

BRUNO BURINI ROBLES ARINE

**MELHORAMENTO DO ÍNDICE DE DETECÇÕES NA
ESPECTROMETRIA GAMA EM AMOSTRAS AMBIENTAIS
USANDO INTELIGÊNCIA ARTIFICIAL**

Sorocaba

2017

BRUNO BURINI ROBLES ARINE

**MELHORAMENTO DO ÍNDICE DE DETECÇÕES NA
ESPECTROMETRIA GAMA EM AMOSTRAS AMBIENTAIS
USANDO INTELIGÊNCIA ARTIFICIAL**

Dissertação apresentada como requisito para a obtenção do título de Mestre em Ciências Ambientais da Universidade Estadual Paulista “Júlio de Mesquita Filho” na Área de Diagnóstico, Tratamento e Recuperação Ambiental

Orientador: Prof. Dr. Antonio Cesar Germano Martins

Sorocaba

2017

PROGRAMA DE PÓS-GRADUAÇÃO em

ciências
ambientais



Ficha catalográfica elaborada pela Biblioteca da Unesp
Instituto de Ciência e Tecnologia – Câmpus de Sorocaba

Arine, Bruno Burini Robles.

Melhoramento do índice de detecções na espectrometria gama em amostras ambientais usando inteligência artificial / Bruno Burini Robles Arine, 2017.

151 f.: il.

Orientador: Antonio Cesar Germano Martins.

Dissertação (Mestrado) – Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Ciência e Tecnologia (Câmpus de Sorocaba), 2017.

1. Inteligência artificial. 2. Espectrometria de raio gama. 3. Raios gama. 4. Meio ambiente. I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Ciência e Tecnologia (Câmpus de Sorocaba). II. Título.

AGRADECIMENTOS

Os agradecimentos são direcionados a:

Prof. Dr. Antonio Cesar Germano Martins, pela excelente orientação, sugestões, correções, paciência em debater idéias até horas depois do expediente, por todo conhecimento que me foi passado e, acima de tudo, pela amizade;

Prof. Dr. Marco Antonio Proença Vieira de Moraes, por transmitir seu conhecimento e experiência de vários anos com espectrometria;

Ana Lúcia, minha esposa, pelo amor e paciência sem limites, por servir de inspiração acadêmica para que eu trilhasse mais esta etapa da minha vida e pela inestimável ajuda na revisão deste texto;

Bárbara Eliodora Leite e Beatriz Nunes Lambiasi, pela indispensável ajuda e colaboração no laboratório de Radiometria;

Marcelo Sartoratto, pelo apoio técnico na preparação dos padrões necessários para o estudo;

Nádia Segre e Ana Cláudia Ferreira Sobral, pelas revisões de última hora;

Toda equipe do Laboratório Radioecológico do Centro Experimental Aramar;

Centro Tecnológico da Marinha, por viabilizar meu crescimento acadêmico;

Universidade Estadual Paulista "Júlio de Mesquita Filho", pela estrutura e ensino de qualidade;

Meus pais, Maria de Lourdes e Djair, por acreditarem em mim e apoiarem minhas escolhas, e por fazerem o impossível para que eu pudesse desfrutar de uma formação de qualidade;

Ruth e Júnior, pela hospitalidade e torcida em todas as horas;

Meu irmão, João Paulo, e meus amigos Vitor, Juliana, Jorge, Carla, Isaías e Jéssica, pelo companheirismo, viagens e momentos de descontração quando precisei.

*We make our world significant by the courage of our questions
and the depth of our answers.
(Carl Sagan)*

RESUMO

A utilização de elementos radioativos se estende hoje a diversos ramos da ciência e tecnologia, exigindo maneiras eficientes, precisas e sustentáveis de monitoramento para a preservação ambiental e a saúde humana. Uma técnica eficiente e não-destrutiva de análise de elementos emissores de raios gama é a espectrometria gama. No entanto, os softwares atuais de análise se deparam com certos empecilhos quando tratam de amostras ambientais, cujas concentrações de radioatividade são próximas do limite de detecção, ou quando certos radionuclídeos de interesse estão sujeitos a interferências. Este trabalho desenvolveu algoritmos de análise através de ferramentas do campo da inteligência artificial, de forma a obter um desempenho superior do que os métodos empregados em softwares clássicos de análise radiométrica. Três classificadores do campo da inteligência artificial foram testados; árvores de decisão (AD), máquinas de vetores de suporte (MVS) e redes neurais artificiais (RNA). Em testes com uma fonte de chumbo-210 de baixa atividade, a MVS obteve acurácia de 0,93, enquanto os métodos tradicionais obtiveram acurácia máxima de 0,73. Em testes com amostras reais de nitrato de uranila, a acurácia de classificação da RNA foi de 0,91, enquanto a acurácia dos métodos tradicionais foi de 0,70. Observou-se que os modelos baseados em inteligência artificial tiveram desempenho superior aos métodos tradicionais em todos os experimentos. Entretanto, esta vantagem diminui à medida que a relação sinal-ruído dos espectros aumenta, se tornando negligível quando a relação sinal-ruído supera o limite de detecção.

Palavras-chave: radiação gama, meio ambiente, inteligência artificial, aprendizado de máquina

ABSTRACT

Usage of radioactive elements is present in several areas of science and technology, requiring efficient, accurate, and sustainable ways of environmental conservation and human health monitoring. Gamma-ray spectrometry is an efficient and non-destructive analysis technique for identification of gamma-ray emitting elements. Nonetheless, current analysis softwares have certain limitations regarding environmental samples, whose radioactivity concentration is often close to the detection limit, or when certain radionuclides of interest are subject to interference. This work aims the development of analysis algorithms based on the field of artificial intelligence in order to achieve better performance than the methods in classical radiometric analysis software. Three artificial intelligence classifiers were tested; decision trees (DT), support vector machines (SVM) and artificial neural networks (ANN). In tests with a low-activity lead-210 source, SVM obtained an accuracy of 0.93, while the traditional methods obtained a maximum accuracy of 0.73. In tests with real samples of uranyl nitrate, the accuracy of ANN classification was 0.91, while the accuracy of the traditional methods was 0.70. It was observed that artificial intelligence-based models performed better than traditional methods in all experiments. However, this advantage decreases as the signal-to-noise ratio of the spectra increases, becoming negligible when the signal-to-noise ratio exceeds the detection limit.

Keywords: gamma radiation, environment, artificial intelligence, machine learning

LISTA DE ILUSTRAÇÕES

Figura 1 – Arranjo experimental de um sistema de espectrometria gama	26
Figura 2 – Curvas de eficiência de um detector HPGe para amostras em diferentes geometrias	28
Figura 3 – Pico de 40K obtido em espectrômetro gama HPGe	28
Figura 4 – Exemplo de resolução dos picos em função da energia segundo programa Genie 2000	29
Figura 5 – Comparação entre espectros obtidos por detectores NaI e HPGe	30
Figura 6 – Espectro gama de 100 g de sedimento contado por 3 horas em detector HPGe	31
Figura 7 – Segunda derivada calculada sobre a função gaussiana	34
Figura 8 – Exemplo de convolução de um espectro	34
Figura 9 – Teste de hipótese	38
Figura 10 – Esquema geral de um modelo de IA	40
Figura 11 – Cenários de generalização de um classificador hipotético	42
Figura 12 – Definição de uma matriz de confusão	43
Figura 13 – Exemplo de curva ROC	45
Figura 14 – Representação de um neurônio biológico	46
Figura 15 – Exemplo de uma única unidade em uma RNA	47
Figura 16 – Função de ativação do tipo sigmoidal	48
Figura 17 – Exemplo de rede neural com mais de uma camada de neurônios	48
Figura 18 – Comparação entre função sigmoidal (esq.) e tangente hiperbólica (dir.)	52
Figura 19 – Exemplo de divisão de classes com um hiperplano	55
Figura 20 – Exemplo de vetores de suporte	58
Figura 21 – Transformação dos dados não-linearmente separáveis	59
Figura 22 – Representação geométrica das variáveis de folga	62
Figura 23 – Exemplo de árvore de decisão para a pergunta devo viajar?	63
Figura 24 – Estrutura genérica de uma árvore de decisão	64
Figura 25 – Relação entre número de nós de uma árvore e erros de classificação	66
Figura 26 – Uma RNA utilizada para identificar radioisótopos	69
Figura 27 – Região do espectro composta por raios-X e gama provenientes de U e Th	70
Figura 28 – Cadeia de decaimento do ^{238}U , exibida até o ^{222}Rn	75
Figura 29 – Etapas de tratamento dos dados, incluindo pré-processamento, extração de parâmetros e redução de dimensionalidade	82
Figura 30 – Validação cruzada do tipo k-dobras sob regime 10x10	85

Figura 31 – Exemplo de espectro artificialmente gerado	87
Figura 32 – Localização da emissão energética da fonte de ^{210}Pb nos espectros adquiridos	89
Figura 33 – Disposição das fontes radioativas no detector	90
Figura 34 – Aparência média dos espectros	91
Figura 35 – Disposição da amostra de urânio no detector	92
Figura 36 – Espectro de uma amostra contendo urânio em proporção isotópica ambiental	94
Figura 37 – Aparência média dos grupos de espectros do cenário A	97
Figura 38 – Comparação entre dois espectros arbitrariamente selecionados	98
Figura 39 – Curva de aprendizado dos modelos de IA em espectros do cenário A	99
Figura 40 – Curvas ROC dos modelos de IA e métodos tradicionais testados em espectros artificiais (Cenário A)	100
Figura 41 – Sumário dos valores de aROC para os diferentes métodos testados em espectros artificiais	103
Figura 42 – Aparência média dos grupos de espectros SNR-BAIXA, SNR-MEDIA e SNR-ALTA	104
Figura 43 – Curvas ROC dos modelos testados em SNR-BAIXA	106
Figura 44 – Matriz de confusão dos modelos testados em SNR-BAIXA	106
Figura 45 – Curvas ROC dos modelos testados em SNR-MEDIA	107
Figura 46 – Curvas ROC dos modelos testados em SNR-ALTA	108
Figura 47 – Sumário dos valores de aROC para os diferentes métodos testados em espectros de ^{210}Pb	108
Figura 48 – Distribuição das contagens brutas na região de interesse dos espectros SNR-BAIXA	110
Figura 49 – Curva de calibração de FWHM em função da energia pelo Genie 2000	111
Figura 50 – Curvas ROC dos modelos testados em espectros de nitrato de uranila	112
Figura 51 – Região de interesse do espectro de urânio	113
Figura 52 – Resultados de diferentes métodos de correção do contínuo aplicados a um espectro de nitrato de uranila	132
Figura 53 – Operador de Ricker	133
Figura 54 – Exemplo de transformada wavelet em um espectro arbitrário	134
Figura 55 – Média dos espectros artificiais utilizados no teste de extração de parâmetros de entrada	135
Figura 56 – Exemplo de parâmetros de entrada dos modelos antes do pré-processamento	137
Figura 57 – Acurácia do modelo em função do número de parâmetros de entrada	138
Figura 58 – Relevância dos parâmetros de entrada	138

Figura 59 – Acurácia de ADs com diversos valores de profundidade e mínimo de amostras para particionamento	140
Figura 60 – Curvas ROC das ADs com diferentes critérios de avaliação de impureza nodal	140
Figura 61 – Desempenho da MVS de acordo com o tipo de núcleo	141
Figura 62 – Curva de aprendizado da MVS em função do tipo de núcleo	142
Figura 63 – Acurácia da MVS com núcleo gaussiano para diversos hiperparâmetros	143
Figura 64 – Erro na validação da MVS em função do valor de C	143
Figura 65 – Curva de aprendizado da RNA	144
Figura 66 – Exemplo de segunda diferença generalizada aplicada a um espectro gama	148
Figura 67 – Probabilidade de detecção de um pico com limiar igual a -3	149
Figura 68 – Exemplo de espectro cujo contínuo foi subtraído pelo método de erosão	150

LISTA DE TABELAS

Tabela 1 – Concentração de atividade de ^{226}Ra em água e sedimento em certas regiões do mundo	36
Tabela 2 – Métricas	44
Tabela 3 – Núcleos frequentemente utilizados	61
Tabela 4 – Fórmulas de avaliação de impureza	65
Tabela 5 – Vantagens e desvantagens dos métodos mais utilizados para análise de urânio	73
Tabela 6 – Probabilidade de emissões gama na região de 92 a 96 keV	76
Tabela 7 – Modelos de IA e hiperparâmetros utilizados neste trabalho	81
Tabela 8 – Estatísticas utilizadas para geração de espectros artificiais	86
Tabela 9 – Características dos grupos de espectros que foram adquiridos	90
Tabela 10 – Concentração de urânio nas amostras preparadas	92
Tabela 11 – Limites de detecção de urânio total na geometria e tempo de contagem utilizada para este experimento	94
Tabela 12 – Desempenho dos modelos no Cenário A (espectros de contínuo e SNR constantes)	98
Tabela 13 – Limiares otimizados dos métodos tradicionais	100
Tabela 14 – Desempenho dos modelos no Cenário A (espectros de contínuo e SNR constantes, com otimização dos limiares dos métodos tradicionais)	100
Tabela 15 – Métricas de desempenho no cenário B (espectros de contínuo variável)	101
Tabela 16 – Desempenho dos modelos no Cenário B (espectros de contínuo e SNR constantes, com otimização dos limiares dos métodos tradicionais)	102
Tabela 17 – Desempenho dos modelos no Cenário C (contínuo inédito e SNR constante)	102
Tabela 18 – Desempenho dos modelos no cenário D (espectros inéditos de contínuo variável)	103
Tabela 19 – Média das contagens totais na região de interesse do grupo SNR-BAIXA	104
Tabela 20 – Tempo de execução dos modelos de aprendizado de máquina	105
Tabela 21 – Métricas de desempenho dos métodos tradicionais de detecção em SNR-BAIXA	105
Tabela 22 – Métricas de desempenho dos métodos tradicionais de detecção em SNR-MEDIA	107
Tabela 23 – Métricas de desempenho dos métodos tradicionais de detecção em SNR-ALTA	109

Tabela 24 – Desempenho dos modelos de IA e dos métodos tradicionais na identificação de amostras com concentração igual ou maior que 300 ug L ⁻¹ de urânio	112
Tabela 25 – Desempenho de uma MVS com diferentes métodos de correção do contínuo de fundo dos espectros	131
Tabela 26 – Desempenho dos modelos de IA com métodos diversos de extração de parâmetros de entrada	135
Tabela 27 – Desempenho da MVS com diversos tipos de núcleo	141

LISTA DE ABREVIATURAS E SIGLAS

AD	Árvore de decisão
AM	Aprendizado de máquina
aROC	<i>Area under the ROC curve</i> (Área sob a curva ROC)
Bq	Becquerel (unidade de medida de atividade radioativa)
CONAMA	Conselho Nacional do Meio Ambiente
CorrBib	Método de correlação com biblioteca
Deriv	Método derivativo
FA	Florestas aleatórias
FWHM	<i>Full width at half maximum</i> (Largura à meia altura)
HPGe	<i>High-purity germanium</i> (Gêrmanio de alta pureza)
L_C	Limite crítico
L_D	Limite de detecção
IA	Inteligência artificial
IAEA	<i>International Atomic Energy Agency</i> (Agência Internacional de Energia Atômica)
ISO	<i>International Organization for Standardization</i> (Organização Internacional para Padronização)
IUPAC	<i>International Union of Pure and Applied Chemistry</i> (União Internacional de Química Pura e Aplicada)
MVS	Máquina de vetores de suporte
P_{AF}	Probabilidade de alarmes falsos
P_D	Probabilidade de detecção
RNA	Rede neural artificial
ROC	<i>Receiver Operating Characteristic</i> (Característica de Operação do Receptor)

SNR	Relação sinal-ruído
TW	Transformada wavelet
TFN	Taxa de falso-negativos
TFP	Taxa de falso-positivos
TVN	Taxa de verdadeiro-negativos
TVP	Taxa de verdadeiro-positivos

SUMÁRIO

1	INTRODUÇÃO	23
2	REVISÃO TEÓRICA	25
2.1	Espectrometria gama	25
2.1.1	Princípios físicos	25
2.1.2	Aquisição do espectro	27
2.1.3	Cálculo de atividade	31
2.1.4	Detecção dos picos	32
2.1.4.1	Rotina de correlação com biblioteca	32
2.1.4.2	Rotina de varredura por métodos derivativos	33
2.1.5	Detecção de picos com baixa relação sinal-ruído	35
2.1.6	Limites mínimos de detecção	36
2.1.6.1	Teste de hipótese pelo limite crítico (ou limite de decisão)	37
2.2	Inteligência artificial	39
2.2.1	Validação dos modelos	41
2.2.2	Subajuste e sobreajuste	42
2.2.3	Métricas de avaliação	42
2.2.3.1	Acurácia	43
2.2.3.2	Área sob curva ROC	44
2.2.4	Redes neurais artificiais	46
2.2.4.1	Visão geral	46
2.2.4.2	Definição	47
2.2.4.3	Treinamento por retro-propagação	49
2.2.4.4	Topologia da rede	51
2.2.5	Máquinas de vetores de suporte	53
2.2.5.1	Definição	54
2.2.5.2	Solução pelo método de Lagrange	55
2.2.5.3	Conjuntos não-linearmente separáveis	59
2.2.5.4	Margens suaves	62
2.2.6	Árvores de decisão	63
2.2.6.1	Definição	64
2.2.6.2	Treinamento	64
2.2.6.3	Soluções para sobreajuste	66
2.2.7	Inteligência artificial aplicado a espectrometria gama	67
2.3	Análise de urânio em efluentes	71

2.3.1	Problemas Ambientais	71
2.3.2	Métodos analíticos para determinação de urânio	73
2.3.3	Espectrometria gama para a análise de urânio	74
2.3.4	Desafios da espectrometria gama para análise de urânio	75
3	OBJETIVOS	79
3.1	Objetivos específicos	79
4	MATERIAIS E MÉTODOS	81
4.1	Implementação dos modelos de aprendizado de máquina	81
4.1.1	Pré-processamento	82
4.2	Métodos tradicionais de detecção	83
4.2.1	Método derivativo	83
4.2.2	Método de correlação por biblioteca	84
4.3	Validação cruzada	84
4.4	Experimento com espectros artificiais	84
4.4.1	Cenário A: Contínuo constante, SNR constante	86
4.4.2	Cenário B: Contínuo variado, SNR variado, picos espúrios	86
4.4.3	Cenário C: Contínuo e SNR com valores inéditos	87
4.4.4	Cenário D: Contínuo variável, SNR maiores	88
4.5	Teste com espectros reais de ²¹⁰Pb	88
4.6	Checagem de estabilização	89
4.7	Aquisição dos espectros	89
4.8	Teste com amostras de urânio	91
4.8.1	Preparação das amostras	92
4.8.2	Região de interesse do espectro	93
4.8.3	Treinamento dos modelos de IA	93
4.8.4	Algoritmos tradicionais de análise	95
5	RESULTADOS E DISCUSSÃO	97
5.1	Testes com espectros artificiais	97
5.1.1	Cenário A: Contínuo constante, SNR constante	97
5.1.2	Cenário B: Contínuo variado, SNR variado, picos espúrios	101
5.1.3	Cenário C: Contínuo inédito	101
5.1.4	Cenário D: Contínuo variável, SNR maiores	102
5.2	Testes com ²¹⁰Pb	104
5.2.1	Sumário dos espectros	104
5.2.2	Teste dos modelos	105
5.2.3	Análise dos resultados	109
5.3	Teste com amostras de urânio	111

6	CONCLUSÃO	115
	REFERÊNCIAS	117
	APÊNDICES	129
	APÊNDICE A – PRÉ-PROCESSAMENTO	131
A.1	Correção do contínuo	131
A.2	Extração de novos parâmetros	133
A.3	Centralização	134
A.4	Seleção de parâmetros	136
	APÊNDICE B – OTIMIZAÇÃO DOS HIPERPARÂMETROS	139
B.1	Árvores de decisão	139
B.2	Máquinas de vetores de suporte	139
B.3	Redes neurais artificiais	144
	ANEXOS	145
	ANEXO A – ALGORITMOS TRADICIONAIS DE DETECÇÃO	147
A.1	Segunda Diferença Generalizada	147
A.2	Identificação de núclídeos por correlação com biblioteca	149

1 INTRODUÇÃO

O homem sempre esteve em contato com a radiação ionizante: certos isótopos radioativos (também chamados de radionuclídeos) sintetizados durante a explosão de estrelas pela galáxia estão presentes na crosta terrestre desde sua formação; raios cósmicos de alta energia do espaço que incidem sobre o planeta não somente nos proporcionam uma considerável dose de radiação ionizante, como também são responsáveis pela constante criação de radionuclídeos na alta atmosfera, como o ^{14}C e o ^7Be (BODANSKY, 2007). Em média, aproximadamente 85% da dose radioativa recebida pela população é oriunda da ingestão de alimentos, do contato com o solo e do próprio ar; o restante da dose tem origem cósmica (UNITED NATIONS, 2010).

Após milênios de um delicado e constante equilíbrio isotópico natural, ações antrópicas aumentaram direta e indiretamente a concentração de radionuclídeos no meio ambiente. Indiretamente, certos processos industriais podem concentrar radionuclídeos que estão presentes naturalmente na matéria-prima, podendo até mesmo expor trabalhadores e indivíduos do público à radiação além de limites regulatórios (SZELES, 1990); diretamente, o uso desenfreado de armamentos atômicos e a imprudência quanto à segurança de instalações nucleares despejou consideráveis concentrações de radionuclídeos artificiais na natureza, sendo estes apenas sintetizados em aceleradores de partículas, assim como em processos de fissão que ocorrem em reatores e explosões nucleares.

Porém, mais de um século após o descobrimento da radioatividade por Henri Becquerel em 1896 (CROWTHER, 1948), o seu uso controlado e bem pesquisado desempenha um importante papel no desenvolvimento humano. Radionuclídeos naturais e antrópicos foram compreendidos e hoje nos fornecem ferramentas para que possamos levar vidas mais produtivas e duradouras, com grande destaque em campos do conhecimento tão diversos quanto o diagnóstico clínico, as ciências biológicas, a produção de energia, o monitoramento ambiental e a segurança nuclear. A exatidão nas medidas de atividade de radionuclídeos é hoje uma condição indispensável para o melhor entendimento do ambiente onde vivemos e para o progresso de diversos processos tecnológicos e científicos. Como exemplo, pode-se citar o uso de materiais radioativos para tratamentos médicos, esterilização de instrumentos, sanitização de alimentos, geração de energia e o monitoramento de processos e materiais (FORD, 2004).

Ao mesmo tempo em que a radioatividade nos agraciou com meios de melhorar a qualidade de vida, ela também nos mostrou quão importante é assegurar seu manuseio de forma cautelosa e pacífica em uma escala global. Visando minimizar o risco de exposição humana a elementos radioativos e à radiação ionizante, uma organização internacional

autônoma, a *International Atomic Energy Agency* (IAEA), foi criada em 1957 para nortear os países detentores de tecnologia nuclear em relação a diretrizes e normas de segurança (FISCHER, 1997), bem como oferecer auxílio aos laboratórios que pretendem aprimorar sua qualidade através de programas de treinamento, inter-comparação laboratorial e materiais de referência.(INTERNATIONAL ATOMIC ENERGY AGENCY, 2015)

Dentre as diretrizes da IAEA, ficou estabelecido que a indústria de energia, organizações militares, centros de pesquisa, hospitais e indústrias em geral, uma vez gerando descargas com concentrações de radionuclídeos de qualquer significância, estão sujeitas a controle regulatório; ou seja, precisam de autorização para funcionamento e devem, obrigatoriamente, ser monitoradas ou estabelecer seus próprios programas de monitoramento ambiental (FORD, 2004).

Portanto, medições precisas de radioatividade permanecem vitais para o uso correto de materiais nucleares e na prevenção contra seus potenciais efeitos deletérios, buscando-se a prevenção de acidentes. Considerando-se o monitoramento ambiental, é importante que se possa detectar concentrações baixas de radionuclídeos, sendo necessária, desta forma, a constante busca por métodos analíticos cada vez mais sensíveis e precisos (INTERNATIONAL ATOMIC ENERGY AGENCY, 1989).

2 REVISÃO TEÓRICA

2.1 Espectrometria gama

Espectrometria gama é a técnica de identificação ou quantificação de radionuclídeos pela análise do espectro de energia produzido por um espectrômetro de radiação gama, sendo este usado em diversas atividades como monitoramento ambiental, medicina nuclear, monitoramento de corrosão de reatores, forense nuclear, teste de materiais, mineralogia, e monitoramento de processos industriais (ORTEC, 2014a).

A Figura 1 apresenta uma montagem típica de espectrometria gama, onde o sistema de detecção se baseia na interação entre raios gama originários da amostra e o material do detector. Tal interação dá origem à produção de cargas elétricas que são adquiridas por uma eletrônica apropriada e geram um espectro com sinais de diferentes amplitudes e são diretamente interpretados por uma eletrônica apropriada.

Ao que se refere à análise de radionuclídeos em amostras ambientais, a espectrometria gama é um dos métodos analíticos mais amplamente utilizados (ZIEGLER, 2002). Caso necessário, a análise pode ser feita de forma não-destrutiva (isto é, isenta de pre-processamentos físico-químicos que modifiquem estruturalmente a amostra) e permite avaliações qualitativas ou quantitativas a respeito de radionuclídeos tanto naturais quanto antrópicos (L'ANNUNZIATA, 2003). Segundo Buchtela (2014), esta técnica é uma escolha viável em análises ambientais se:

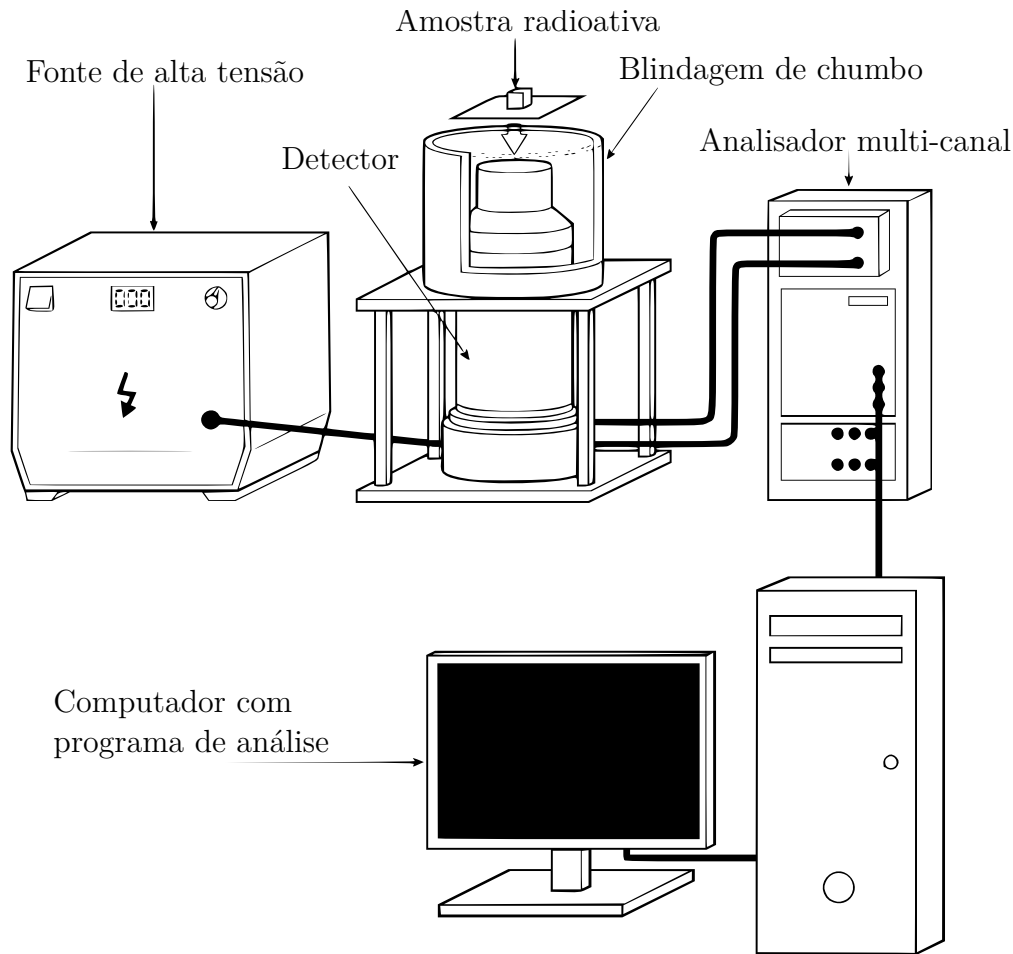
- procedimentos de separação radioquímica são indesejáveis devido a seu alto custo e complexidade e;
- se os elementos a serem analisados emitem radiação gama em abundância apropriada para a técnica e dentro dos limites de energia do detector.

2.1.1 Princípios físicos

A espectrometria gama é possível devido às consequências da interação entre os raios gama originários da amostra e certos materiais especiais. São dois os tipos de detectores gama utilizados para espectrometria: cintiladores sólidos e semicondutores (KNOLL, 2010).

O princípio básico dos cintiladores sólidos é a utilização de um material capaz de emitir luz sob incidência de radiação (L'ANNUNZIATA, 2003). O material cintilante mais comumente utilizado neste tipo de instrumento é o cristal de iodeto de sódio dopado com tálio ou NaI(Tl) (PEREZ-ANDUJAR; PIBIDA, 2004). Os fótons produzidos pelo material

Figura 1 – Arranjo experimental de um sistema de espectrometria gama



Fonte: Mustoe-Playfair (2011)

cintilante são direcionados para um tubo fotomultiplicador. Este tubo é capaz de produzir elétrons na incidência de luz sobre sua superfície. Os fotoelétrons, como são chamados, são direcionados por uma série de dínodos através da aplicação de alta voltagem (KNOLL, 2010). Os dínodos são placas de um material especial cuja incidência de um elétron produz uma cascata de outros elétrons. Estes novos elétrons incidem sobre os próximos dínodos, e assim sucessivamente. Ao final do processo, os elétrons produzidos são coletados por um ânodo e interpretados por instrumentos digitais.

Os detectores gama que utilizam semicondutores funcionam através da ionização destes materiais sob incidência de raios gama (L'ANNUNZIATA, 2003). Diversos cristais são utilizados como semicondutores, sendo o germânio de alta pureza (*high-purity germanium* ou HPGe) o material mais utilizado (KNOLL, 2010). Quando a radiação gama oriunda de uma amostra é absorvida por estes cristais semicondutores, um par elétron-lacuna é formado. Estes pares são transportados sob um campo elétrico externo para os eletrodos onde um pulso é gerado com amplitudes que são diretamente dependentes do tipo de interação (efeito fotoelétrico, espalhamento Compton ou produção de pares) e a

energia do fóton incidente. Tais cargas elétricas são amplificadas, digitalizadas, filtradas e interpretadas pelo software do computador (L'ANNUNZIATA, 2003).

2.1.2 Aquisição do espectro

A primeira etapa da cadeia de análise na espectrometria gama é a aquisição do espectro. Após a amostra ser posicionada nas proximidades do detector, aos sinais gerados pelo sistema são atribuídos valores de energia que são posteriormente registrados por um analisador digital multicanal. Cada evento é registrado em um canal de energia correspondente. O espectro gama é efetivamente um histograma de contagens de emissões gama que foram observadas pelo sistema (GILMORE, 2008). Usualmente, detectores HPGe são capazes de detectar raios gama de 30 a 3000 keV, que são registrados ao longo de espectros com 8192 canais (CANBERRA, 2003).

