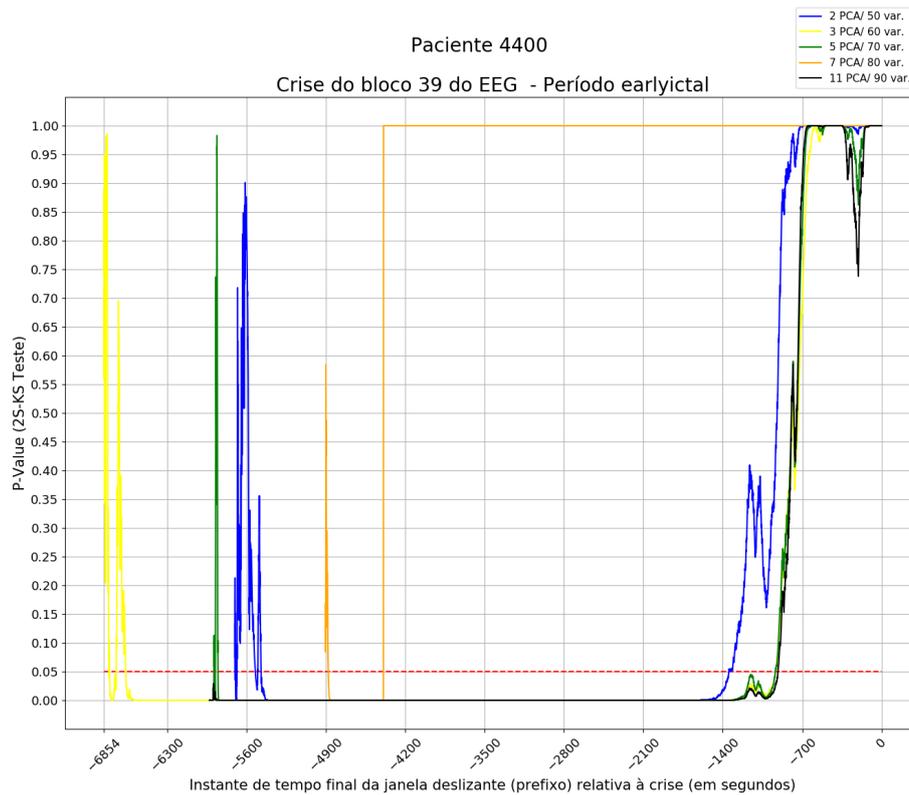


Figura 5.15: Paciente 4400 - Crise do bloco nº 39 do EEG

Dentre os modelos de testes, apenas o modelo com maior número de componentes principais na PCA e por conseguinte, o modelo com maior variância acumulada, 90% foi o único que identificou apenas um único período antecipatório, cuja longitude ficou bem abaixo do que foi aferido pelos demais modelos, como pode-se observar a seguir:

- modelo 50% de variância:
 1. -5702.5 a -5698
 2. -5697 a -5696
 3. -5682.5 a -5521
 4. -5499.5 a -5470.5
 5. -1347 a -0.5
- modelo 60% de variância:
 1. -6854.5 a -6810.5
 2. -6809 a -6806
 3. -6747 a -6663

4. -936 a -0.5
- modelo 70% de variância:
 1. -5886 a -5881.5
 2. -5879.5
 3. -5873.5 a -5850.5
 4. -936.5 a -0.5
 - modelo 80% de variância:
 1. -4902.5 a -4877.5
 2. -4391 a -0.5
 - modelo 90% de variância:
 1. -913.5 a -0.5

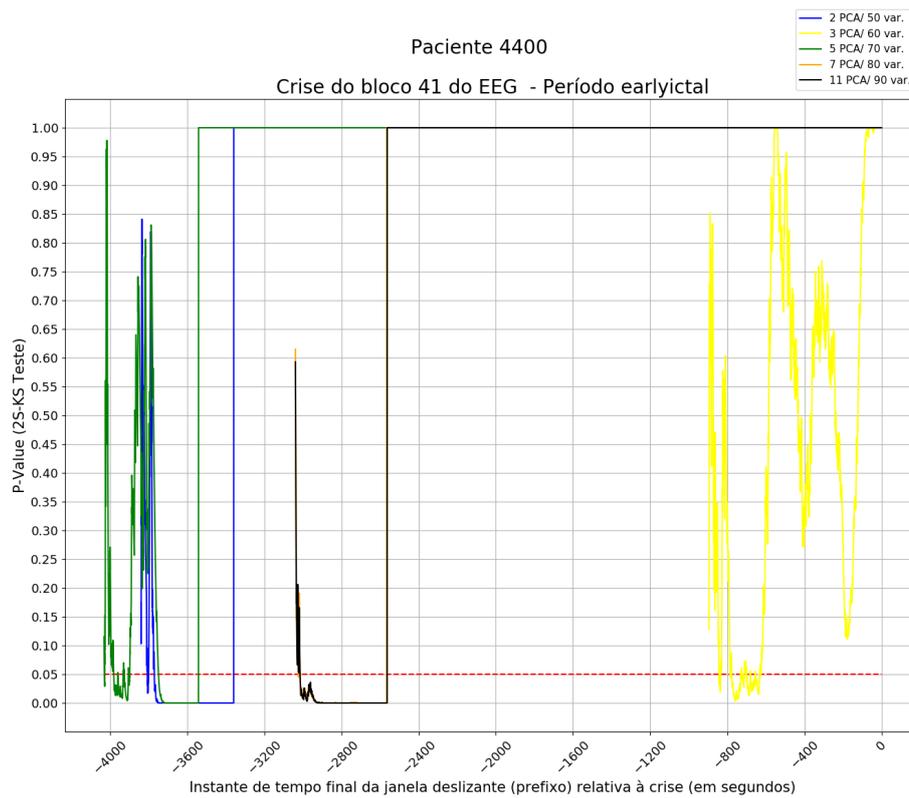


Figura 5.16: Paciente 4400 - Crise do bloco nº 41 do EEG

Nessa crise, à exceção do modelo de teste com 60% de variância acumulada, todos os modelos tiveram um comportamento semelhante, caracterizado pela existência de um período antecipatório longo, imediatamente anterior ao início da crise, no qual os

prefixos avaliados no teste estatístico 2S-KS Test resultaram em valores de p-value igual a 1.0, ou seja, em que a confiança para rejeitar a hipótese de que tratavam de sinais oriundos de uma mesma distribuição foi zero. Por outro lado, os períodos antecipatórios identificados por meio do modelo com 60% de variância cobriram um intervalo de tempo bem menor se comparado ao intervalo coberto pelos demais modelos. Os resultados obtidos com os modelos de teste, exceto o modelo com 60% de variância, estão listados a seguir.

- modelo 50% de variância:

1. -3839 a -3807
2. -3799.5 a -3770.5
3. -3358.5 a -0.5

- modelo 70% de variância:

1. -4031 a -4029.5
2. -4027 a -3981
3. -3959.5
4. -3932
5. -3930.5 a -3926.5
6. -3904.5 a -3903.5
7. -3897.5 a -3749
8. -3541 a -0.5

- modelo 80% de variância:

1. -3039 a -3022.5
2. -3021.5 a -3015
3. -2565.5 a -0.5

- modelo 90% de variância:

1. -3039 a -3015
2. -2563 a -0.5

Em relação ao modelo de teste com 60% de variância, este apresentou um comportamento muito irregular, com um total de dezoito períodos antecipatórios, porém muitos destes compostos por apenas um ou dois instantes de tempo. Assim, esses períodos podem ser agrupados em quatro intervalos de tempo, como demonstrado abaixo:

- modelo 60% de variância:

1. $[-895, -777]$
2. $[-730.5, -680]$
3. $[-656, -654.5]$
4. $[-632, -0.5]$

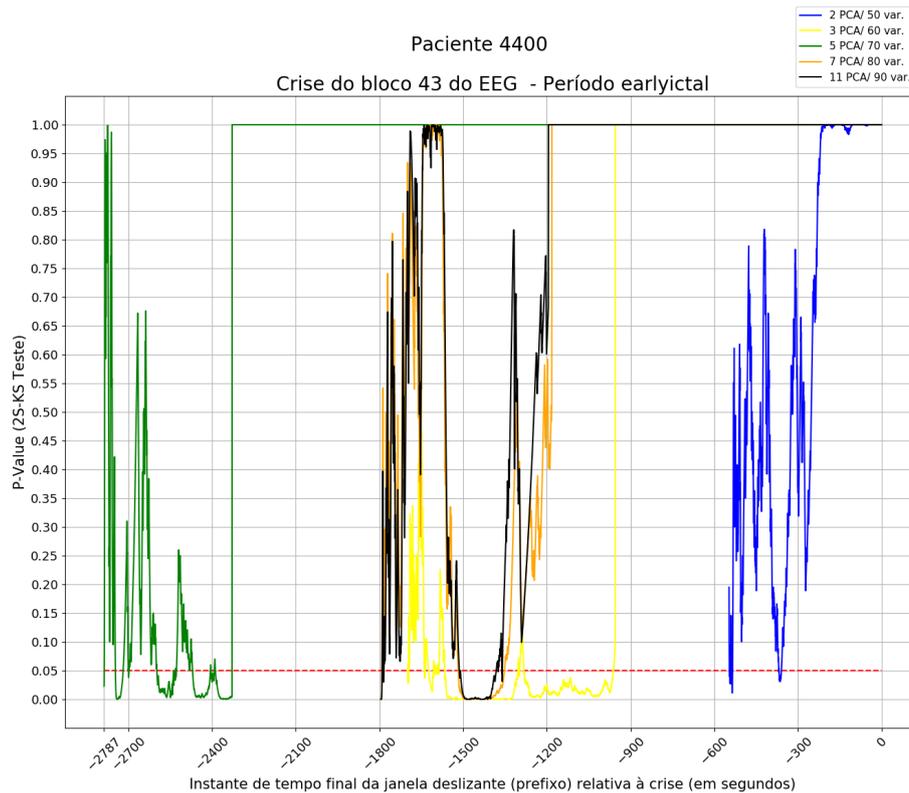


Figura 5.17: Paciente 4400 - Crise do bloco n^o 43 do EEG

Nessa crise, podemos notar que os modelos de teste com 60%, 80% e 90% obtiveram resultados parecidos e que os modelos com 50% e 70% apresentaram resultados distintos entre si e em relação aos modelos citados anteriormente. Cabe destacar o fato que os períodos antecipatórios mais distantes do início da crise foram obtidos através do modelo com 70% e que os períodos antecipatórios mais próximos da crise foram identificados pelo modelo com 50%. A seguir, os resultados mencionados são demonstrados:

- modelo 50% de variância:

1. -547 a -546
2. -545 a -544.5

3. -543
4. -541.5
5. -540.5 a -539
6. -537.5
7. -534 a -369
8. -368
9. -359 a -0.5

• modelo 60% de variância:

1. -1699
2. -1697.5
3. -1695.5 a -1639.5
4. -1636.5
5. -1635.5 a -1624.5
6. -1607
7. -1606 a -1603
8. -1600.5 a -1598.5
9. -1597 a -1595
10. -1592 a -1588.5
11. -1586 a -1566.5
12. -1304 a -1300
13. -1297 a -1282.5
14. -961.5 a -0.5

• modelo 70% de variância:

1. -2786 a -2746.5
2. -2718.5 a -2700.5
3. -2699.5
4. -2697 a -2598
5. -2530
6. -2527.5 a -2470
7. -2405.5 a -2404

8. -2391.5 a -2388
 9. -2328 a -0.5
- modelo 80% de variância:
 1. -1790.5 a -1518
 2. -1351.5 a -0.5
 - modelo 90% de variância:
 1. -1790.5 a -1786
 2. -1783.5 a -0.5

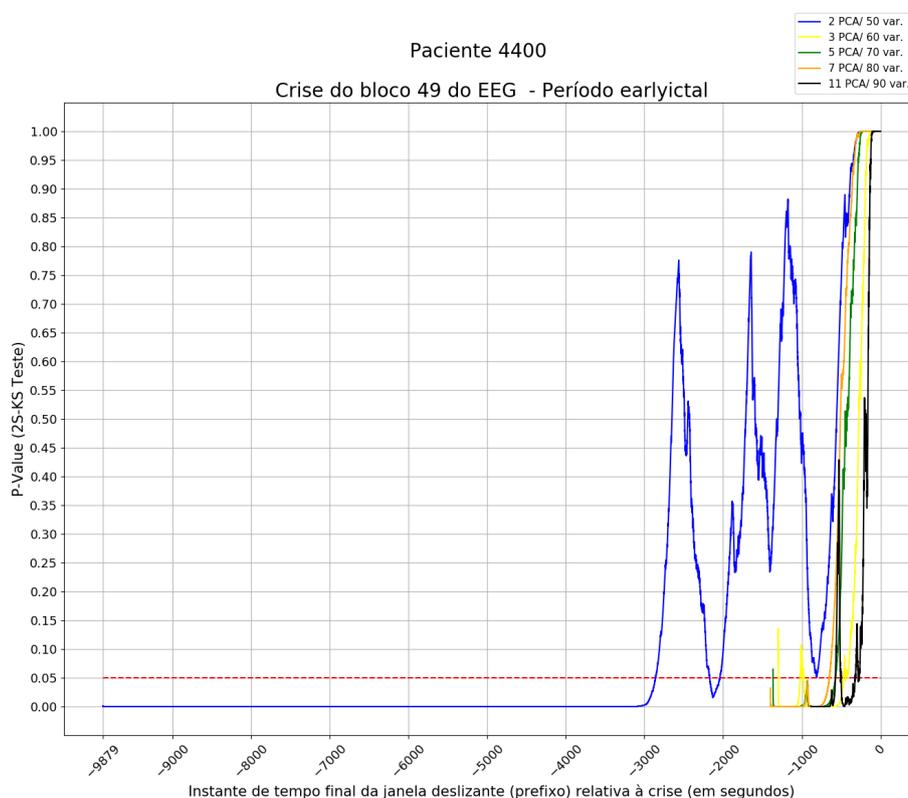


Figura 5.18: Paciente 4400 - Crise do bloco n^o 49 do EEG

A se destacar o fato de que três dos cinco modelos de testes apresentaram poucos períodos antecipatórios nos seus resultados. São eles: modelos com 50% e 80% (um período); modelo com 80% (dois períodos). Por outro lado, levando-se em consideração os intervalos de tempo nos quais os períodos antecipatórios estão contidos, temos o modelo com 50% com o intervalo mais distante da crise, iniciando no instante de tempo de -2860.5 segundos antes do início da crise. Os resultados obtidos são listados a seguir:

- modelo 50% de variância:

1. -2860.5 a -0.5

- modelo 60% de variância:

1. -1303 a -1299.5

2. -1020 a -994

3. -992.5 a -988

4. -986 a -977

5. -975.5

6. -470 a -447

7. -442 a -0.5

- modelo 70% de variância:

1. -1367

2. -573 a -0.5

- modelo 80% de variância:

1. -652 a -0.5

- modelo 90% de variância:

1. -574 a -509

2. -505 a -504.5

3. -503.5 a -501.5

4. -324.5 a -323.5

5. -320 a -318.5

6. -317.5 a -285

7. -282.5 a -282

8. -281

9. -280 a -275

10. -272.5 a -0.5

Análise

Os modelos de teste, em regra, variaram bastante de comportamento na caracterização de períodos antecipatórios, o que não tornaria simples a tarefa de indicar um deles como modelo de teste a ser adotado para futuras crises a serem analisadas pela metodologia de caracterização de períodos antecipatórios de crise. No entanto, como dentre todos os modelos, o modelo com variância de 70% foi o que menos variou de forma significativa na avaliação dos intervalos de tempo que continham períodos antecipatórios, este modelo seria o indicado para ser usado nas futuras crises, caso só pudesse ser usado um modelo de teste para tal.

5.2.5 Paciente N^o 4500

Atributo	Valor
Idade à época do EEG	49 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	27/03/2010
Duração da coleta de dados (EEG)	116 horas
Número de blocos (EEG)	72 blocos
Primeira crise epilética	27 anos
Sexo	Masculino
Número de crises analisadas	4
Tipos de crises analisadas	Focais complexas
Padrões de ondas	Tetas rítmicas Alfas rítmicas

Tabela 5.9: Atributos do paciente n^o 4500.

Em termos de modelos de experimentos referente aos dados do paciente n^o 4500, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	Nº de componentes	Nº de <i>concepts</i>
50%	52.89%	3	198
60%	60.34%	4	223
70%	72.18%	6	230
80%	80.88%	9	245
90%	90.93%	15	238

Tabela 5.10: Detecção de *concepts* do paciente n^o 4500.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 4500, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

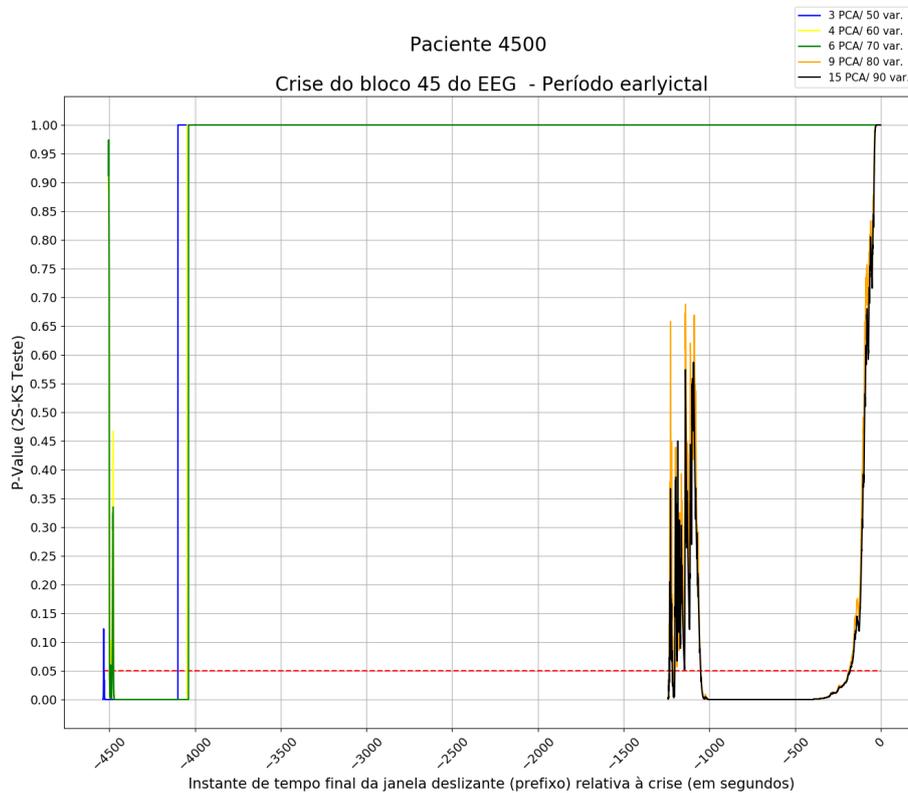


Figura 5.19: Paciente 4500 - Crise do bloco n^o 45 do EEG

Nessa crise, os modelos de testes com 50% 60% e 70% tem um comportamento similar, como por exemplo, identificando um período antecipatório de longa duração. Da mesma forma, os modelos de testes com 80% e 90% também identificaram um períodos antecipatórios similares, tanto em quantidade, quanto em duração destes. A seguir, os resultados obtidos com os cinco modelos de teste nessa referida crise:

- modelo 50% de variância:

1. -4533 a -4532
2. -4099.5 a -0.5

- modelo 60% de variância:

1. -4504.5 a -4499
2. -4493.5 a -4490
3. -4485 a -4474.5
4. -4045.5 a -0.5

- modelo 70% de variância:

1. -4504.5 a -4499
2. -4492.5
3. -4483 a -4475.5
4. -4037.5 a -0.5

- modelo 80% de variância:

1. -1234 a -1212.5
2. -1211.5
3. -1201.5 a -1050
4. -185 a -0.5

- modelo 90% de variância:

1. -1232.5 a -1232
2. -1231 a -1215.5
3. -1213.5
4. -1201 a -1050.5
5. -179 a -178.5
6. -177.5 a -0.5

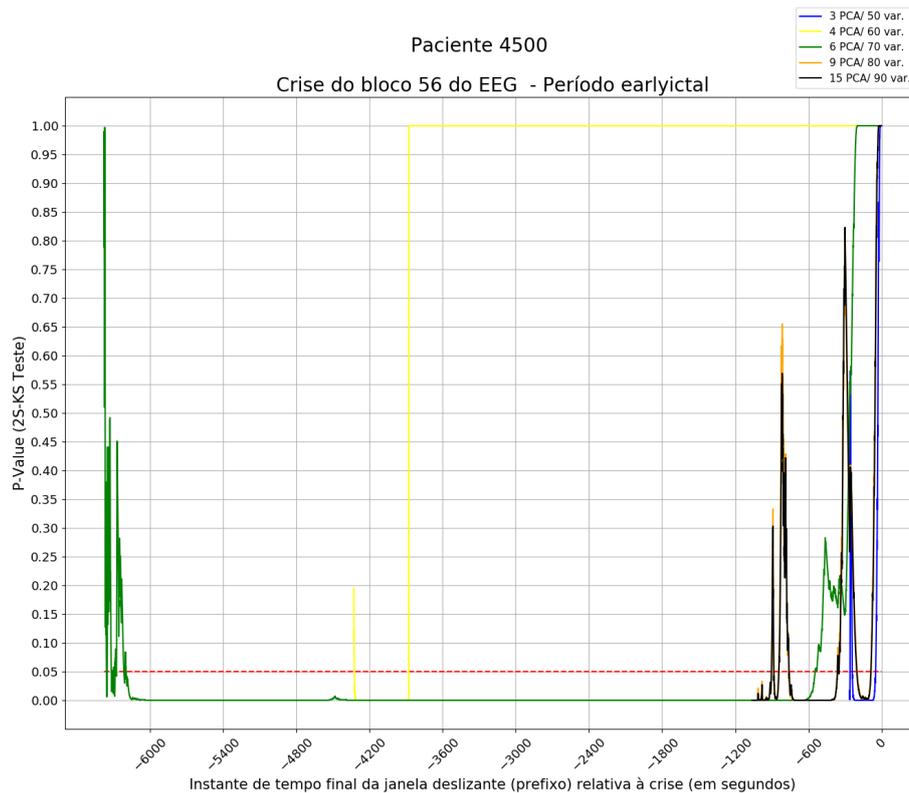


Figura 5.20: Paciente 4500 - Crise do bloco nº 56 do EEG

Nesse caso, percebemos quatro comportamentos distintos dos modelos de testes. Os modelos com 50%, 80% e 90% de variância acumulada (PCA) identificaram períodos antecipatórios em curtos intervalos de tempo, principalmente o modelo com 50%. Já os modelos com 60% e 70% caracterizaram períodos antecipatórios em maior número e com maior duração, mais notadamente o modelo com 70% de variância. Listamos a seguir os períodos antecipatórios obtidos nessa crise:

- modelo 50% de variância:
 1. -258.5 a -241
 2. -48.5 a -0.5
- modelo 60% de variância:
 1. -4327.5 a -4323.5
 2. -3877 a -0.5
- modelo 70% de variância:
 1. -6375.5 a -6356
 2. -6349.5 a -6316

3. -6309.5
4. -6295.5 a -6295
5. -6284.5 a -6278.5
6. -6277.5 a -6275.5
7. -6273.5 a -6213.5
8. -6212.5
9. -6602 a -6194.5
10. -545 a -0.5

● modelo 80% de variância:

1. -903 a -884
2. -844 a -761
3. -363 a -0.5

● modelo 90% de variância:

1. -903 a -884.5
2. -842.5 a -759.5
3. -362 a -351
4. -349 a -203.5
5. -87 a -0.5

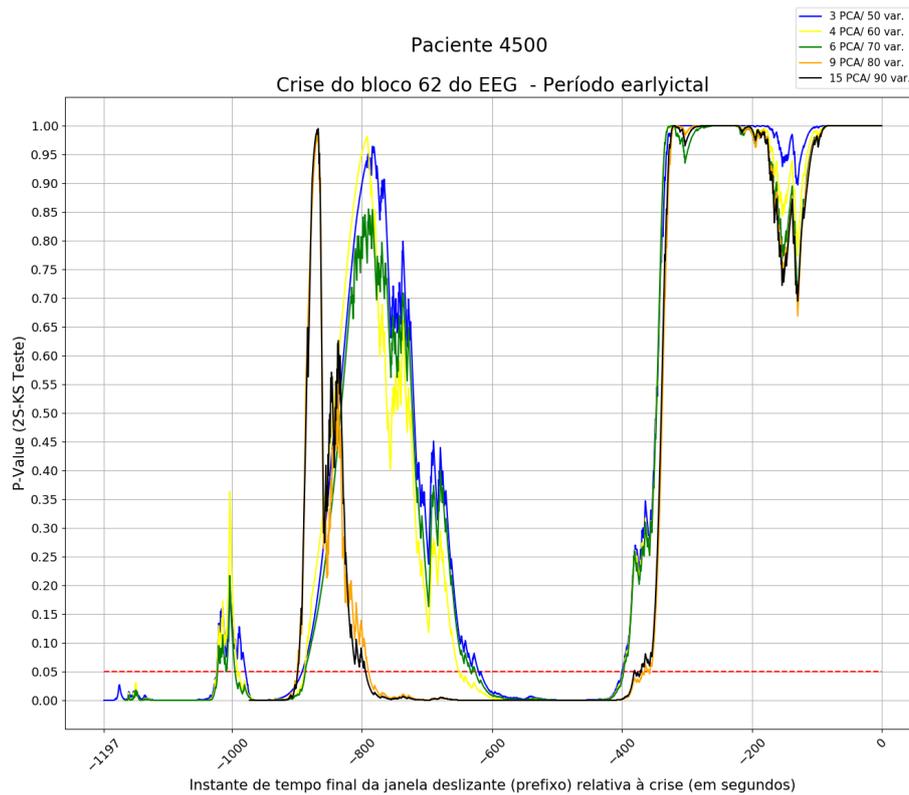


Figura 5.21: Paciente 4500 - Crise do bloco nº 62 do EEG

Nesse crise os modelos com 50%, 60% e 70% de variância mostraram comportamento semelhante na avaliação dos períodos antecipatórios. Da mesma forma, os modelos com 80% e 90% identificaram períodos antecipatórios em intervalos de tempo similares. Os comportamentos dos modelos de teste são apresentados a seguir:

- modelo 50% de variância:
 1. -1023.5 a -979.5
 2. -892 a -0.5
- modelo 60% de variância:
 1. -1023 a -992
 2. -990.5
 3. -888.5 a -0.5
- modelo 70% de variância:
 1. -1022 a -995.5
 2. -887 a -0.5

- modelo 80% de variância:

1. -899.5 a -790
2. -365 a -359
3. -357 a -0.5

- modelo 90% de variância:

1. -899 a -794.5
2. -380.5 a -380
3. -371.5 a -371
4. -370 a -0.5

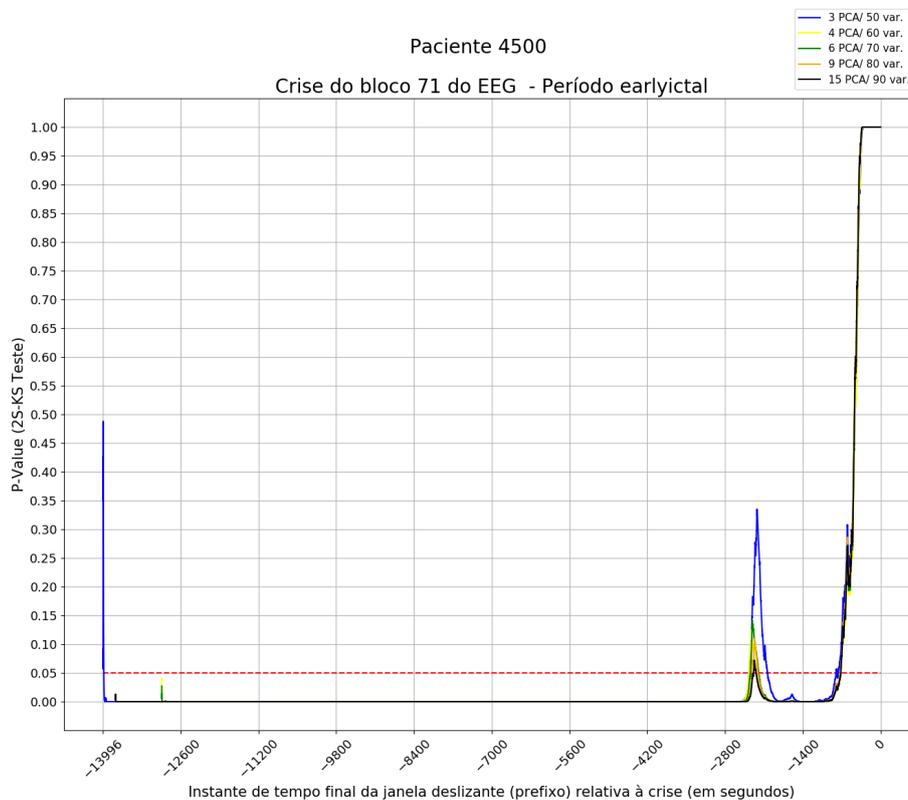


Figura 5.22: Paciente 4500 - Crise do bloco nº 71 do EEG

Os experimentos executados sobre a crise do bloco nº 71 mostram que todos modelos identificaram períodos antecipatórios parecidos em termos de quantidade e duração dos mesmos. Porém, uma à exceção foi observada nos resultados do modelo com 50%, pois este foi o único modelo que identificou um período antecipatório distante de todos os demais períodos aferidos, considerando ele mesmo e os demais modelos de teste. A lista de resultados seguinte explicita o que analisamos em relação a essa crise:

- modelo 50% de variância:
 1. -13995.5 a -13981.5
 2. -2366.5 a -0.5
- modelo 60% de variância:
 1. -2371 a -2224.5
 2. -717 a -0.5
- modelo 70% de variância:
 1. -2340
 2. -2339 a -2193.5
 3. -719 a -0.5
- modelo 80% de variância:
 1. -2313.5 a -2206.5
 2. -722
 3. -721 a -0.5
- modelo 90% de variância:
 1. -2284.5 a -2239
 2. -717 a -0.5

Análise

No caso desse paciente, dois comportamentos foram notados em relação aos modelos de teste usados: os modelos com limiar de variância acumulada de 60% e 70% tenderam a identificar intervalos maiores de períodos antecipatórios, enquanto os modelos com limiar de 80% e 90% tenderam ao oposto. Assim, em caso de se optar por um modelo de teste para futuras crises, qualquer um desses quatro modelos seria uma boa opção, sendo que a escolha por um deles ficaria condicionada a decisão de se tentar caracterizar períodos maiores, aceitando possíveis resultados falso positivos ou não.