A eficiência na detecção de raios gama (isto é, razão entre eventos observados e eventos ocorridos) não é constante ao longo do espectro de energia de radiação. A atenuação da radiação gama incidente no material do detector é diretamente dependente da o detector¹, número atômico dos elementos que compõem a amostra e energia da radiação incidente (KNOLL, 2010). Desta forma, contagens de amostras de diferentes materiais, geometrias e volumes devem ser corrigidas através de diferentes curvas de eficiência (Figura 2).

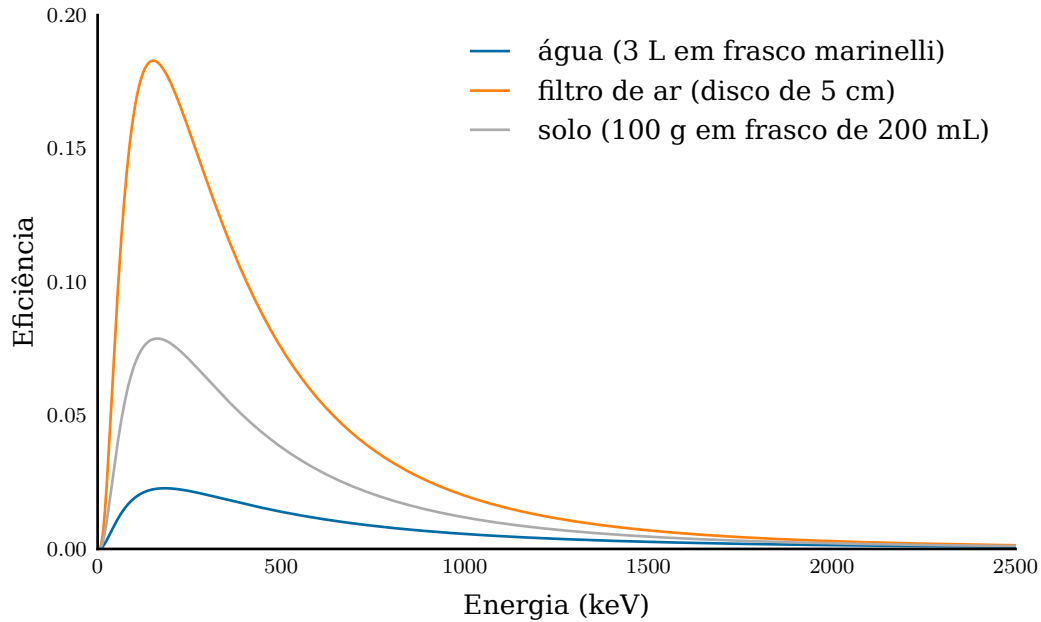
O tempo de aquisição (também chamado de tempo de contagem) é escolhido de acordo com a estatística do evento que se deseja observar, levando em consideração a eficiência do detector para aquele tipo de amostra, radionuclídeos de interesse, interferências e limites de detecção determinados pelo laboratório ou pela legislação vigente.

Devido à atividade dos elementos radioativos na amostra, os registros das emissões gama que são observadas pelo detector formarão picos de contagens. O centroide destes picos, isto é, a energia média das emissões daquelas regiões com maior concentração de contagens, corresponde à energia de emissão dos elementos radioativos (L'ANNUNZIATA, 2003).

Em um sistema hipoteticamente ideal, onde a energia dos raios gama fosse totalmente absorvida pelo material do detector e onde não houvesse interferências eletrônicas, as emissões registradas no espectro se assemelhariam a linhas (KNOLL, 2010). No entanto, devido a propriedades da estrutura da matéria e imperfeições nos componentes eletrônicos do detector, as contagens das emissões são dispersas, produzindo picos que se assemelham a curvas gaussianas com um suave caudeamento (*tailing*) no lado de menor energia. Um exemplo pode ser observado na Figura 3.

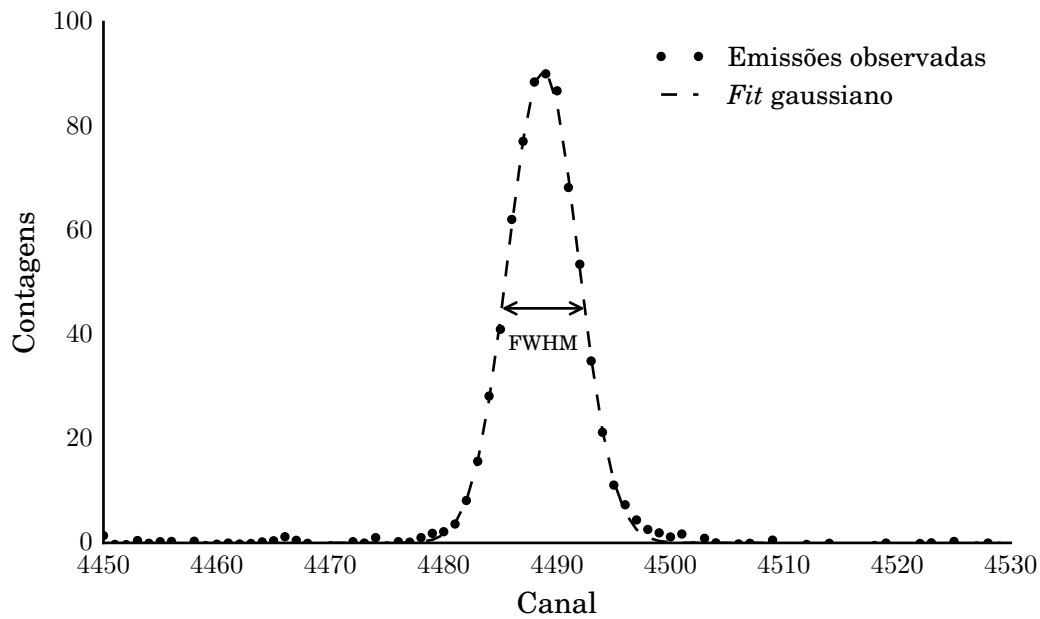
¹ Este efeito ocorre devido à atenuação causada pela atmosfera que permeia o detector e a amostra. No vácuo, a redução da eficiência em relação à distância estaria correlacionada com a abertura do ângulo do feixe de emissões e não com a atenuação (JÄDERSTRÖM et al., 2015).

Figura 2 – Curvas de eficiência de um detector HPGe para amostras em diferentes geometrias



Fonte: autoria própria

Figura 3 – Pico de 40K obtido em espectrômetro gama HPGe



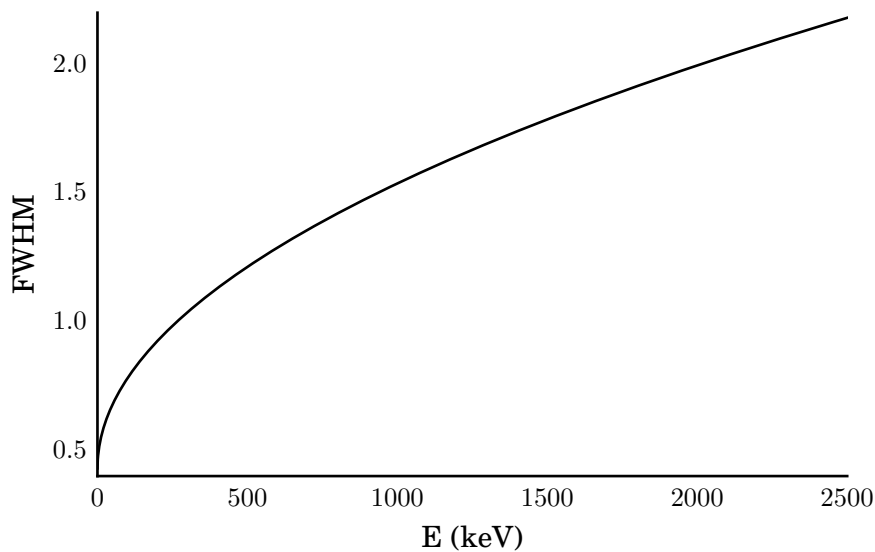
Fonte: autoria própria

A resolução de um pico é normalmente avaliada no campo da espectrometria gama através de sua largura à meia altura (*full width at half maximum* ou FWHM), que está relacionada à largura σ de uma curva gaussiana de tal forma que $\text{FWHM} \approx 2,355 \sigma$. O FWHM dos picos não é constante, pois a produção e coleção de cargas, assim como ruído eletrônico, é diretamente dependente da energia da emissão gama (GILMORE, 2008). O programa Genie 2000 (CANBERRA, 2006) define a resolução dos picos em função da energia como:

$$\text{FWHM} = a + b\sqrt{E}, \quad (2.1)$$

onde a e b são parâmetros arbitrários característicos do sistema e E é a energia. O resultado de uma curva gerada pela Equação 2.1 é mostrado na Figura 4.

Figura 4 – Exemplo de resolução dos picos em função da energia segundo programa Genie 2000



Fonte: autoria própria

Esta equação foi criticada por Gilmore (2008), afirmando que a equação sugerida por Debertin e Helmer (1988) gera resultados mais precisos por levar em conta que os parâmetros relacionados à produção e coleção de cargas no sistema eletrônico do detector variam de forma quadrática:

$$\text{FWHM} = \sqrt{a^2 + b^2 E} \quad (2.2)$$

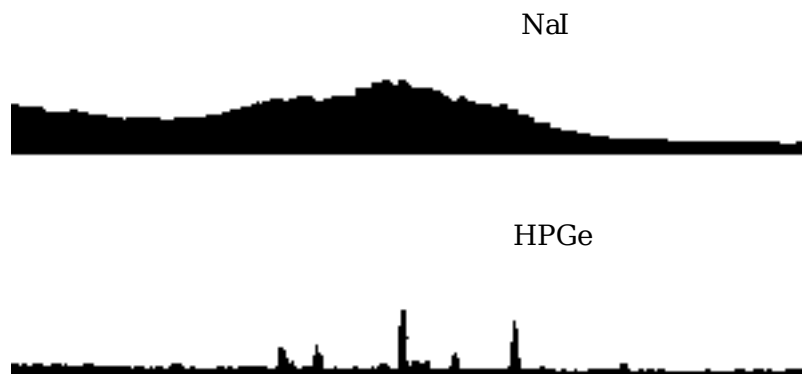
Ainda, Gilmore (2008) concluiu que um termo adicional à equação de Debertin e Helmer (1988) deveria ser considerado, este relativo ao ruído eletrônico. Sua hipótese foi testada com êxito em espectros de referência fornecidos por programas de inter-

comparação promovidos pela IAEA (GILMORE, 2008). Sendo assim, a forma completa da curva de FWHM em função da energia é dada por:

$$\text{FWHM} = \sqrt{a^2 + b^2 E + c^2 E^2} \quad (2.3)$$

Para a produção de um único fóton no detector de cintilação sólida cujo material cintilante é o iodeto de sódio (NaI), é necessário aproximadamente 100 eV de radiação incidente. Em contrapartida, a formação de um par elétron-lacuna no semiconductor dos detectores HPGe ocorre a cada 3 eV de radiação absorvida (KNOLL, 2010). Esta propriedade confere aos detectores HPGe um poder de resolução consideravelmente maior (KEYSER; TWOMEY; UPP, 2005), como pode ser observado na Figura 5.

Figura 5 – Comparação entre espectros obtidos por detectores NaI e HPGe



Fonte: ORTEC (2014b)

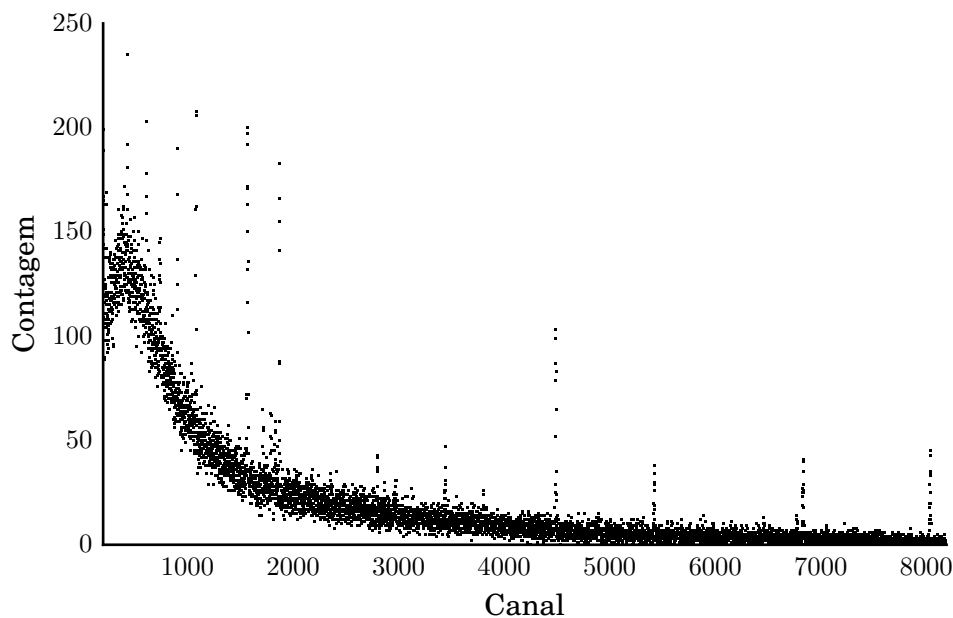
Apesar da baixa resolução, os detectores de cintilação sólida possuem maior eficiência de detecção do que os semicondutores devido à maior absorção de raios gama do iodo ($Z = 53$) em relação ao germânio ($Z = 32$) e ao volume efetivo dos detectores de cintilação sólida, que são normalmente maiores que os semicondutores (SHULTIS; FAW, 2007). Por esta razão, ainda são convenientes para propósitos ambientais, caso o radionuclídeo a ser analisado emita numa região sem interferentes próximos no espectro (PEREZ-ANDUJAR; PIBIDA, 2004). Contudo, detectores HPGe são capazes de trabalhar com espectros mais densamente povoados, como é geralmente o caso de espectros de amostras contendo radionuclídeos naturais. Portanto, os detectores HPGe são atualmente o instrumento de escolha para análises de espectrometria gama de alta resolução (KHANDAKER, 2011).

Durante a aquisição, além das emissões gama da amostra, serão observadas também emissões provenientes do ambiente (MURRAY et al., 1987). Por isto, o trabalho com amostras de baixa atividade demandam salas preparadas para este tipo de análise

(BRODZINSKI et al., 1985), incluindo, entre outros cuidados, uma estrutura capaz de suportar blindagens espessas de chumbo envolvendo os detectores.

Além das emissões gama do ambiente, sinais espúrios surgem no espectro resultantes do ruído eletrônico do próprio sistema (GEORGIEV; GAST, 1993), ou de eventos nucleares que ocorrem dentro do próprio material do detector (SEMKOW et al., 2002). O conjunto de todas estas contagens indesejadas (ou seja, que não tem origem e não estão relacionadas com a amostra em questão) compõem o contínuo de fundo (*background*) do espectro.

Figura 6 – Espectro gama de 100 g de sedimento contado por 3 horas em detector HPGe



Fonte: autoria própria

A Figura 6 mostra um exemplo de espectro gama típico obtido a partir de uma amostra de matriz ambiental. A soma do *background* com os picos gerados a partir das emissões gama das amostras compõem então o espectro gama.

2.1.3 Cálculo de atividade

Após a obtenção do espectro gama da amostra, a estimativa da concentração de atividade de um determinado radionuclídeo é dada pela seguinte equação²:

$$a = \frac{n_A - \frac{t_B}{t_A} n_B}{\epsilon \cdot p \cdot t_A \cdot V}, \quad (2.4)$$

² Esta equação não leva em conta correções de decaimento da amostra durante análise, auto-atenuação e efeito cascata.

onde n_A é a contagem líquida de emissões do radionuclídeo de interesse na sua região de emissão no espectro da amostra, sem contabilizar as contagens do *background*. n_B representa a contagem líquida do mesmo radionuclídeo (se houver) em um branco da mesma geometria e o tempo de aquisição dos espectros da amostra e do branco são dadas por t_A e t_B respectivamente, ϵ é a eficiência do detector naquela energia, p representa a probabilidade de emissão gama durante um decaimento do radionuclídeo de interesse e V é a quantidade de amostra (volume ou massa).

2.1.4 Detecção dos picos

Em um processo automatizado, a busca por picos é uma etapa básica da cadeia da análise e se encontra em todos os programas voltados para a espectrometria gama (BLAAUW et al., 1997). Só depois da busca por picos ser efetuada é que suas contagens líquidas serão contabilizadas para o cálculo de atividade. Esta busca pode ser realizada através de uma rotina de correlação com bibliotecas de referência ou por uma rotina de varredura.

2.1.4.1 Rotina de correlação com biblioteca

A biblioteca de referência é uma relação tabular de radionuclídeos de interesse, contendo suas respectivas meias-vidas, energias, probabilidades de emissões gama para cada evento de decaimento, além de outros dados relevantes para, não somente a qualificação, como posterior quantificação dos elementos presentes na amostra (CANBERRA, 2006).

A rotina de correlação itera sobre as energias listadas na biblioteca as quais se deseja detectar no espectro. Em cada iteração, é calculada a área da região sobre a qual cada pico listado, se presente, se localizaria. Os limites da região onde o pico estaria posicionado são calculados a partir da curva de calibração que relaciona o FWHM e a energia (L'ANNUNZIATA, 2003). Geer (2004) sugere que a menor incerteza obtida no cálculo de área dos picos ocorre quando a região de interesse se limita ao centróide do pico $\pm 1,25 \cdot \text{FWHM}$.

De acordo com o manual da CANBERRA (2006), o algoritmo “*Library Correlation NID*” presente no *software* Genie 2000 define uma região de interesse no espectro para o radionuclídeo contido em sua biblioteca e remove o contínuo de fundo através de um processo chamado de “erosão de picos” (*peak erosion*). Este método de eliminação do contínuo é uma variação do método proposto por Ryan et al. (1988) e conhecido como SNIP (*Sensitive Nonlinear Iterative Clipping Peak method*), o qual suaviza iterativamente o espectro de acordo com uma janela de tamanho proporcional à resolução esperada do pico no canal iterado. Esta suavização é subtraída do espectro original, resultando

no pico isolado do radionuclídeo supostamente presente na amostra. Este é comparado posteriormente com as calibrações de energia, formato e eficiência do sistema.

Estabelecidos os limites, a área do pico é calculada e reportada, podendo este valor ser, inclusive, negativo. Alguns programas comerciais, como o Genie 2000, determinam se o pico é suficientemente significativo antes deste ser reportado. O teste de significância se baseia no teste de χ^2 entre o formato do pico observado e o formato do pico esperado (CANBERRA, 2006).

A vantagem da rotina de correlação com bibliotecas é o potencial de detecção de picos malformados, pequenos e mesclados com outros picos interferentes. (L'ANNUNZIATA, 2003).

2.1.4.2 Rotina de varredura por métodos derivativos

Segundo Gilmore (2008), o método mais conhecido e utilizado em rotinas de varredura de espectros gama é atribuído a Mariscotti (1967), sendo este consagrado pelo *software* SAMPO (KOSKELO; AARNIO; ROUTTI, 1981), amplamente utilizado na década de 80, e incorporado em programas modernos como o Genie 2000 (CANBERRA, 2006) e GammaVision (ORTEC, 2009).

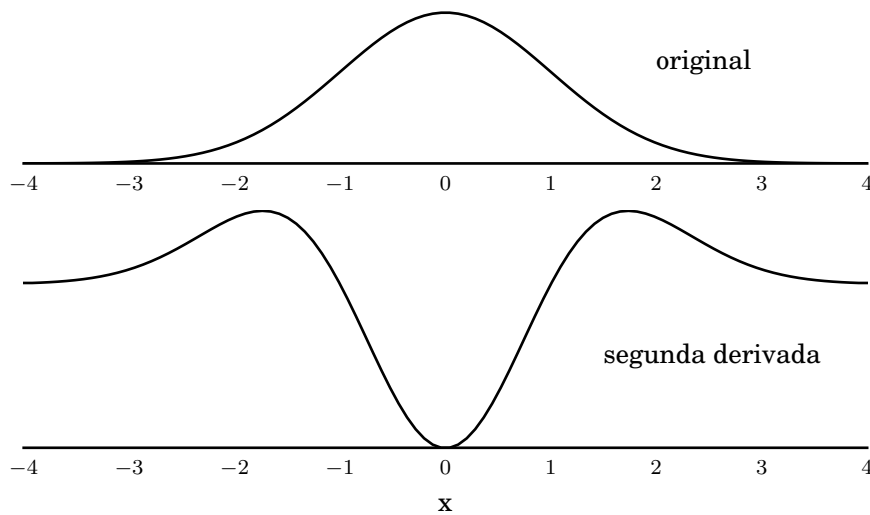
Tanto a primeira quanto a segunda derivada possuem propriedades relevantes para detectar a presença de um pico (O'HAVER et al., 1982). Quando se calcula a primeira derivada da função que descreve uma curva aproximadamente gaussiana, o ponto de cruzamento desta curva com a abscissa (onde há uma mudança de sinal) corresponde ao centroide da curva gaussiana. A segunda derivada produz uma curva cujo mínimo também corresponde ao centroide da curva gaussiana (HUANG; PARRISH, 1984). Um exemplo é mostrado na Figura 7.

Um efeito interessante da diferenciação de sinais correspondentes a picos é a direta correlação entre a amplitude das derivadas e a largura dos picos (CRAVEN; FAIRHURST; SUTCLIFFE, 1988). Esta propriedade permite que sinais altos e estreitos, geralmente correspondendo aos analitos de interesse, sejam separados do contínuo largo, raso e levemente curvado, como ocorre na espectrometria gama. Uma vez que a inclinação em áreas extensas sem picos é praticamente nula, o método derivativo também possui capacidade de ignorar a altura do contínuo.

Este método pode ser utilizado em diversos campos onde a detecção de picos em espectros é relevante, e não se restringe à espectrometria gama, sendo também bastante utilizado em espectroscopia na química analítica (PHILLIPS; MARLOW, 1976)(HUANG; PARRISH, 1984) (FERREE; SHANNON, 2001).

O programa HypermetPC (RÉVAY et al., 2001) utiliza uma variação deste método, mas que também pode ser considerado como um método derivativo. Trata-se da

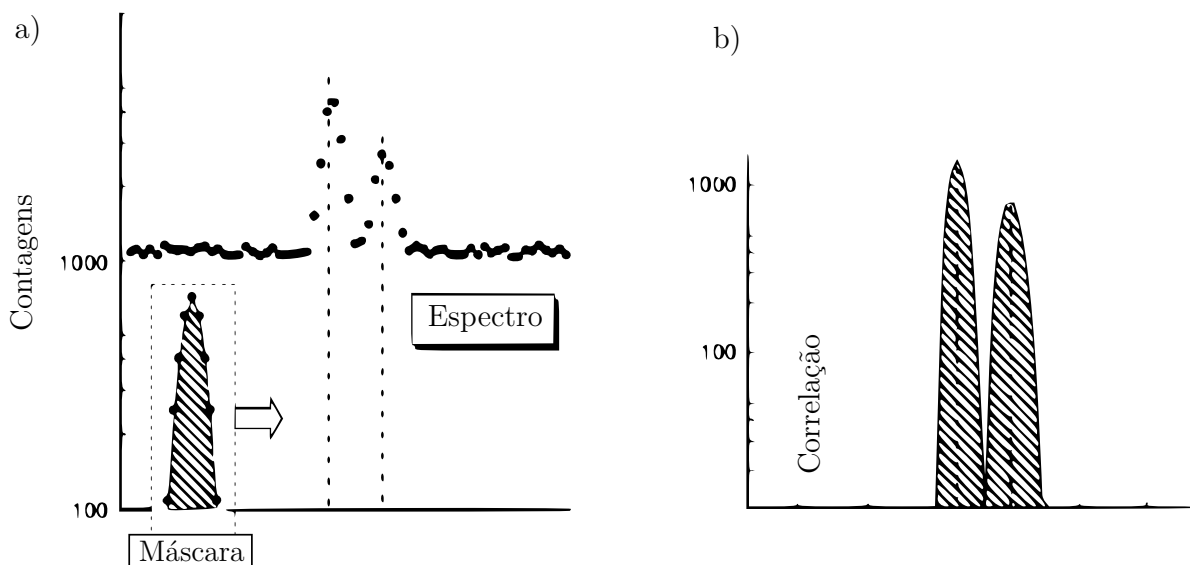
Figura 7 – Segunda derivada calculada sobre a função gaussiana



Fonte: autoria própria

convolução do espectro através de uma máscara de 1 dimensão cuja soma seja zero. À medida que a máscara percorre o espectro, esta muda de formato para que sua janela positiva seja da mesma largura dos picos que poderiam estar naquele valor de energia (Figura 8).

Figura 8 – Exemplo de convolução de um espectro



Fonte: Gilmore (2008)

Legenda: (a) a busca por picos é feita percorrendo uma máscara pelo espectro. (b) Os pontos de maior correlação indicam a presença de picos.

Tanto o método baseado na segunda derivada quanto a convolução de área-zero tem a propriedade de ignorar a altura do ruído de fundo de espectros.

2.1.5 Detecção de picos com baixa relação sinal-ruído

As emissões radioativas provenientes do decaimento dos átomos são eventos independentes e aleatórios, sendo que a probabilidade de se observar um número específico destes eventos é regida pela distribuição binomial. A distribuição de Poisson pode ser utilizada como aproximação para mais de 10 contagens e tempos de contagem muito menores do que as meias-vidas dos radionuclídeos de interesse (SITEK; CELLER, 2015). Sendo assim, a incerteza σ associada a uma média μ de eventos nucleares observados é igual ao desvio padrão da distribuição de Poisson, que é dada por

$$\sigma = \sqrt{\mu}. \quad (2.5)$$

Uma das formas de se avaliar a qualidade de um determinado sinal é através da razão entre sua intensidade e o ruído que se sobrepõe ao sinal. Esta medida é chamada de “relação sinal-ruído” (ou SNR). Em grandezas não-negativas tais como eventos nucleares (SCHROEDER, 2000), é possível relacionar a SNR e incerteza de medição através da seguinte expressão:

$$\text{SNR} = \frac{\mu}{\sigma} = \frac{\mu}{\sqrt{\mu}} = \frac{n_A}{\sqrt{n_A + 2N_B}}, \quad (2.6)$$

Onde n_A é a contagem líquida do pico e N_B é a contagem bruta do contínuo sob o pico.

Na identificação de picos em espectrometria gama, problemas surgem quando a SNR é muito baixa. Apesar do analista experiente ser capaz de visualizar picos de baixa intensidade com certa naturalidade, métodos matemáticos de detecção necessitam uma sequência de canais com contagens estatisticamente significativas e acima de um determinado limiar para que os picos sejam detectados (GILMORE, 2008). Este empecilho é bastante comum em análises ambientais, nas quais radionuclídeos naturais e antrópicos em amostras de rotina se encontram muito próximos dos limites mínimos de detecção, sobretudo quando a amostra é analisada *in natura*.

Tal condição é acentuada em amostras ambientais líquidas sem pré-concentração. A ausência de partículas adsorventes em matrizes aquáticas faz com que suas quantidades de radionuclídeos por massa de amostra sejam bastante menores do que matrizes sólidas como solos, sedimentos e cinzas. Um exemplo deste fato pode ser observado na Tabela 1, onde concentrações de radionuclídeos são comparadas em água e sedimento coletadas nos mesmos locais.

Tabela 1 – Concentração de atividade de ^{226}Ra em água e sedimento em certas regiões do mundo

Água (Bq L ⁻¹)	Sedimento (Bq kg ⁻¹)	Local	Autor
0,01-0,08	0,1-3000	Ontario, Canadá	Clulow et al. (1998)
0,0001-0,0002	17-200	Alexandria, Egito	El-Saharty (2013)
0,005-0,049	30-321	Iaduguda, Índia	Sethy et al. (2011)

Fonte: autoria própria

A detecção de picos pelo método de varredura depende que o analista estabeleça um valor de corte *a priori*. Todo pico que produza um sinal com significância menor do que este valor é descartado. Caso este valor de corte esteja acima do limite almejado pelo analista, um grande número de falso-negativos pode ser gerado (L'ANNUNZIATA, 2003). Por outro lado, o aumento da sensibilidade do método através da diminuição do valor de corte inevitavelmente elevará a quantidade de alarmes falsos.

Desta forma, um sistema de análises de rotina não pode depender exclusivamente do *software*, havendo necessidade da presença de um analista treinado para avaliar minuciosamente o resultados obtidos da espectrometria (GILMORE, 2008).

2.1.6 Limites mínimos de detecção

Uma abordagem básica sobre limites mínimos de detecção levanta basicamente duas questões: a) “o sinal detectado é significativo?” e b) “qual é a magnitude mínima do sinal o qual se possa detectar com confiança?”.

Currie (1999) afirma que, durante décadas, não houve consenso sobre as respostas a essas perguntas. Diferentes definições levavam a resultados com até três ordens de grandeza de diferença (CURRIE, 1968). Apesar de algumas organizações ainda proporem definições e métodos variados para se calcular estes limites, podendo levar a resultados variados (SANAGI et al., 2009), acadêmicos da IUPAC³ e ISO⁴ se uniram em 1995 para tentar estabelecer e padronizar a base teórica de limites mínimos de detecção, tratando de três definições básicas: um “limite de decisão” onde se pode decidir se o sinal observado indica ou não a ocorrência do efeito físico; um “limite de detecção” no qual o resultado através de um método analítico é confiável e um “limite de determinação” no qual um procedimento seja preciso o suficiente para gerar um resultado quantitativo confiável (CURRIE, 1999).

³ *International Union of Pure and Applied Chemistry* (ou União Internacional de Química Pura)

⁴ *International Organization for Standardization* (ou Organização Internacional para Padronização)

2.1.6.1 Teste de hipótese pelo limite crítico (ou limite de decisão)

Para análises qualitativas, o limite crítico permite a decisão *a posteriori* se um sinal S observado na análise representa ou não a presença do analito. O teste estatístico de hipótese para esta decisão está sujeito a dois tipos de erro: decidir que o analito está presente quando não está (erro tipo I ou α) ou falhar em decidir que o analito está presente quando não está (erro tipo II ou β). McCroan e Keller (2006) afirmam que o teste de hipótese é a abordagem mais comum para análises qualitativas no campo dos métodos analíticos nucleares.

Partindo da definição de Currie (1968) (revisada pelo próprio autor em 1999), seja um “branco” uma substância idêntica à amostra em quantidade, geometria e elementos interferentes, exceto que o branco não contém o analito de interesse. Seja N_B a contagem bruta e n_B a contagem líquida de emissões de um branco. Uma série de medições deste branco deverão produzir contagens líquidas com uma média $\mu_B \approx \mu_0 = 0$ e desvio padrão $\sigma_0 = \sqrt{N_B}$. O termo μ_0 representa a média das contagens líquidas de um branco.

O máximo valor aceitável para α e β , juntamente com σ_0 , estabelecem o “limite crítico” (L_C). Análises cujos sinais líquidos são maiores do que este limite indicam que o analito está presente com intervalo de confiança de $1 - \alpha$. Matematicamente, L_C é dado por

$$L_C = k_\alpha \sigma_0, \quad (2.7)$$

onde k_α representa os valores da abscissa da distribuição normal padrão para o nível de probabilidade $1 - \alpha$ mono-caudal. Esta definição de L_C é análoga ao “limite de não-deteção” (CANADIAN STANDARDS ASSOCIATION, 2010), “limite crítico” (AMERICAN NATIONAL STANDARDS INSTITUTE, 1996) e “limiar de decisão” (INTERNATIONAL STANDARDS ORGANIZATION, 2010). Uma representação gráfica desta definição é dada na Figura 9.

Analogamente, L_D é chamado de “limite de detecção” e se baseia em L_C . O limite de detecção é um parâmetro usado para estimar *a priori* um nível aceitável de β para erros tipo II e o desvio padrão σ_0 . L_D é definida por

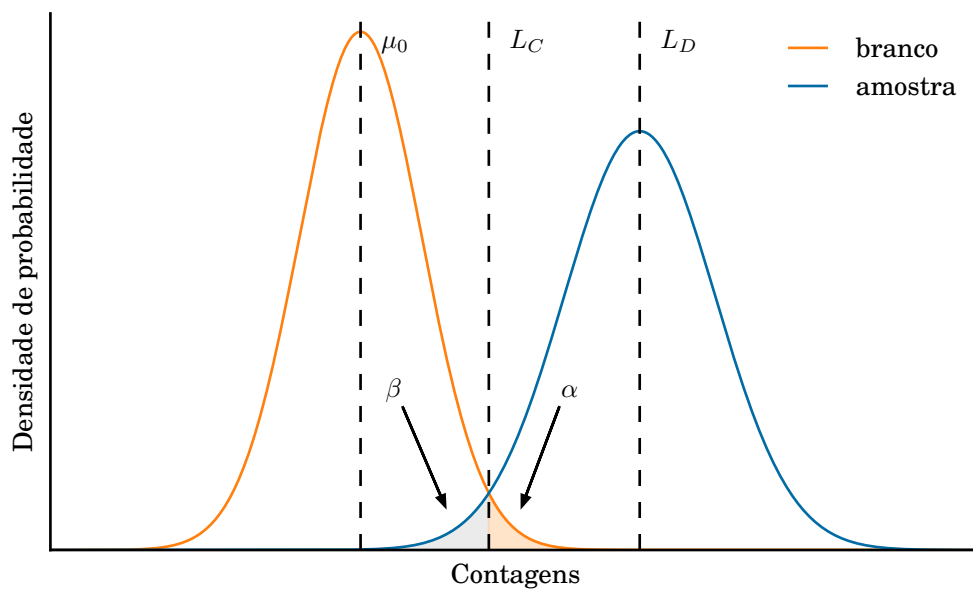
$$L_D = L_C + k_\beta \sigma_0, \quad (2.8)$$

onde k_β representa os valores da abscissa da distribuição normal padrão para o nível de probabilidade β mono-caudal. O limite de detecção, colocado desta maneira, é equivalente ao “limite de garantia de pureza” definido por Kaiser (1965) e “mínima atividade detectável” por Altshuler e Pasternack (1963), sendo este um jargão ainda usado

atualmente no campo da Radiometria (DONE; IOAN, 2016)(LEE et al., 2016)(ZAGNI et al., 2016).

De acordo com Gilmore (2008), o limite de detecção tenta responder à pergunta “qual é a menor quantidade possível de se detectar com confiança?”, servindo como um limite superior de confiança para o limite crítico. Normalmente, quando um analito não é detectado pela técnica radiométrica (isto é, seu resultado é possivelmente menor que L_C), então sugere-se que o resultado seja reportado como “ $< L_D$ ” (INTERNATIONAL STANDARDS ORGANIZATION, 2010).

Figura 9 – Teste de hipótese



Fonte: autoria própria

Legenda: O teste de hipótese é um teste estatístico para determinar se há evidência suficiente em um conjunto de observações para inferir que certa condição é válida. Segundo Currie (1968), observações que sejam menores que o limite de detecção podem não ser detectados e valores acima serão detectados. Em contrapartida, observações acima do limite crítico podem indicar a presença do analito, mas abaixo deste limite, não há evidências da presença deste.

Quando se trata de medições radiométricas, não é possível saber o real valor de σ_0 , pois o contínuo de fundo do branco e da amostra não serão necessariamente iguais. A presença de elementos diversos na amostra fazem com que o contínuo de fundo seja mais elevado devido à absorção incompleta das emissões gama. Portanto, é possível obter apenas uma estimativa do *background* da amostra através do contínuo sob o pico do analito de interesse. Esta estimativa, portanto, dependerá do *software* utilizado e da maneira que este define as regiões onde um pico se inicia e termina.

Seja $n_A = N_A - N_B$ a contagem líquida de um pico (total de sinais observados

menos *background*). Considerando que a probabilidade de decaimento de um elemento radioativo é regido pela distribuição de Poisson (CANNIZZARO et al., 1978), tem-se que o número de emissões observadas é igual à sua própria variância. Então,

$$\sigma_A^2 = n_A^2 + N_B^2 + 2 * \sigma_B^2 \quad (2.9)$$

Desta forma, baseando-se na Equação 2.7, o limite crítico e limite de detecção para análises radioativas é dado por:

$$L_C = k_\alpha \sqrt{2N_B} \quad (2.10)$$

$$L_D = L_C + \frac{k_\beta^2}{2} \left(1 + \sqrt{1 + \frac{4L_C}{k_\beta^2} + \frac{4L_C^2}{k_\alpha^2 k_\beta^2}} \right) \quad (2.11)$$

Para o caso específico onde $k_\alpha = k_\beta = k$, a Equação 2.11 é reduzida para a forma

$$L_D = k^2 + 2L_C = k^2 + 2k_\alpha \sqrt{2N_B}. \quad (2.12)$$

Segundo Currie (1968), a região analítica localizada entre L_C e L_D naturalmente retornará um número razoável de falso-positivos e falso-negativos, preferindo-se utilizar valores acima de L_D para tais decisões qualitativas.

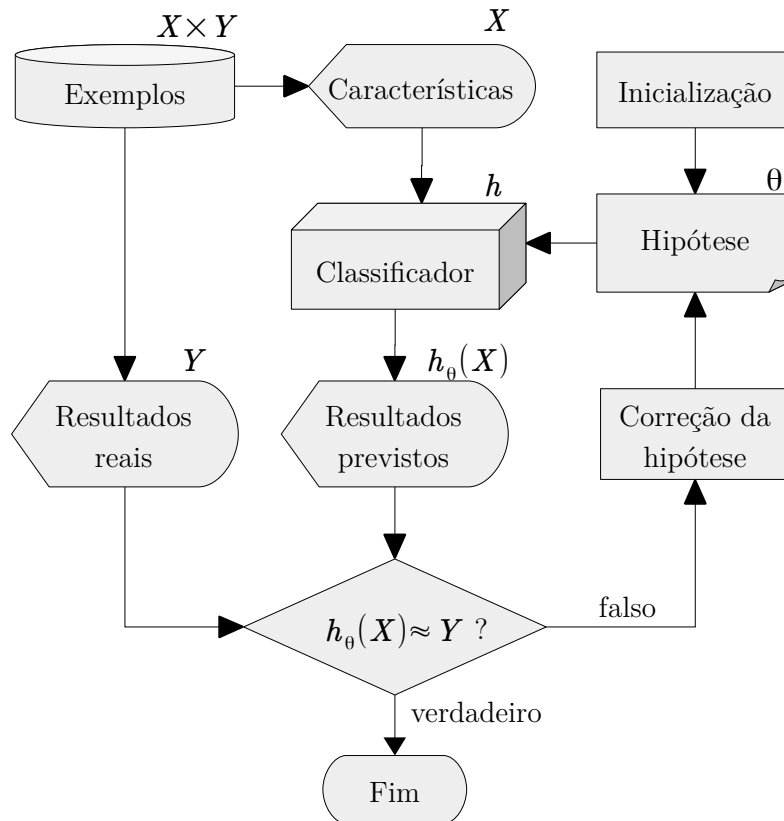
2.2 Inteligência artificial

Inteligência artificial (IA) é o nome dado a um campo específico da ciência da computação, o qual se dedica, sobretudo, à elaboração e aperfeiçoamento de técnicas e algoritmos que permitam ao computador aperfeiçoar seu desempenho em alguma tarefa, de maneira assistida ou independente de intervenção humana (LANGLEY, 1996). Resumidamente, a IA é a ciência de fazer com que computadores “aprendam” sem serem explicitamente programados. Um esquema geral de um modelo classificador de IA é mostrado na Figura 10.

Entre suas aplicações, destacam-se o reconhecimento de padrões, modelagens, previsão de séries temporais, mecanismos de busca, e várias outras tarefas onde a delineação e programação de algoritmos baseados em regras são demasiadamente complexas para a natureza do problema (WERNICK et al., 2010).

Os algoritmos de IA são categorizados de acordo com o tipo de variável de saída: “regressão”, quando se deseja prever uma variável de saída com valores contínuos; ou “classificação”, quando se deseja prever o grau de pertinência à uma ou várias classes. Em

Figura 10 – Esquema geral de um modelo de IA



Fonte: autoria própria

relação ao modelo de aprendizado empregado, se destacam principalmente o aprendizado supervisionado e não-supervisionado.

Aprendizado supervisionado: utilizando para o treinamento um conjunto de observações rotuladas (eventos com resultados conhecidos, ou rotulações com base no conhecimento de especialistas). No aprendizado supervisionado, deseja-se que o algoritmo atribua corretamente rótulos a observações inéditas (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

Aprendizado não-supervisionado: dado um conjunto não-rotulado de observações, espera-se que o sistema seja capaz de descobrir grupos de indivíduos similares por técnicas de “agrupamento” (*clustering*). Alternativamente, este tipo de IA pode ser usado para estimativa de distribuição dos dados, ou para a projeção dos dados em um espaço de menor dimensionalidade para fins de visualização (BISHOP, 2006).

Para o caso do aprendizado supervisionado, que é o foco deste trabalho, seja um conjunto $\mathbf{x} \in \mathbb{R}^n$ formado por m exemplos rotulados, onde para cada elemento \mathbf{x} , conhece-se seu resultado y e os n parâmetros que o caracterizam. Logo, o i -ésimo exemplo deste conjunto é um vetor $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})$, $i = 1, \dots, m$. As observações descritas por $(\mathbf{x}, y) \in X \times Y$ são definidas como observações independentes de uma variável aleatória

$Z = (X, Y)$ (DENIS, 2000).

Seja P a lei de probabilidade de Z . Se X for um conjunto discreto, a probabilidade de se observar (x, y) pode ser escrita simplesmente como $P(x, y)$. Pode-se igualmente definir as probabilidades marginais $P(x) = \sum_{y \in Y} P(x, y)$ e $P(y) = \sum_{x \in X} P(x, y)$. O objetivo do aprendizado supervisionado, portanto, é encontrar uma função $h_\theta(\mathbf{x}) : X \rightarrow Y$ cujo erro $E(h_\theta) = P(y \neq h_\theta(\mathbf{x}))$ seja o menor possível (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Os parâmetros θ da função h_θ são chamados de “hiper-parâmetros”. Na prática, os hiper-parâmetros caracterizam e dão forma aos modelos de IA. Eles podem configurar, por exemplo, as constantes de regularização, os tipos de funções de ativação e número de camadas de neurônios (no caso de redes neurais artificiais), margens de folga de classificação (no caso de máquinas de vetores de suporte) etc.

Este processo de indução utilizando o conjunto X de amostras é conhecido como “treinamento”. Depois de treinado, o modelo de IA aceita como entrada novos valores \mathbf{x}' e, através da função h_θ , realiza previsões y' .

2.2.1 Validação dos modelos

Espera-se que o processo de indução elaborado pelo modelo de IA durante a fase de treinamento seja aplicável a dados que ainda não o foram apresentados. Esta capacidade de extrapolação é chamada de “generalização” (LORENA; CARVALHO, 2007).

Para avaliar o poder de generalização de um modelo de IA, é comum utilizar uma métrica de avaliação sobre um conjunto de dados que não teve contato com o modelo de IA durante a fase de treinamento. Para isto, em vez de se utilizar todo o conjunto de dados disponível para o treinamento, um subconjunto é separado aleatoriamente para servir como conjunto de dados ainda não-observados (“conjunto de teste”). Este método é chamado de “validação cruzada”.

A validação cruzada pode ser realizada dividindo-se o conjunto de treinamento por uma proporção fixa dos dados disponíveis entre conjunto de treinamento e conjunto de teste e, opcionalmente, repetindo o mesmo particionamento repetidas vezes de forma aleatória. Este método é chamado de “retenção” (*holdout*). Normalmente, utiliza-se 2/3 dos dados disponíveis para treinamento e o restante para os testes (KOHAVI, 1995).

Quando se dispõe de um número insuficiente de dados para a separação entre treinamento e teste, costuma-se utilizar um método chamado “ k -dobras” (*k-folds*). Primeiramente, os dados são particionados em k segmentos de tamanhos semelhantes. A partir daí, são realizadas k iterações de treinamento e teste de tal forma que durante cada iteração, um segmento diferente de dados é mantido de fora para o teste, enquanto os $k - 1$ segmentos restantes são utilizados para o treinamento (REFAEILZADEH; TANG; LIU, 2009). No entanto, não existe um número de dobras para o qual a incerteza da

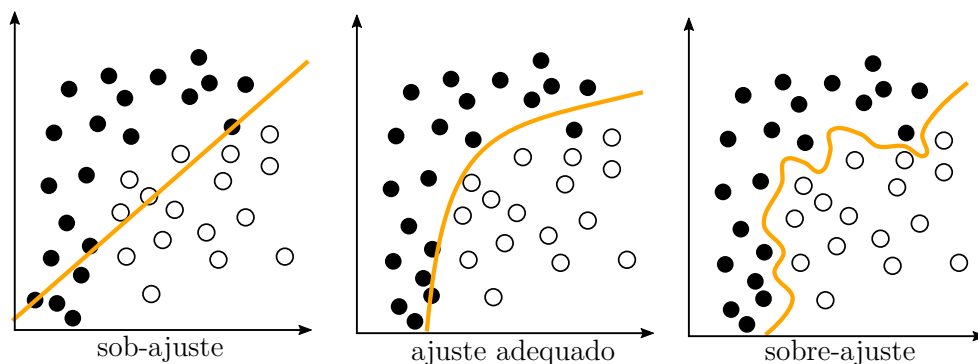
performance do modelo não seja subestimada (BENGIO; GRANDVALET, 2004).

A validação pode ser instável caso haja um número insuficiente de amostras de treino em relação à complexidade do problema e ao número de parâmetros por amostras. Quando isto ocorre, pequenas variações nos conjunto de treinamento podem resultar em saídas radicalmente diferentes. A quantidade necessária de amostras de treinamento pode variar de acordo com o grau de correlação entre si, ruído das amostras etc. Como regra geral, para um modelo que utilize n parâmetros de entrada, recomenda-se dispor não menos de $m = n^2$ amostras no conjunto de treinamento (HASSANAT et al., 2014).

2.2.2 Subajuste e sobreajuste

Diz-se que há “subajuste” quando determinado modelo tem um desempenho ruim durante o aprendizado do conjunto de treinamento. Normalmente, este sintoma é causado devido à utilização de um modelo simples demais para o problema em questão, incapaz de capturar a relação entre os exemplos de entrada X e os valores alvos Y (HAYKIN, 1994).

Figura 11 – Cenários de generalização de um classificador hipotético



Fonte: autoria própria

Em contrapartida, alguns modelos podem mostrar excelente desempenho quanto à classificação dos dados de treinamento, mas são incapazes de classificar corretamente os dados de validação. Este sintoma é chamado de “sobreajuste”, normalmente causado pelo excesso de complexidade do modelo de IA em relação ao problema em questão. Isto pode ocorrer quando o modelo incorpora o próprio “ruído” dos dados de treinamento na indução dos resultados previstos. Exemplos de sob e sobreajuste são mostrados na Figura 11.

2.2.3 Métricas de avaliação

Diversas são as métricas de avaliação disponíveis para que se possa julgar o desempenho de determinado algoritmo, como coeficiente de correlação (ou r^2) e erro quadrático

médio (*mean squared error* ou MSE) para tarefas de regressão ou acurácia para tarefas de classificação (POWERS, 2011). As métricas possuem principalmente três propósitos:

- Comparação de desempenho entre diferentes modelos;
- Otimização do modelo escolhido através da busca dos melhores hiper-parâmetros;
- Autoaperfeiçoamento. Algumas técnicas de IA envolvem a mistura (ou *ensemble*) de diversos modelos diferentes. Esta combinação é realizada de tal forma que os modelos que obtiverem melhor métrica de avaliação terão um peso maior no resultado final (DIETTERICH, 2000).

2.2.3.1 Acurácia

Normalmente, classificadores traduzem um valor de significância ou probabilidade sobre uma dada previsão e, a partir da escolha de um limiar, as converte para um valor binário, ou seja, todos os valores que se encontram acima do limiar são classificadas positivamente (POLLÁN, 2011). Desta forma, uma variável contínua entre 0 e 1 pode ser traduzida para “sim” ou “não”, “presente” ou “ausente”, “maligno” ou “benigno” etc.

A Figura 12 mostra as métricas fundamentais que podem ser obtidas após a limitização de um conjunto de amostras. A representação de erros e acertos como mostrada na Figura 12 é chamada de “matriz de confusão” e pode existir tanto para classificadores binários quanto para classificadores multiclases. A partir destas métricas básicas, é possível calcular diversas outras estatísticas úteis para a avaliação de desempenho (Tabela 2).

Figura 12 – Definição de uma matriz de confusão

		resultado da previsão	
		P	N
valor real	P	Verdadeiro Positivos (VP)	Falso Negativos (FN)
	N	Falso Positivos (FP)	Verdadeiro Negativos (VN)

A acurácia é uma métrica frequentemente utilizada na comparação entre classificadores quando se deseja trabalhar com uma métrica de fácil compreensão (neste caso, simplesmente a taxa de classificações corretas) e quando se está interessado no desempenho de um classificador com um limiar previamente escolhido.

Tabela 2 – Métricas

Métrica	Fórmula
Taxa de verdadeiro-positivos (TVP)	$VP/(VP+FN)$
Taxa de falso-positivos (TFP)	$FP/(FP+VN)$
Taxa de verdadeiro-negativos (TVN)	$VN/(VN+FP)$
Acurácia	$(VP+VN)/(P+N)$
Taxa de erro	1-acurácia
Probabilidade de alarmes falsos (P_{AF})	TFP
Probabilidade de detecção (P_D)	1-TVN

Fonte: Altman e Bland (1994)

2.2.3.2 Área sob curva ROC

Ocasionalmente, um classificador pode ser bastante adequado para uma tarefa de classificação, mas a escolha inadequada do limiar compromete sua acurácia. Neste caso, é sensato aferir o desempenho de um classificador ao longo de vários limiares. Para isto, existem as “métricas ordenadas”, como é o caso da curva característica de operação do receptor (*receiver operating characteristic curve*) ou ROC.

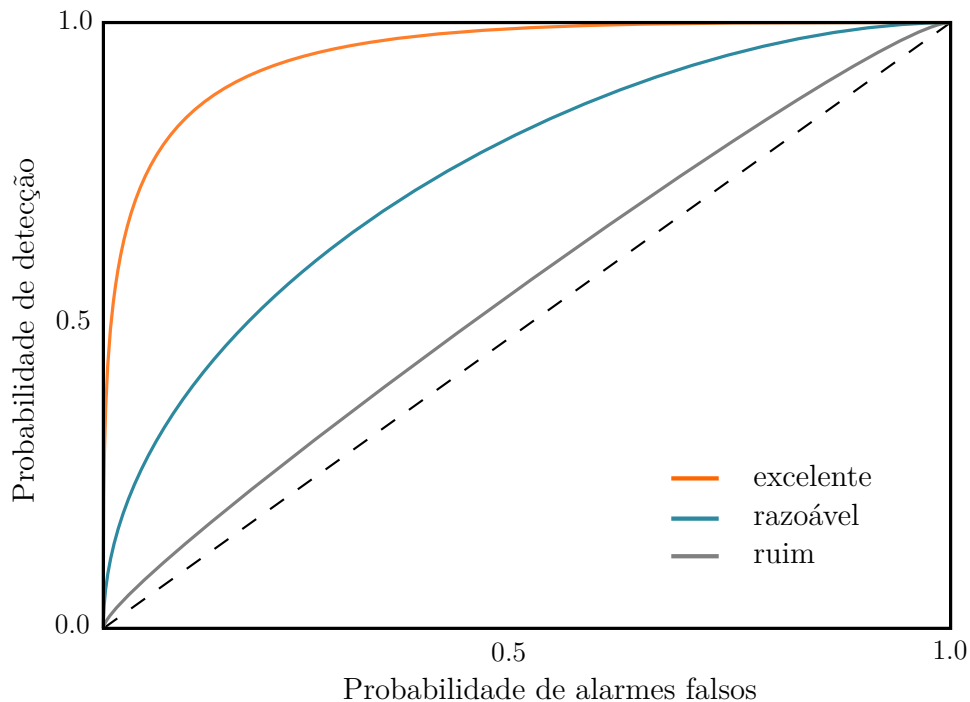
As curvas ROC são construídas a partir do índice de alarmes falsos em relação ao índice de detecção para diferentes limiares (HANLEY; MCNEIL, 1982). Desta forma, se é possível ter uma visão geral do classificador e do balanço que o modelo oferece entre sensibilidade e especificidade, sem necessidade da determinação de um limiar fixo. Um exemplo de curva ROC é mostrado na Figura 13.

O desempenho de um classificador pode ser sumarizado através da área sob a curva ROC (*area under the curve*) ou aROC. Um classificador aleatório produziria uma diagonal no gráfico com $aROC = 0,5$. O classificador aleatório é normalmente utilizado como base para se avaliar quão úteis são os modelos testados em relação a pura sorte. Deste modo, um bom algoritmo classificador deve ter um valor de aROC próxima de 1, enquanto um classificador praticamente aleatório deve se aproximar de 0,5.

As curvas ROC podem ser utilizadas não somente com classificadores que retornam um valor de probabilidade ou significância entre $[0, 1]$, mas com qualquer algoritmo ou modelo cujos resultados possam ser ordenados em um *ranking*.

Ao se trabalhar com classificadores binários, é comum a utilização da curva ROC sobretudo no campo da teoria de detecção de sinais (BRADLEY, 1997). No entanto, esta métrica vem ganhando destaque também no campo da inteligência artificial (FAWCETT, 2006), tendo sido sugerida como uma melhor alternativa à medida de acurácia (HUANG; LING, 2005) por ser resistente à prevaência de classes no conjunto. Por exemplo, para um conjunto de amostras que é predominantemente positiva ou negativa, um classificador que sempre retorna 1 ou 0 obterá uma alta acurácia. Em termos de aROC, sua pontuação

Figura 13 – Exemplo de curva ROC



Fonte: autoria própria

Legenda: o classificador com desempenho excelente exibe uma curva ROC que tende a se aproximar do canto superior esquerdo, quando o índice de alarmes falsos é mínima e o índice de detecções positivas é ótima. Em contrapartida, um classificador ruim tende a se aproximar da linha central tracejada, que representaria a utilização de valores aleatórios como classificação.

seria próxima de 0,5.

No entanto, Lobo, Jiménez-Valverde e Real (2008) criticam o uso indiscriminado da aROC como métrica única de avaliação, argumentando que a aROC leva em consideração regiões extremas da curva ROC cujos limiares não teriam nenhuma serventia em aplicações práticas. Por exemplo, um pesquisador dificilmente se interessaria em um limiar classificatório que permitisse uma probabilidade de detecção (P_D) de 0,9, mas probabilidade de alarmes falsos (P_{AF}) de aproximadamente 1,0. No entanto, este limiar é incorporado no cálculo da aROC com mesma importância de um limiar mais apropriado.

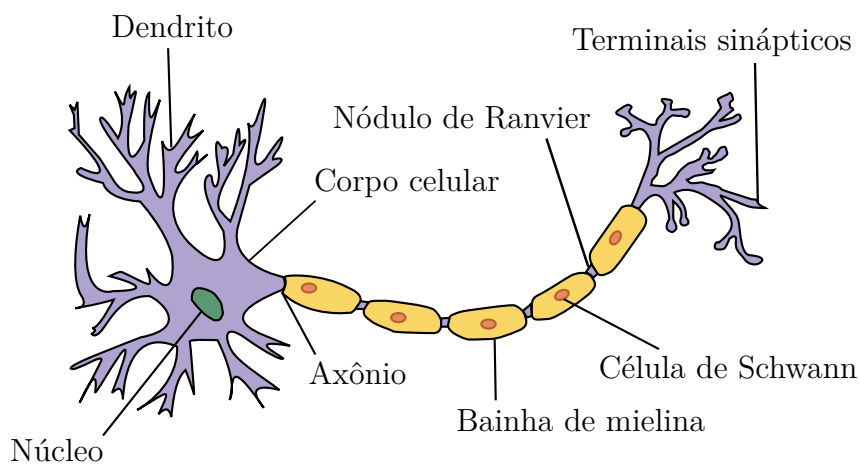
Além disso, tanto a acurácia quanto a aROC consideram que P_D e P_{AF} possuem o mesmo peso. Em algumas aplicações, pode-se desejar avaliar classificadores sob um ponto de vista conservador, onde o custo de um falso-negativo é maior do que o custo de um falso-positivo, tal como em diagnósticos médicos e análises de monitoramento ambiental.

2.2.4 Redes neurais artificiais

2.2.4.1 Visão geral

Redes neurais artificiais (*artificial neural networks*) ou RNAs são modelos de IA, cujas estruturas se assemelham às redes neurais biológicas. As RNAs são formadas por um conjunto interconectado de “nós” ou “unidades”, cuja função é análoga à do neurônio (RUSSELL, 1991). A Figura 14 representa o esquema genérico de um neurônio biológico. Sua contra-parte artificial, modelada matematicamente como uma unidade em uma rede artificial, é mostrado na Figura 15.

Figura 14 – Representação de um neurônio biológico



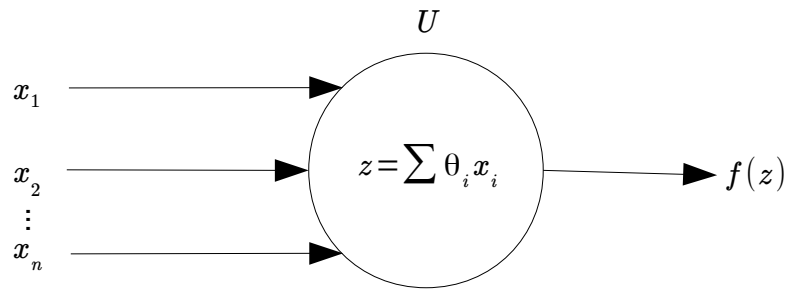
Fonte: Jarosz (2009)

O neurônio possui um corpo celular no qual estão ligadas diversas conexões de entrada chamadas de “dendritos”, responsáveis por receber impulsos elétricos. De acordo com o modelo simplificado de Hodgkin e Huxley (1952), o neurônio é uma unidade computacional que realiza internamente uma integração dos sinais elétricos de entrada e emite pulsos elétricos discretos de saída (*spikes*) pelo axônio até os dendritos de outros neurônios, caso a integração dos valores de entrada ultrapassem determinado limiar de intensidade (ARCAS; FAIRHALL; BIALEK, 2001).

Sem considerar a complexidade de uma célula biológica, a unidade de uma RNA pode ser comparado ao neurônio, embora apenas no sentido de que este é uma unidade logística que computa valores de saída de acordo com os valores de entrada (BISHOP, 2006).

As RNAs são uma espécie de “aproximadores” universais, caso um número suficiente de unidades estejam presentes (HORNIK; STINCHCOMBE; WHITE, 1989). Diferente de estimadores estatísticos, as RNA aproximam funções, ou melhor, valores de entrada se relacionam com valores de saída sem um modelo matemático explícito (KANESKI et al., 1996).

Figura 15 – Exemplo de uma única unidade em uma RNA



Fonte: autoria própria

RNAs são utilizadas em quimometria para análises espectrais, sendo comum RNAs com desempenhos melhores do que qualquer outro método de IA, dada sua robustez à não-linearidade dos dados quando sua topologia é corretamente projetada (ALLEN; ELLIS; PETRIDIS, 2010).

2.2.4.2 Definição

Seja $x = (x_1, x_2, \dots, x_n)$ um vetor de n valores de entrada apresentadas à unidade U e $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ é um vetor de pesos atribuídos às conexões do vetor de entrada $x^{(i)}$ à unidade U , resultando em uma soma $z \in \mathbb{R}$, tal que:

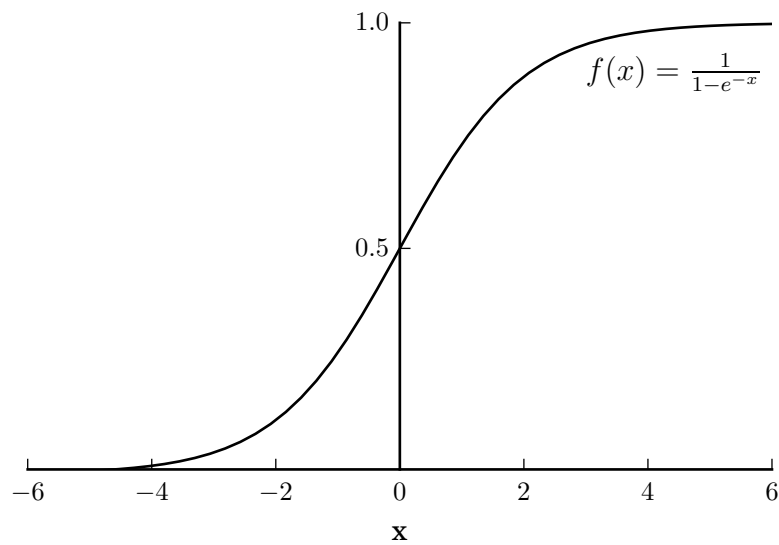
$$z = \sum \theta_i x_i \quad (2.13)$$

Os pesos em uma RNA podem ser tanto negativos quanto positivos, fornecendo desta forma influências excitatórias ou inibitórias para cada entrada (ZUPAN; GASTEIGER, 1993).

Ao resultado z , é aplicada uma função de ativação $f : \mathbb{R} \rightarrow \mathbb{R}$. O estado da unidade U é dado, então, por um valor $a \in \mathbb{R}$, que representa o valor de ativação de U . A imagem da função f , no entanto, é restrita a um intervalo que é caracterizado de acordo com a função de ativação utilizada, que pode ser linear, sigmoideal, tangencial hiperbólica, ou qualquer outro tipo de função crescente. A Figura 16 apresenta um exemplo de função de ativação do tipo sigmoideal, bastante utilizada nas RNAs (BISHOP, 2006).