5.2.6 Paciente N^o 4600

Atributo	Valor
Idade à época do EEG	16 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	05/04/2010
Duração da coleta de dados (EEG)	114 horas
Número de blocos (EEG)	81 blocos
Primeira crise epilética	14 anos
Sexo	Masculino
Número de crises analisadas	5
Tipos de crises analisadas	Focais complexas
	Não classificadas
Padrões de ondas	Tetas rítmicas
	Deltas rítmicas

Tabela 5.11: Atributos do paciente n^o 4600.

Em termos de modelos de experimentos referente aos dados do paciente n^o 4600, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	N ^o de componentes	N ^o de <i>concepts</i>
50%	53.23%	4	222
60%	65.15%	6	223
70%	73.27%	8	215
80%	80.02%	10	227
90%	90.80%	15	232

Tabela 5.12: Detecção de *concepts* do paciente n^o 4600.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 4600, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

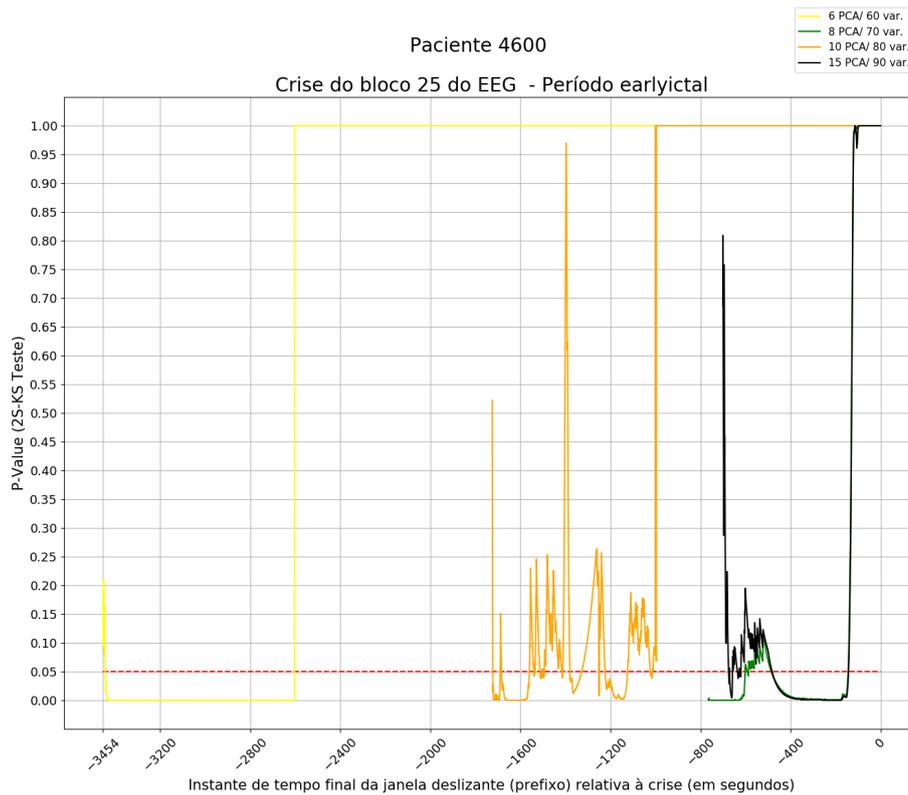


Figura 5.23: Paciente 4600 - Crise do bloco nº 25 do EEG

Nessa crise, o modelo de teste com 50% de variância acumulada não verificou nenhuma mudança no *concept* dos dados, ou seja, não houve mudança no comportamento dos dados ao longo do sinal de EEG da crise desse bloco, por isso conclui-se que não houve caracterização de período antecipatório, ou que todo o período analisado seria um enorme período antecipatório. Decidiu-se nessa pesquisa optar pela primeira opção.

Os demais modelos de testes conseguiram identificar períodos antecipatórios, ao contrário do modelo com 50%. Dentre esses modelos, os com 70% e 90% apresentaram períodos antecipatórios de duração menor em comparação aos modelos com 60% e 80% de variância. Os resultados obtidos foram os seguintes:

- modelo 60% de variância:
 1. -3453.5 a -3443.5
 2. -2602.5 a -0.5
- modelo 70% de variância:
 1. -601.5 a -592.5
 2. -586.5 a -485.5
 3. -144 a -0.5

- modelo 80% de variância:

1. -1725 a -1723
2. -1690
3. -1689 a -1684
4. -1560 a -1540.5
5. -1538.5 a -1536
6. -1534.5 a -1513
7. -1502 a -1501
8. -1499.5 a -1498.5
9. -1493.5 a -1416.5
10. -1412 a -1380.5
11. -1328.5 a -1251
12. -1250 a -1224.5
13. -1124 a -1023
14. -1013.5 a -0.5

- modelo 90% de variância:

1. -701 a -675.5
2. -672
3. -658 a -636
4. -627.5 a -625
5. -620 a -482
6. -142.5 a -0.5

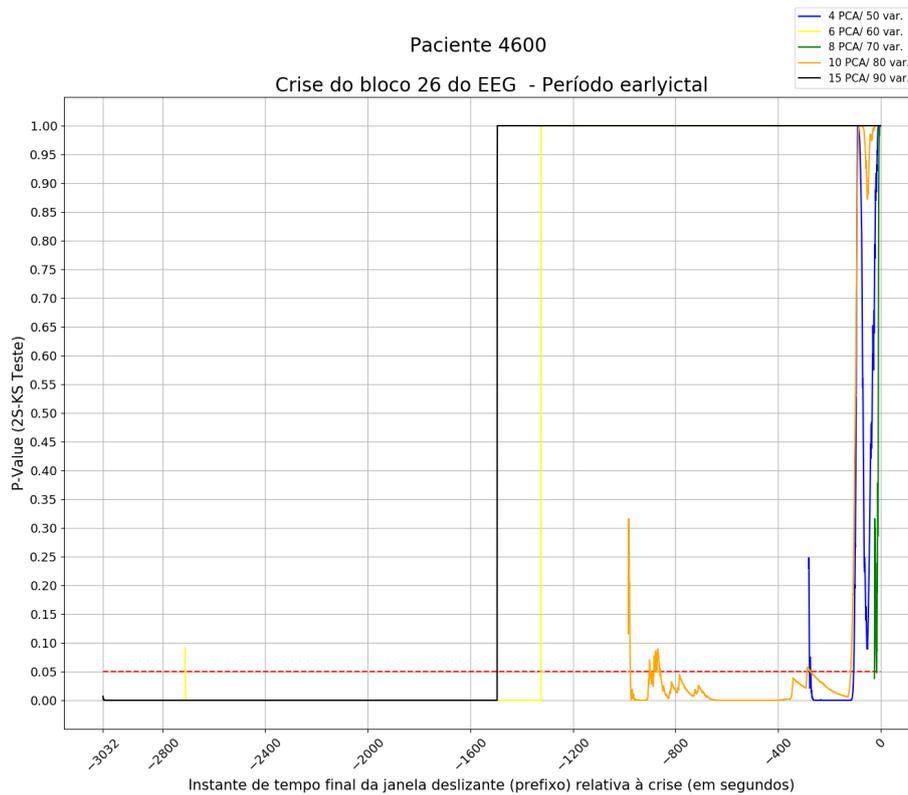


Figura 5.24: Paciente 4600 - Crise do bloco nº 26 do EEG

Dois comportamentos podem ser notados ao analisar os resultados dos modelos de teste para essa crise: dois modelos (60% e 90%) com comportamento mais estável, com a caracterização de praticamente um período antecipatório com final no instante -0.5 e três modelos (50%, 70% e 80%) com comportamento variável em termos de p -values (2S-KS Test). Cabe ainda ressaltar que o modelo com 70% de variância caracterizou dois períodos antecipatórios separados por um único instante de tempo (-15.5), que unidos configuram um curto intervalo. Em termos de quantidade de períodos antecipatórios e intervalos de tempo abrangidos por esses períodos caracterizados, não são notadas correlações evidentes entre os resultados dos modelos de teste, como pode ser observado a seguir:

- modelo 50% de variância:
 1. -280.5 a -277.5
 2. -276
 3. -275 a -273.5
 4. -105 a -0.5

- modelo 60% de variância:

1. -2712.5
2. -1323.5 a -0.5

- modelo 70% de variância:

1. -24 a -16
2. -15 a -0.5

- modelo 80% de variância:

1. -984 a -976.5
2. -901.5 a -898.5
3. -884.5 a -873.5
4. -872.5 a -861.5
5. -859.5 a -859
6. -287 a -273
7. -116.5 a -0.5

- modelo 90% de variância:

1. -1495 a -0.5

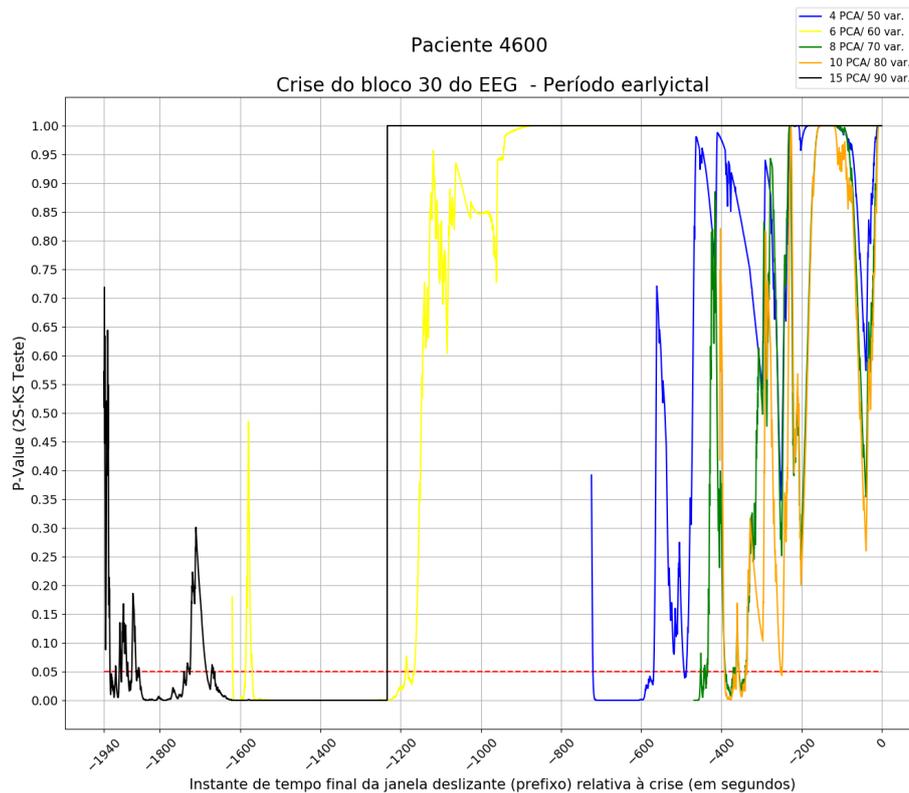


Figura 5.25: Paciente 4600 - Crise do bloco nº 30 do EEG

Nessa crise, os modelos de teste com 70% e 80% de variância acumulada apresentaram resultados semelhantes considerando o intervalo de tempo nos quais os períodos antecipatórios foram identificados. Além disso, destaca-se que os modelos de teste com 50%, 60% e 90% não possuem similaridade na caracterização de períodos antecipatórios. A seguir, os períodos antecipatórios obtidos com todos os modelos de teste, exceto o modelo com 90% de variância:

- modelo 50% de variância:
 1. -724 a -720.5
 2. -568.5 a -493
 3. -487 a -0.5
- modelo 60% de variância:
 1. -1620 a -1619
 2. -1586.5 a -1570.5
 3. -1187.5 a -1182.5
 4. -1167.5
 5. -1165.5 a -0.5

- modelo 70% de variância:

1. -452 a -450.5
2. -443
3. -441.5 a -441
4. -435.5 a -435
5. -434 a -390.5
6. -370 a -369.5
7. -366.5 a -366
8. -363 a -356
9. -340 a -0.5

- modelo 80% de variância:

1. -403.5 a -392
2. -362 a -356.5
3. -339 a -337
4. -336 a -252
5. -251 a -250.5
6. -248.5 a -0.5

Um caso particular, se comparado aos demais modelos de teste, foi o modelo com 90% de variância. Este modelo caracterizou dezesseis períodos antecipatórios, dos quais muitos encontram-se separados por poucos instantes de tempo. Considerando esses instantes mencionados como *outliers*, pode-se concluir que os dezesseis períodos caracterizados estão contidos em três intervalos de tempo, a saber:

- modelo 90% de variância:

1. $[-1939.5, -1852]$
2. $[-1740, -1667]$
3. $[-1233, -0.5]$

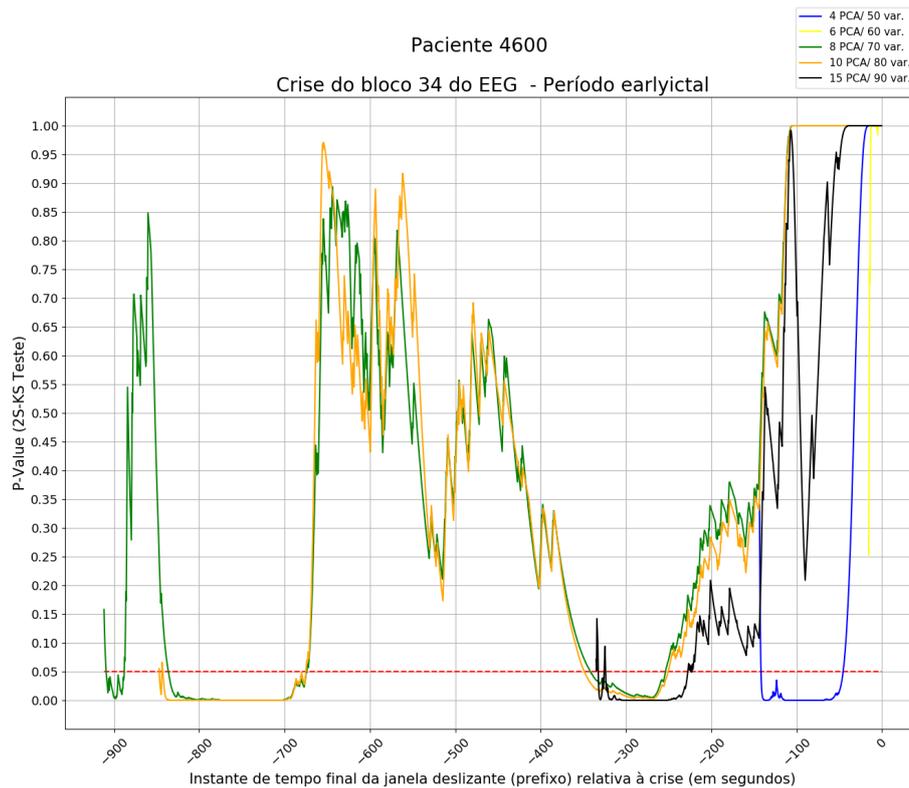


Figura 5.26: Paciente 4600 - Crise do bloco nº 34 do EEG

Os resultados dessa crise mostraram que os modelos de teste com 50% e 60% de variância acumulada identificaram períodos antecipatórios muito próximos do início da crise (o modelo com 50% ainda identificou mais um período, menor e um pouco mais afastado). Também apresentaram semelhanças os resultados dos modelos com 70% e 80% de variância acumulada, através de períodos antecipatórios em intervalos de tempo parecidos. Por fim, o modelo de teste com 90% de variância teve resultado distinto dos demais modelos, como pode ser conferido abaixo:

- modelo 50% de variância:

1. -143.5 a -142
2. -45.5 a -0.5

- modelo 60% de variância:

1. -15.5 a -0.5

- modelo 70% de variância:

1. -911.5 a -909.5
2. -887.5 a -863.5

3. -674.5
4. -673.5 a -342
5. -253.5 a -0.5

● modelo 80% de variância:

1. -847
2. -843.5 a -843
3. -674.5 a -349
4. -250.5 a -0.5

● modelo 90% de variância:

1. -335 a -332.5
2. -325 a -324
3. -227 a -223.5
4. -222.5 a -0.5

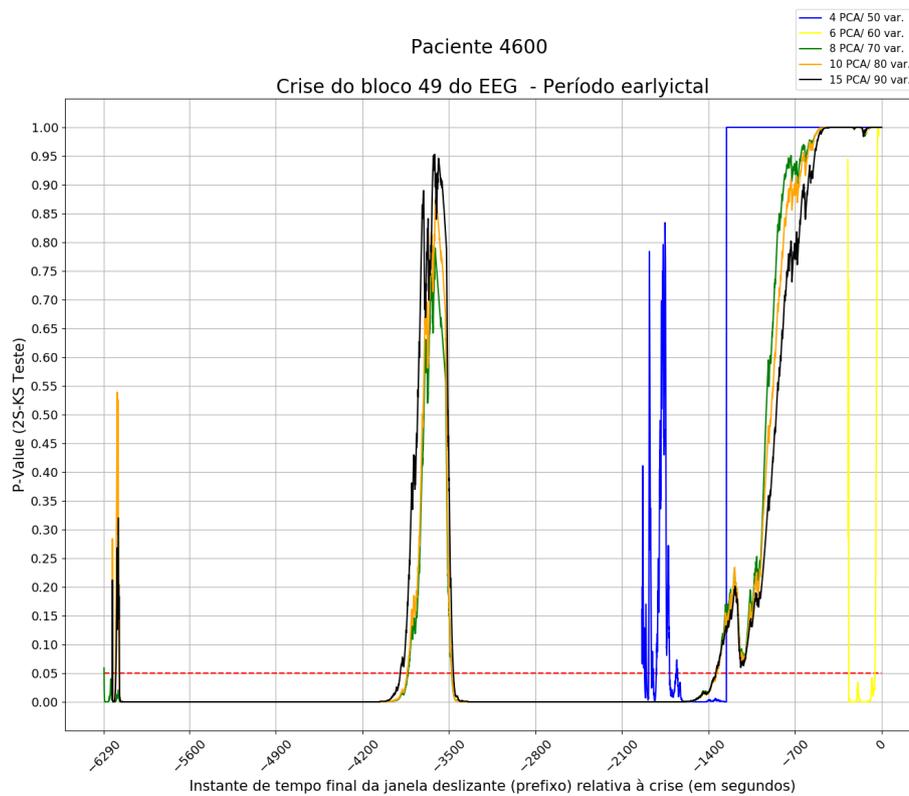


Figura 5.27: Paciente 4600 - Crise do bloco nº 49 do EEG

Nessa crise, os modelos de teste com 80% e 90% de variância acumulada apresentaram resultados praticamente idênticos, enquanto que dos três períodos antecipatórios caracterizados pelo modelo com 70% de variância, os dois primeiros são muito semelhantes com os dois primeiros períodos antecipatórios identificados pelos modelos com 80% e 90% de variância acumulada, o que indica uma maior confiabilidade nesses períodos citados como de fato antecipadores da crise epilética em questão. Tal conclusão fica evidenciada a seguir, com a apresentação dos resultados desses três modelos de teste:

- modelo 70% de variância:

1. -6289.5
2. -3830.5 a -3478.5
3. -1337 a -0.5

- modelo 80% de variância:

1. -6223 a -6221
2. -6197 a -6166
3. -3838 a -3477
4. -1335 a -0.5

- modelo 90% de variância:

1. -6223 a -6221
2. -6193.5 a -6165
3. -3893 a -3464
4. -1341 a -0.5

Em relação modelo com 60% de variância acumulada identificou dois períodos antecipatórios perto do início a crise, sendo que o primeiro inicia-se muito próximo, sendo estes:

- modelo 60% de variância:

1. -273.5 a -266.5
2. -54 a -0.5

Por fim, o modelo de teste com 50% de variância acumulada (PCA) identificou treze períodos antecipatórios, porém muitos deles separados por curtos instantes de tempo. Dessa forma, o resultado desse modelo de teste mostra que os períodos caracterizados formam cinco intervalos de períodos antecipatórios contendo poucos *outliers*, como demonstrado a seguir:

- modelo 50% de variância:

1. $[-1937.5, -1901]$
2. $[-1884, -1847.5]$
3. $[-1823, -1715]$
4. $[-1660.5, -1651]$
5. $[-1255, -0.5]$

Análise

Notou-se nas crises desse paciente que o comportamento dos modelos de teste variou de acordo com cada crise analisada. De todo modo, em termos de definição de um modelo de teste para ser usado em futuras crises, um bom candidato seria o modelo com 80% de variância acumulada na PCA, pois dentre todos os modelos usados, este foi o que apresentou um grau de variabilidade menor em termos do tamanho dos períodos antecipatórios aferidos.

5.2.7 Paciente N^o 5800

Atributo	Valor
Idade à época do EEG	26 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	29/05/2010
Duração da coleta de dados (EEG)	110 horas
Número de blocos (EEG)	67 blocos
Primeira crise epilética	7 anos
Sexo	Feminino
Número de crises analisadas	7
	Focais simples
Tipos de crises analisadas	Focais complexas
	Não classificadas
Padrões de ondas	Alfas rítmicas

Tabela 5.13: Atributos do paciente n^o 5800.

Em termos de modelos de experimentos referente aos dados do paciente n^o 5800, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	N ^o de componentes	N ^o de <i>concepts</i>
50%	58.09%	3	205
60%	63.95%	4	197
70%	73.35%	6	211
80%	81.39%	9	232
90%	91.05%	15	235

Tabela 5.14: Detecção de *concepts* do paciente n^o 5800.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 5800, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

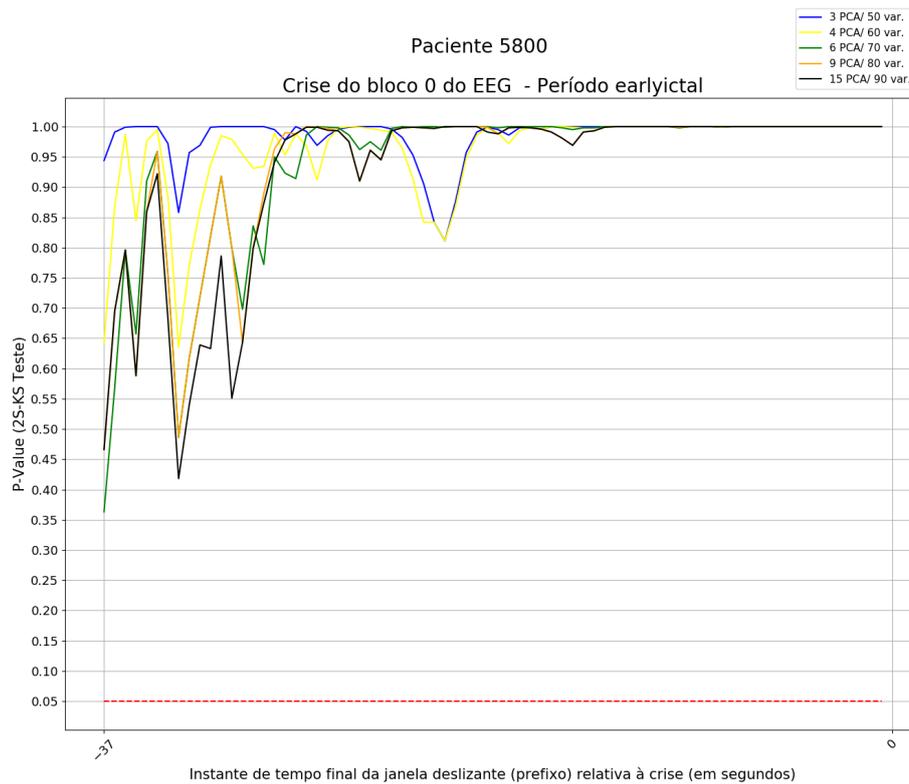


Figura 5.28: Paciente 5800 - Crise do bloco n^o 0 do EEG

Embora os modelos de teste tenham usado diferentes quantidades de componentes principais na PCA, todos eles caracterizaram um único e mesmo período antecipatório, compreendido entre os instantes -37 e -0.5 . Tal fato explica-se, pois

como essa crise ocorreu no início da coleta de dados do paciente, o segmento do EEG analisado pela metodologia de caracterização de períodos antecipatórios de crises epiléticas apresentada neste trabalho.

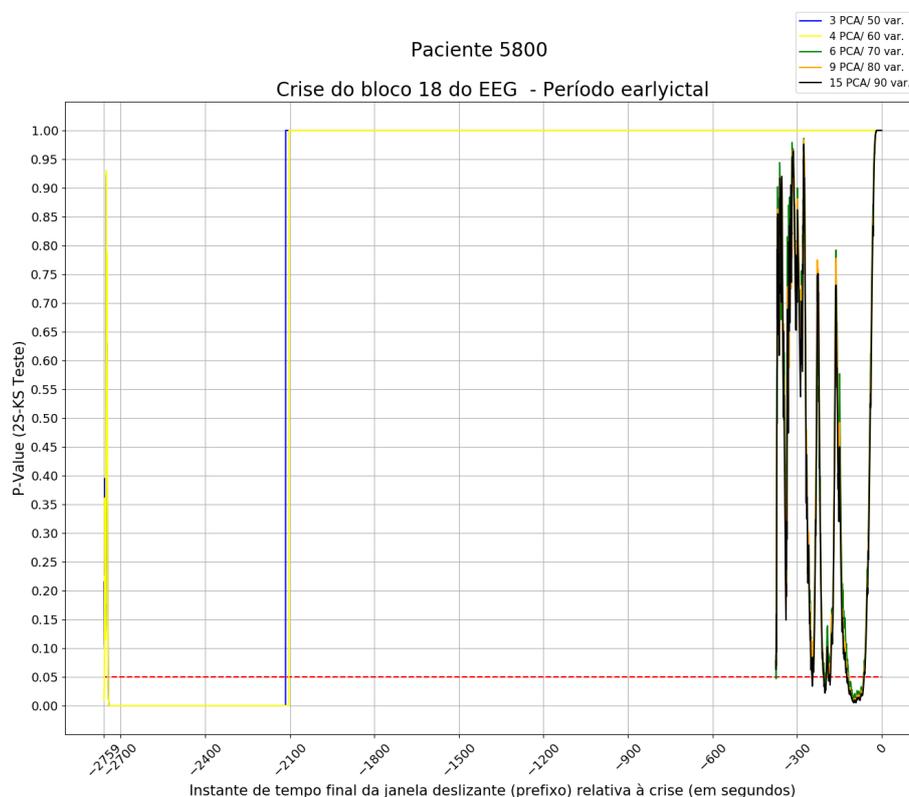


Figura 5.29: Paciente 5800 - Crise do bloco nº 18 do EEG

Nessa crise os modelos com menos variância acumulada (PCA), 50% e 60%, caracterizaram períodos antecipatórios maiores em duração em relação aos demais modelos (70%, 80% e 90%), evidenciando nesse caso uma correlação entre o aumento de variância acumulada/número de componentes principais na PCA e uma caracterização de períodos antecipatórios mais curtos. Os resultados obtidos são listados a seguir:

- modelo 50% de variância:

1. -2759 a -2744.5
2. -2114.5 a -0.5

- modelo 60% de variância:

1. -2757.5 a -2745
2. -2103 a -0.5

- modelo 70% de variância:

1. -375
2. -374 a -203
3. -197.5 a -117.5
4. -64 a -0.5

- modelo 80% de variância:

1. -375 a -246.5
2. -245 a -206
3. -196 a -123
4. -122 a -121
5. -119.5
6. -63 a -0.5

- modelo 90% de variância:

1. -375 a -247.5
2. -244.5 a -206
3. -196 a -189
4. -186.5
5. -185.5 a -185
6. -181.5 a -124
7. -62 a -0.5

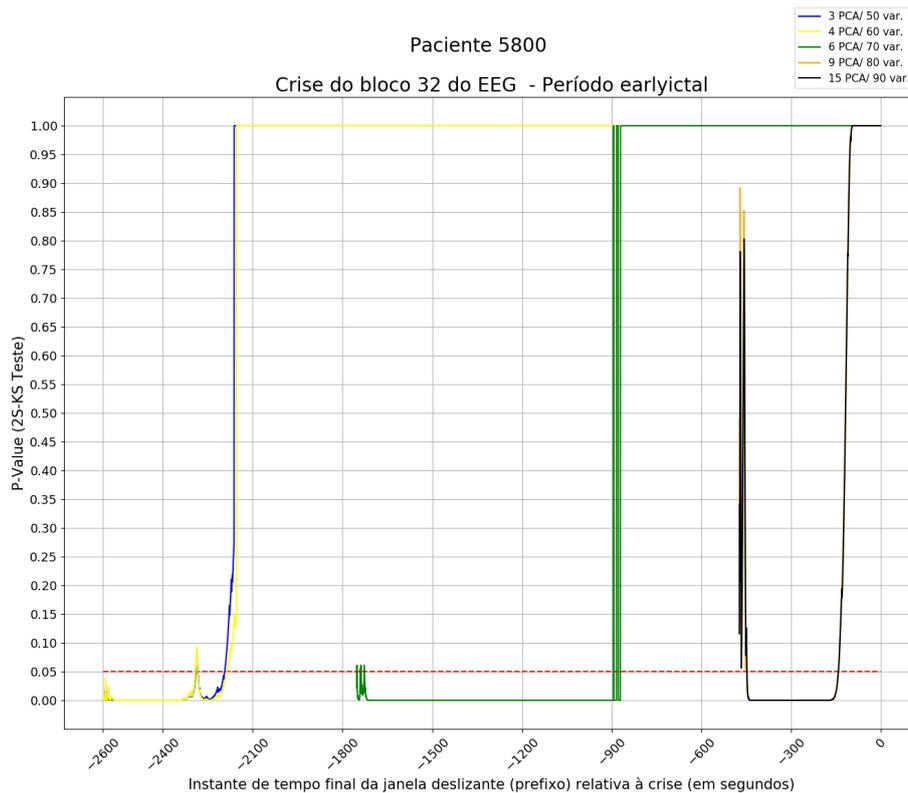


Figura 5.30: Paciente 5800 - Crise do bloco nº 32 do EEG

Nesse caso, observa-se o mesmo comportamento dos modelos na crise anterior (bloco nº 18): resultados de caracterização variando de detecções com maior duração temporal a detecções com menor duração, de acordo com o aumento da variância acumulada (PCA)/número de componentes nos modelos, iniciando do modelo com 50% até o modelo com 90%, passando pelos modelos com 60%, 70% e 80%. Nota-se que os modelos com 50% e 60% tiveram resultados semelhantes, assim como os modelos 80% e 90% apresentaram resultados similares entre si. Por fim, o modelo com 70% identificou períodos antecipatórios mais longos que os modelos com 80% e 90% e menos longos que os modelos com 50% e 60%. A seguir, os resultados obtidos com os modelos de teste nessa crise:

- modelo 50% de variância:
 1. -2289.5 a -2284.5
 2. -2193.5 a -0.5
- modelo 60% de variância:
 1. -2291 a -2284.5
 2. -2186 a -0.5

- modelo 70% de variância:

1. -1751.5
2. -1740.5
3. -1738
4. -1727 a -1726.5
5. -893
6. -883.5 a -878
7. -870 a -0.5

- modelo 80% de variância:

1. -472.5 a -448
2. -141.5 a -0.5

- modelo 90% de variância:

1. -472.5 a -447.5
2. -141.5 a -0.5

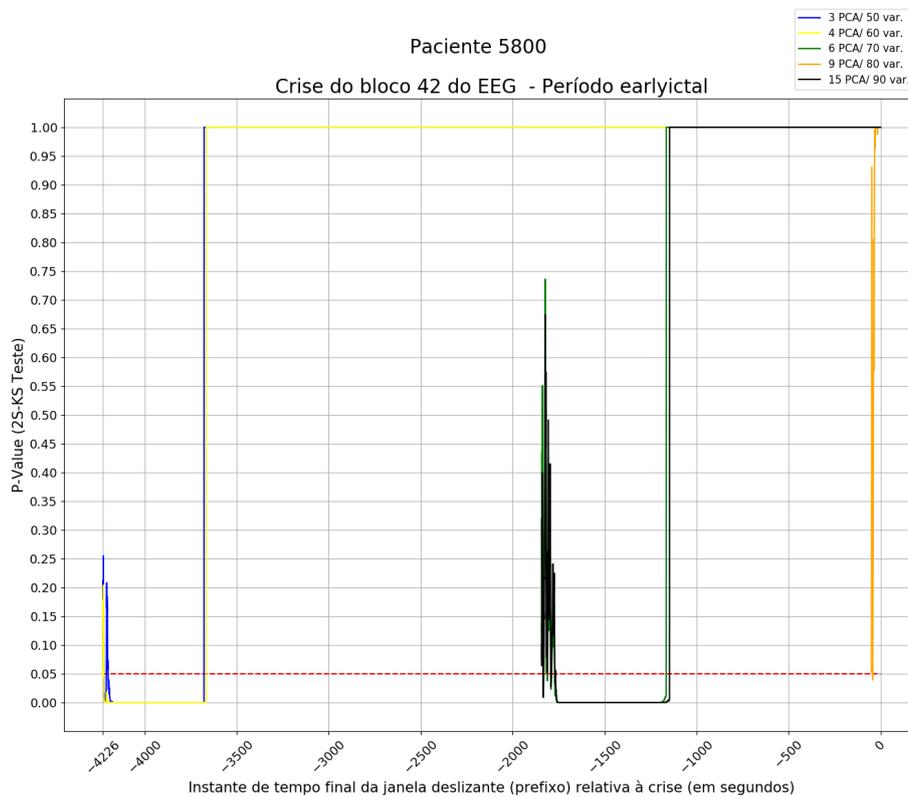


Figura 5.31: Paciente 5800 - Crise do bloco n^o 42 do EEG

Como nas duas crises anteriores (blocos nº 18 e 32), os resultados obtidos nesse caso também relacionam os modelos com menos variância a períodos antecipatórios mais longos. De maneira análoga, modelos com mais variância apresentaram períodos antecipatórios mais curtos, com exceção nessa crise para o modelo com 80% de variância acumulada, cujos períodos antecipatórios estão contidos em um curto intervalo de tempo próximo ao início da crise em questão. Os períodos antecipatórios obtidos com os modelos de teste são os seguintes:

- modelo 50% de variância:

1. -4226 a -4222.5
2. -4207.5 a -4199
3. -3677 a -0.5

- modelo 60% de variância:

1. -4226 a -4222.5
2. -4224.5 a -4222.5
3. -3666 a -0.5

- modelo 70% de variância:

1. -1842 a -1834.5
2. -1831.5 a -1812
3. -1810 a -1792.5
4. -1788 a -1772
5. -1165 a -0.5

- modelo 80% de variância:

1. -50 a -44
2. -43 a -0.5

- modelo 90% de variância:

1. -1842 a -1834.5
2. -1830 a -1792.5
3. -1789 a -1770.5
4. -1768
5. -1767
6. -1147.5 a -0.5

Análise

O comportamento dos modelos de teste ficou bem definido no caso desse paciente, mantendo uma mesma tendência em todas as crises analisadas. Isto posto, para a escolha de um modelo teste para a caracterização de períodos antecipatórios de futuras crises, caso o objetivo fosse ser mais restritivo na caracterização, então a escolha ficaria entre os modelos com 70%, 80% e 90% de variância acumulada na PCA. De maneira análoga, se a intenção fosse ser mais generalista, então a opção seria ou o modelo com 50%, ou o modelo com 60% de variância acumulada.

5.2.8 Paciente N^o 5900

Atributo	Valor
Idade à época do EEG	27 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	19/06/2010
Duração da coleta de dados (EEG)	107 horas
Número de blocos (EEG)	77 blocos
Primeira crise epilética	13 anos
Sexo	Feminino
Número de crises analisadas	3
Tipos de crises analisadas	Focais complexas
	Não classificadas
Padrões de ondas	Tetas rítmicas
	Betas rítmicas

Tabela 5.15: Atributos do paciente n^o 5900.

Em termos de modelos de experimentos referente aos dados do paciente n^o 5900, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	Nº de componentes	Nº de <i>concepts</i>
50%	64.15%	2	195
60%	64.15%	2	195
70%	70.08%	3	200
80%	82.04%	6	239
90%	90.30%	11	233

Tabela 5.16: Detecção de *concepts* do paciente nº 5900.

Nesse paciente os modelos de teste com limiares de variância acumulada (PCA) iguais a 50% e 60% atingiram esses valores de variância com o mesmo número de componentes principais (2). Dessa forma, os modelos citados apresentam sempre os mesmos resultados nas crises epiléticas analisadas deste paciente. Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente nº 5900, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

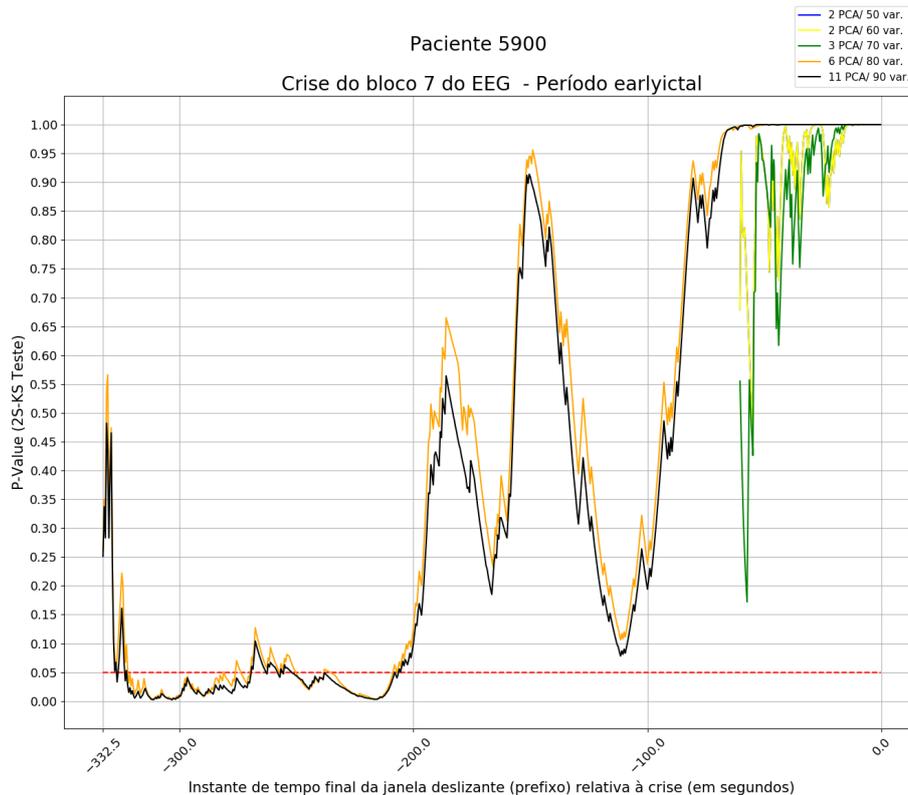


Figura 5.32: Paciente 5900 - Crise do bloco nº 7 do EEG

Os modelos de teste com variância de 50%, 60% e 70% identificaram um único período antecipatório no intervalo entre -60.5 e -0.5 segundos. Nota-se que embora

o modelo com 70% contasse com três componentes principais e os modelos com 50% e 60% com apenas duas, o resultado obtido com aquele modelo foi exatamente o mesmo obtido com estes dois últimos. Em relação aos modelos com 80% e 90% de variância, estes apresentaram resultados muito similares, inclusive com a ocorrência de um mesmo período antecipatório em ambos. A seguir, listamos os resultados obtidos nessa crise com os modelos com 80% e 90%:

- modelo 80% de variância:

1. -332.5 a -322
2. -281
3. -276 a -274
4. -269.5 a -257.5
5. -256.5 a -250.5
6. -238 a -235.5
7. -208.5 a -207.5
8. -206 a -0.5

- modelo 90% de variância:

1. -332.5 a -327
2. -326 a -323.5
3. -322.5
4. -269 a -263
5. -262 a -258.5
6. -256.5 a -256
7. -255 a -252
8. -208
9. -206 a -0.5

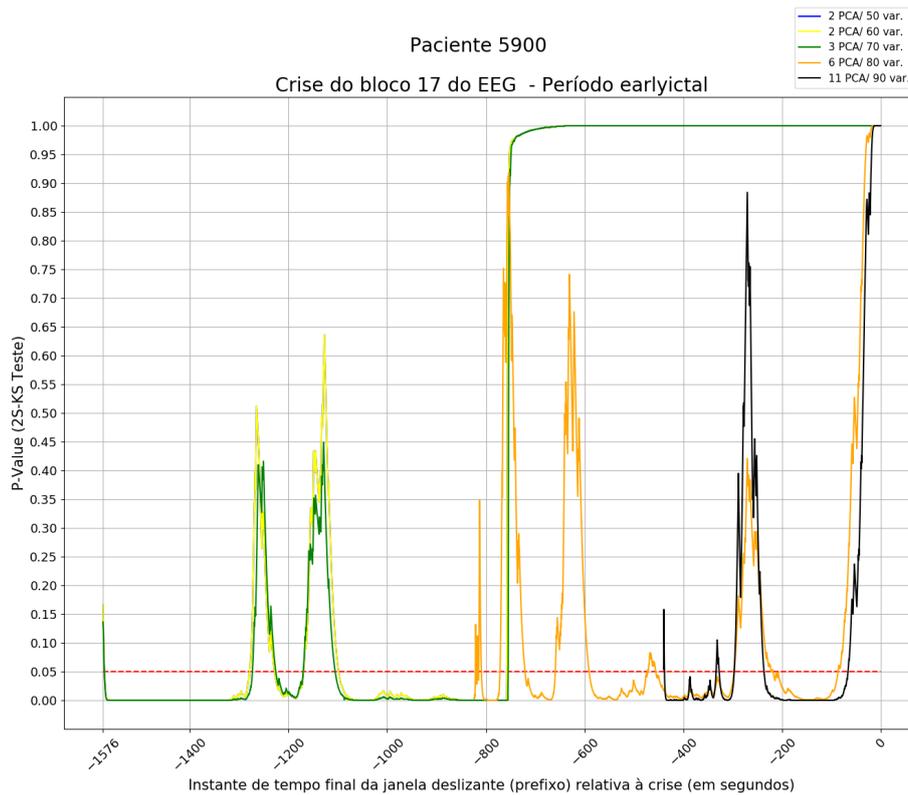


Figura 5.33: Paciente 5900 - Crise do bloco nº 17 do EEG

Assim como na crise anterior (bloco nº 7), o modelo com variância de 70% apresentou resultados correlatos com os resultados obtidos com os modelos com 50% e 60%, embora não idênticos. Por outro lado, os modelos com 80% e 90% identificaram períodos antecipatórios em intervalos de tempo menores que os outros modelos, sendo que o resultado de ambos não apresenta nenhuma similaridade aparente entre si. A seguir, listamos os períodos antecipatórios obtidos com os modelos de teste, lembrando que os modelos com 50% e 60%, por terem o mesmo número de componentes principais (PCA), possuem os mesmos períodos:

- modelos 50% e 60% de variância:

1. -1576 a -1573.5
2. -1279 a -1228.5
3. -1169.5 a -1099
4. -757.5 a -0.5

- modelo 70% de variância:

1. -1576 a -1573.5
2. -1273 a -1226

3. -1170 a -1106

4. -755 a -0.5

• modelo 80% de variância:

1. -821.5 a -819.5

2. -818 a -816

3. -814.5 a -809.5

4. -771.5 a -721

5. -659.5 a -591.5

6. -472.5

7. -471.5 a -456

8. -297 a -218.5

9. -86.5 a -0.5

• modelo 90% de variância:

1. -440 a -438.5

2. -333.5 a -327.5

3. -297 a -237

4. -66 a -0.5

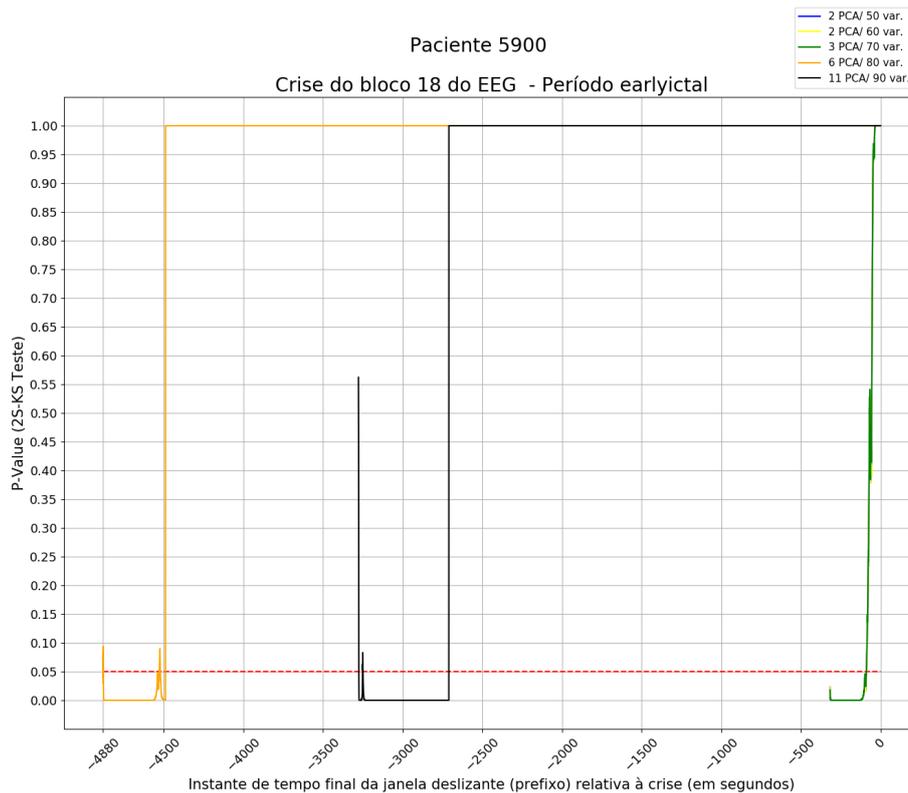


Figura 5.34: Paciente 5900 - Crise do bloco nº 18 do EEG

De maneira análoga às duas crises anteriores (blocos nº 7 e 17), os modelos com 50% e 60%, de mesma quantidade de componentes principais (PCA), caracterizaram períodos antecipatórios similares aos que foram caracterizados com o modelo com 70% de variância. Em contraponto, os modelos com 80% e 90%, embora não tenham apresentado resultados semelhantes entre si, tiveram em comum o fato de terem identificados períodos antecipatórios mais longos que os outros modelos de teste. A seguir, os resultados obtidos nessa crise:

- modelos 50% e 60% de variância:

1. -88.5 a -0.5

- modelo 70% de variância:

1. -89.5 a -0.5

- modelo 80% de variância:

1. -4879.5

2. $-4878.5.5$ a -4878

3. -4527 a -4520.5

4. -4487.5 a -0.5
- modelo 90% de variância:
 1. -3277 a -3275
 2. -3251 a -3249
 3. -2709.5 a -0.5

Análise

Considerando o fato de que os modelos com 50%, 60% e 80% de variância acumulada apresentaram resultados similares em todas as crises analisadas, conclui-se que qualquer um deles seria uma boa opção para ser usado como modelo de teste na caracterização de períodos antecipatórios de crises futuras desse paciente.

5.2.9 Paciente N^o 6300

Atributo	Valor
Idade à época do EEG	33 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	07/09/2010
Duração da coleta de dados (EEG)	139 horas
Número de blocos (EEG)	85 blocos
Primeira crise epilética	14 anos
Sexo	Masculino
Número de crises analisadas	3
Tipos de crises analisadas	Focais complexas
	Secundariamente generalizadas
	Não classificadas
Padrões de ondas	Alfa rítmicas
	Tetas rítmicas

Tabela 5.17: Atributos do paciente n^o 6300.

Em termos de modelos de experimentos referente aos dados do paciente n^o 6300, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	Nº de componentes	Nº de <i>concepts</i>
50%	53.87%	3	224
60%	63.92%	4	232
70%	70.28%	5	235
80%	81.87%	8	228
90%	90.94%	13	216

Tabela 5.18: Detecção de *concepts* do paciente n^o 6300.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 6300, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

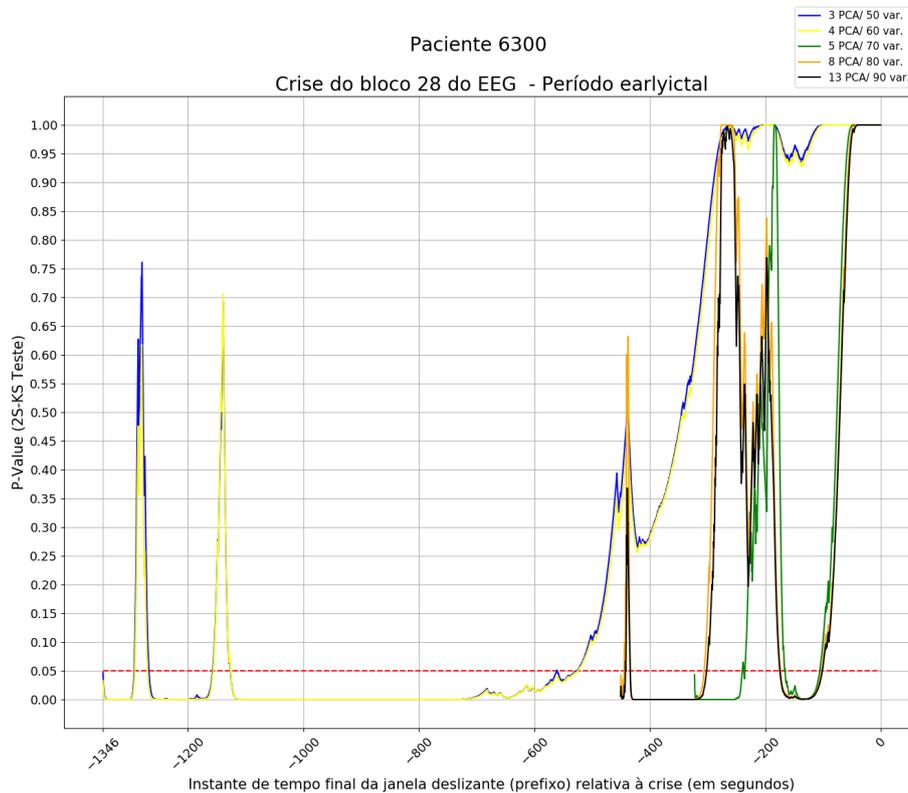


Figura 5.35: Paciente 6300 - Crise do bloco n^o 28 do EEG

Nessa crise os modelos com menos variância acumulada (PCA), 50% e 60%, caracterizaram períodos antecipatórios maiores em duração em relação aos demais modelos (70%, 80% e 90%), evidenciando nesse caso uma correlação entre o aumento de variância acumulada/número de componentes principais na PCA e uma caracterização de períodos antecipatórios mais curtos. Os resultados obtidos são listados a seguir:

• modelo 50% de variância:

1. -2759 a -2744.5
2. -2114.5 a -0.5

• modelo 60% de variância:

1. -2757.5 a -2745
2. -2103 a -0.5

• modelo 70% de variância:

1. -375
2. -374 a -203
3. -197.5 a -117.5
4. -64 a -0.5

• modelo 80% de variância:

1. -375 a -246.5
2. -245 a -206
3. -196 a -123
4. -122 a -121
5. -119.5
6. -63 a -0.5

• modelo 90% de variância:

1. -375 a -247.5
2. -244.5 a -206
3. -196 a -189
4. -186.5
5. -185.5 a -185
6. -181.5 a -124
7. -62 a -0.5

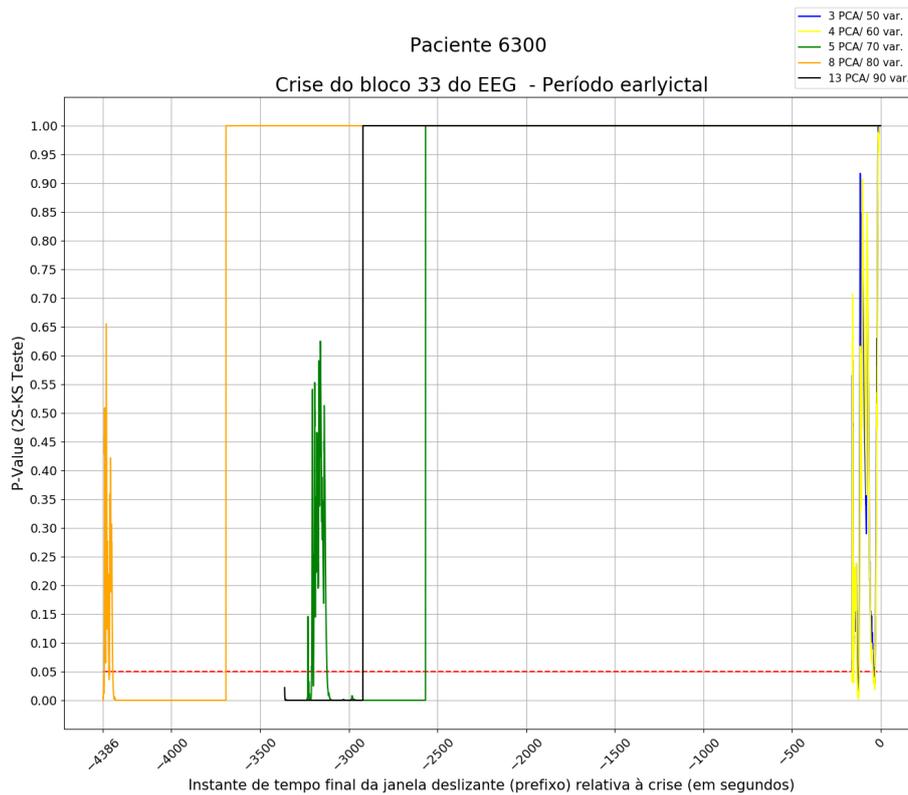


Figura 5.36: Paciente 6300 - Crise do bloco nº 33 do EEG

Os resultados obtidos com os modelos de teste nessa crise mostram que os modelos com 50% e 60% de variância acumulada (PCA) caracterizaram períodos antecipatórios muito similares e curtos em duração, enquanto os demais modelos (70%, 80% e 90%) identificaram períodos de maior duração, embora sem apresentar qualquer semelhança entre seus resultados. Um detalhe a observar é que o único modelo que apresentou apenas um período antecipatório foi o modelo com 90% de variância, que identificou tal período no intervalo de tempo entre -2919.5 e -0.5 . Os resultados dos demais modelos são mostrados a seguir:

- modelo 50% de variância:

1. -161.5 a -152.5
2. -151 a -131
3. -121.5 a -40
4. -37.5 a -36.5
5. -31 a -0.5

- modelo 60% de variância:

1. -160.5 a -152.5

2. -151 a -133.5

3. -120 a -43

4. -29.5 a -0.5

• modelo 70% de variância:

1. -3230.5 a -3229

2. -3209 a -3201

3. -3198 a -3122.5

4. -2566.5 a -0.5

• modelo 80% de variância:

1. -4378.5 a -4353

2. -4351 a -4305.5

3. -4349.5 a -4331

4. -3692 a -0.5

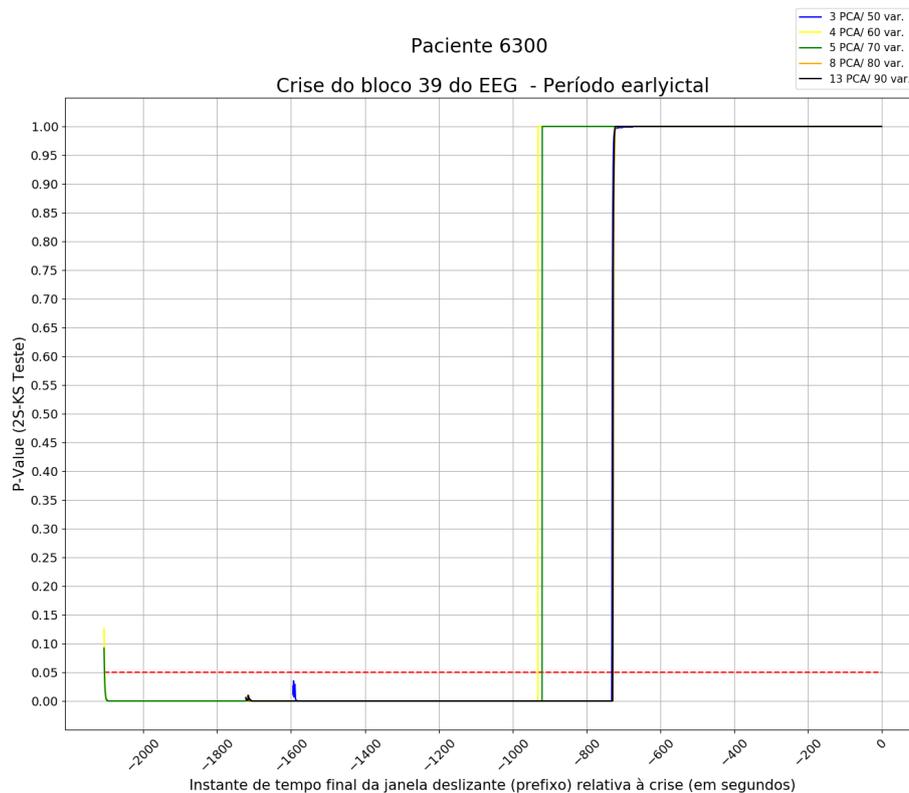


Figura 5.37: Paciente 6300 - Crise do bloco nº 39 do EEG

Nessa crise os modelos com 50%, 80% e 90% de variância acumulada caracterizaram apenas um período antecipatório cada um, em um intervalo de tempo muito

semelhante: -731 a -0.5 (modelo com 50%), -727.5 a -0.5 (modelo com 80%) e -728.5 a -0.5 (modelo com 90%). Do mesmo modo, os modelos com 60% e 70% de variância apresentaram comportamentos parecidos entre si, pois ambos caracterizaram dois períodos antecipatórios em intervalos de tempo similares. Nota-se que não houve flutuações nos valores de *p-values* resultantes da aplicação do 2S-KS Test durante a caracterização. A seguir, os períodos antecipatórios obtidos com os modelos com 60% e 70% de variância acumulada (PCA) são listados:

- modelo 60% de variância:

1. -2106.5 a -2105
2. -931 a -0.5

- modelo 70% de variância:

1. -2106.5 a -2105.5
2. -919.5 a -0.5

Análise

Desconsiderando a primeira crise analisada, nas demais crises o modelo de teste com limiar de 50% de variância acumulada seria o escolhido para a caracterização em crises futuras, caso a intenção fosse adotar a premissa de se buscar períodos antecipatórios menores, evitando falsos positivos. Por outro lado, se a premissa fosse o oposto, então o modelo com limiar de 70% de variância seria o modelo a ser adotado como modelo de teste na caracterização de períodos antecipatórios de novas crises.

5.2.10 Paciente N^o 6600

Atributo	Valor
Idade à época do EEG	44 anos
Número de eletrodos (EEG)	27
Data da coleta de dados (EEG)	23/10/2010
Duração da coleta de dados (EEG)	106 horas
Número de blocos (EEG)	62 blocos
Primeira crise epilética	20 anos
Sexo	Feminino
Número de crises analisadas	3
Tipos de crises analisadas	Focais simples
	Focais complexas
Padrões de ondas	Alfa rítmicas
	Tetas rítmicas

Tabela 5.19: Atributos do paciente n^o 6600.