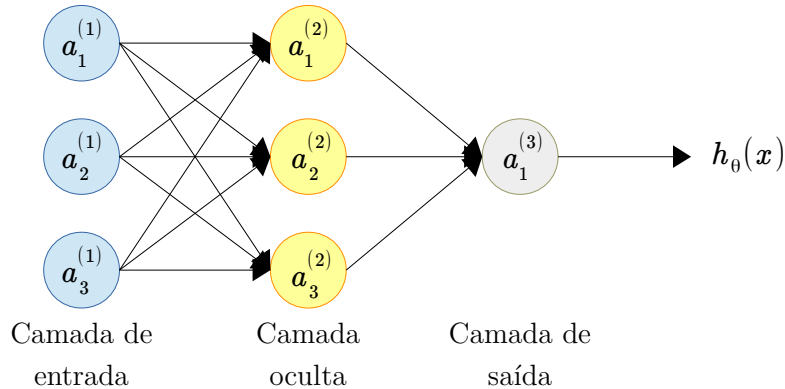
As RNAs não se limitam, no entanto, à utilização de apenas uma unidade. A Figura 17 mostra um exemplo de uma RNA com três camadas e três unidades em cada camada. Neste caso, a saída de uma unidade tipicamente se torna a entrada de outra unidade da próxima camada. As unidades da última camada, chamada de “camada de saída”, retornam um ou mais valores através de uma combinação dos valores de ativação recebidas da penúltima camada.

Figura 16 – Função de ativação do tipo sigmoidal



Fonte: autoria própria

Figura 17 – Exemplo de rede neural com mais de uma camada de neurônios



Fonte: autoria própria

Na RNA da Figura 17, a entrada dos dados para a rede se dá pela primeira camada (azul), cujos valores de ativação são denotados por $a^{(1)} = (a_1^{(1)}, a_2^{(1)}, a_3^{(1)})$. No exemplo da Figura 17, a primeira camada possui três unidades e é chamada de “camada de entrada”. Para esta camada em especial, $a^{(1)} = x$.

A segunda camada (amarela), representada por $a^{(2)} = (a_1^{(2)}, a_2^{(2)}, a_3^{(2)})$, é chamada de “camada oculta” ou “camada intermediária”, uma vez que os valores transitórios computados nas unidades desta camada servem apenas de suporte para a estimativa do resultado final, e não representam o resultado final em si. Os componentes de $a^{(2)}$ são dados

pela função de ativação aplicada à combinação das unidades da camada anterior, tal que

$$\begin{aligned} a_1^{(2)} &= f(\theta_{11}^{(1)} a_1^{(1)} + \theta_{12}^{(1)} a_2^{(1)} + \theta_{13}^{(1)} a_3^{(1)}) \\ a_2^{(2)} &= f(\theta_{21}^{(1)} a_1^{(1)} + \theta_{22}^{(1)} a_2^{(1)} + \theta_{23}^{(1)} a_3^{(1)}) \\ a_3^{(2)} &= f(\theta_{31}^{(1)} a_1^{(1)} + \theta_{32}^{(1)} a_2^{(1)} + \theta_{33}^{(1)} a_3^{(1)}) \end{aligned} \quad (2.14)$$

A camada de saída (cinza) retorna a função hipótese $h_\theta(x)$, que é também uma função não-linear de ativação aplicada à combinação das unidades da camada anterior, tal que

$$h_\theta(x) = a_1^{(3)} = f(\theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)}) \quad (2.15)$$

Sendo assim, $a^{(j)}$ representa os valores de ativação da camada j e $a_i^{(j)}$ representa o valor da i -ésima unidade da camada j . A matriz de parâmetros que controla o mapeamento da camada j para a camada $j + 1$ é dado por $\theta^{(j)}$.

Logo, a saída $h_\theta(x)$ de uma RNA com L camadas, pesos dados pela matriz θ e um vetor x de parâmetros de entrada é calculada pelo seguinte algoritmo:

Algoritmo 1 Alimentação direta

```

 $a^{(1)} = x$ 
para  $l = 1, \dots, L - 1$  faça
   $z^{(l+1)} = \theta^{(l)} a^{(l)}$ 
   $a^{(l+1)} = g(z^{(l+1)})$ 
fim para
 $h_\theta(x) = a^{(L)}$ 

```

Este algoritmo é chamado de “alimentação direta” (*feedforward*), uma vez que as unidades reagem com estímulos diferentes de acordo com os valores de entradas e tais estímulos são propagados unidirecionalmente até a saída da rede (BISHOP, 2006).

2.2.4.3 Treinamento por retro-propagação

O treinamento da RNA é realizado na tentativa de solucionar um problema de otimização, no qual os parâmetros θ da rede são ajustados de forma a minimizar uma função de custo, que normalmente, em problemas de regressão, é a soma do quadrado dos erros entre a saída $h_\theta(x)$ e o valor esperado y para o exemplo x do conjunto de treinamento.

Para problemas de classificação, o erro entre a saída da rede e o valor esperado é calculado pela função de entropia cruzada (*cross-entropy*). Seja o conjunto de treinamento formado por m exemplos, $\{(y^{(i)}, x^{(i)})\}_{i=1}^m$, onde x e y são os vetores de entrada e suas saídas

esperadas respectivamente, e h_θ o conjunto de pesos das conexões que mapeiam x a y . A estimativa do erro da RNA classificadora é dada por

$$\text{Erro}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)), & \text{se } y = 1 \\ -\log(1 - h_\theta(x)), & \text{se } y = 0 \end{cases} \quad (2.16)$$

Simplificando a fórmula acima, é possível escrevê-la da seguinte maneira:

$$\text{Erro}(h_\theta(x), y) = y \log h_\theta(x) + (1 - y) \log(1 - h_\theta(x)) \quad (2.17)$$

Sendo assim, a função de custo J para em função de θ é dada por

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log h_\theta(x^{(i)})_k + (1 - y_k^{(i)}) \log(1 - h_\theta(x^{(i)})_k) \quad (2.18)$$

onde $h_\theta(x^{(i)})_k$ é a k -ésima unidade da camada de saída.

Para que ocorra o aprendizado da RNA, deve-se minimizar a função de custo J . O problema de minimização normalmente é solucionado por variações do método do gradiente descendente. Este método, aplicado a diversas camadas de uma RNA, é chamado de “retro-propagação”, uma vez que se utiliza a diferença entre o vetor de valores de saída e o vetor de resultados esperados para ajustar os pesos das conexões da RNA, da última até a primeira camada (LECUN et al., 2012).

Sejam as camadas da RNA denominadas por $l = 1, 2, \dots, L$. O “erro” presente na unidade j da camada l é dado pelo algoritmo de retro-propagação:

Algoritmo 2 Retro-propagação

$$\delta^{(L)} = a^{(L)} - y$$

para $l = L - 1, \dots, 2$ **faça**

$$\delta^{(l)} = (\theta^{(l)})^T \times \delta^{(l+1)} \odot f'(z^{(l)})$$

fim para

O próximo passo é calcular o gradiente de $J(\theta)$ para que, através de algoritmos apropriados de otimização, seja possível encontrar o mínimo da função de custo. A estimativa do gradiente pode ser realizada iterativamente através do algoritmo de alimentação direta e o algoritmo de retro-propagação.

Durante a busca pelo mínimo de $J(\theta)$, a cada iteração, o valor da função de custo deve ser menor do que na iteração anterior. Após um número suficientemente grande de iterações, $J(\theta)$ convergirá, indicando que o mínimo da função foi encontrado e os pesos das conexões na RNA representados por θ foram otimizados para o melhor desempenho possível da rede.

Algoritmo 3 Gradiente de $J(\theta)$

```

 $\Delta^{(l)} = 0$  para  $l = 1, \dots, L - 1$ 
para  $i = 1, \dots, m$  faça
   $a^{(1)} = x^{(i)}$ 
  Estimar  $a^{(l)}$  para  $l = 2, \dots, L$  usando alimentação direta
  Estimar  $\delta^{(l)}$  para  $l = L, \dots, 2$  usando retro-propagação
   $\Delta^{(l)} = \Delta^{(l)} + \delta^{(l+1)}(a^{(l)})^T$  para  $l = L - 1, \dots, 1$ 
fim para
 $\frac{\partial}{\partial \theta^{(l)}} J(\theta) = \frac{1}{m} \Delta^{(l)}$  para  $l = 1, \dots, L - 1$ 

```

2.2.4.4 Topologia da rede

O conjunto de configurações (funções de ativação, número de camadas ocultas, número de unidades, pesos iniciais das conexões, regularização etc), são chamados de hiper-parâmetros e são as características que definem a topologia de uma rede neural.

Na maioria dos casos, não existe uma regra para determinar o melhor número de unidades e camadas ocultas por unidade, sem que haja o treinamento de diversas RNAs com topologias distintas. Este número depende de uma série de características do problema em questão, tais como:

- número de unidades de entrada e saída;
- tamanho do conjunto de treino;
- quantidade de ruído estatístico do conjunto de treino;
- complexidade da classificação a ser aprendida;
- tipos de funções de ativação empregadas nas camadas ocultas;
- algoritmo de treinamento;
- regularização.

Se a RNA tiver unidades ocultas em excesso, é possível obter um índice baixo de erros durante o treinamento da rede, mas um índice alto de erros na fase de validação devido à alta variância (também chamado de *overfitting*) (GEMAN; BIENENSTOCK; DOURSAT, 1992).

Quando há apenas duas camadas ocultas na rede, Huang (2003) recomenda que o número ótimo de unidades da primeira camada seja $\sqrt{(k+2)m} + 2\sqrt{\frac{m}{k+2}}$ e da segunda camada seja $k\sqrt{\frac{m}{k+2}}$, onde m é o número de exemplos do conjunto de treinamento e k é o número de unidades de saída.

No entanto, o número de unidades e camadas em uma RNA se limita apenas à capacidade de processamento do *hardware*. Le (2013), por exemplo, modelou uma RNA para reconhecimento facial cuja estrutura continha milhões de unidades e um bilhão de conexões. O treinamento da rede se estendeu por três dias em um aglomerado de 1000 ser-

vidores, totalizando 16000 CPUs. Izhikevich e Edelman (2008) estruturaram uma RNA da mesma ordem de grandeza de um cérebro humano para estudar o circuito talamocortical do cérebro de mamíferos, cuja execução durou 50 dias em um aglomerado de 27 servidores para gerar um minuto de simulação. Sua RNA era composta de 100 bilhões de unidades e 10^{15} conexões sinápticas.

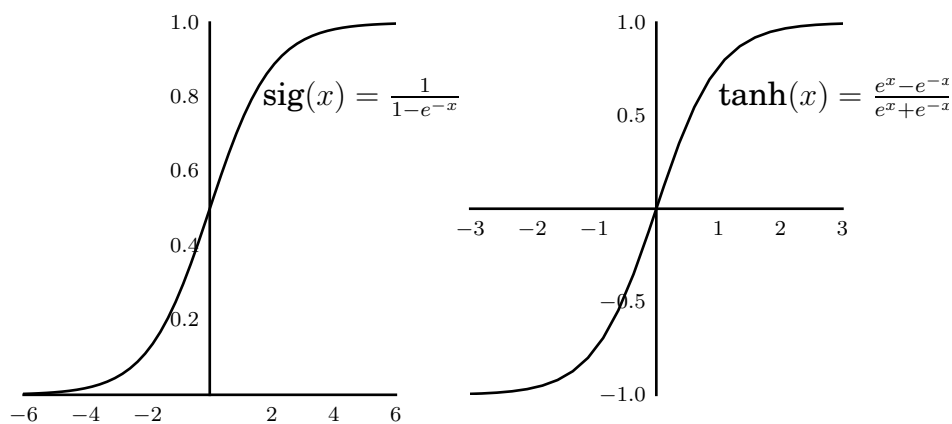
Haykin (1994) recomenda que, para minimizar o tempo de aprendizado, conjuntos cujos parâmetros de entrada com média diferente de zero devem ser evitados. Em relação ao vetor x de entrada da rede, que é distribuída aos neurônios da camada de entrada, a centralização dos parâmetros é feita através da expressão

$$x_j^{(i)} = \frac{x_j^{(i)} - \bar{x}_j}{\sigma_j}, \quad (2.19)$$

onde $x_j^{(i)}$ é o j -ésimo parâmetro centralizado da i -ésima amostra do conjunto de amostras, \bar{x}_j é a média do j -ésimo parâmetro de todas as amostras do conjunto, e σ_j o respectivo desvio padrão. A aplicação desta expressão sobre o vetor x de entrada faz com que este passe a ter média igual a zero e desvio padrão de aproximadamente 1.

Entretanto, quando o modelo de RNA possui mais de uma camada, o problema de médias não-nulas persiste nas camadas ocultas. Se a função de ativação for assimétrica, como é o caso da função sigmoideal, a saída de cada unidade é restrita ao intervalo $[0, 1]$. Tal escolha de função de ativação introduz um influência sistemática (*bias*) nas unidades localizadas além da primeira camada da rede. A solução para este problema é a utilização de uma função de ativação antissimétrica, tal como a tangente hiperbólica (Figura 18).

Figura 18 – Comparação entre função sigmoideal (esq.) e tangente hiperbólica (dir.)



Fonte: autoria própria

Através da função tangente hiperbólica, a saída de cada unidade pode assumir tanto valores positivos quanto negativos no intervalo de $[-1, 1]$, havendo maior possibili-

dade da média se aproximar de zero. Se a rede possuir um grande número de conexões, o aprendizado por retro-propagação com funções de ativação antissimétricas é capaz de atingir convergência mais rapidamente do que um processo similar com funções de ativação não-simétricas (LECUN; KANTER; SOLLA, 1991).

A redução da dimensionalidade dos valores de entrada é usualmente recomendado devido ao alto custo computacional das redes neurais. Caso isto não seja possível, um número grande de unidades de entrada torna as redes neurais pouco ideais para o uso em tempo real, apesar de serem particularmente eficientes em problemas de reconhecimento de padrões. Este problema pode ser contornado através da criação de sistemas onde os cálculos executados pela rede ocorrem em nível de *hardware* em vez de *software*.

2.2.5 Máquinas de vetores de suporte

As máquinas de vetores de suporte (*support vectors machines*) ou MVS são modelos de aprendizado supervisionado baseados na teoria estatística de Vapnik e Chervonenkis (1964) e são uma generalização dos classificadores lineares. Após sua primeira implementação, as MVSs se popularizaram rapidamente devido à sua capacidade de se trabalhar com dados de grandes dimensões, número reduzido de hiper-parâmetros, sólido embasamento teórico e bons resultados práticos (SCHÖLKOPF; SMOLA, 2002).

As MVSs podem ser utilizadas para a resolução de problemas tanto de classificação quanto de regressão. Em relação à classificação, MVS podem ser utilizadas para k classes, implementando-se k classificadores binários (VAPNIK, 1998).

Para o caso de classificação binária de um conjunto de amostras, a finalidade do modelo MVS é encontrar um hiperplano que separe as amostras entre duas classes distintas (FRANCOEUR, 2010). Uma concepção desta formalidade é exibida na Figura 19. A linha contínua central representa o hiperplano divisor, e as linhas tracejadas são chamadas de “vetores de suporte”.

Evidentemente, há uma infinidade de hiperplanos válidos que resolveriam a tarefa de classificação. Entretanto, segundo Vapnik (1998), o objetivo da MVS é encontrar o hiperplano com a maior margem possível entre as amostras mais próximas de cada classe. Este é um problema que admite apenas uma solução (HASTIE; TIBSHIRANI; FRIEDMAN, 2009), tornando o MVS um classificador bastante interessante em relação a classificadores estocásticos, pois o resultado será sempre igual para o mesmo conjunto de treinamento (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). Além disso, a maximização das margens faz do MVS um classificador com boa robustez a *outliers* e erros de generalização (JORDAAN, 2004).

2.2.5.1 Definição

Seja um conjunto de treinamento dado por $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}\}$, onde cada exemplo do conjunto é dado pelo vetor \mathbf{x} que possui n parâmetros ($x^{(i)} \in \mathbb{R}^n$). A classe a qual a amostra $x^{(i)}$ pertence é representada por $y^{(i)} \in \{-1, 1\}$.

Dispondo de m amostras na forma $(\mathbf{x}^{(i)}, y^{(i)}) \in \mathbb{R}^n \times \{-1, 1\}$, $i = 1, \dots, m$, e supondo que estes sejam linearmente separáveis, diz-se que existe um hiperplano em \mathbb{R}^n tal que todos os dados pertencentes à classe 1 se encontram de um lado do hiperplano, enquanto aqueles da classe -1 se encontram do lado oposto. Segundo Bishop (2006), a equação geral de um hiperplano em \mathbb{R}^n é dada por

$$\mathbf{w} \cdot \mathbf{x} + b = 0, \quad (2.20)$$

Mohri, Rostamizadeh e Talwalkar (2012) salientam que, geometricamente, os elementos do par (\mathbf{w}, b) representam, respectivamente, o vetor ortogonal e a constante de parametrização dos hiperplanos presentes em \mathbb{R}^n . Alguns autores como Burbidge e Buxton (2001) se referem a \mathbf{w} como “vetor de peso” e b como “tendência” (ou $-b$ como “limiar”).

A partir da Equação 2.20, é possível gerar uma infinidade de hiperplanos, bastando multiplicar \mathbf{w} e b por constantes arbitrárias. Partindo deste princípio, define-se como “hiperplano canônico” aquele cuja distância até o ponto mais próximo de cada uma das classes satisfaz a seguinte relação (LORENA; CARVALHO, 2007):

$$|\mathbf{w} \cdot \mathbf{x}^{(i)} + b| = 1. \quad (2.21)$$

A distância d de uma amostra $x^{(i)} \in \mathbb{R}^n$ ao hiperplano separador é definida por

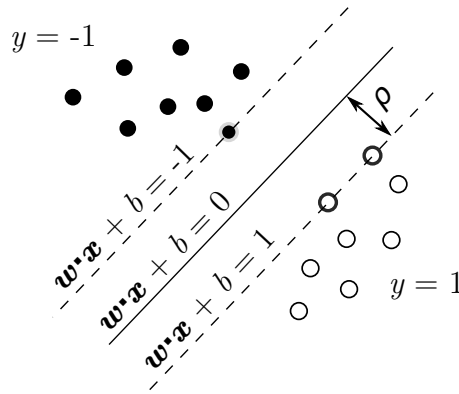
$$d^{(i)} = \frac{|\mathbf{w} \cdot \mathbf{x}^{(i)} + b|}{\|\mathbf{w}\|}. \quad (2.22)$$

Uma vez estabelecidas estas relações, procura-se maximizar a margem ρ entre o hiperplano e as amostras. Em outras palavras, deseja-se que todos os pontos do conjunto de treinamento estejam a uma distância $d \geq \rho$. Esta condição de contorno é definida por

$$\frac{|\mathbf{w} \cdot \mathbf{x}^{(i)} + b|}{\|\mathbf{w}\|} \geq \rho, \quad i = 1, \dots, m. \quad (2.23)$$

A Figura 19 mostra um exemplo de hiperplano que separa duas classes de dados. É possível observar que as bandas marginais são também hiperplanos e que estes são paralelos ao hiperplano separador. Sendo assim, ambos admitem o mesmo vetor ortogonal \mathbf{w} .

Figura 19 – Exemplo de divisão de classes com um hiperplano



Fonte: autoria própria

A maximização da margem do hiperplano canônico é equivalente à minimização de $\|\mathbf{w}\|$ ou $\frac{1}{2}\|\mathbf{w}\|^2$. Logo, para o caso linearmente separável de classificação, o modelo de MVS é estruturado através do seguinte problema de otimização:

$$\begin{aligned} & \underset{\mathbf{w}, b}{\text{minimizar}} && \frac{1}{2}\|\mathbf{w}\|^2 \\ & \text{sujeito a} && y^{(i)}(\mathbf{w} \cdot \mathbf{x}^{(i)} + b) \geq 1, \quad i = 1, \dots, m. \end{aligned} \quad (2.24)$$

A função objetiva da Equação 2.24 possui as características necessárias para a otimização convexa: um caso de programação quadrática com condições de contorno na forma de desigualdades (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Tal otimização admite apenas uma solução, sendo esta uma propriedade favorável e importante que não se faz presente em todos os algoritmos de IA (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). Algoritmos para a otimização deste tipo de problema são extensivamente utilizados tanto em softwares comerciais quanto de código livre (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

2.2.5.2 Solução pelo método de Lagrange

Para a otimização através do Método dos Multiplicadores de Lagrange, Hastie, Tibshirani e Friedman (2009) afirmam ser conveniente reescrever a equação canônica do hiperplano da seguinte maneira:

$$-(y^{(i)}(\mathbf{w} \cdot \mathbf{x}^{(i)} + b) - 1) \leq 0, \quad i = 1, \dots, m \quad (2.25)$$

Desta forma, a função lagrangiana, a qual introduz as variáveis $\alpha^{(i)} \geq 0$ associadas às m condições, pode ser definida, de acordo com Hastie, Tibshirani e Friedman (2009),

para todo $\mathbf{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$ e $\boldsymbol{\alpha} \in \mathbb{R}_+^m$, gerando o seguinte problema primal:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} - \sum_{i=1}^m \alpha^{(i)} (y^{(i)} (\mathbf{w} \cdot \mathbf{x}^{(i)} + b) - 1). \quad (2.26)$$

A partir desta função primal, deve-se agora calcular a função objetiva do problema dual. Este problema é solucionado maximizando $\boldsymbol{\alpha}$ ao mesmo tempo que se minimiza \mathbf{w} e b (LORENA; CARVALHO, 2007), o que corresponde ao ponto onde a derivada da lagrangiana em relação às variáveis da primal é nula (FRANCOEUR, 2010). Tem-se então que

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^m \alpha^{(i)} y^{(i)} \mathbf{x}^{(i)} = 0 \quad (2.27)$$

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = - \sum_{i=1}^m \alpha^{(i)} y^{(i)} = 0, \quad (2.28)$$

de tal forma que tais equações possam ser reescritas como

$$\mathbf{w} = \sum_{i=1}^m \alpha^{(i)} y^{(i)} \mathbf{x}^{(i)} \quad (2.29)$$

$$\sum_{i=1}^m \alpha^{(i)} y^{(i)} = 0. \quad (2.30)$$

Utilizando Equação 2.29 e Equação 2.30 para reescrever Equação 2.26 unicamente em função das variáveis duais, tem-se que

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimizar}} \quad L(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2} \mathbf{w} \cdot \mathbf{w} - \sum_{i=1}^m \alpha^{(i)} (y^{(i)} (\mathbf{w} \cdot \mathbf{x}^{(i)} + b) - 1) \\ &= \sum_{i=1}^m \alpha^{(i)} - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} \mathbf{x}^{(i)} \cdot \mathbf{x}^{(j)}). \end{aligned} \quad (2.31)$$

É necessário adicionar a condição de contorno $\sum_{i=1}^m \alpha^{(i)} y^{(i)} = 0$ para se assegurar de que o problema de maximização corresponda ao mínimo da lagrangiana. De fato, uma vez que a condição de contorno $\mathbf{w} = \sum_{i=1}^m \alpha^{(i)} y^{(i)} \mathbf{x}^{(i)}$ é automaticamente satisfeita pela construção da função, nada garante que a condição $\sum_{i=1}^m \alpha^{(i)} y^{(i)} = 0$ seja respeitada

(FRANCOEUR, 2010). Desta forma, temos o problema dual a ser resolvido passa a ser (HASTIE; TIBSHIRANI; FRIEDMAN, 2009):

$$\begin{aligned}
 \underset{\boldsymbol{\alpha}}{\text{maximizar}} \quad & \sum_{i=1}^m \alpha^{(i)} - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} \mathbf{x}^{(i)} \cdot \mathbf{x}^{(j)}) \\
 \text{sujeito a} \quad & \sum_{i=1}^m \alpha^{(i)} y^{(i)} = 0, \quad i = 1, \dots, m \\
 & \alpha^{(i)} \geq 0, \quad i = 1, \dots, m.
 \end{aligned} \tag{2.32}$$

A solução deste problema de otimização será portanto um vetor $\boldsymbol{\alpha}^*$. Segundo Mohri, Rostamizadeh e Talwalkar (2012), as propriedades das condições impostas determinam que o problema primal e dual sejam equivalentes. A vantagem de se utilizar a forma dual do problema de otimização é a maior simplicidade das restrições e a possibilidade de representar o problema de otimização através do produto escalar entre amostras do conjunto de treinamento. Isto diminui o custo computacional, sobretudo em casos complexos de classificação onde a separação linear dos exemplos de treinamento não é possível sem o mapeamento do conjunto a um espaço com maior número de dimensões (LORENA; CARVALHO, 2007).

A função $\mathbf{w} \cdot \mathbf{w}$ é convexa, contínua e derivável, a restrição $-(y^{(i)}(\mathbf{w} \cdot \mathbf{x}^{(i)} + b) - 1) < 0$ é uma função linear e o domínio do problema é \mathbb{R}^n . Estas características respeitam as condições de Karush-Kühn-Tucker, portanto este problema de otimização possui uma apenas uma solução ótima (BOYD; VANDENBERGHE, 2004). Particularmente, ela respeita a condição complementar de Karush-Kühn-Tucker (BISHOP, 2006), tal que

$$\alpha^{*(i)}(y^{(i)}(\mathbf{w}^* \cdot \mathbf{x}^{(i)} + b^*) - 1) = 0 \tag{2.33}$$

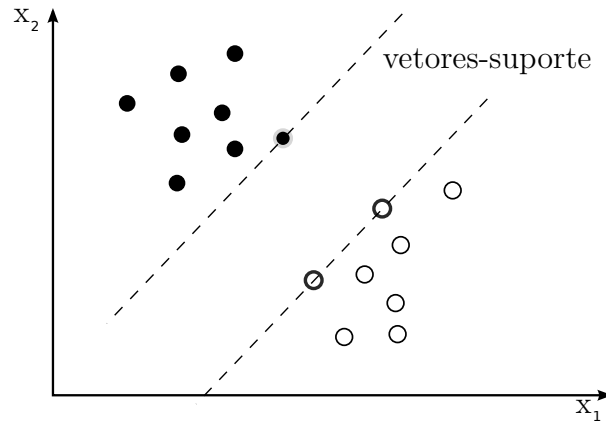
Nesta equação, α^* representa a solução ótima do problema dual e (\mathbf{w}^*, b^*) representa a solução do problema primal. Segundo Francoeur (2010), esta condição implica que se $y^{(i)}(\mathbf{w}^* \cdot \mathbf{x}^{(i)} + b^*) - 1 \neq 0$, então $\alpha^{(i)} = 0$. Consequentemente, os únicos casos onde $\alpha^{(i)}$ não podem ser nulos são aqueles onde $y^{(i)}(\mathbf{w}^* \cdot \mathbf{x}^{(i)} + b^*) - 1 = 0$; isto é, aqueles onde

$$y^{(i)}(\mathbf{w}^* \cdot \mathbf{x}^{(i)} + b^*) = 1. \tag{2.34}$$

Desta forma, os únicos pontos onde $y^{(i)}(\mathbf{w}^* \cdot \mathbf{x}^{(i)} + b^*) - 1 = 0$ são aqueles que se situam sobre a margem. A partir daí, observa-se que apenas estes pontos sobre a margem podem ter $\alpha^{(i)}$ não-nulos. Estes pontos são chamados de “vetores de suporte”.

A razão deste nome é tal que estes são os únicos pontos úteis para determinar o hiperplano ótimo de classificação. Uma vez que o vetor de coeficientes do hiperplano é dado por $\mathbf{w} = \sum_{i=1}^m \alpha^{(i)} y^{(i)} \mathbf{x}^{(i)}$, todos os pontos que não passam pela margem não

Figura 20 – Exemplo de vetores de suporte



Fonte: autoria própria

contribuem para a determinação do hiperplano, uma vez que $\alpha^{(i)}$, para estes casos, seria nulo. Os vetores de suporte podem, então, serem vistos como os pontos que contém toda a informação necessária para a resolução do problema.

Dadas as restrições do problema primal, se um exemplo $x^{(k)}$ se encontra sobre a margem, então diz-se que

$$y^{(k)} = (\mathbf{w} \cdot \mathbf{x}^{(k)} + b) = 1. \quad (2.35)$$

Sendo assim, o cálculo de b se torna trivial:

$$b = \frac{1}{y^{(k)}} - \mathbf{w} \cdot \mathbf{x}^{(k)} = y^{(k)} - \sum_{i=1}^m \alpha^{(i)} y^{(i)} \mathbf{x}^{(i)} \cdot \mathbf{x}^{(k)}, \quad y^{(k)} \in \{-1, 1\}. \quad (2.36)$$

Desta forma, a resolução do problema dual permite a construção do hiperplano canônico separando os exemplos com a maior margem possível e, conseqüentemente, a classificação de dados de acordo com suas características denotadas por \mathbf{x} .

Sob a hipótese de que os exemplos são linearmente separáveis, utiliza-se a seguinte função para a classificação dos exemplos:

$$H_{\mathbf{w},b}(\mathbf{x}) = \text{sinal}(\mathbf{w} \cdot \mathbf{x} + b), \quad (2.37)$$

onde

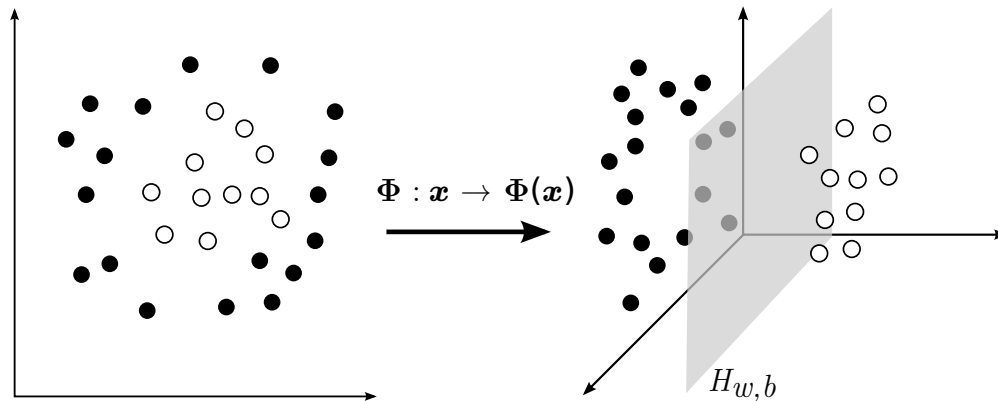
$$\text{sinal}(\mathbf{w} \cdot \mathbf{x} + b) = \begin{cases} -1 & \text{se } \mathbf{w} \cdot \mathbf{x} + b < 0 \\ 0 & \text{se } \mathbf{w} \cdot \mathbf{x} + b = 0 \\ 1 & \text{se } \mathbf{w} \cdot \mathbf{x} + b > 0 \end{cases} \quad (2.38)$$

Isto significa que um hiperplano H definido por (\mathbf{w}, b) classifica corretamente um ponto $\mathbf{x}^{(i)}$ quando $\mathbf{w} \cdot \mathbf{x}^{(i)} + b$ tem o mesmo sinal de $y^{(i)}$ (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

2.2.5.3 Conjuntos não-linearmente separáveis

Apesar do problema original ser dado dentro de um espaço de dimensões finitas, é comum acontecer das classes a serem discriminadas não serem separáveis linearmente dentro deste espaço. Por essa razão, algoritmos de MVS são implementados de tal forma que o espaço original de dimensões finitas (espaço de entrada ou *input space*) seja mapeado para um espaço com quantidade maior de dimensões (espaço de características ou *feature space*) (CORTES; VAPNIK, 1995).

Figura 21 – Transformação dos dados não-linearmente separáveis



Fonte: autoria própria

Legenda: Conjunto de dados não-linearmente separáveis no espaço de entrada são mapeados até um espaço de maior dimensão, onde a separação linear se torna possível.

Em espaços com numerosas dimensões, a separação dos dados pode ocorrer de maneira linear e, portanto, com menor custo computacional. Desta forma, os modelos MVS se mostram extremamente interessantes no uso de análises espectrais, já que estes proporcionam dados de entrada de alta dimensionalidade na grande maioria dos casos (CLARKE et al., 2008).

Seja $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^r$, $\mathbf{x} \rightarrow \Phi(\mathbf{x})$ a transformação aplicada aos exemplos do conjunto de treinamento para torná-los linearmente separáveis, com r como a dimensão do espaço de características. Bastante frequentemente, $r > n$, o que significa que a transformação leva os exemplos a um espaço de dimensão superior a fim de melhor os separar (FRANCOEUR, 2010).

Para encontrar o hiperplano separador, procede-se da mesma maneira como no caso dos dados linearmente separáveis, mas substituindo $\mathbf{x}^{(i)}$ por $\Phi(\mathbf{x}^{(i)})$. Assim, o objetivo

passa a ser a resolução do seguinte problema:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimizar}} \quad \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{sujeito a} \quad y^{(i)}(\mathbf{w} \cdot \Phi(\mathbf{x}^{(i)}) + b) \geq 1, \quad i = 1, \dots, m. \end{aligned} \quad (2.39)$$

Da mesma maneira, o novo problema dual passa a ser

$$\begin{aligned} & \underset{\alpha}{\text{maximizar}} \quad \sum_{i=1}^m \alpha^{(i)} - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} \Phi(\mathbf{x}^{(i)}) \cdot \Phi(\mathbf{x}^{(j)})) \\ & \text{sujeito a} \quad \sum_{i=1}^m \alpha^{(i)} y^{(i)} = 0, \quad i = 1, \dots, m \\ & \quad \alpha^{(i)} \geq 0, \quad i = 1, \dots, m. \end{aligned} \quad (2.40)$$

A função de decisão associada ao problema passa a se tornar, conseqüentemente,

$$H_{\mathbf{w},b}(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^m (\alpha^{(i)} y^{(i)} \Phi(\mathbf{x}^{(i)})) \cdot \Phi(\mathbf{x}) + b \right), \quad (2.41)$$

onde

$$b = y^{(k)} - \sum_{i=1}^m \alpha^{(i)} y^{(i)} \Phi(\mathbf{x}^{(i)}) \Phi(\mathbf{x}^{(k)}), \quad (2.42)$$

com $\Phi(\mathbf{x}^{(k)})$ sendo uma amostra sobre o vetor de suporte no espaço de características. Se a transformação utilizada é apropriada, a resolução de um dos problemas (primal ou dual) permite a estimativa de um hiperplano separador não-linear com a maior margem possível.

No entanto, o mapeamento das amostras a espaços de maiores dimensões levantam certos problemas; além de ser necessário escolher a transformação correta, falta lhes aplicar a todos os dados, após efetuar os cálculos com os dados transformados, isto é, no espaço de características. Desta forma, como a dimensão deste espaço é frequentemente maior que o espaço de entrada, os cálculos requeridos podem se tornar extremamente longos de se efetuar.

Nota-se que quando o problema se apresenta em sua forma dual, os dados do conjunto de treinamento são simplesmente o produto escalar de outros dados do mesmo conjunto. Ocorre o mesmo com a função dual de decisão. Isto leva à definição de uma função chamada de “núcleo” (*kernel*):

$$K : (\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \rightarrow \Phi(\mathbf{x}^{(i)}) \cdot \Phi(\mathbf{x}^{(j)}) \quad (2.43)$$

Esta função toma como entrada dois pontos no espaço de entradas e calcula seu produto escalar no espaço de características. A vantagem de tal função é que ela permite o cálculo do produto escalar entre amostras sem a necessidade do mapeamento prévio destas ao espaço de características (FRANCOEUR, 2010). Este cálculo pode ser feito diretamente a partir dos dados no espaço de entrada. Graças ao conceito de núcleo, é possível reescrever o problema dual da seguinte maneira:

$$\begin{aligned} & \underset{\alpha}{\text{maximizar}} && \sum_{i=1}^m \alpha^{(i)} - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})) \\ & \text{sujeito a} && \sum_{i=1}^m \alpha^{(i)} y^{(i)} = 0, \quad i = 1, \dots, m \\ & && \alpha^{(i)} \geq 0, \quad i = 1, \dots, m. \end{aligned} \tag{2.44}$$

A função de decisão associada ao problema torna-se, conseqüentemente,

$$H_{\mathbf{w},b}(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^m (\alpha^{(i)} y^{(i)} K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})) + b \right), \tag{2.45}$$

onde

$$b = y^{(k)} - \sum_{i=1}^m \alpha^{(i)} y^{(i)} K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}), \tag{2.46}$$

sendo $x^{(k)}$ um vetor de suporte.

Vale lembrar que desta maneira, uma vez que a função núcleo é contínua, a transformação $\Phi(\mathbf{x})$ não teria partes nulas, nem no problema, nem na aplicação da solução. Por conseqüência, graças à função núcleo, não é necessário efetuar a transformação sobre os dados. Esta função permite então se realizar todos os cálculos sem ter de se preocupar com a dimensão do espaço de características.

Tabela 3 – Núcleos frequentemente utilizados

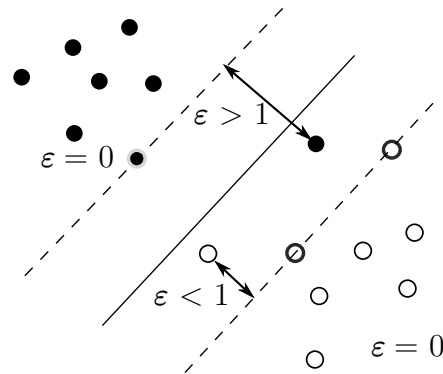
Nome	Núcleo
Linear	$K(x^{(i)}, x^{(j)}) = (x^{(i)} \cdot x^{(j)})$
Polinomial de grau d	$K(x^{(i)}, x^{(j)}) = (x^{(i)} \cdot x^{(j)} + 1)^d$
Gaussiana (ou base radial)	$K(x^{(i)}, x^{(j)}) = \exp\left(\frac{\ x^{(i)} - x^{(j)}\ }{2\sigma^2}\right)$
Multi-quadrática inversa	$K(x^{(i)}, x^{(j)}) = \frac{1}{\sqrt{(x^{(i)} - x^{(j)}) \cdot (x^{(i)} - x^{(j)}) + b}}$

Fonte: Francoeur (2010)

2.2.5.4 Margens suaves

Conjuntos de treinamento frequentemente possuem *outliers* que impedem a separação perfeita entre duas classes de dados. Segundo Lorena e Carvalho (2007), este problema pode ser contornado usando-se margens suaves em vez de margens rígidas, introduzindo uma variável de folga ε (Figura 22). Assim, a restrição à qual a Equação 2.24 está sujeita passa a ser dada por $y^{(i)}(\mathbf{w} \cdot \mathbf{x}^i + b) \geq 1 - \varepsilon^{(i)}, \forall i \in [1, m]$.

Figura 22 – Representação geométrica das variáveis de folga



Fonte: autoria própria

Utilizando esta nova restrição, a Equação 2.24 passa a ser rescrita da seguinte forma:

$$\begin{aligned}
 & \underset{\mathbf{w}, b, \varepsilon}{\text{minimizar}} && \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \varepsilon_i \\
 & \text{sujeito a} && y^{(i)}(\mathbf{w} \cdot \mathbf{x}^{(i)} + b) \geq 1 - \varepsilon_i, \quad i = 1, \dots, m \\
 & && \varepsilon_i \geq 0
 \end{aligned} \tag{2.47}$$

onde C é uma constante que representa a penalização por dados mal classificados. O índice de classificações erradas será inversamente proporcional a C , e sua escolha tem grande impacto no modelo de MVS. Na prática, diversos modelos são construídos com diferentes valores de C e, em seguida, o melhor valor é escolhido (FRANCOEUR, 2010).

Uma vez que ε representa a margem de erro de um ponto mal classificado, $\sum_{i=1}^m \varepsilon^{(i)}$ serve como um limite no número de erros de treinamento (LORENA; CARVALHO, 2007). A resolução do problema primal e dual é realizada de maneira análoga ao problema com margens rígidas.

Sendo α^* a solução do problema dual e $(\mathbf{w}^*, b^*, \varepsilon^*)$ a solução da forma primal, $\varepsilon^{*(i)}$

são calculadas através da seguinte equação:

$$\varepsilon^{*(i)} = \max \left\{ 0, 1 - y^{(i)} \sum_{k=1}^m y^{(k)} \alpha^{*(k)} \mathbf{x}^{(j)} \cdot \mathbf{x}^{(i)} + b^* \right\}, \quad (2.48)$$

onde a variável b^* é calculada a partir das condições de Karush-Kühn-Tucker, reescritas da seguinte forma:

$$\alpha^{*(i)} (y^{(i)} (\mathbf{w}^* \cdot x^{(i)} + b^*) - 1 + \varepsilon^{*(i)}) = 0 \quad (2.49)$$

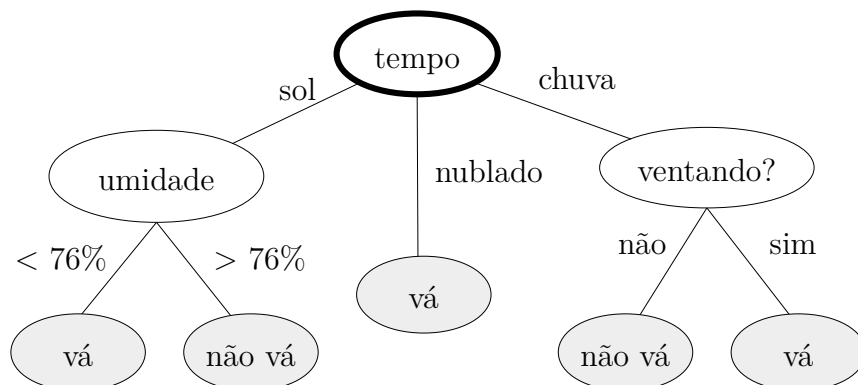
$$(C - \alpha^{*(i)}) \varepsilon^{*(i)} = 0. \quad (2.50)$$

Tais condições implicam que se um ponto é tal que $\varepsilon^{*(i)} \neq 0$, então este se encontra do lado errado do hiperplano separador. Conseqüentemente, todos os pontos para os quais $\varepsilon^{*(i)} = 0$ se encontram do lado correto do hiperplano separador (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

2.2.6 Árvores de decisão

Árvores de decisão (ou AD) são modelos de IA não-paramétricos, capazes de computar um mapa de regras através de uma hierarquia de testes sobre os valores de um conjunto de parâmetros, e prever um valor numérico contínuo ou um item contido em um conjunto discreto de conclusões (QUINLAN, 1986). A Figura 23 apresenta um exemplo de árvore de decisão.

Figura 23 – Exemplo de árvore de decisão para a pergunta devo viajar?



Fonte: Monard e Baranauskas (2003)

A classificação de uma amostra se inicia pela raiz da AD, descendo por seus nós pela direita quando a resposta a uma questão no nó é positiva, ou esquerda quando

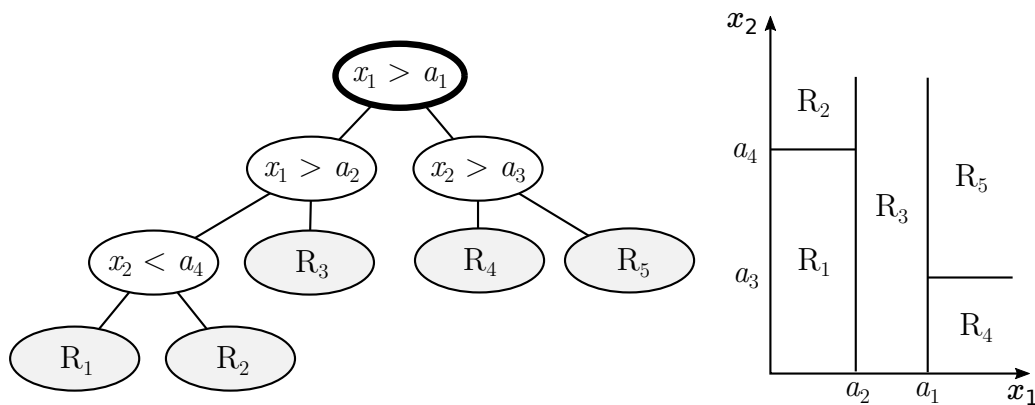
negativa. Ao atingir uma folha, rotula-se a amostra com o valor desta folha (BREIMAN et al., 1984).

As ADs são um dos poucos modelos de IA capazes de resolver tarefas de classificação com mais de dois rótulos possíveis, sem necessidade de implementação de múltiplas instâncias de classificadores binários. Apesar das ADs não terem um desempenho tão bom quanto outros algoritmos atuais de IA, como a MVS ou RNA, são consideravelmente mais rápidas e sua estrutura é de fácil visualização e compreensão (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

2.2.6.1 Definição

Seja um conjunto de treinamento dado por $\{(\mathbf{x}^{(i)}, y^{(i)})\}$, onde \mathbf{x} representa uma amostra do conjunto que possui n parâmetros ($\mathbf{x}^{(i)} \in \mathbb{R}^n$) e $y^{(i)}$ a classe a qual a amostra $x^{(i)}$ pertence, tal que $y = 1, \dots, k$, sendo k o número total de rótulos. A solução da AD é um particionamento retangular do espaço de entrada em k conjuntos disjuntos, R_1, \dots, R_k , tal que o valor previsto de y é j se $x \in R_j$, $j = 1, \dots, k$ (LOH, 2011). A Figura 24 mostra um exemplo de espaço de entrada bidimensional, com parâmetros x_1 e x_2 .

Figura 24 – Estrutura genérica de uma árvore de decisão



Fonte: Mohri, Rostamizadeh e Talwalkar (2012)

Legenda: considerando um espaço de entrada com dois atributos, a figura mostra como a estrutura de uma AD (esq.) é representada no plano cartesiano (dir.), sendo que as elipses brancas representam os nós, a elipse em negrito representa o nó raiz e as elipses cinzas representam as folhas da árvore.

2.2.6.2 Treinamento

A topologia de uma AD envolve, além da escolha dos parâmetros de entrada, o número máximo de nós, a profundidade da árvore e os limiares para cada nó. A determinação da topologia ótima a fim de minimizar os erros de classificação se torna computacionalmente impossível devido ao grande número combinatório de soluções (BISHOP, 2006).

Segundo Mohri, Rostamizadeh e Talwalkar (2012), uma maneira de contornar este problema durante a estruturação de uma AD é utilizando uma heurística simples chamada de “algoritmo voraz” (*greedy algorithm*), que consiste em realizar particionamentos localmente ótimos em cada nó da AD (MURTHY; SALZBERG, 1995).

Através deste algoritmo, a implementação da AD se inicia com somente um nó (“raiz”), a qual funciona como uma folha cujo valor inicial é a classe que detém a maioria das amostras do conjunto de treinamento. Em seguida, o melhor atributo é escolhido para um novo nó a ser questionado e particionado. A escolha deste atributo é feita através de fórmulas que retornam o grau de relevância daquele atributo para a classificação das amostras (MONARD; BARANAUSKAS, 2003). Normalmente, a escolha do parâmetro ótimo é realizada através de uma busca exaustiva, atribuindo a cada um dos parâmetros todos os seus valores possíveis dentro do domínio do problema; o parâmetro questionado que faz com que a AD retorne o menor erro (também chamado de “impureza”) é selecionado (BISHOP, 2006).

Logo, seja $P_l(n)$ a fração de pontos de classe $l \in [1, k]$ em um nó n . Baseando-se nesta relação, as três medidas mais comuns de impureza F são mostradas na tabela Tabela 4.

Tabela 4 – Fórmulas de avaliação de impureza

Nome	Fórmula
Classificação incorreta	$1 - \max_{l \in [1, k]} P_l(n)$
Entropia	$-\sum_{l=1}^k P_l(n) \log_2 P_l(n)$
Índice Gini	$\sum_{l=1}^k P_l(n)(1 - P_l(n))$

Fonte: Mohri, Rostamizadeh e Talwalkar (2012)

A cada rodada, um nó n_t é particionado de acordo com a questão q_t . O par (n_t, q_t) é estimado de tal maneira que a impureza $F(n)$ do nó seja minimizada. A diminuição da impureza após particionamento do nó n baseado na questão q é definida por

$$\tilde{F}(n, q) = F(n) - [\eta(n, q)F(n_-(n, q)) + (1 - \eta(n, q))F(n_+(n, q))], \quad (2.51)$$

onde $n_+(n, q)$ é o descendente direito de n após particionamento, $n_-(n, q)$ o descendente esquerdo e $\nu(n, q)$ a fração de amostras na região definida por n que são movidas para $n_-(n, q)$. A impureza total das folhas $n_-(n, q)$ e $n_+(n, q)$ é dada por $\eta(n, q)F(n_-(n, q)) + (1 - \eta(n, q))F(n_+(n, q))$ (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012)

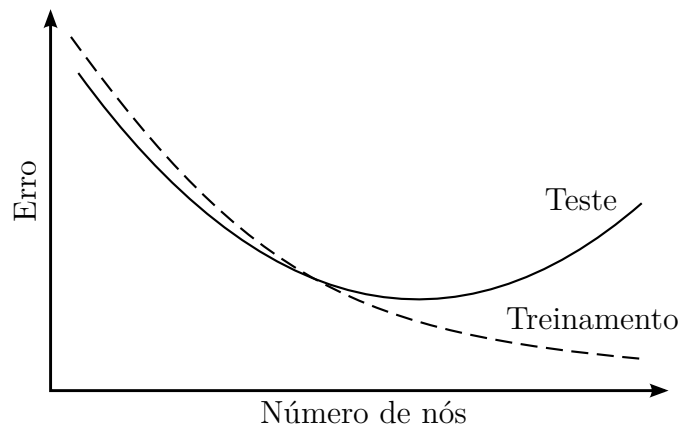
Este procedimento recursivo ocorre até uma determinada condição de parada do algoritmo, a qual está relacionada com a profundidade e precisão do classificador produzido. Esta condição é atingida quando o número de pontos por folha chegou a um número pequeno demais para futuro particionamento, quando o grau de pureza dos nós atin-

giu um determinado limiar, ou quando alguma heurística similar é obedecida (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

2.2.6.3 Soluções para sobreajuste

Após a construção da árvore utilizando o algoritmo voraz, todas as amostras do conjunto de treinamento estão possivelmente sobreajustadas (ESPOSITO et al., 1997). O sobreajuste é um dos maiores desafios quando se implementa uma AD. Se não forem impostos limites, a árvore pode chegar a uma acurácia de 100% no conjunto de treinamento, simplesmente porque, no pior dos cenários, pode acabar tão complexa a ponto de criar uma folha por amostra. A Figura 25 mostra a relação entre a complexidade de uma AD e sua tendência a sobreajustes.

Figura 25 – Relação entre número de nós de uma árvore e erros de classificação



Fonte: Monard e Baranauskas (2003)

O número mínimo de amostras para particionamento de um nó é um dos parâmetros que pode ser utilizado para controle de sobreajustes. Valores maiores previnem que um modelo aprenda parâmetros muito específicos a uma amostra selecionada pela árvore. Contudo, se este valor for alto demais, a AD pode se tornar propensa a subajustar os dados. Da mesma forma, a profundidade máxima da árvore se correlaciona com a complexidade do modelo. Uma profundidade maior leva a AD a se ramificar mais frequentemente, chegando a um número muito grande de folhas.

Alternativamente, a árvore pode ser simplificada por um algoritmo chamado “poda”. A árvore T resultante do algoritmo voraz é podada a fim de minimizar a função objetiva definida pelas restrições de generalização, composta pela soma do erro empírico de classificação e um termo de complexidade expressada em termos do tamanho \tilde{T} da árvore, que equivale ao conjunto de folhas de T :

$$G_{\lambda}(T) = \sum_{n \in \tilde{T}} |n|F(n) + \lambda|\tilde{T}|. \quad (2.52)$$

Nesta equação, $\lambda \geq 0$ é o parâmetro de regularização determinando a relação inversa entre classificações incorretas e a complexidade da árvore (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). Um exemplo de como a complexidade da AD impacta em seu potencial de generalização é mostrado na Figura 25.

Para toda árvore T' , define-se $\hat{R}(T')$ o erro empírico total $\sum_{n \in \tilde{T}'} |n|F(n)$. Procura-se então uma subárvore T_λ de T que minimize G_λ e que possua o menor tamanho possível.

Segundo Mohri, Rostamizadeh e Talwalkar (2012), é possível mostrar que T_λ admite uma única solução. Para determinar T_λ , o seguinte método de poda é empregado, o qual define uma sequência de subárvores $T^{(0)}, \dots, T^{(n)}$: começa-se com uma árvore $T^{(0)} = T$ e para $i = 1, \dots, n - 1$, define-se $T^{(i+1)}$ a partir de $T^{(i)}$ colapsando o nó interno n' de $T^{(i)}$, ou seja, substituindo a raiz da subárvore em n' com uma folha (o que equivale à combinação das regiões de todas as folhas dominadas por n'). Então, n' é escolhido de tal forma que, ao se colapsar, ocorra o menor aumento possível de $\hat{R}(T^{(i)})$ por nó, gerando a uma árvore menor $r(T^{(i)}, n')$ definida por

$$r(T^{(i)}, n') = \frac{|n'|F(n') - \hat{R}(T')}{|\tilde{T}'| - 1}, \quad (2.53)$$

onde n' é um nó interno de $T^{(i)}$. Este procedimento continua até que a árvore $T^{(n)}$ obtida possua apenas um nó.

O problema do sobreajuste nas ADs também pode ser mitigado através de uma mistura (*ensemble*) de árvores. Este método é chamado de “floresta aleatória” (*random forest*) ou FA. Neste, amostragens com reposição são feitas repetidas vezes a partir do conjunto de treinamento, onde cada amostragem é utilizada para o treinamento de uma AD diferente. Uma vez que a estrutura das ADs são bastante sensíveis aos dados de treinamento, cada árvore criada neste processo terá ramificações distintas. A combinação dos resultados através da moda, média, ou mediana dos resultados da mistura geralmente é mais robusta que a previsão de um modelo apenas.

Através desta implementação, as FAs podem ser úteis para a identificação dos parâmetros de entrada mais relevantes para o modelo. A relevância de um parâmetro é calculada pela redução total normalizada do critério utilizado para aquele parâmetro (LOUPPE et al., 2013).

2.2.7 Inteligência artificial aplicado a espectrometria gama

Segundo Wunsch, Fink e Jondral (2015), algoritmos convencionais de detecção de picos não incorporam detalhes sobre o formato dos picos para julgar se o sinal está ou não presente. Os modelos de IA podem preencher esta lacuna através do reconhecimento de padrões.

Algoritmos de análise envolvendo IA são rotineiramente usados dentro do campo da química analítica, mais especificamente no domínio da quimometria, para a interpretação de espectros de espectrometria de massa, espectroscopia do infravermelho e espectroscopia Raman (ALLEN; ELLIS; PETRIDIS, 2010). Contudo, até o momento, as ferramentas disponíveis de IA não estão incorporadas no campo da espectrometria gama, embora alguns estudos tenham mostrado resultados promissores.

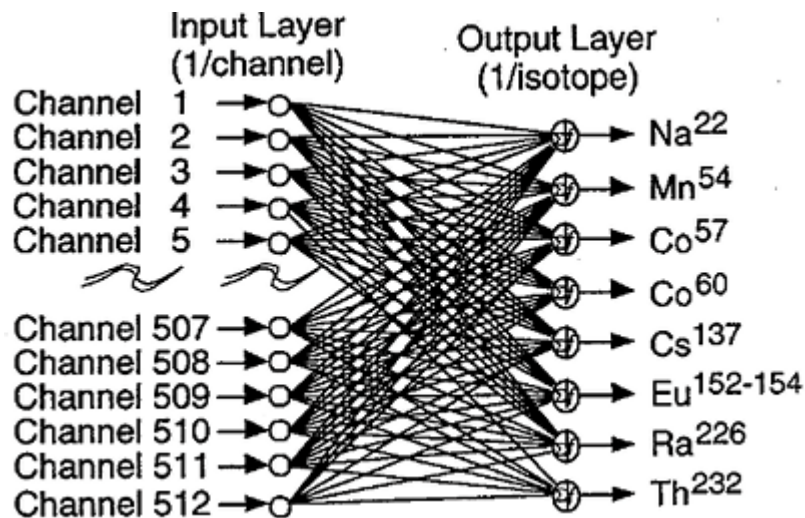
Um dos primeiros trabalhos usando IA na espectrometria gama foi apresentado por Fukuda e Kitamura (1994). Com a necessidade de resultados rápidos para o campo da segurança nuclear, o custo computacional para que os *softwares* de análise operassem em tempo real era proibitivo. A ideia de utilizar um modelo de IA compatível com o poder computacional disponível era promissora, e a escolha do modelo de IA foi um algoritmo chamado Rede Neural Associativa Linear (*Linear Associative Neural Network*) ou RNAL.

Segundo Zgurovsky e Zaychenko (2016), uma RNAL funciona de maneira análoga a uma RNA. Quando se utiliza parâmetros de entrada $x^{(i)}$, a saída desejada deve ser $y' = y(i)$. No entanto, na RNAL, quando o parâmetro de entrada é $x^{(i)} + \varepsilon$, então $y' = y(i) + \varepsilon$. Quando um vetor de entrada o qual não foi utilizado para o treinamento é apresentado à RNAL, a saída da rede retorna um resíduo. Fukuda e Kitamura (1994) utilizou este tipo de rede de forma que os dados de entrada fossem os 512 canais de um espectrômetro NaI e sua saída fosse a probabilidade da presença de um espectro não pertencente ao conjunto de treinamento. Para o treinamento da rede, os autores utilizaram aproximadamente 60 espectros de radionuclídeos naturais gerados artificialmente pelo programa EGS4, capaz de simular interações entre energia e matéria com razoável precisão pelo método Monte-Carlo (NELSON; HIRAYAMA; ROGERS, 1985). Quando o espectro continha traços de radionuclídeos antrópicos, a RNAL foi capaz de identificar sua presença. No entanto, os autores não publicaram os resultados quantitativos e parecem não ter repetido o experimento utilizando espectros gerados a partir de análises reais.

Um estudo análogo foi realizado por Keller e Kouzes (1994), utilizando RNALs para interpretação de espectros de um detector NaI. A rede elaborada aceitava como entrada os 512 canais do detector, mas os neurônios de saída indicavam a presença ou ausência de radionuclídeos individualmente. Mais especificamente, a rede possuía um neurônio de saída dedicado para o estado de cada um dos seguintes radionuclídeos: ^{22}Na , ^{54}Mn , ^{57}Co , ^{60}Co , ^{137}Cs , ^{152}Eu , ^{154}Eu , ^{226}Ra e ^{232}Th . O esquema geral de sua rede pode ser visto na figura Figura 26.

A conclusão de Keller e Kouzes (1994) é semelhante à de Fukuda e Kitamura (1994): RNALs são eficazes no reconhecimento de padrões de espectros gama, porém espectros representativos de cada radionuclídeo que compõem o conjunto de treinamento devem ser perfeitamente livre de interferentes. Caso contrário, o desempenho da rede será comprometido, embora servindo para a detecção de anomalias. Este princípio foi utilizado

Figura 26 – Uma RNA utilizada para identificar radioisótopos



Fonte: Keller e Kouzes (1994)

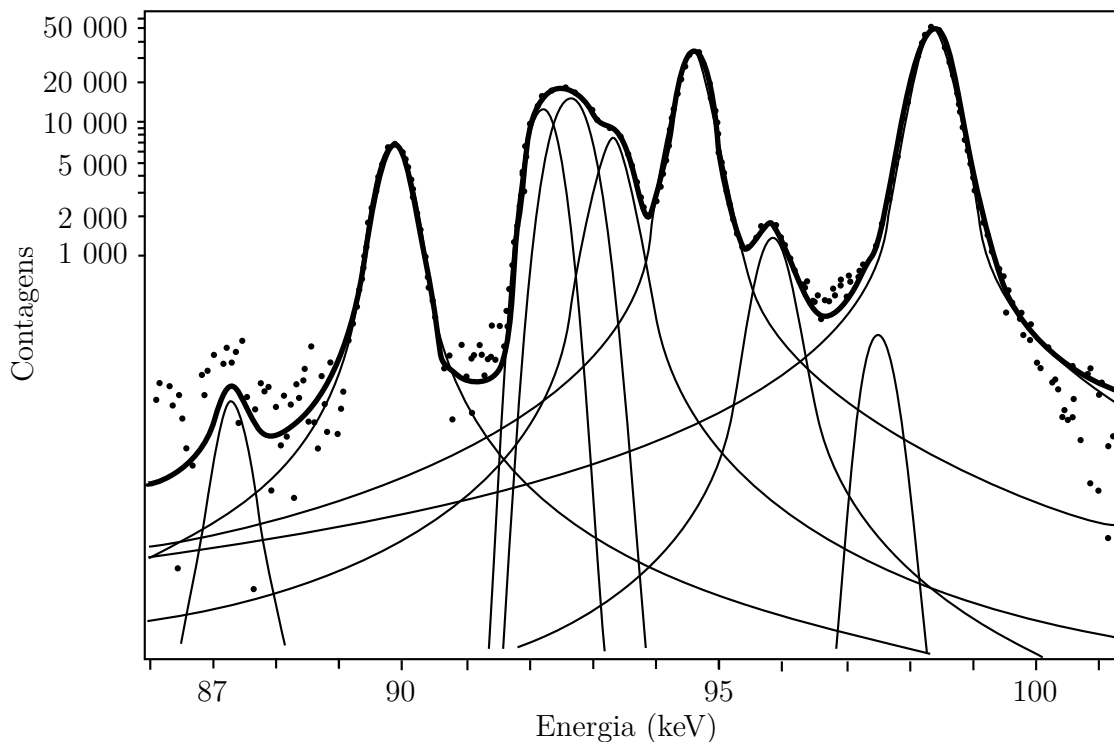
muitos anos depois por Varley et al. (2015), que propuseram algumas maneiras de detecção de anomalias em espectros gama, com foco na aplicação dessas técnicas na prevenção de atentados e tráfico de material nuclear.

Após os resultados de Keller e Kouzes (1994), o trabalho de Vigneron et al. (1996) levou o uso de RNAs em espectrometria gama um passo adiante. Em vez de um espectrômetro de 512 canais, os testes foram realizados com espectrômetros HPGe de alta resolução com 4096 canais. Além disso, os espectros eram obtidos de amostras reais, não mais gerados por um simulador. O estudo também é inédito em relação à natureza da tarefa que, em vez de um problema de classificação, foi proposta uma regressão na qual a RNA calcularia o índice de enriquecimento de urânio em pastilhas combustíveis.