Em termos de modelos de experimentos referente aos dados do paciente n^o 6600, foram identificadas as seguintes quantidades de *concepts*:

Limiar	Variância acumulada	N ^o de componentes	N ^o de <i>concepts</i>
50%	51.51%	4	148
60%	61.81%	6	141
70%	70.41%	8	140
80%	81.64%	12	140
90%	91.21%	17	138

Tabela 5.20: Detecção de *concepts* do paciente n^o 6600.

Segue o resultado da análise sobre a parte *earlyictal* dos *concepts de interesse* do paciente n^o 6600, considerando que para cada crise analisada há um *concept de interesse* associado a ela.

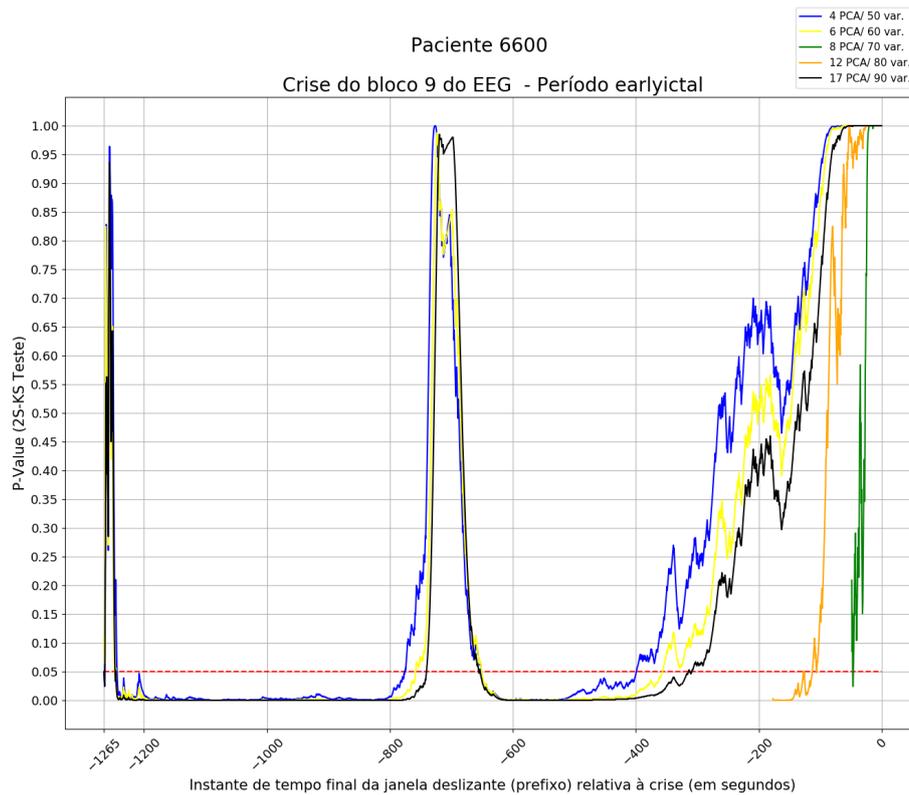


Figura 5.38: Paciente 6600 - Crise do bloco nº 9 do EEG

Nessa crise os modelos com variância acumulada (PCA) em 50%, 80% e 90% caracterizaram períodos antecipatórios similares, mais longos se comparados aos períodos antecipatórios obtidos através dos modelos com 60% e 70% de variância. Os resultados destes dois últimos não possuem similaridade, exceto pelo fato de que os períodos antecipatórios caracterizados por ambos cobrirem um intervalo curto de tempo. A seguir, os resultados obtidos são demonstrados:

- modelo 50% de variância:

1. -1263.5 a -1244
2. -776
3. -775 a -651.5
4. -399 a -0.5

- modelo 60% de variância:

1. -1265 a -1248
2. -1247
3. -1246
4. -758 a -651

5. -356.5 a -0.5

• modelo 70% de variância:

1. -49 a -47.5

2. -46 a -0.5

• modelo 80% de variância:

1. -127.5

2. -113 a -107

3. -106 a -0.5

• modelo 90% de variância:

1. -1265

2. -1263.5 a -1247

3. -1246

4. -740.5 a -655.5

5. -654.5 a -653.5

6. -314.5 a -312.5

7. -309 a -0.5

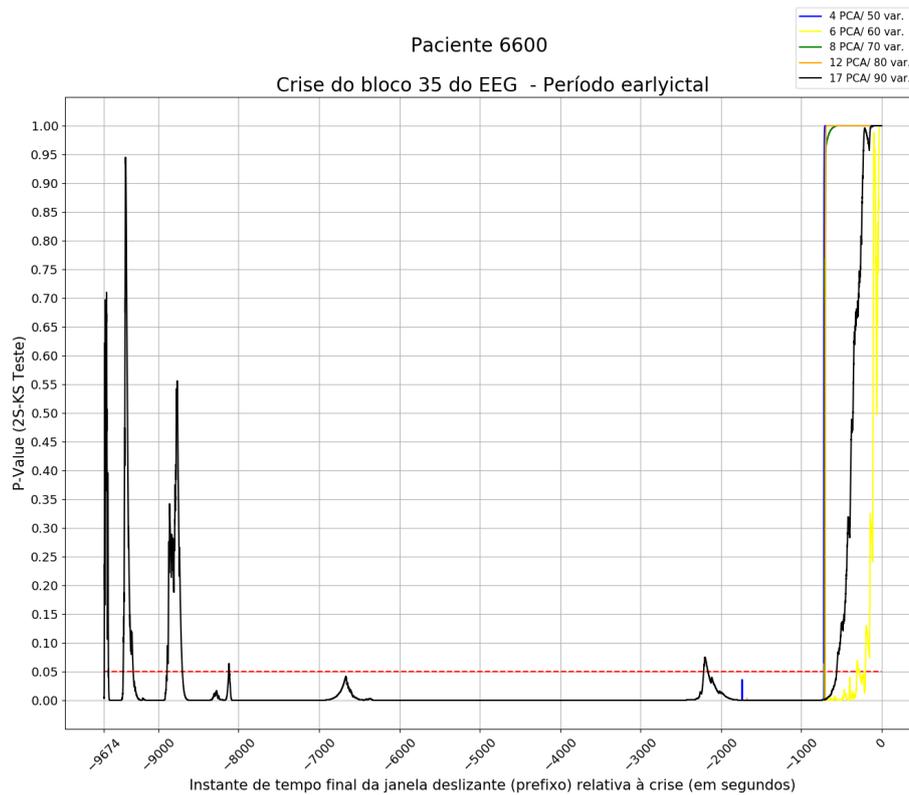


Figura 5.39: Paciente 6600 - Crise do bloco nº 35 do EEG

Nessa crise, os modelos com variância acumulada (PCA) em 50%, 70% e 80% tiveram resultados semelhantes ao caracterizar um único período antecipatório em um intervalo de tempo similar: -720 a -0.5 , -707 a -0.5 e -702 a -0.5 , respectivamente. Por outro lado, os modelos com 60% e 90% de variância identificaram um número maior de períodos antecipatórios, sem no entanto apresentar alguma similaridade em termos de resultados, o que pode ser observado a seguir:

- modelo 60% de variância:

1. -714.5
2. $-307.5.5$ a -286.5
3. -283 a -279.5
4. -278.5 a -274.5
5. -200.5 a -0.5

- modelo 90% de variância:

1. -9671.5 a -9620
2. -9618.5
3. -9435.5 a -9312

4. -8890.5 a -8700
5. -8124 a -8115.5
6. -2214 a -2160
7. -559 a -0.5

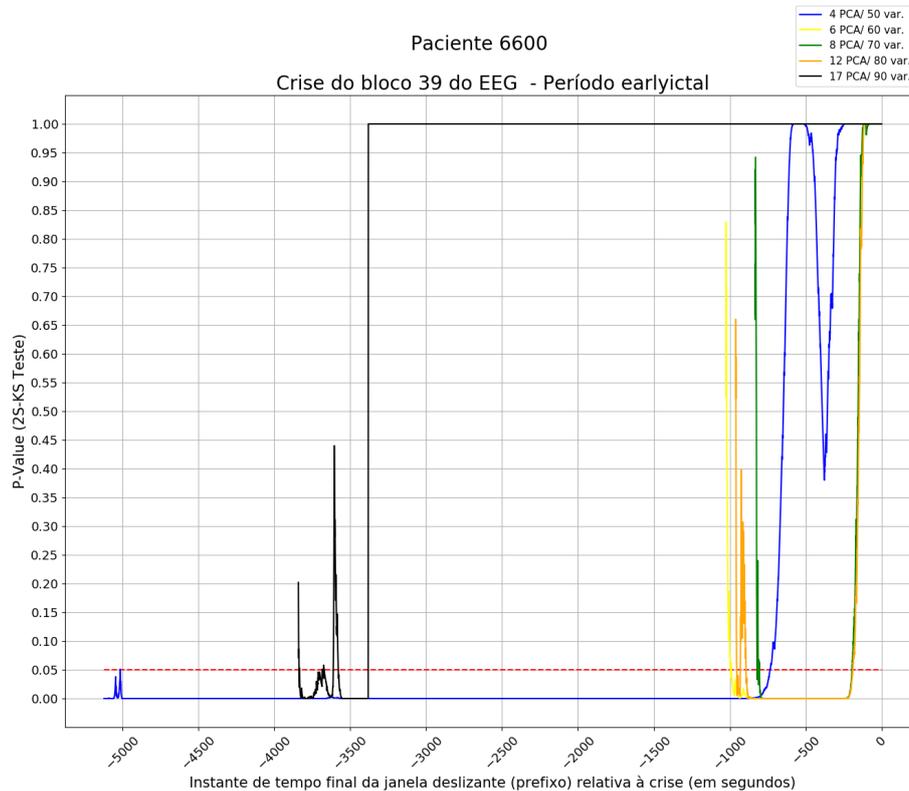


Figura 5.40: Paciente 6600 - Crise do bloco n^o 39 do EEG

Nessa crise os modelos com 60%, 70% e 80% de variâncias apresentaram identificaram períodos antecipatórios em intervalos de tempo próximos, embora não exatamente iguais. Os períodos caracterizados estão contidos em dois intervalos de tempo, ambos de curta duração, sendo um próximo ao início da crise (a menos de 200 segundos desta) e outro mais afastado. Já o modelo com 50% apresentou um período antecipatório de longa duração e outro período formado por um único instante de tempo, a mais de 5000 segundos do início da crise em questão. Por fim, o modelo com 90% de variância identificou seis períodos antecipatórios que juntos correspondem a um intervalo de tempo de longa duração. A seguir, apresentamos os resultados obtidos:

- modelo 50% de variância:

1. -5015

2. -735.5 a -0.5
- modelo 60% de variância:
 1. -1025.5 a -995.5
 2. -991.5 a -989.5
 3. -193.5 a -0.5
 - modelo 70% de variância:
 1. -833.5 a -821
 2. -819.5
 3. -818.5 a -813
 4. -810.5
 5. -809.5 a -809
 6. -807 a -805.5
 7. -195.5 a -0.5
 - modelo 80% de variância:
 1. -961.5 a -955
 2. -930.5 a -893.5
 3. -190 a -0.5
 - modelo 90% de variância:
 1. -3841.5 a -3837.5
 2. -3834
 3. -3684.5 a -3684
 4. -3675 a -3671.5
 5. -3611.5 a -3578.5
 6. -3381 a -0.5

Análise

Nas crises desse paciente, todos os modelos obtiveram resultados semelhantes na caracterização dos períodos antecipatórios, exceto o modelo com limiar de 90% de variância acumulada na PCA. Assim, qualquer um dos demais modelos (50%, 60%, 70% e 80%) poderia ser usado em futuras crises como modelo de teste na avaliação de períodos antecipatórios.

Capítulo 6

Conclusões

Nesta Seção apresentamos um resumo analítico do trabalho que foi apresentado nesta tese, as principais contribuições e propostas de melhorias na metodologia para caracterização de crises epiléticas que serão desenvolvidas a posteriori, como trabalhos futuros.

6.1 Resumo

De forma geral, a metodologia proposta no presente trabalho foi a de identificar intervalos de tempo nos quais poderiam ser caracterizados períodos que seriam capazes de antecipar a ocorrência de uma crise epilética, baseando-se na investigação do chamado *raw data*, i.e., os dados na forma como eles foram obtidos com o eletroencefalograma (EEG), sem fazer uso de nenhuma técnica de processamento de sinal usual (*transformada de Wavelet, Fourier*, dentre outras). A ideia por trás dessa premissa foi a de desenvolver algo inovador e menos custoso em termos de tempo e requisitos de memória e armazenamento. Por isso, a opção foi a de não utilizar transformações robustas sobre os dados, usando-os em sua forma original.

Inicialmente, a metodologia busca formar o conjunto de dados de entrada a partir da segmentação do *raw data* (EEG) em entradas referente ao volume de dados coletados em 1 (um) segundo pelo EEG, considerando uma sobreposição de 0.5 (meio) segundo entre as entradas. Tais entradas teriam a dimensão de $1024 \times n$, sendo n o número de eletrodos efetivamente usado na coleta dos sinais neuronais dos pacientes durante o exame de EEG. Por motivo de redução da dimensionalidade inicial das entradas, um fator de compressão ψ foi aplicado sobre elas, reduzindo a dimensionalidade das entradas para $1 \times n$, algo sempre desejável quando se lida com dados de alta dimensionalidade.

Em seguida, adota-se a abstração desses dados de dimensão $1 \times n$ como vetores no espaço R^n do conjunto de entradas e a aplicação da análise das principais componentes desses dados (PCA), tomando por base um limiar estabelecido para o

valor desejado em termos da variância acumulada por componente. Neste trabalho, optou-se por utilizar cinco limiares distintos: 50%, 60%, 70%, 80% e 90% da variância acumulada. Assim, são disponibilizadas alternativas de ajuste da metodologia para cada caso pontual (paciente/crise) analisado. Como efeito colateral benéfico, nesse estágio da metodologia, os dados são novamente reduzidos em termos de dimensionalidade, passando de $1 \times n$ para $1 \times p$, sendo p o número de componentes principais com as quais o limiar de variância acumulada por componente foi atingido.

Por fim, os dados sofrem um processo de unidimensionalização, no qual as distâncias *euclidianas* entre pares de dados (vetores de entrada) no espaço R^p obtido anteriormente são calculadas, gerando assim um sinal unidimensional que representa as distâncias aferidas. Sobre esse sinal é então aplicado o detetor de *concept drifts* ADWIN e através dele o “conceito de interesse” é identificado. Por “conceito de interesse” entenda-se o segmento do sinal de distância no qual estão contidos os elementos que representam dados que referem-se aos sinais do EEG obtidos durante a ocorrência de uma determinada crise epilética (estágio *ictal*). Assim, os períodos antecipatórios de uma crise epilética são caracterizados a partir de uma análise estatística sobre o “conceito de interesse” correlacionado àquela crise. O teste estatístico adotado para esse fim foi o teste biamostrais Kolmogorov-Smirnov (2S-KS Test), que mostrou-se eficiente em avaliar a semelhança entre a parte *interictal* que antecede a parte *ictal* do “conceito de interesse” e os seus prefixos. A essa parte do “conceito de interesse” foi dado o nome de *earlyictal*. Cada prefixo do *earlyictal* identificado como estatisticamente semelhante ao próprio *earlyictal* pelo 2S-KS Test foi então considerado parte de um período antecipatório da crise analisada.