O método para determinação do enriquecimento de U através de espectrometria gama é bem estabelecido e se dá normalmente pelo cálculo da área do pico do produto de decaimento do ^{238}U , o $^{234\text{m}}\text{Pa}$, de 1001 keV e o pico do ^{235}U de 143 keV (YÜCEL; DIKMEN, 2009). Ambos os picos são de baixa intensidade e, portanto, quando é sabido *a priori* que as amostras contém baixa concentração de U, picos alternativos são utilizados, como o pico de 63 keV do ^{234}Th e o pico de 186 keV do ^{235}U . No entanto, ^{226}Ra emite também um pico de 186 keV, sendo este radionuclídeo um produto de decaimento do próprio ^{238}U (EBAID; EL-MONGY; ALLAM, 2005). Uma alternativa seria o cálculo de área através de regressões não-lineares dos picos de raios gama (curvas gaussianas) e raios-X (curvas Voigt) causadas por auto-fluorescência, tanto do ^{238}U quanto do ^{235}U , em torno da região de 93 keV. No entanto, esta região é conturbada por dezenas de interferentes, como é possível observar na Figura 27.

A proposta de Vigneron et al. (1996) foi utilizar o poder de reconhecimento de padrões das RNAs para interpretar a região do espectro entre 83 e 103 keV como entrada

Figura 27 – Região do espectro composta por raios-X e gama provenientes de U e Th



Legenda: imagem de um algoritmo especializado em ajustes não-lineares às curvas presentes na região de 93 keV de espectros gama gerados por amostras contendo urânio.

Fonte: Ruellan et al. (1996)

de uma RNA de tal forma que esta retornasse, no neurônio de saída, o valor de enriquecimento de amostras de U. A quantidade de espectros reais obtidas pelos autores a partir de padrões de U, com valores certificados de enriquecimento entre 0,7% e 10%, era muito pequena (aproximadamente 60 espectros) em relação ao número de parâmetros passados para a RNA (aproximadamente 40 canais). Desta forma, foi realizada uma redução de dimensionalidade dos dados através de um técnica chamada Análise de Componente Principal (*Principal Component Analysis*) ou ACP, a qual aplica uma transformação capaz de mapear dados de um espaço de muitas dimensões para outro espaço de menor número. Assim, a camada de entrada da rede foi reduzida para apenas 6 unidades, enquanto a camada oculta possuía 3 unidades.

Os resultados de Vigneron et al. (1996) foram aceitáveis, com desvio de no máximo 3% entre os valores calculados. Os autores concluem que a determinação de enriquecimento de U por RNAs são eficientes, rápidas e apropriadas para quando não se tem modelos matemáticos definidos *a priori* para a região de análise. Vale ressaltar, no entanto, que os autores utilizaram amostras com concentrações de urânio consideravelmente altas para o experimento e a RNA foi testada com os próprios dados do conjunto de treinamento. Isto faz com que o desempenho do modelo normalmente seja superestimado.

Da mesma forma, Sharma et al. (2012) foram bem-sucedidos nas análises de ^{226}Ra em espectros obtidos em sondagens profundas de antigos locais de contaminação, cuja geometria de grandes dimensões (grandes porções de solo subterrâneo em campo aberto) seria extremamente complexa de se reproduzir com contêineres e padrões analíticos.

A utilização de inteligência artificial não se restringe somente aos dados brutos das análises espectrométricas, mas aos resultados já processados de algum método analítico estabelecido. Andrello, Appoloni e Guimarães (2003) classificaram solos a partir dos resultados de análise de ^{226}Ra , ^{238}U , ^{235}U , ^{40}K , ^{134}Cs , ^{137}Cs , ^{232}Th e ^7Be através de uma rede neural artificial. Ainda nesta linha de trabalho, vários outros estudos são citados por Dragovic e Onjia (2007), nos quais solos são classificados a partir de resultados de análises químicas e reconhecimento de padrões. Kanevski et al. (1996) afirmam que dados ambientais e ecológicos possuem tendências altamente complexas e variáveis em diferentes escalas espaciais e temporais. Tais propriedades promovem uma maior complexidade na análise e interpretação dos resultados. Tendências não-lineares, como as presentes neste ramo da ciência, podem ser abordadas de diversas maneiras: krigagem universal, krigagem residual (CRESSIE, 1986), análise de tendência de superfície ou modelagens matemáticas (VENKATRAM, 1988). No entanto, modelagens como estas são particularmente trabalhosas pela necessidade de um modelo de dispersão pelo meio em que está inserido o contaminante.

2.3 Análise de urânio em efluentes

2.3.1 Problemas Ambientais

O minério de urânio está naturalmente presente em grande parte da crosta terrestre em condições e proporções adequadas que permitem sua utilização e exploração para fins comerciais (PRADO, 2007). Sua concentração em solo brasileiro varia de 0 a 2,3 ppm (PÉREZ et al., 1998), mas também pode ser encontrado em água superficial e subterrânea, ar, plantas e animais em quantidades variadas, também tipicamente na ordem partes por milhão (OLIVEIRA, 1999).

O elemento químico urânio, com número atômico 92, pertence à série dos actínídeos na tabela periódica. Seu isótopo mais comum é o ^{238}U , com meia-vida de 4,5 bilhões de anos, compondo em torno de 99,3% do urânio encontrado na natureza (BRENNECKA et al., 2010). O segundo isótopo mais comum é o ^{235}U , que compõe apenas 0,7% do elemento encontrado naturalmente, além de possuir uma meia-vida consideravelmente menor de 0,7 bilhões de anos (SCHÖN; WINKLER; KUTSCHERA, 2004). Tanto o ^{235}U quanto o ^{238}U são precursores de uma vasta gama de outros radionuclídeos naturais, gerados através de uma cadeia de decaimento, no qual os isótopos de urânio se fissionam sucessivamente através de decaimento alfa e decaimento beta até o ponto de formação de um elemento

estável (chumbo).

O principal uso do urânio no setor civil é o abastecimento de usinas nucleares de energia (OLIVEIRA, 1999). Dada a natureza radioativa deste elemento, um quilograma de ^{235}U pode teoricamente produzir 20 terajoules de energia, assumindo fissão completa. Este valor equivale a aproximadamente 1500 toneladas de carvão (EMSLEY, 2003). Por este motivo, o minério de urânio é processado e enriquecido, isto é, submetido ao processo de separação isotópica para elevar o teor de ^{235}U , de maior atividade específica, em relação ao ^{238}U . No Brasil, o urânio está presente no setor de mineração sobretudo nas minas do município de Caetité na Bahia (LAMEGO et al., 2003); no processo de enriquecimento por centrifugação no Centro Tecnológico da Marinha em São Paulo (FAGUNDES; MORAES, 2009); e na geração de energia através das usinas nucleares de Angra I e Angra II.

Nas aplicações energéticas, o risco de contaminação ambiental de urânio está presente em todo o ciclo do combustível, desde a sua mineração até seu enriquecimento, processamento e utilização do mesmo em forma de pastilhas. Dos subprodutos gerados durante o ciclo, além daqueles produzidos dentro do reator nuclear, a maioria possui características de contaminação semelhantes à do urânio.

A contaminação de urânio no meio ambiente pode ocorrer não somente devido ao processo do ciclo do combustível, mas também por outras atividades antrópicas. Segundo Mangini et al. (1979), a fosforita, que é uma rocha sedimentar com elevado teor de urânio, pode ser usada na fabricação de fertilizantes fosfatados e facilmente adsorvidos nas camadas superiores do solo, tornando-se fontes comuns de contaminação por urânio no ambiente.

O consumo de água potável é responsável por aproximadamente 85% da ingestão total de urânio pela população (SINGH et al., 2013). Havendo contaminação interna de urânio, este acumula-se sobretudo nos ossos e nos rins. Embora o urânio possua uma concentração de radioatividade específica bem conhecida, muitos reconhecem que sua toxicidade química prevalece sobre os possíveis danos radiológicos em caso de contaminação interna em concentrações naturais (CAMARGO; MAZZILLI, 1998)(BABU et al., 2008)(SINGH et al., 2013). No entanto, caso a concentração do contaminante seja maior, o risco radiológico passa a se tornar consideravelmente mais relevante (BABU et al., 2008). Tal risco é agravado pelo fato do urânio ser um elemento radioativo de meia-vida longa.

A *International Commission on Radiological Protection*⁵ recomenda um limite máximo de urânio em água de $0,0019 \text{ mg L}^{-1}$, enquanto a Organização Mundial de Saúde recomenda um máximo de $0,015 \text{ mg L}^{-1}$ em nascentes destinadas ao consumo (WORLD HEALTH ORGANIZATION, 2011), assim como as normas do CONAMA, que também adota o valor de $0,015 \text{ mg L}^{-1}$ para consumo humano (CONSELHO NACIONAL DO

⁵ Comissão Internacional sobre Proteção Radiológica, ou ICRP

MEIO AMBIENTE, 2008).

2.3.2 Métodos analíticos para determinação de urânio

Atualmente existem diversos métodos para a determinação de urânio em amostras nas mais variadas matrizes, sendo, em geral, o primeiro passo para qualquer um deles, a separação do urânio de outros elementos interferentes através de precipitação, eletrolise ou coluna de troca iônica (ROZMARIC; IVSIĆ; GRAHEK, 2009). Os métodos mais relevantes estão relacionados na Tabela 5.

Tabela 5 – Vantagens e desvantagens dos métodos mais utilizados para análise de urânio

(continua)

Método de análise	Vantagens	Desvantagens
Ativação de nêutrons	Resultados precisos; alta resolução e sensibilidade; baixíssimo limite de detecção. (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)	Método caro; necessita de uma fonte de nêutrons; destruição da amostra; produção de rejeitos radioativos (MISDAQ; BOURZIK, 2002).
Fluorescência de raios-X	Análises rápidas (normalmente alguns minutos); boa resolução espectral. (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)	Extremamente sensível a interferências na composição da matriz da amostra, alto limite de detecção (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)
TIMS ⁶	Altíssima precisão; baixíssimo limite de detecção (ALEXANDRA, 2006)	Equipamento caro; tediosa preparação da amostra; análises demandam ambientes e cuidados especiais; não apropriado para monitoração biológica em larga escala; destruição da amostra (ALEXANDRA, 2006)
Espectrometria alfa	Baixo limite de detecção, dependendo da preconcentração e tempo de contagem; equipamento de baixo custo	Destruição da amostra; necessita separação; tediosa preparação da amostra; longo tempo de análise

⁶ Thermal ionization mass spectrometry

(conclusão)

Método de análise	Vantagens	Desvantagens
Espectrofotometria	Baixo limite de detecção (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)	Necessita altíssima quantidade de insumos; necessita separação radioquímica (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)
Fluorimetria	Baixo limite de detecção; sistema de detecção de baixo custo (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)	Necessita vários insumos; preparação das amostras é complexa; pode necessitar separação radioquímica dependendo da matriz (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)
KPA ⁷	Necessita pouca quantidade de amostra; baixo limite de detecção (ALEXANDRA, 2006)	Alto custo do equipamento; sensível a interferentes na matriz; reagentes caros e específicos
Gravimetria	Baixo custo; resultados precisos; análise rápida (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)	Altíssimo limite de detecção; utiliza insumos; necessita separação (INTERNATIONAL ATOMIC ENERGY AGENCY, 1992)

Fonte: autoria própria

2.3.3 Espectrometria gama para a análise de urânio

Como é possível observar na Tabela 5, as principais técnicas de análise de urânio em amostras podem apresentar um alto custo financeiro pela ampla utilização de insumos de laboratório, ou pelo gerenciamento e manuseio de resíduos gerados – algo que vai contra a crescente tendência no meio científico de buscar técnicas analíticas mais sustentáveis, ou seja, econômica e ecologicamente mais viáveis.

Uma das grandes vantagens da técnica de espectrometria gama é seu potencial de substituir procedimentos de separação radioquímica por medidas não-destrutivas. Além disso, é possível medir emissores gama diretamente da amostra original sem a necessidade de preparações químicas, permitindo tanto a identificação qualitativa quanto determinação quantitativa dos radionuclídeos em uma amostra (INTERNATIONAL ATOMIC ENERGY AGENCY, 1989). A possibilidade de evitar a destruição das amostras ou a utilização de insumos de laboratório reduz o impacto ambiental das análises (GREDILLA et al., 2016).

⁷ Kinetic phosphorescence analysis

2.3.4 Desafios da espectrometria gama para análise de urânio

Dentro do contexto ambiental, inúmeros são os problemas que emergem ao tentar-se analisar urânio em uma amostra através da espectrometria gama.

Primeiramente, o isótopo ^{238}U , que compõe quase inteiramente a fração de urânio total na amostra, possui baixa probabilidade de emitir raios gama durante o decaimento. Portanto, a medição de sua atividade radioativa acontece indiretamente através da medição de alguns de seus produtos de decaimento (YÜCEL; CETINER; DEMIREL, 1998). Uma relação desses produtos, comumente chamados de “filhos” do ^{238}U , está representada na Figura 28.

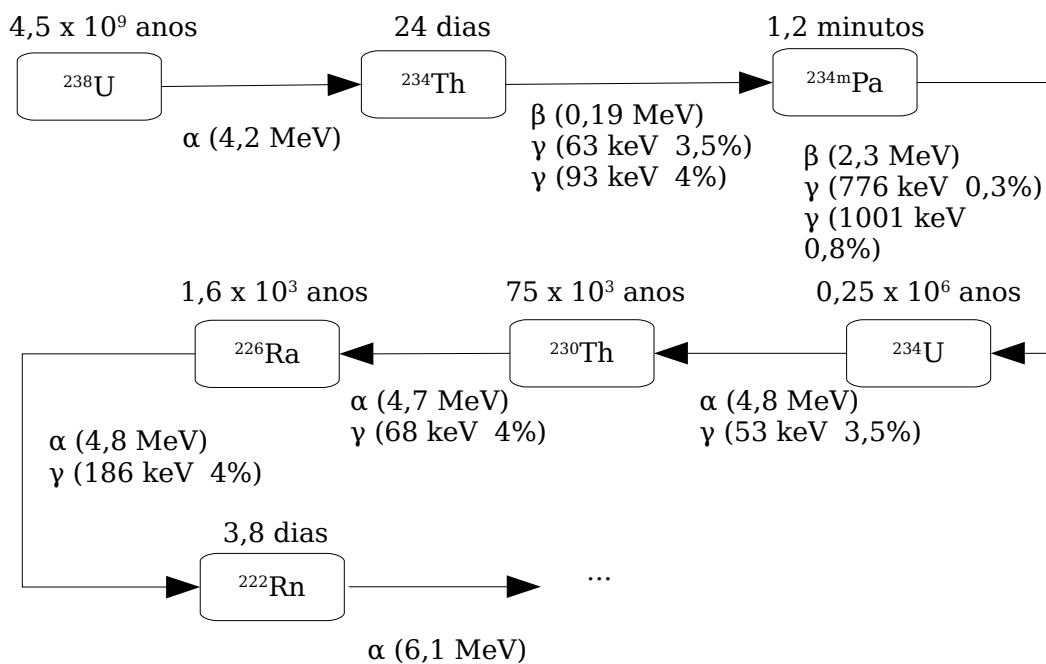


Figura 28 – Cadeia de decaimento do ^{238}U , exibida até o ^{222}Rn

Fonte: Haynes (2014)

Os radionuclídeos filhos do ^{238}U que normalmente são escolhidos para quantificação são o ^{234}Th e $^{234\text{m}}\text{Pa}$, uma vez que estes: emitem radiação gama; têm meia-vida consideravelmente mais curta e estão no topo da cadeia de decaimento, havendo portanto maior probabilidade de estarem em equilíbrio de atividade com o ^{238}U .

A escolha entre o ^{234}Th ou $^{234\text{m}}\text{Pa}$ (ou ambos) depende das condições em que se encontra a amostra. A medição através do elemento protactínio é a mais indicada apenas quando há uma concentração considerável de urânio na amostra, uma vez que as linhas de energia pertencentes ao $^{234\text{m}}\text{Pa}$ são pouco abundantes, ou seja, há uma baixa probabilidade de raios gama serem emitidos do núcleo durante o decaimento: 0,3% e 0,8% para as linhas de 776 keV e 1001 keV respectivamente (GARCÍA-TALAVERA, 2003).

Caso a linha de energia do $^{234\text{m}}\text{Pa}$ não seja estatisticamente razoável, o ^{234}Th passa

a ser o elemento escolhido para quantificação. As linhas de energia emitidas pelo ^{234}Th ocupam a janela de baixa energia do espectro, que normalmente possui dezenas de outras emissões provenientes de processos físicos diversos que ocorrem externa e internamente no detector (GARCÍA-TALAVERA, 2003).

Tabela 6 – Probabilidade de emissões gama na região de 92 a 96 keV

E (keV)	Origem	Probabilidade
92,37	^{234}Th (γ)	2,42%
92,79	^{234}Th (γ)	2,39%
93,35	^{228}Ac ($\text{ThK}_{\alpha 1}$)	0,32%
93,35	^{235}U ($\text{ThK}_{\alpha 1}$)	5,81%
93,35	$\text{ThK}_{\alpha 1}$ (auto-fluorescência)	-
93,90	^{227}Th (γ)	1,37%
94,67	^{234}Pa + $^{234\text{m}}\text{Pa}$ ($\text{UK}_{\alpha 2}$)	0,16%
94,67	$\text{UK}_{\alpha 2}$ (auto-fluorescência)	-
94,85	^{226}Ra ($\text{RnK}_{\beta 1}$)	0,17%
94,85	^{224}Ra ($\text{RnK}_{\beta 1}$)	0,08%
94,85	^{223}Ra ($\text{RnK}_{\beta 1}$)	8,7%

Fonte: García-Talavera (2003)

Tal fenômeno causa uma deformação no formato dos picos pertencentes ao ^{234}Th , causando dificuldade no processo de detecção e identificação por parte dos algoritmos de análise.

Além do problema causado pela própria natureza dos radionuclídeos em questão, há também o fato de que amostras com baixa concentração de urânio (em torno de 1 ppm) produzem picos de intensidades próximas ao limite de detecção do sistema, fato este que resulta em uma quantidade improdutiva de falso-positivos e falso-negativos (GILMORE, 2008).

Somam-se a essas adversidades o estado atual de desenvolvimento dos softwares de análise de espectrometria gama. Um dos softwares mais utilizados para espectrometria gama na área nuclear é o Genie 2000, fabricado pela Canberra Industries (ZAHN; GENEZINI; MORALLES, 2009). Mesmo liderando a preferência entre instituições acadêmicas e industriais da área nuclear, seus principais algoritmos de busca e análise de picos em espectro são limitados e não fazem uso do potencial computacional de sistemas atuais; como exemplo, seu principal algoritmo de varredura de picos é uma versão readaptada de algoritmos implementados há mais de 30 anos (CANBERRA, 2006).

Tendo em vista os entraves presentes nas análises de amostras com concentração de urânio próximas aos limites de detecção dos sistemas de espectrometria gama, algoritmos de análise mais capazes e que façam uso dos recursos computacionais atuais se mostram extremamente interessantes. Uma das possibilidades para lidar com os padrões problemá-

ticos nos espectros é a utilização de técnicas de aprendizado de máquina (KELLER et al., 1996).

3 OBJETIVOS

A espectrometria gama é uma técnica não-destrutiva e eficiente, como também, em certos aspectos, mais sustentável do que outras técnicas para análise de radionuclídeos. No entanto, os algoritmos tradicionais para radiometria gama têm desempenho comprometido quando se trabalham com amostras ambientais cuja concentração de radionuclídeos é naturalmente baixa, ou quando se deseja detectar emissores gama problemáticos como o urânio. Em instituições onde efluentes radioativos são gerados, uma detecção rápida e precisa dos radionuclídeos de interesse é necessária para prevenir possíveis impactos ambientais.

Deste modo, o objetivo geral deste trabalho foi elaborar um algoritmo para detecção de radionuclídeos em espectros gama caracterizados pela baixa relação sinal-ruído, como normalmente ocorre em amostras ambientais de baixa radioatividade. A ideia básica é que modelos de inteligência artificial utilizassem o histórico de análises anteriores, assim como informações espectrais descartadas por algoritmos de detecção convencionais, para identificar radionuclídeos presentes em amostras com resolução pobre e intenso ruído de fundo.

3.1 Objetivos específicos

Primeiramente, foi avaliada a viabilidade de abordagens baseadas em inteligência artificial para a interpretação de espectros gerados artificialmente, de forma a simular diversos cenários que poderiam ocorrer na análise pelo sistema de espectrometria gama.

Em seguida, pretendeu-se utilizar espectros reais para este estudo de viabilidade. Foi estabelecido um ambiente controlado onde uma fonte de radiação gama com emissão mono-energética permaneceu na vizinhança de um detector para um determinado número de aquisições de espectro e, em seguida, o mesmo número de aquisições foi realizada sem a presença da fonte radioativa. Esperou-se que os modelos de aprendizado de máquina obtivessem um desempenho melhor do que os algoritmos clássicos para a identificação de espectros e com menor número de positivos e negativos falsos.

Uma vez que se obteve sucesso na etapa anterior, o próximo objetivo foi validar os modelos de inteligência artificial através da detecção de urânio em concentração acima de um determinado limite arbitrariamente estabelecido neste trabalho.

4 MATERIAIS E MÉTODOS

4.1 Implementação dos modelos de aprendizado de máquina

Os classificadores escolhidos para este trabalho foram dois modelos paramétricos (MVS e RNA) e um não-paramétrico (AD). Os modelos foram implementados neste trabalho na linguagem de programação Python versão 2.7.10, utilizando a biblioteca Scikit-Learn 0.18, que contém um extenso conjunto de algoritmos voltados para o desenvolvimento de sistemas de inteligência artificial (PEDREGOSA et al., 2011). Os modelos de inteligência artificial (IA) foram executados em processador AMD 3.0 GHz de 6 núcleos com sistema operacional Linux de 64 bits (Fedora 23). Os hiperparâmetros utilizados para a configuração dos modelos são mostrados na Tabela 7.

Tabela 7 – Modelos de IA e hiperparâmetros utilizados neste trabalho

Modelo	Hiperparâmetro	Valor
AD	Número mínimo para particionamento	10%
	Profundidade máxima	7
	Critério de impureza	Gini
	Mínimo de amostras por folha	1
	Número máx. de parâmetros avaliados por nó	\sqrt{m}
	Impureza mínima para particionamento	10^{-7}
MVS	C (parâmetro de penalização)	1.0
	Núcleo	linear
RNA	Neurônios na primeira camada oculta	64
	Neurônios na segunda camada oculta	12
	Função de ativação	tanh
	Termo de regularização	6.0
	Otimizador das conexões	Adam*
	Número máximo de iterações	2000
	Embaralhar amostras a cada iteração	Sim
	ε (parâmetro de escala do otimizador)	10^{-8}

Fonte: autoria própria

Legenda: * Kingma e Ba (2014)

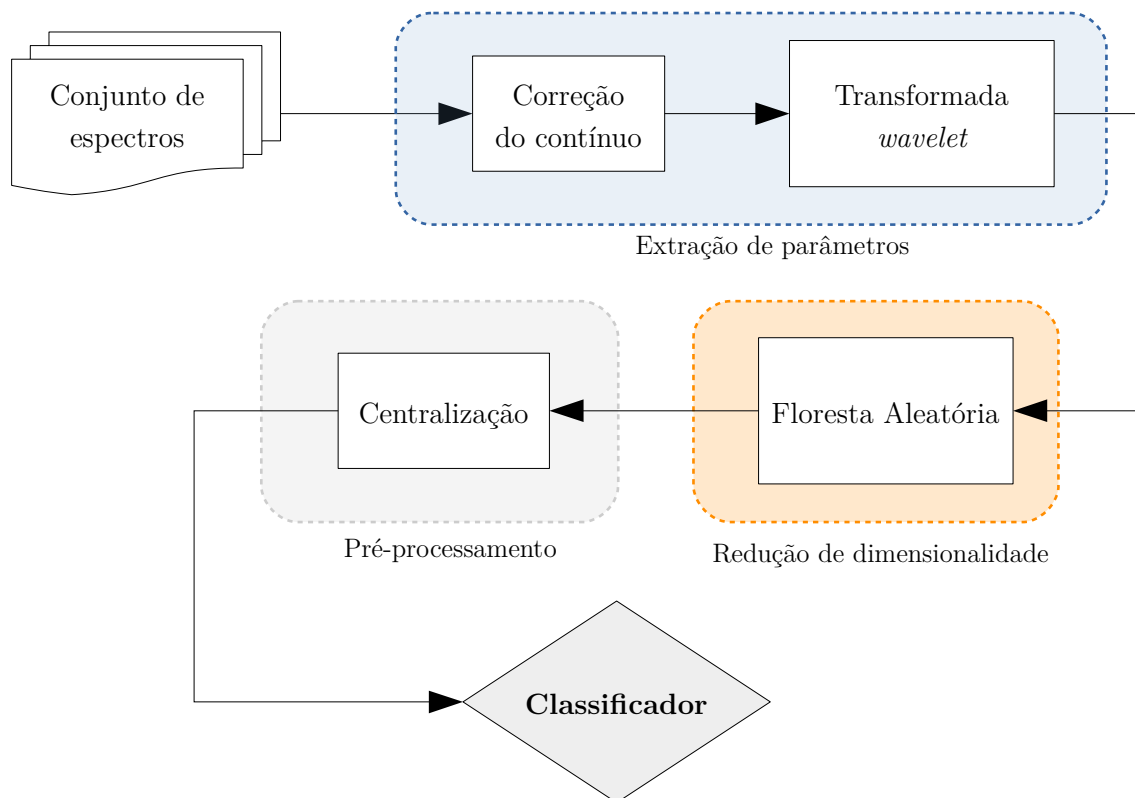
A escolha destes hiperparâmetros se baseou na busca exaustiva por combinações de valores que minimizassem a métrica de acurácia através de validação cruzada por k -dobras com $k = 10$. Os espectros utilizados para a realização dos testes foram artificialmente

gerados contendo picos com SNR entre 2 e 4 dB. Detalhes sobre o procedimento da otimização dos hiperparâmetros são encontrados no Apêndice B.

4.1.1 Pré-processamento

A Figura 29 mostra uma visão geral do processamento dos espectros até a etapa de classificação.

Figura 29 – Etapas de tratamento dos dados, incluindo pré-processamento, extração de parâmetros e redução de dimensionalidade



Fonte: autoria própria

As diferentes intensidades do contínuo nos espectros, que podem ser provenientes de interferências diversas nas amostras, prejudicam o desempenho dos modelos de AM caso não sejam corrigidas. Como etapa inicial de pré-processamento, o contínuo dos espectros foi estimado através de um ajuste não-linear assimétrico e, em seguida, subtraído dos espectros originais.

Em seguida, extraiu-se novos parâmetros de entrada a partir dos espectros originais, com a premissa de fornecer informações extras sobre os espectros aos modelos de IA. A extração de novos parâmetros foi realizada através de uma transformada *wavelet* contínua utilizando o operador de Ricker, variando-se a escala do operador entre os valores de 1 a 4.

A terceira etapa do pré-processamento, antes da seleção dos parâmetros mais relevantes, envolveu a centralização dos parâmetros de entrada em torno de suas médias de acordo com a Equação 2.19.

A redução de dimensionalidade do modelo foi feita através da seleção dos parâmetros mais relevantes de entrada, descartando os parâmetros redundantes que não contribuem para o desempenho do modelo de IA. Para esta tarefa, utilizou-se uma floresta aleatória de 100 ADs para análise da importância relativa dos parâmetros de entrada e seleção dos 21 parâmetros mais relevantes. Este número foi escolhido baseando-se no formato dos picos a serem analisados, de forma que os modelos de IA tivessem a oportunidade de utilizar como parâmetros de entrada o centróide de um pico e 10 canais à esquerda e direita. Esta região deve encompassar $1,25 \times \text{FWHM}$ mais 4 ou 5 canais de contínuo de fundo como referência.

Maiores detalhes sobre cada etapa do processo de pré-processamento se encontram no Apêndice A.

4.2 Métodos tradicionais de detecção

Como referência, os modelos de IA foram comparados com alguns métodos tradicionais de detecção utilizados na espectrometria gama: método derivativo e método por correlação com biblioteca. Os parâmetros de configuração para os métodos tradicionais (limiares, limites de sensibilidade etc.) foram inicialmente aqueles sugeridos pelo manual do *software* Genie 2000 (CANBERRA, 2006). Posteriormente, os limiares foram ajustados e otimizados de forma que os métodos testados tivessem a maior acurácia possível.

4.2.1 Método derivativo

A técnica de varredura de picos baseada no método derivativo (Deriv) foi a chamada “segunda diferença generalizada”, proposta por Mariscotti (1967), aprimorada por Koskelo, Aarnio e Routti (1981) e utilizando as modificações e constantes propostas por CANBERRA (2006). Os valores de limiares utilizados foram os mesmos sugeridos no manual do Genie 2000, uma vez que este *software* possui uma implementação deste método. Diferente do método de correlação por biblioteca, que retorna um valor de probabilidade entre 0 e 1 indicando a presença do pico em questão, a implementação original do método derivativo retorna apenas o valor de significância do pico identificado, que corresponde à amplitude do sinal resultante da segunda diferenciação do espectro em relação ao desvio padrão. Este valor pode variar no intervalo $[0, +\infty]$. Deste modo, optou-se por fazer uma pequena modificação no algoritmo para que um valor de probabilidade fosse calculado entre 0 e 1. Esta modificação não alterou fundamentalmente o funcionamento do método, funcionando apenas como complemento para que se pudesse obter uma estatística apro-

priada para comparações com os modelos de IA através da curva ROC. Os detalhes da implementação do algoritmo e de sua modificação se encontram em Anexos seção A.1.

A configuração deste algoritmo de detecção depende do valor limiar da segunda derivada escolhido pelo analista (sendo -3.0 o padrão sugerido pelo *software*), acima do qual todo sinal é ignorado e acima do qual todo sinal é considerado como pico encontrado no espectro.

4.2.2 Método de correlação por biblioteca

O algoritmo implementado para o método de correlação com biblioteca (CorrBib) se encontra no trabalho de Russ (2007). Este algoritmo exige que o espectro possua curvas de calibração de energia e formato do pico bastante precisas, portanto se aferiu semanalmente a estabilidade do sistema, realizando novas calibrações com uma fonte de ¹⁵²Eu sempre que necessário.

A decisão sobre a presença ou ausência dos picos relacionados na biblioteca são feitas calculando-se a correlação entre as contagens do espectro e a curva gaussiana esperada para os picos naquela região, caso presentes. O limiar de correlação utilizado foi de 0.80, como sugerido pelo manual do *software* Genie 2000. Detalhes sobre a implementação do algoritmo podem ser encontrados em Anexos seção A.2.