6.2 Principais Contribuições

Considerando-se a complexidade da tarefa de caracterização de períodos antecipatórios de crises epiléticas, devido à própria natureza dos eventos epiléticos (imprevisibilidade, variedade dos tipos de crises, caráter heterogêneo das síndromes epiléticas, etiologia diversa etc), a metodologia proposta neste trabalho apresentou resultados interessantes em termos de caracterização de tais períodos antecipatórios na maior parte dos testes efetuados. Os pontos a serem destacados como as principais contribuições dessa pesquisa de doutorado são:

1. o uso dos dados do eletroencefalograma em sua forma original (*raw data*), evitando aumentar a complexidade da arquitetura da metodologia com a adoção de técnicas robustas de processamento de dados;
2. a representação dos dados eletroencefalográficos de alta dimensionalidade de uma forma mais simplificada, representando cada segundo de informação com

um único valor por eletrodo, formando um conjunto de entradas de dados de dimensionalidade reduzida, sem perda significativa de informação;

3. a não fixação de um limiar único de variância acumulada na PCA, durante a fase de representação do conjunto de entradas de maneira vetorial no espaço R^n , o que permitiu uma análise individualizada de cada crise epilética estudada, de acordo com o desempenho dos cinco modelos de teste adotados (limiares de variância acumulada de 50%, 60%, 70%, 80% e 90%);
4. representação unidimensional de um sinal multidimensional como o sinal do EEG, através das transformações desenvolvidas neste trabalho: compressão dos dados, representação vetorial dos dados a partir de um número p de suas componentes principais, unidimensionalização dos dados de p dimensões a partir do cálculo das distâncias dos vetores de dados em relação aos seus predecessores imediatos;
5. utilização combinada da técnica de inteligência computacional para detecção de mudanças no conceito dos dados chamada ADWIN e o teste estatístico biamostrado não paramétrico de Kolmogorov-Smirnov, para identificar os potenciais períodos antecipatórios de crises epiléticas.

6.3 Trabalhos Futuros

Em termos de trabalhos a serem efetuados posteriormente visando aprimorar a metodologia de caracterização de períodos antecipatórios de crises epiléticas, há pelo menos duas ideias a serem desenvolvidas. Uma delas é a de verificar a possibilidade de haver períodos antecipatórios que não estejam compreendidos dentro do chamado período *earlyictal* (dados *interictais* do EEG precedendo dados *ictais* em um mesmo *concept*). Isto se daria pela verificação da existência de *concepts* anteriores ao *concept* que possui dados referentes ao período *ictal* (ocorrência de crise), mas com as mesmas características dele, o que configuraria a ocorrência do padrão **recorrente** de *concept drift*. Além disso, um segundo objetivo seria a verificação das chamadas crises epiléticas subliminares, as quais não manifestam sintomas externos e por isso, passam despercebidas pelos profissionais médicos e pelos portadores de epilepsia.

A outra ideia seria testar o uso da metodologia em forma de um comitê, no qual haveria uma instância da metodologia para cada uma das cinco regiões do cérebro: frontal, temporal, central, occipital e parietal. Dessa forma, poderíamos investigar a ocorrência de períodos antecipatórios separadamente, por região cerebral. O objetivo aqui seria buscar uma caracterização mais refinada e também verificar se um período

antecipatório seria um evento generalizado (presente no cérebro como um todo) ou localizado (em alguma região ou regiões específicas).

Por último, mas não menos importante, verificar se os resultados obtidos com essa metodologia de fato conseguem ser úteis na predição de crises epiléticas. Tal ação se daria pelo uso das informações acerca de períodos antecipatórios de crise como base para a rotulação de possíveis períodos *preictais* (períodos que sinalizam que uma crise epilética é iminente) nos dados eletroencefalográficos dos pacientes. Torna-se então interessante investigar diferentes modelos de predição, principalmente os pertencentes à área da inteligência computacional, e.g., redes neurais artificiais, máquinas de de vetores de suporte (SVM), regressões, árvores de regressões (XGBoost), *random forests* etc.

Referências Bibliográficas

- [1] ORGANIZATION, W. H. “Epilepsy”. fev. 2022. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/epilepsy>>.
- [2] NGUYEN, R., TELLEZ-ZENTENO, J. “Injuries in epilepsy: a review of its prevalence, risk factors, type of injuries and prevention”, *Neurology international*, v. 1, n. 1, pp. 72–78, nov. 2009. ISSN: 0960-1643.
- [3] RIDSDALE, L., CHARLTON, J., ASHWORTH, M., et al. “Epilepsy mortality and risk factors for death in epilepsy: a population-based study”, *British Journal of General Practice*, v. 61, n. 586, pp. e271–e278, maio 2011. ISSN: 0960-1643. doi: 10.3399/bjgp11X572463.
- [4] IHLE, M., FELDWISCH-DRENTROP, H., TEIXEIRA, C. “EPILEPSIAE - a European Epilepsy Database”, *Computer Methods and Programs in Biomedicine*, v. 106, n. 3, pp. 127–138, jun. 2012.
- [5] KLATT, J., FELDWISCH-DRENTROP, H., IHLE, M., et al. “The EPILEPSIAE Database: an extensive electroencephalography database of epilepsy patients”, *Epilepsia*, v. 53, n. 9, pp. 1669–1676, jun. 2012.
- [6] IHLE, M., SCHELTER, B., TIMMER, J., et al. “The World’s Largest Epilepsy database: content and structure”, *Recent Advances in Predicting and Preventing Epileptic Seizures*, pp. 278–291, jun. 2013.
- [7] BANDARABADI, M., RASEKHI, J., TEIXEIRA, C. A., et al. “Optimal preictal period in seizure prediction”. In: *IWBBIO*, Granada, Spain, abr. 2014.
- [8] FISHER, R. S., ACEVEDO, C., ARZIMANOGLU, A., et al. “A Practical Clinical Definition of Epilepsy”, *Epilepsia: official journal of the International League Against Epilepsy*, v. 55, n. 4, pp. 475–482, abr. 2014. ISSN: 0013-9580. doi: 10.1111/epi.12550.
- [9] PANAYIOTOPOULOS, C. P. “Reflex Seizures and Reflex Epilepsies”. In: *The Epilepsies - Seizures, syndromes and management*, cap. 1, Oxfordshire, UK, Bladon Medical Publishing, 2005. ISBN: 1-904218-34-2.

- [10] OKUDAN, Z. V., ÖZKARA, C. “Reflex Epilepsy: triggers and management strategies”, *Neuropsychiatric Disease and Treatment*, v. 14, pp. 327–337, jan. 2018.
- [11] WHO. *Epilepsy: a public health imperative*. Geneva, World Health Organization, 2019.
- [12] GUERREIRO, C. A. M., GUERREIRO, M. M., CENDES, F., et al. *Epilepsia*. São Paulo, SP, Brasil, Editora Lemos, 2000. ISBN: 85-7450.
- [13] JR., J. E. “A Proposed Diagnostic Scheme for People with Epileptic Seizures and with Epilepsy”, *Epilepsia: official journal of the International League Against Epilepsy*, v. 42, n. 6, pp. 796–803, jun. 2001. doi: doi:10.1046/j.1528-1157.2001.10401.x.
- [14] BHALLA, D., GODET, B., DRUET-CABANAC, M., et al. “Etiologies of Epilepsy: a comprehensive review”, *Expert Review of Neurotherapeutics*, v. 11, n. 6, pp. 861–876, 2011. doi: 10.1586/ern.11.51. PMID: 21651333.
- [15] FISHER, R. S. “The New Classification of Seizures by the International League Against Epilepsy 2017”, *Current Neurology and Neuroscience Reports*, v. 17, n. 6, abr. 2017.
- [16] GAO, H., SANDER, J. W., XIAO, Y., et al. “A Comparison between the 1981 and 2017 International League Against Epilepsy Classification of Seizure Types based on an Outpatient Setting”, *Epileptic Disorders*, v. 20, n. 4, pp. 257–264, ago. 2018.
- [17] L.LOCKMAN. “Absence, Myoclonic, and Atonic Seizures”, *Pediatric clinics of North America*, v. 36, n. 2, pp. 331–341, abr. 1989. ISSN: 0031-3955. doi: 10.1016/s0031-3955(16)36652-4.
- [18] GOLDSTEIN, J., SLOMSKI, J. “Epileptic Spasms: a variety of etiologies and associated syndromes”, *Journal of Child Neurology*, v. 23, n. 4, pp. 407–414, abr. 2008.
- [19] APPLETON, R. E., PANAYIOTOPOULOS, C. P., ACOMB, B. A., et al. “Eyelid Myoclonia with Typical Absences: an epilepsy syndrome”, *Journal of Neurology, Neurosurgery & Psychiatry*, v. 56, n. 12, pp. 1312–1316, dez. 1993. doi: 10.1136/jnnp.56.12.1312.
- [20] CHEIN, H., KOUBEISSI, M. Z. “Electroencephalography in Epilepsy Evaluation”, *Continuum: Lifelong Learning in Neurology (Minneapolis)*, v. 25, n. 2, pp. 431–453, abr. 2019.

- [21] NOACHTAR, S., RÉMI, J. “The Role of EEG in Epilepsy: A Critical Review”, *Epilepsy & Behavior*, v. 15, pp. 22–33, fev. 2009. doi: 10.1016/j.yebeh.2009.02.035.
- [22] FERNÁNDEZ-TORRE, J. L. “Interictal EEG”. In: *Atlas of Epilepsies*, pp. 701–712, London, Springer London, 2010. ISBN: 978-1-84882-128-6. doi: 10.1007/978-1-84882-128-6_101.
- [23] ABDI, H., WILLIAMS, L. “Principal Component Analysis”, *Interdisciplinary Reviews - Computational Statistics*, v. 2, n. 4, pp. 387–515, jul. 2010.
- [24] LANGE, K. *Numerical Analysis for Statisticians*. New York, NY, Springer New York, 2010. ISBN: 978-1-4419-5944-7. doi: 10.1007/978-1-4419-5945-4_9.
- [25] DUA, D., GRAFF, C. “UCI Machine Learning Repository”. 2017. Disponível em: <<https://archive.ics.uci.edu/ml/datasets/iris>>.
- [26] BASSEVILLE, M., NIKIFOROV, I. V. *Detection of Abrupt Changes: theory and application*. Upper Saddle River, NJ, USA, Prentice-Hall, Inc., abr. 1993.
- [27] GURTIN, M. E. “The Impact of Changing Populations on Classifier Performance”. In: *In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, p. 367–371, California, San Diego, USA, ago. 1999.
- [28] KADWE, Y., SURYAWANSHI, V. “A Review on Concept Drift”, *IOSR Journal of Computer Engineering*, v. 17, n. 1, pp. 20–26, jan. 2015.
- [29] GAMA, J., ZLIOBAITÉ, I., BIFET, A., et al. “A Survey on Concept Drift Adaption”, *ACM Computing Surveys*, v. 46, n. 4, pp. 1–37, abr. 2014. doi: 10.1093/biomet/41.1-2.100.
- [30] NAMITHA, K., KUMAR, G. S. “Learning in the Presence of Concept Recurrence in Data Stream Clustering”, *Journal of Big Data Volume*, v. 7, n. 75, set. 2020. doi: 10.1186/s40537-020-00354-1.
- [31] PAGE, E. S. “Continuous Inspection Schemes”, *Biometrika*, v. 41, n. 1–2, pp. 100–115, jun. 1954. doi: 10.1145/2523813.
- [32] SANIGA, E. M., DAVIS, D., MCWILLIAMS, T., et al. “Statistical CUSUM Designs with Minimum Sampling Cost”, *Quality Technology & Quantitative Management*, v. 15, n. 4, pp. 475–483, 2018. doi: 10.1080/16843703.2016.1226708.

- [33] SAKAMOTO, Y., FUKUI, K., GAMA, J., et al. “Concept Drift Detection with Clustering via Statistical Change Detection Methods”. In: *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*, pp. 37–42, 2015. doi: 10.1109/KSE.2015.19.
- [34] NAMITHA, K., KUMAR, G. S. “Learning in the Presence of Concept Recurrence in Data Stream Clustering”, *Journal of Big Data*, v. 7, n. 75, set. 2020. doi: 10.1186/s40537-020-00354-1.
- [35] GAMA, J., MEDAS, P., CASTILLO, G., et al. “Learning with Drift Detection”. In: *Advances in Artificial Intelligence – SBIA 2004, 17th Brazilian Symposium on Artificial Intelligence*, pp. 286–295, São Luís, Maranhão, Brazil, set. 2004.
- [36] BIFET, A., GAVALDÀ, R. “Learning from Time-Changing Data with Adaptive Windowing”. In: *Seventh SIAM International Conference on Data Mining*, pp. 443–448, Minneapolis, Minnesota, USA, abr. 2007.
- [37] BIFET, A., HOLMES, G., PFAHRINGER, B., et al. “Fast Perceptron Decision Tree Learning from Evolving Data Streams”. In: *M. Zaki, J. Yu, B. Ravindran, V. Pudi (Eds.), Advances in Knowledge Discovery and Data Mining, Vol. 6119 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg*, pp. 299–310, Minneapolis, Minnesota, USA, 2010. Disponível em: http://dx.doi.org/10.1007/978-3-642-13672-6_30.
- [38] PRATT, J., GIBBONS, J. D. “Kolmogorov-Smirnov Two-Sample Tests”. In: *Concepts of Nonparametric Theory*, pp. 318–344, New York, NY, Springer New York, 1981. ISBN: 978-1-4612-5931-2. doi: 10.1007/978-1-4612-5931-2_7. Disponível em: https://doi.org/10.1007/978-1-4612-5931-2_7.
- [39] PIANOSI, F., WAGENER, T. “A Simple and Efficient Method for Global Sensitivity Analysis based on Cumulative Distribution Functions”, *Environmental Modelling Software*, v. 67, pp. 1–11, maio 2015. ISSN: 1364-8152. doi: <https://doi.org/10.1016/j.envsoft.2015.01.004>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364815215000237>.
- [40] DREW, J. H., GLEN, A. G., LEEMIS, L. M. “Computing the Cumulative Distribution Function of the Kolmogorov–Smirnov Statistic”, *Computational Statistics Data Analysis*, v. 34, n. 1, pp. 1–15, jul. 2000. ISSN:

0167-9473. doi: [https://doi.org/10.1016/S0167-9473\(99\)00069-9](https://doi.org/10.1016/S0167-9473(99)00069-9). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167947399000699>>.

[41] KEOGH, E., MUEEN, A. “Curse of Dimensionality”. In: *Encyclopedia of Machine Learning and Data Mining*, pp. 314–315, Boston, MA, Springer US, 2017. ISBN: 978-1-4899-7687-1. doi: 10.1007/978-1-4899-7687-1_192. Disponível em: https://doi.org/10.1007/978-1-4899-7687-1_192>.

[42] LIBERTI, L., LAVOR, C., MACULAN, N., et al. “Euclidean Distance Geometry and Applications”, *SIAM Review*, v. 56, n. 1, pp. 3–69, 2014. doi: 10.1137/120875909. Disponível em: <https://doi.org/10.1137/120875909>>.