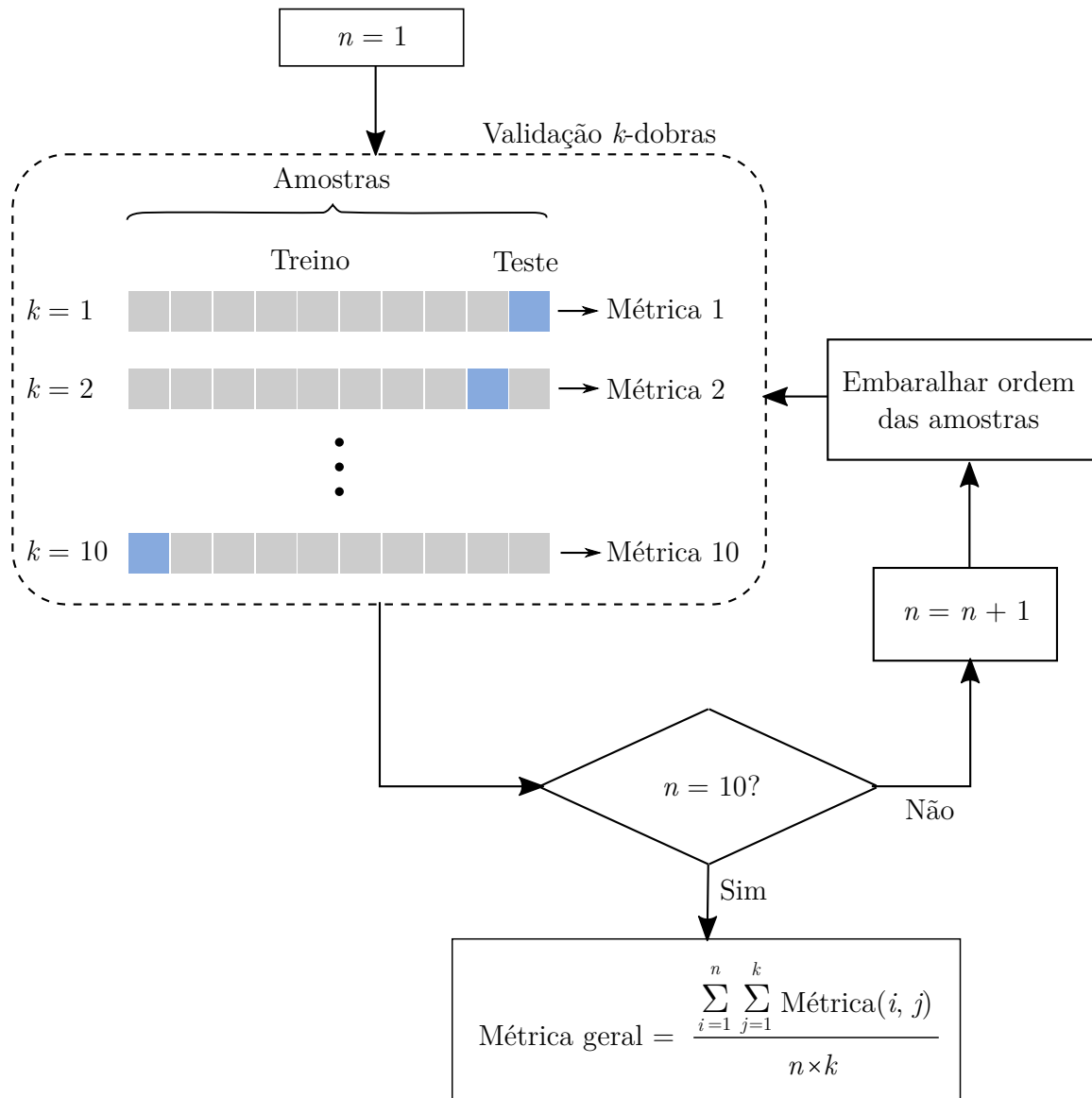
4.3 Validação cruzada

Seguindo as recomendações de Dietterich (1997), todos os experimentos deste trabalho foram realizados sob regime de validação cruzada do tipo k -dobras com $k = 10$ repetidas 10 vezes, totalizando 100 rodadas. O diagrama da Figura 30 apresenta as etapas da validação cruzada. Os testes de hipótese para averiguar se um método foi estatisticamente melhor do que outro foi o teste t pareado, porém com $df = 10$ para controlar a taxa de erro tipo I e a repetibilidade dos experimentos (BOUCKAERT, 2003).

4.4 Experimento com espectros artificiais

Como forma de avaliar a viabilidade de usar modelos de IA na detecção de picos em espectros gama, um programa em Python foi desenvolvido para que simulasse, de forma simplificada, espectros adquiridos por um sistema de espectrometria dentro de vários cenários diferentes. A vantagem dos espectros simulados é a possibilidade de se estudar casos específicos que seriam impossíveis sem a geração de uma grande quantidade de amostras radioativas. Por exemplo, para o cenário onde há 1000 espectros de treino com picos de intensidades diferentes entre si, a geração deste mesmo número de amostras seria inviável. Entretanto, o uso de espectros simulados permitirá que se faça

Figura 30 – Validação cruzada do tipo k-dobras sob regime 10x10



Fonte: Autoria Própria

Legenda: Na validação cruzada do tipo k-dobras, as amostras disponíveis são separadas em k porções (ou dobras) com tamanhos iguais. Efetua-se então k rodadas de treino, onde em cada rodada, uma dobra diferente é segregada para testes. Para se obter uma estatística menos enviesada, recomenda-se repetir este procedimento diversas vezes, embaralhando a ordem das amostras em cada repetição.

uma avaliação prévia dos modelos e as melhores estratégias para maximização de seus respectivos desempenhos, deixando as amostras reais para a validação final. As contagens de um pico simulado são distribuídas de maneira gaussiana em torno do canal escolhido para o centróide do pico. Apesar dos picos de espectros gama possuírem leve caudamento na parte de baixa energia, a função gaussiana é uma razoável aproximação do efeito real (HELMER; LEE, 1980).

A Tabela 8 resume as estatísticas utilizadas para a criação dos espectros de cada cenário.

Tabela 8 – Estatísticas utilizadas para geração de espectros artificiais

Variável	Cenário A	Cenário B	Cenário C	Cenário D
SNR (dB)	3	1 a 3	1 a 3	1 a 6
Fundo (cont./canal)	40	40 a 80	200 a 300	40 a 80
Espectros gerados	1000	1000	1000	1000
Fração com pico principal	50%	50%	50%	50%
Fração com picos espúrios	0%	5%	0%	0%

Fonte: autoria própria

A definição de SNR utilizada é dada pela seguinte expressão:

$$\text{SNR} = 10 \times \log_{10} \frac{n}{\sqrt{n + 2B}}, \quad (4.1)$$

onde n representa a área do pico de interesse (contagem bruta do pico subtraindo-se o contínuo) e B representa a contagem bruta do contínuo sob a área do pico.

4.4.1 Cenário A: Contínuo constante, SNR constante

Neste cenário, foram criados 1000 espectros que consistiam de um vetor nulo com 100 canais, no qual um contínuo de fundo foi somado (média de 40 contagens por canal). As contagens de fundo foram geradas através de um amostrador de números aleatórios da distribuição de Poisson, tal como ocorre em eventos nucleares. Em 50% dos espectros gerados, foi somado um pico de SNR = 3 dB e área calculada para esta SNR. A posição de cada contagem do pico foi distribuída em torno do canal 50, sendo que o desvio de cada contagem em relação ao centróide foi aleatoriamente amostrado por uma distribuição normal.

4.4.2 Cenário B: Contínuo variado, SNR variado, picos espúrios

O contínuo de um espectro surge devido ao efeito de retroespalhamento de emissões gama de maior energia e sua intensidade é proporcional à taxa de emissões gamas de diversas energias observadas pelo detector. Desta forma, amostras contendo diferentes

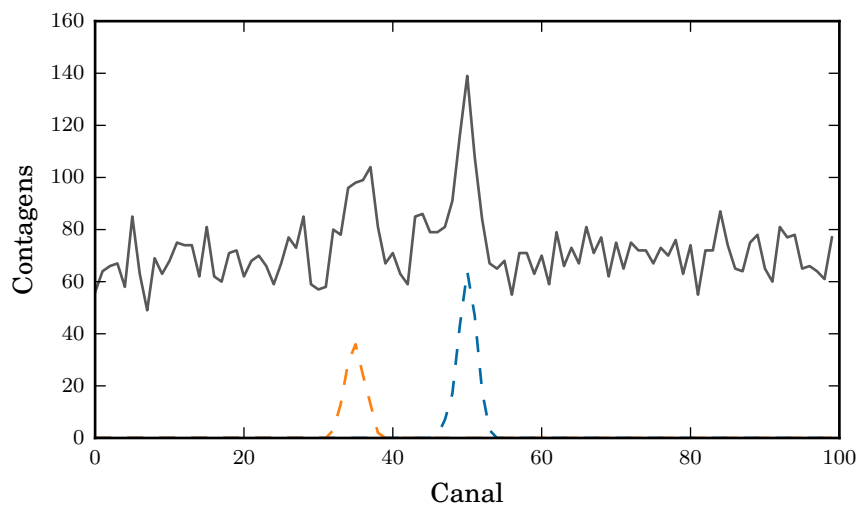
concentrações de radionuclídeos e interferentes naturais que variam de acordo com a sazonalidade, mostrarão variação na altura do contínuo do espectro.

Neste cenário, tentou-se aproximar os espectros simulados a situações reais e com um número maior de empecilhos aos modelos. O algoritmo gerou contínuo de fundo de altura variável de espectro para espectro. A média da contagem de fundo por canal variou uniformemente por espectro (entre 40 e 80 contagens). Foram gerados 1000 espectros de forma que, aleatoriamente, 50% havia picos com SNR variando entre 1 e 3 dB. Reescrevendo a Equação 4.1 para estimar a área do pico em função de SNR e B , temos

$$n = 10^{0,1\text{SNR}} + \frac{\sqrt{8 \times B \times 10^{0,2 \times \text{SNR}} + 10^{0,4 \times \text{SNR}}}}{2}. \quad (4.2)$$

A 5% dos espectros foram adicionados picos espúrios, simulando a presença de radionuclídeos inesperados nas amostras. Estes tiveram valores aleatórios entre 100 e 200 contagens, embora não se sobrepondo ao pico de interesse de classificação. Um exemplo é mostrado na Figura 31.

Figura 31 – Exemplo de espectro artificialmente gerado



Fonte: autoria própria

Legenda: A linha azul tracejada representa o pico principal, o qual deverá ser classificados pelos modelos. A linha laranja tracejada representa picos espúrios que surgiram intencionalmente em uma pequena parcela dos espectros gerados. Os espectros artificialmente gerados possuem contínuo de fundo com contagens amostradas de uma distribuição de Poisson. A média desta distribuição variou uniformemente entre cada espectro, assim como a intensidade dos picos nos espectros que os continham.

4.4.3 Cenário C: Contínuo e SNR com valores inéditos

Este cenário teve como objetivo testar a capacidade dos modelos de classificar corretamente espectros que com intensidades de picos e contínuos que não foram vistos

durante a fase de treinamento. Para isto, 1000 novos espectros foram gerados mas utilizados somente na fase de teste dos modelos. O treinamento ocorreu com os espectros do cenário anterior. Nestes espectros, o contínuo variou entre 200 e 300 contagens por canal.

4.4.4 Cenário D: Contínuo variável, SNR maiores

A quarta etapa do experimento testou a capacidade dos modelos quando depurados com picos de intensidades maiores. Em condições reais, espera-se que um bom classificador seja capaz de classificar tão bem ou melhor espectros que contenham picos maiores do que os utilizados para a fase de treinamento. Nesta fase, o SNR dos picos variou entre 1 e 6 dB.

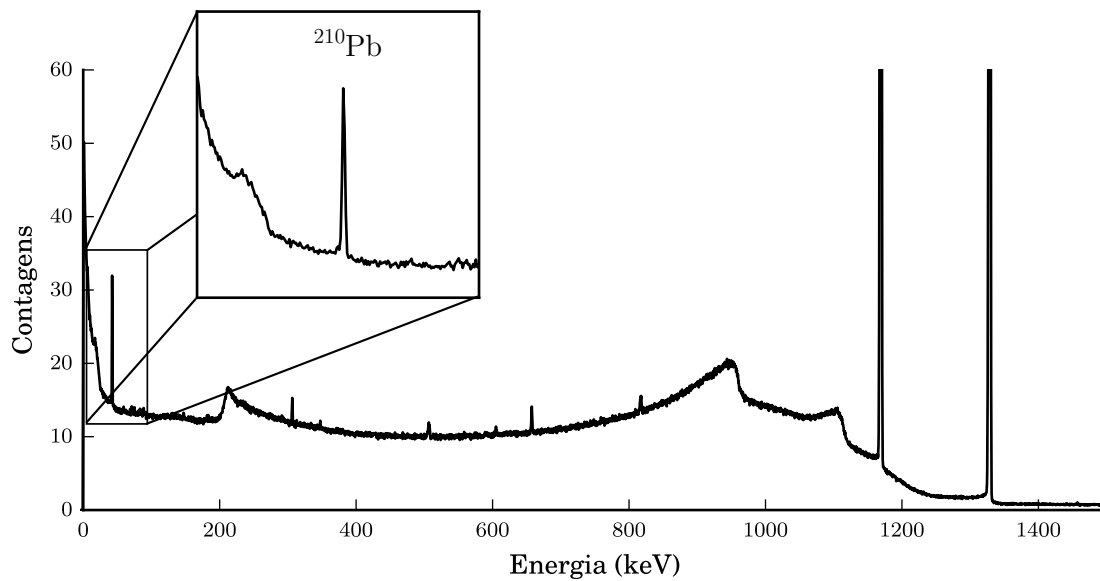
4.5 Teste com espectros reais de ^{210}Pb

Para se certificar de que os modelos de IA testados anteriormente em espectros artificiais são também apropriados para a mesma tarefa em espectros reais, adquiriu-se milhares de espectros de uma fonte radioativa real através de um espectrômetro gama de alta resolução.

Uma fonte pontual de ^{210}Pb de 635 ± 12 Bq fornecida pelo Instituto de Radioproteção e Dosimetria (IRD, Rio de Janeiro-RJ) foi colocada em contato com a janela superior do detector. A fonte de ^{210}Pb foi escolhida por ser um radionuclídeo mono-energético e, dada sua baixa energia de 46,5 keV e baixa probabilidade de emissão de 4,2% (HELMER; LEUN, 2000), produzir um pico em uma região particularmente ruidosa do espectro, disputando espaço com ruído eletrônico inerente ao detector e raios-x da blindagem e do ambiente (Figura 32).

Uma fonte pontual de ^{60}Co (Amersham) de 976 ± 59 Bq foi intencionalmente deixada dentro da blindagem do detector durante todo o experimento para que o contínuo sob o pico de ^{210}Pb se intensificasse. Desta forma, a razão sinal-ruído é drasticamente reduzida. O intuito da elevação do contínuo de fundo é tornar a detecção de picos mais difícil e providenciar um cenário, embora controlado, mais próximo das reais dificuldades nas análises ambientais. Apesar das linhas de emissões do ^{60}Co (1173 keV e 1332 keV) estarem distantes da emissão única do ^{210}Pb (46,5 keV), a faixa de baixa energia do espectro se torna rapidamente preenchida pela radiação de frenagem (*Bremsstrahlung*) oriunda da interação entre raios gama do ^{60}Co e a matéria. A disposição das fontes no detector é mostrada na Figura 33.

Os parâmetros de entrada dos modelos de IA foram a região de interesse em torno do pico de ^{210}Pb , que foi limitada em 100 canais em torno da emissão de 46,5 keV.

Figura 32 – Localização da emissão energética da fonte de ^{210}Pb nos espectros adquiridos

Fonte: autoria própria

Legenda: O espectro exibido é composto pela média de centenas de espectros de ^{210}Pb adquiridos para este trabalho. O ^{210}Pb é um emissor monoenergético cujo raio gama possui 46,5 keV, localizado quase no limite da detecção de baixa energia do sistema. Os dois picos proeminentes à direita do espectro são emissões da fonte de ^{60}Co , intencionalmente deixada próxima ao detector para a geração de radiação de frenagem na parte de menor energia do espectro.

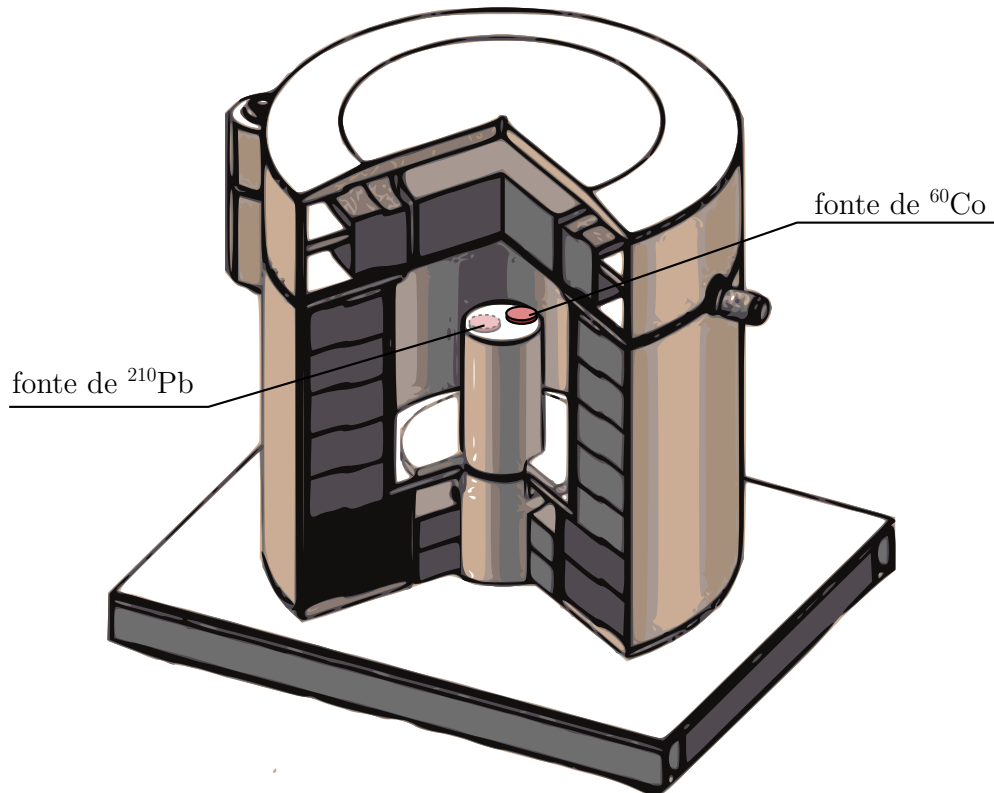
4.6 Checagem de estabilização

O detector utilizado para o experimento foi monitorado durante um mês antes do início e durante o experimento. Uma fonte pontual de ^{241}Am foi colocada próxima ao detector e seu espectro foi adquirido por 500 segundos semanalmente, tendo a resolução e posição dos picos registrados. O ^{241}Am emite radiação gama de 59,9 keV, que é suficientemente próxima à energia de emissão do ^{210}Pb e, portanto, é apropriado para avaliar o comportamento do equipamento na região de baixa energia. O propósito da aferição semanal foi garantir que o formato do espectro permaneceu estável, não sofrendo desvios maiores que 1% durante todo o decorrer do experimento.

4.7 Aquisição dos espectros

O espectro de um branco (detector sem a presença da fonte de ^{210}Pb , mas com a fonte de ^{60}Co) foi adquirido por 55000 segundos para avaliação da taxa de contagem de fundo na região de interesse do espectro. A taxa registrada foi de 0,33 cps. Em seguida, um espectro de 55000 segundos foi adquirido para avaliar a taxa de contagem líquida do pico de ^{210}Pb , que foi registrada em 0,17 cps. A eficiência do detector calculada para esta

Figura 33 – Disposição das fontes radioativas no detector



Fonte: autoria própria

Legenda: Um detector de germânio hiperpuro (HPGe) foi utilizado para o experimento. Para ter controle sobre interferências no espectro, optou-se por manter a blindagem do detector fechada e utilizar uma fonte de cobalto como fonte de interferência.

energia e posição da amostra é

$$\varepsilon = \frac{0,17 \text{ cps}}{635 \text{ Bq}} = 0,00021. \quad (4.3)$$

Com estes dados, foram escolhidos três limites de tempo distintos para que fossem registrados picos de diferentes SNR. Os picos adquiridos tiveram as características mostradas na Tabela 9.

Tabela 9 – Características dos grupos de espectros que foram adquiridos

Grupo	n	Tempo (s)	Área (cont.)	SNR (dB)	L_C (Bq)	L_D (Bq)
SNR-BAIXA	2000	96	17	3	645	1438
SNR-MÉDIA	140	150	26	4	507	1111
SNR-ALTA	600	270	48	5	388	846

Fonte: autoria própria

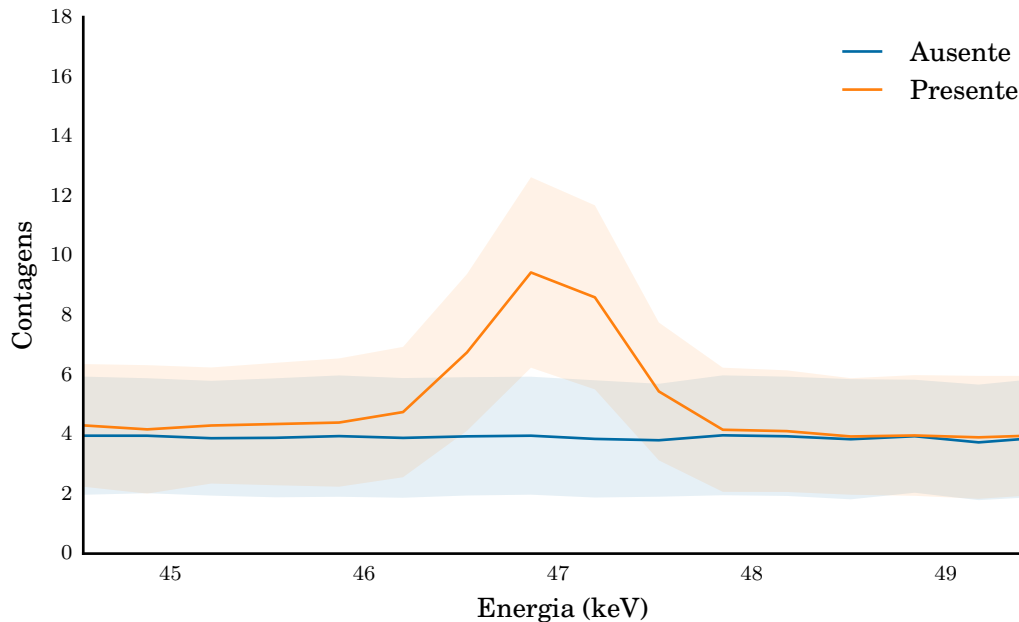
Um algoritmo foi elaborado para adquirir automaticamente os espectros continuamente, sendo que metade destes foram adquiridos com a fonte de ^{210}Pb e ^{60}Co (classe

“presente”) e a outra metade somente com a fonte de ^{60}Co (classe “ausente”).

A aquisição dos espectros de ^{210}Pb SNR-BAIXA ($n = 2000$) ocorreu ao longo de 20 dias, enquanto SNR-MÉDIA ($n = 140$) e SNR-ALTA ($n = 600$) ao longo de 15 dias. A aquisição ocorreu em dias não-consecutivos devido à rotina do laboratório utilizado para o experimento. A função canal-energia do sistema de espectrometria foi re-calibrada com uma fonte de ^{152}Eu sempre que uma nova bateria de espectros de ^{210}Pb fosse adquirida.

A aparência média dos espectros pertencentes à classe “presente” e “ausente”, adquiridos pelo detector gama na presença e ausência de uma fonte de ^{210}Pb , é sumarizada na Figura 34.

Figura 34 – Aparência média dos espectros



Fonte: autoria própria

Legenda: O gráfico sumariza a região de interesse dos 2000 espectros SNR-BAIXA adquiridos pelo espectrômetro gama, onde a linha central representa a média e a área esmaecida ao seu redor representa o desvio padrão das contagens em cada canal da região de interesse. Esta região abrange todo o pico (centróide $\pm 1,25$ FWHM) e uma margem com ± 3 canais de fundo. Devido à baixa contagem, os altos índices de incerteza de cada classe se sobrepõem.

4.8 Teste com amostras de urânio

Para a validação do método de detecção de radionuclídeos de interesse proposto neste trabalho, decidiu-se realizar testes em espectros provenientes de amostras reais acrescidas de uma determinada alíquota de padrão de $\text{UO}_2(\text{NO}_3)_2$ (nitrato de uranila). O objetivo do teste é avaliar a capacidade dos modelos de IA de identificarem a concentração de urânio maior ou igual a $300 \mu\text{g L}^{-1}$, um limiar escolhido arbitrariamente.

4.8.1 Preparação das amostras

Foram preparados quatro *beakers* do tipo marinelli contendo 3 L de água destilada. Diluiu-se diferentes alíquotas de padrão de nitrato de urânio de tal forma que os *beakers* tivessem a seguinte configuração (Tabela 10):

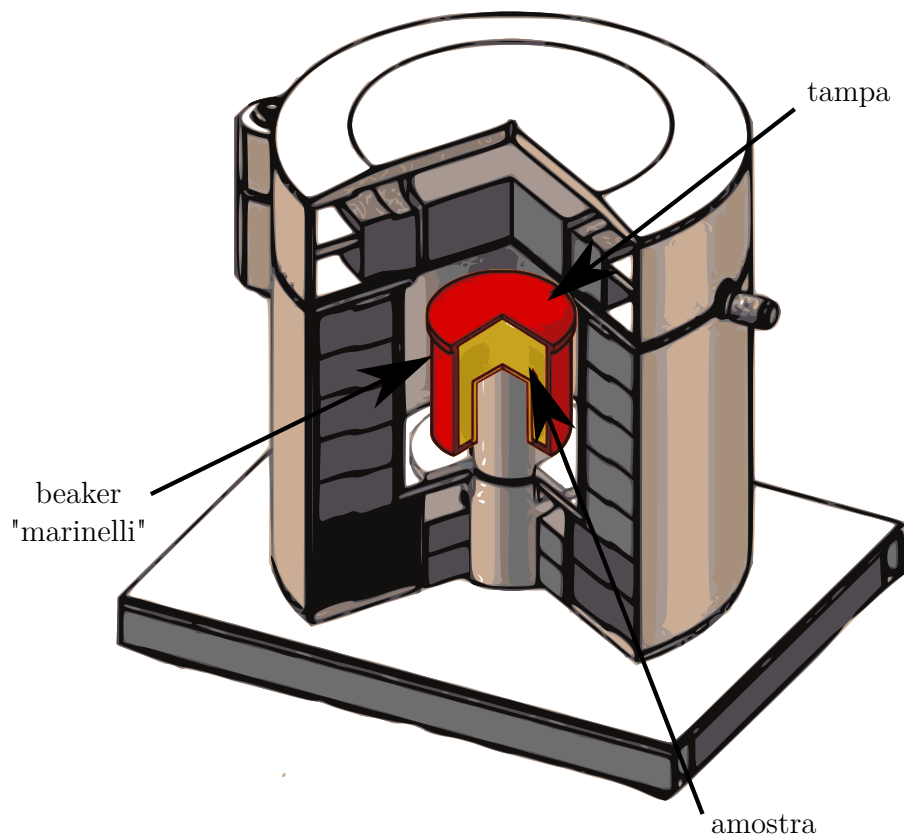
Tabela 10 – Concentração de urânio nas amostras preparadas

Identificação da Amostra	Concentração ($\mu\text{g L}^{-1}$)
Branco	0
Amostra A	300
Amostra B	600
Amostra C	900

Fonte: autoria própria

A figura Figura 35 mostra o esquema da posição da amostra no detector. As amostras foram analisadas durante 5000 segundos no espectrometro. As contagens foram repetidas 25 vezes para cada amostra.

Figura 35 – Disposição da amostra de urânio no detector



Fonte: autoria própria

Legenda: Um detector de germânio hiperpuro (HPGe) foi utilizado para o experimento e a amostra foi contida em um beaker do tipo marinelli de 3 L apropriado para amostras líquidas.

Assim como no experimento com ^{210}Pb , o detector foi monitorado durante um mês antes do início e durante o experimento com uma fonte pontual de ^{241}Am colocada próxima ao detector.

4.8.2 Região de interesse do espectro

O radionuclídeo predominante em amostras de urânio é o ^{238}U , compondo cerca de 99,2% dos isótopos caso a amostra possua proporção isotópica ambiental, isto é, caso a amostra seja de urânio natural não enriquecido.

Para amostras naturais de urânio, a quantificação pela espectrometria gama não é realizada através do ^{238}U , uma vez que este radionuclídeo não é um emissor abundante de radiação gama, mas através dos produtos de seu decaimento: ^{234}Th e $^{234\text{m}}\text{Pa}$. Tais radionuclídeos emitem radiação gama durante o decaimento (53,6 keV e 1001 keV). Este método é apropriado desde que a amostra permaneça em repouso durante tempo suficiente para alcançar equilíbrio secular. Na física nuclear, diz-se que um radionuclídeo produto de decaimento de outro radionuclídeo está em equilíbrio secular quando a quantidade de isótopos radioativos permanece constante devido à sua taxa de produção (isto é, devido ao decaimento do radionuclídeo originário) ser igual à taxa de decaimento. Considera-se que os radionuclídeos estejam em equilíbrio secular a partir de aproximadamente 6 meias-vidas do radionuclídeo produzido. O padrão de urânio utilizado foi fabricado há 1 ano, portanto os radionuclídeos produzidos pelo urânio estão em equilíbrio secular.

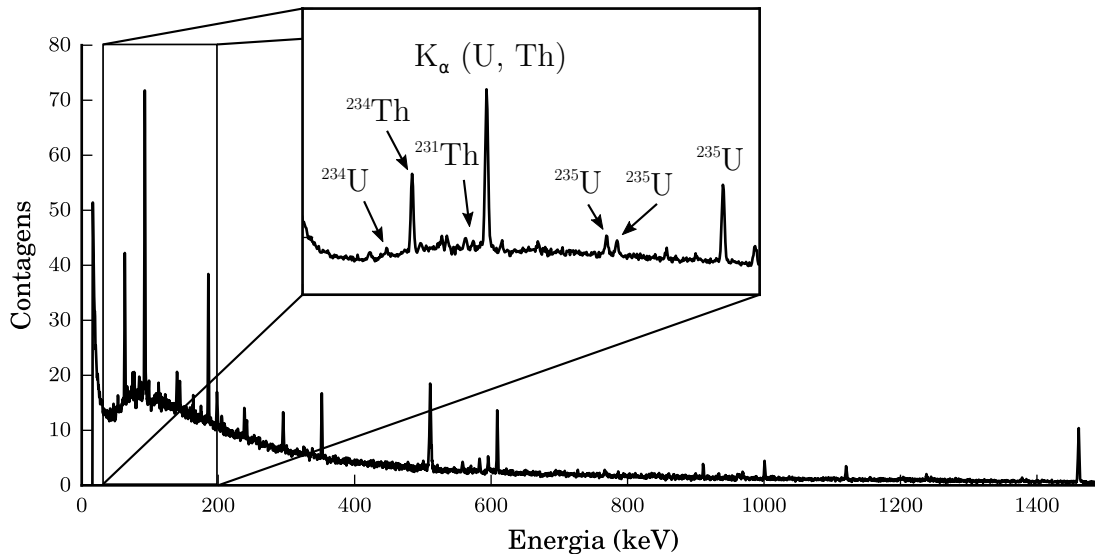
Conhecendo-se a proporção isotópica da amostra, é possível determinar a concentração de urânio através da análise de ^{235}U , que é um excelente emissor gama em 185,7 keV, que é a região de maior eficiência do detector. Porém, a determinação de ^{238}U é feita somente em último caso, pois este pico do ^{235}U sofre interferência direta do ^{226}Ra (186 keV), que também é um produto de decaimento do ^{238}U . As emissões gama do ^{235}U em outras energias possuem baixa intensidade e são úteis apenas em concentrações relativamente altas de urânio na amostra. Portanto, o cálculo de atividade do ^{238}U foi feito através do pico de baixa energia do ^{234}Th , uma vez que o pico de alta energia do $^{234\text{m}}\text{Pa}$ esteve abaixo do limite de detecção em todas as contagens realizadas. Desta forma, a região de interesse do espectro de urânio se limitou à faixa de energia entre 40 e 200 keV (Figura 36).

O L_D de urânio nas amostras calculado por diferentes radionuclídeos é mostrado na Tabela 11.

4.8.3 Treinamento dos modelos de IA

Em vez de utilizar as amostras A, B e C para o treinamento dos modelos, resolveu-se gerar 500 espectros artificiais através da interpolação entre a Amostra C e o branco

Figura 36 – Espectro de uma amostra contendo urânio em proporção isotópica ambiental



Fonte: autoria própria

Legenda: Amostras contendo urânio na proporção isotópica emitem raios gama de diversas energias, graças ao ^{235}U e aos produtos formados na cadeia de decaimento do ^{238}U como o ^{234}Th e $^{234\text{m}}\text{Pa}$. Além disso, a região de 90 keV possui uma alta densidade de emissões energéticas provenientes da auto-fluorescência (K_{α}) de U, Th e alguns de seus produtos de decaimento.

Tabela 11 – Limites de detecção de urânio total na geometria e tempo de contagem utilizada para este experimento

Radionuclídeo analisado	L_D (Bq L ⁻¹)	L_D (μg L ⁻¹)
^{235}U	0,3	750
^{234}Th	9,3	800
$^{234\text{m}}\text{Pa}$	103	2700

Fonte: autoria própria

em proporções variadas. A premissa desta ação foi simular espectros de amostras com concentrações de urânio que variassem uniformemente entre 0 e 900 μg L⁻¹.

A utilização de espectros brutos como dados de entrada faz com que a dimensionalidade do problema em questão seja consideravelmente alta, já que o analisador multi-canal onde o espectro gama é armazenado possui 8192 elementos. Portanto, fez-se necessário simplificar o sistema através da redução de dimensões utilizando os valores de importância de parâmetros estimados pela mistura de ADs, de tal forma que apenas 30 dos elementos mais importantes fossem selecionados como parâmetros de entrada.

4.8.4 Algoritmos tradicionais de análise

A análise dos espectros pelos métodos tradicionais foi feita por algoritmos implementados em Python, baseando-se no trabalho de Koskelo, Aarnio e Routti (1981) (método derivativo) e Russ (2007) (correlação com biblioteca), utilizando valores de limiares otimizados obtidos nos testes com cenários artificiais. Os valores de acurácia, probabilidade de detecção e probabilidade de alarmes falsos foram estimados a partir da curva ROC.

5 RESULTADOS E DISCUSSÃO

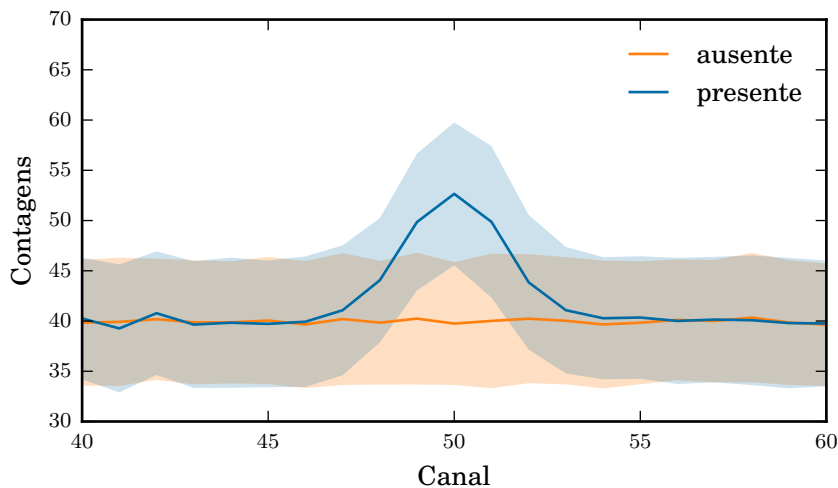
5.1 Testes com espectros artificiais

5.1.1 Cenário A: Contínuo constante, SNR constante

O cenário A constitui o teste dos modelos de IA com 1000 espectros, os quais 50% possuem apenas ruído (classe negativa) com 40 ± 6 contagens por canal e os outros 50% possuem ruído mais um pico de área = 42 ± 6 contagens (classe positiva), equivalendo a um SNR de 5 dB.

A Figura 37 mostra a aparência dos espectros da classe positiva e negativa. É possível distinguir visualmente as duas classes quando se avalia suas aparências médias. Entretanto, convém observar que estes espectros possuem intencionalmente baixa SNR e elevada incerteza, conferindo compreensível dificuldade de discriminação entre um pico e seu contínuo mesmo para um analista treinado (Figura 38).

Figura 37 – Aparência média dos grupos de espectros do cenário A

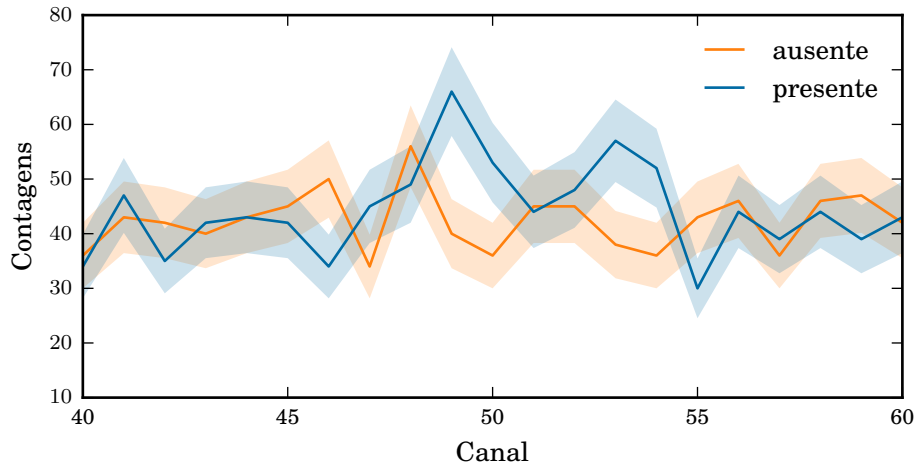


Fonte: autoria própria

Legenda: as áreas coloridas representam o intervalo entre ± 1 desvio padrão das contagens dos canais para cada grupo de espectros.

O desempenho dos modelos de IA nestes espectros é mostrado na Tabela 12. Apesar da dificuldade visual em se encontrar picos nos espectros do cenário A, contínuo constante entre uma amostra e outra, assim como a intensidade constante do pico principal, faz com que os modelos de IA prevejam corretamente a classe dos espectros em boa parte dos casos.

Figura 38 – Comparação entre dois espectros arbitrariamente selecionados



Fonte: autoria própria

Legenda: Exemplo de espectros de classe ausente (laranja) e presente (azul) escolhidos aleatoriamente. A região esmaecida em torno da linha contínua representa a incerteza de contagem em cada canal. A dificuldade em se encontrar visualmente o pico na região de interesse (em torno do canal 50) se torna evidente quando a relação sinal-ruído se torna extremamente baixa.

Tabela 12 – Desempenho dos modelos no Cenário A (espectros de contínuo e SNR constantes)

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.95 ± 0.02	0.88 ± 0.04	0.88 ± 0.05	0.12 ± 0.06
MVS	0.97 ± 0.01	0.91 ± 0.03	0.92 ± 0.04	0.11 ± 0.05
RNA	0.97 ± 0.02	0.90 ± 0.02	0.91 ± 0.03	0.10 ± 0.05
Deriv	0.81 ± 0.02	0.52 ± 0.01	0.02 ± 0.03	0.00 ± 0.00
CorrBib	0.84 ± 0.03	0.60 ± 0.02	0.20 ± 0.03	0.02 ± 0.02

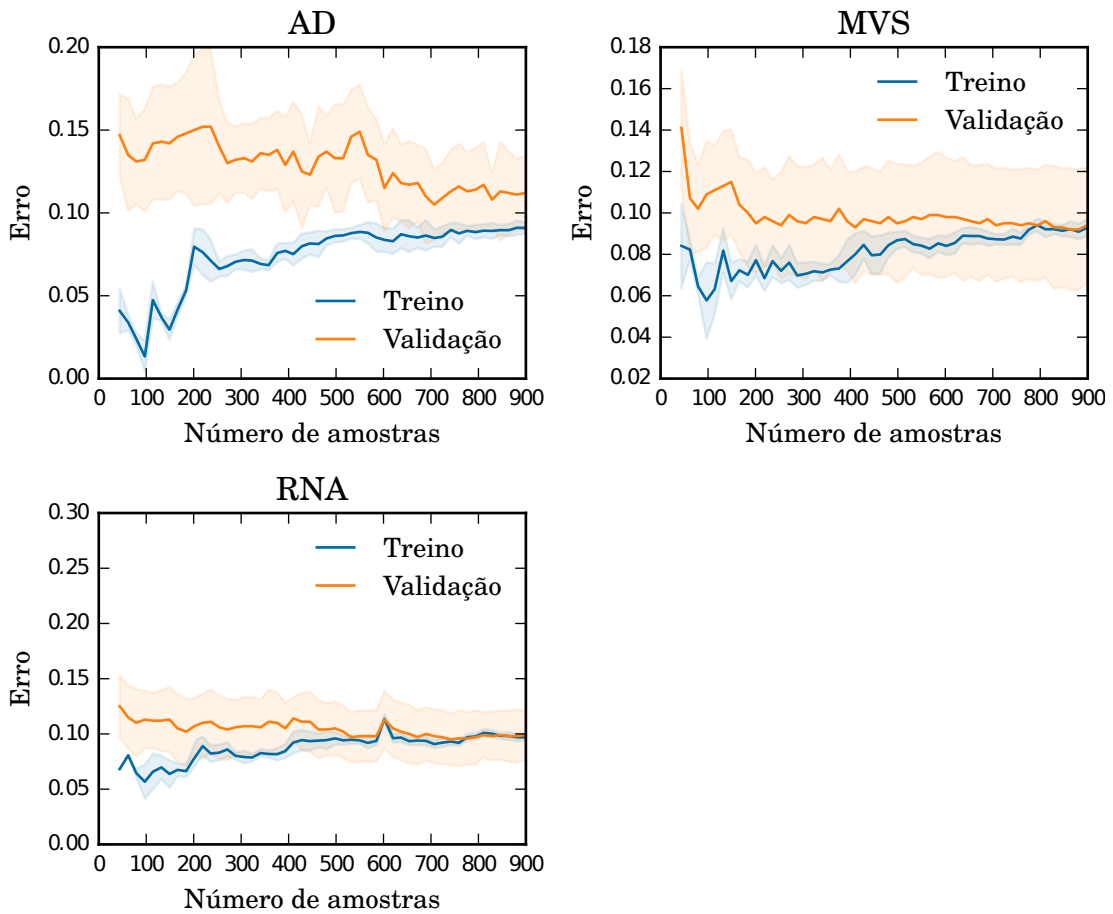
Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

A taxa de erros de aprendizagem de acordo com o tamanho do conjunto de amostras é mostrado na Figura 39. Percebe-se que a taxa de erros de treinamento é menor do que a taxa de erro de validação, mas que ambas convergem para o mesmo ponto em aproximadamente 900 amostras.

Observa-se pela Tabela 12 que o desempenho dos modelos de IA foram superiores em relação aos métodos tradicionais quando se comparam os valores de aROC. Entretanto, os valores de acurácia dos métodos tradicionais são demasiadamente menores que seus respectivos valores aROC. Isto indica que, em contraste com os modelos de IA, os limiares propostos para os métodos tradicionais não foram ajustados *ad hoc* para que os algoritmos obtivessem acurácia máxima. Esta hipótese pode ser confirmada através da Figura 40, onde é possível constatar que os limiares utilizados não se localizam no melhor ponto

Figura 39 – Curva de aprendizado dos modelos de IA em espectros do cenário A



Fonte: autoria própria

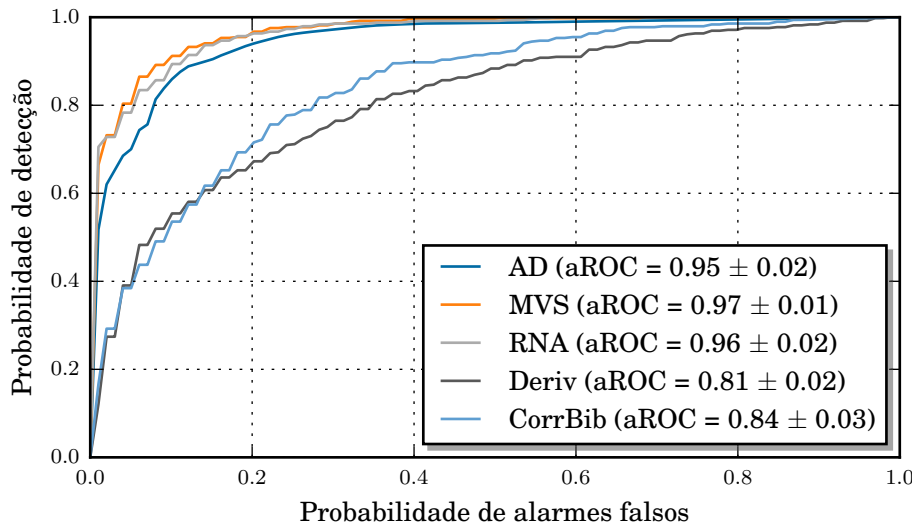
possível da curva.

A otimização do limiar de um método através da curva ROC pode ser feito obtendo-se o ponto onde a curva ROC esteja mais próximo do canto superior esquerdo. Ao se maximizar P_D e $1 - P_{AF}$, conseqüentemente se maximiza a acurácia do método. A Tabela 13 apresenta novos valores de limiares através da otimização *post hoc* utilizando o conjunto de treinamento para a otimização. A Tabela 14 mostra como seriam as métricas dos métodos tradicionais com os novos limiares.

Nota-se que o desempenho dos métodos tradicionais, mesmo depois de otimizados, ainda são inferiores aos modelos de IA em todas as métricas utilizadas neste trabalho.

Todavia, a otimização dos limiares dos métodos tradicionais desvirtua a comparação de desempenho entre métodos tradicionais e inteligência artificial, pois a otimização de hiperparâmetros através de um conjunto de treinamento é, fundamentalmente, uma abordagem de aprendizado de máquina.

Figura 40 – Curvas ROC dos modelos de IA e métodos tradicionais testados em espectros artificiais (Cenário A)



Fonte: autoria própria

Legenda: os pontos vermelhos representam o valor de corte determinado pelos limiares propostos pelo manual do software onde os métodos tradicionais foram propostos. Os pontos verdes são os valores de corte os quais estão mais próximos do canto superior esquerdo do quadro.

Tabela 13 – Limiares otimizados dos métodos tradicionais

Método	Limiar sugerido	Limiar otimizado	Acurácia original	Acurácia otimizada
Deriv	-3.0	-1.47	0.55 ± 0.02	0.72 ± 0.05
CorrBib	0.8	0.53	0.60 ± 0.02	0.76 ± 0.05

Fonte: autoria própria

Legenda: os limiares sugeridos pelos artigos originais ou manual do software onde estes algoritmos são implementados servem para casos generalizados, mas devem ser ajustados para cenários específicos caso se deseje maximizar a acurácia do método. Na tabela, os limiares sugeridos e otimizados são mostrados ao lado da acurácia antes e depois da otimização.

Tabela 14 – Desempenho dos modelos no Cenário A (espectros de contínuo e SNR constantes, com otimização dos limiares dos métodos tradicionais)

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.95 ± 0.02	0.88 ± 0.04	0.88 ± 0.05	0.12 ± 0.06
MVS	0.97 ± 0.01	0.91 ± 0.03	0.92 ± 0.04	0.11 ± 0.05
RNA	0.97 ± 0.02	0.90 ± 0.02	0.91 ± 0.03	0.10 ± 0.05
Deriv	0.81 ± 0.04	0.72 ± 0.05	0.60 ± 0.08	0.16 ± 0.06
CorrBib	0.84 ± 0.03	0.76 ± 0.03	0.77 ± 0.05	0.25 ± 0.07

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

5.1.2 Cenário B: Contínuo variado, SNR variado, picos espúrios

Neste cenário, foram criados 1000 espectros, dentre os quais 50% possuíam um pico principal com SNR variando entre 1 e 3 dB. Em 5% dos espectros, um pico espúrio de intensidade variando entre 1 e 3 dB foi colocado em posições aleatórias dos espectros. O desempenho dos diferentes métodos de detecção é exibido na Tabela 15.

Tabela 15 – Métricas de desempenho no cenário B (espectros de contínuo variável)

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.90 ± 0.03	0.81 ± 0.04	0.81 ± 0.06	0.19 ± 0.08
MVS	0.93 ± 0.02	0.84 ± 0.04	0.82 ± 0.05	0.14 ± 0.04
RNA	0.92 ± 0.02	0.84 ± 0.04	0.83 ± 0.05	0.15 ± 0.04
Deriv	0.74 ± 0.03	0.51 ± 0.01	0.02 ± 0.02	0.00 ± 0.00
CorrBib	0.79 ± 0.04	0.55 ± 0.02	0.10 ± 0.04	0.01 ± 0.01

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos.

Como é possível observar pela Tabela 15, todos os modelos obtiveram uma boa acurácia na tentativa de classificar os espectros artificiais em relação à existência do pico principal. A AD foi o modelo com pior desempenho (0,81). Não foi possível observar um desempenho estatisticamente diferente entre a MVS e RNA ($p > 0,05$). A acurácia reduzida da AD é resultante tanto do alto índice de alarmes falsos (0,19) em comparação com a MVS e RNA (0,14 e 0,15) quanto do índice de detecção (0,81 vs. 0,82 e 0,83 respectivamente).

De acordo com as métricas mostradas na Tabela 15, os limiares utilizados nos métodos tradicionais foram novamente incompatíveis com a tarefa de classificação. As métricas de desempenho com limiares otimizados para os métodos tradicionais são mostrados na Tabela 16.

5.1.3 Cenário C: Contínuo inédito

Esta etapa consistiu em testar o desempenho dos modelos em relação à classificação de espectros cujos contínuos fossem de magnitude não observadas pelos modelos durante a fase de treinamento. Os picos gerados nos espectros mantiveram a mesma SNR daquelas observadas pelos modelos. A premissa deste cenário é simular a classificação de espectros cujos contínuos tenham aumentado devido à presença de radionuclídeos interferentes de alta energia. Os resultados do teste são mostrados na Tabela 17.

Os modelos de IA mantiveram todos um desempenho similar aos cenários anteriores. Porém, os métodos tradicionais demonstraram uma melhora de desempenho em

Tabela 16 – Desempenho dos modelos no Cenário B (espectros de contínuo e SNR constantes, com otimização dos limiares dos métodos tradicionais)

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.90 ± 0.03	0.81 ± 0.04	0.81 ± 0.06	0.19 ± 0.08
MVS	0.93 ± 0.02	0.84 ± 0.04	0.82 ± 0.05	0.14 ± 0.04
RNA	0.92 ± 0.02	0.84 ± 0.04	0.83 ± 0.05	0.15 ± 0.04
Deriv	0.76 ± 0.06	0.75 ± 0.05	0.65 ± 0.06	0.15 ± 0.05
CorrBib	0.79 ± 0.04	0.72 ± 0.05	0.74 ± 0.06	0.30 ± 0.07

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

Tabela 17 – Desempenho dos modelos no Cenário C (contínuo inédito e SNR constante)

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.92 ± 0.01	0.86 ± 0.01	0.90 ± 0.01	0.19 ± 0.01
MVS	0.94 ± 0.01	0.86 ± 0.01	0.90 ± 0.01	0.17 ± 0.01
RNA	0.94 ± 0.01	0.85 ± 0.01	0.89 ± 0.01	0.18 ± 0.01
Deriv	0.82 ± 0.01	0.73 ± 0.01	0.64 ± 0.02	0.18 ± 0.01
CorrBib	0.85 ± 0.01	0.78 ± 0.01	0.75 ± 0.01	0.19 ± 0.01

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos. Os limiares dos métodos tradicionais foram otimizados para este cenário.

relação ao cenário B, mas desempenho similar ao cenário A.

5.1.4 Cenário D: Contínuo variável, SNR maiores

No cenário D, foram gerados 1000 espectros com contínuos variáveis, como o cenário B, mas com adição de picos em 50% cuja SNR variou entre 1 e 6 dB. A premissa deste teste é avaliar a capacidade dos modelos de classificarem corretamente espectros com concentrações inesperadamente altas do radionuclídeo. Um bom classificador deve obter bom ou melhor desempenho com picos de valores extremos, mesmo que estes não tenham sido observados pelos modelos durante a fase de treinamento. A Tabela 18 exhibe os resultados do teste.

Como é possível observar, a acurácia dos modelos foi maior de maneira geral, assim como aROC. Isto indica que os modelos estão aptos a classificarem picos extremos com razoável grau de confiança.

A Figura 41 sumariza os valores de aROC de todos os métodos testados nos dife-

Tabela 18 – Desempenho dos modelos no cenário D (espectros inéditos de contínuo variável)

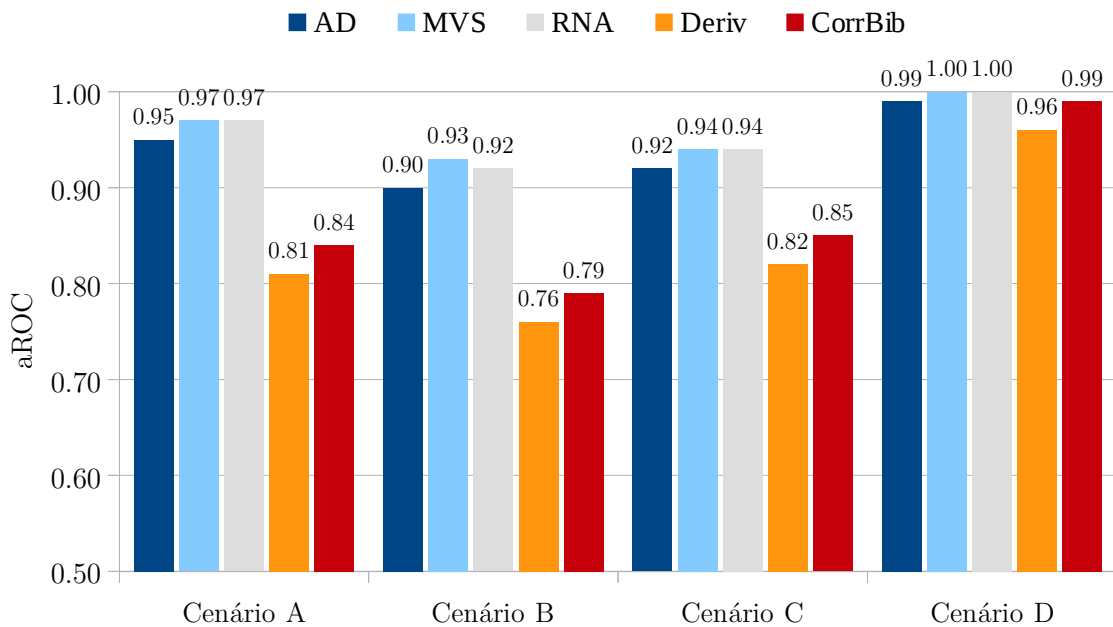
Método	aROC	Acurácia	P_D	P_{AF}
AD	0.99 ± 0.00	0.95 ± 0.01	1.00 ± 0.00	0.11 ± 0.01
MVS	1.00 ± 0.00	0.96 ± 0.01	1.00 ± 0.00	0.09 ± 0.01
RNA	1.00 ± 0.00	0.95 ± 0.01	1.00 ± 0.00	0.11 ± 0.01
Deriv	0.96 ± 0.01	0.90 ± 0.01	0.95 ± 0.01	0.15 ± 0.01
CorrBib	0.99 ± 0.00	0.88 ± 0.01	1.00 ± 0.00	0.24 ± 0.01

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

rentes cenários criados artificialmente.

Figura 41 – Sumário dos valores de aROC para os diferentes métodos testados em espectros artificiais



Fonte: autoria própria

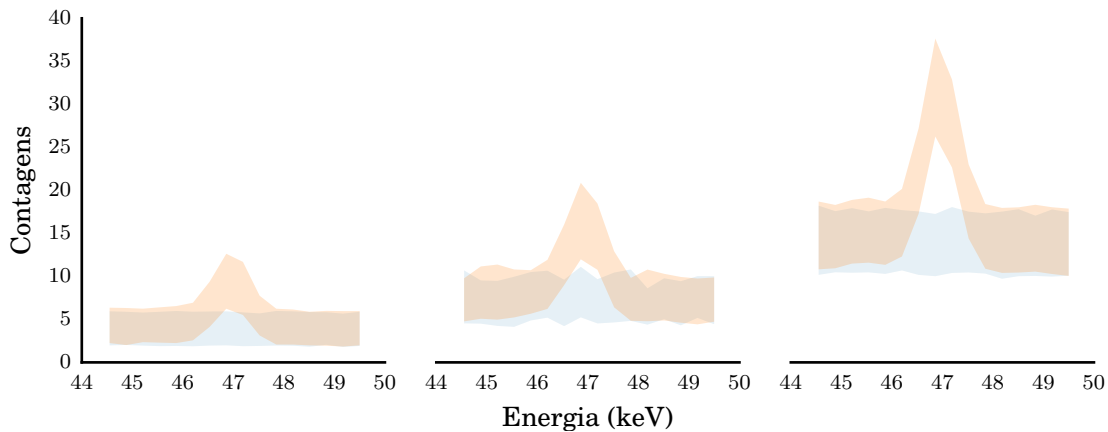
Legenda: Os modelos de IA mostraram desempenho melhor do que os métodos tradicionais em todos os cenários artificiais propostos. No cenário D, onde picos com SNR sensivelmente maiores foram artificialmente criados, o desempenho de todos os métodos foram estatisticamente equivalentes.

5.2 Testes com ^{210}Pb

5.2.1 Sumário dos espectros

A Figura 42 demonstra lado a lado como o aumento da relação SNR causa uma segregação mais perceptível entre as classes de espectros “ausente” e “presente”. Ressalta-se que, apesar das áreas dos picos serem gradativamente maiores em cada um dos grupos, a razão entre área e tempo de contagem permanece constante, uma vez que ela é proporcional à atividade do radionuclídeo observado.

Figura 42 – Aparência média dos grupos de espectros SNR-BAIXA, SNR-MEDIA e SNR-ALTA



Fonte: autoria própria

Legenda: as áreas coloridas representam o intervalo entre ± 1 desvio padrão das contagens dos canais para cada grupo de espectros. O sinal-ruído de um pico pode ser mudado apenas através do tempo de contagem, mesmo quando a intensidade de um sinal (neste caso, a atividade ou contagens por segundo de um radionuclídeo) é constante.

A estatística de contagem da região do ^{210}Pb é mostrada na Tabela 19.

Tabela 19 – Média das contagens totais na região de interesse do grupo SNR-BAIXA

Grupo	Fração do LD (%)	Classe	Contagens Totais	Contagens Líquidas
SNR-BAIXA	70	Ausente	19,5	0,0
		Presente	35,0	16,8
SNR-MÉDIA	85	Ausente	38,2	0,0
		Presente	62,2	26,1
SNR-ALTA	100	Ausente	69,8	0,0
		Presente	117,0	49,6

Fonte: autoria própria

Legenda: O limite de detecção foi estimado de tal forma que o índice máximo de falso-negativos e falso-positivos fosse igual a 0,05.

5.2.2 Teste dos modelos

O tempo total de execução da fase de treinamento e teste para os três grupos de espectros foi de pouco mais de 2 segundos para cada modelo. O treinamento da RNA foi ligeiramente maior. Os cálculos de todos os algoritmos foram realizados exclusivamente pela CPU em vez da GPU. A Tabela 20 detalha o tempo de execução de cada modelo e algoritmo testado.

Tabela 20 – Tempo de execução dos modelos de aprendizado de máquina

Método avaliado	Duração do treinamento (s)	Duração do teste (s)
AD	0,653	0,071
MVS	0,476	0,008
RNA	1,426	0,001
CorrBib	-	3,150
Deriv	-	0,012

Fonte: autoria própria

Legenda: Considerou-se treinamento a etapa seguinte ao pré-processamento dos espectros.

A Tabela 21 resume as métricas obtidas pelos modelos testados em SNR-BAIXA. O algoritmo CorrBib mostrou maior acurácia em relação a Deriv (0,73 vs. 0,67). A aROC de Deriv indica que o algoritmo, de uma forma geral e considerando todos os limiares possíveis, possui desempenho inferior a CorrBib (0,76 vs. 0,79).

Tabela 21 – Métricas de desempenho dos métodos tradicionais de detecção em SNR-BAIXA

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.96 ± 0.01	0.90 ± 0.02	0.89 ± 0.03	0.08 ± 0.02
MVS	0.97 ± 0.01	0.93 ± 0.02	0.92 ± 0.03	0.07 ± 0.02
RNA	0.97 ± 0.01	0.92 ± 0.01	0.91 ± 0.02	0.07 ± 0.01
Deriv	0.76 ± 0.04	0.67 ± 0.03	0.70 ± 0.06	0.37 ± 0.03
CorrBib	0.79 ± 0.02	0.73 ± 0.02	0.77 ± 0.05	0.32 ± 0.03

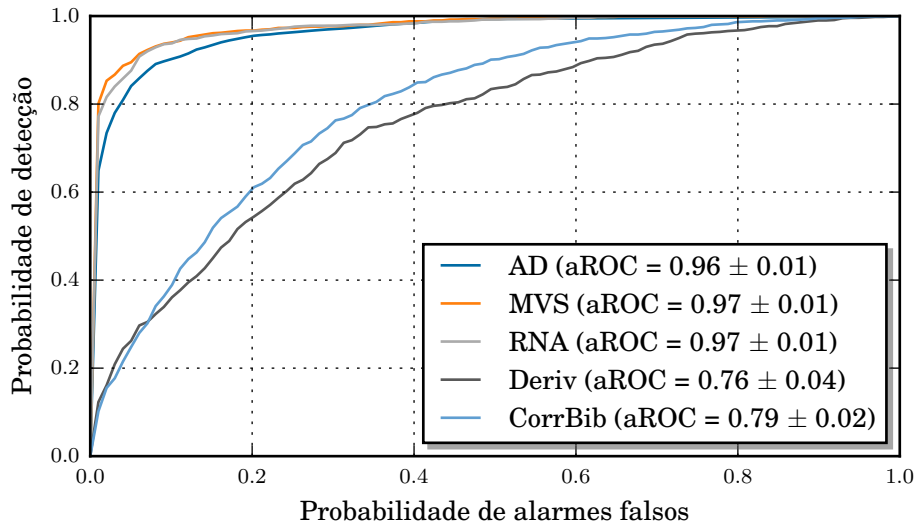
Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

Apesar de CorrBib possuir boa vantagem sobre Deriv, os três modelos de IA foram superiores a ambos métodos tradicionais. Nota-se pela curva ROC (Figura 43) que para toda P_{AF} , a P_D de todos os modelos de IA é maior. A Figura 44 discrimina em detalhes os erros e acertos de cada algoritmo.

Em relação ao grupo de espectros SNR-MEDIA, a diferença de desempenho entre os modelos de IA e os métodos tradicionais foi menor. No entanto, os modelos de IA ainda

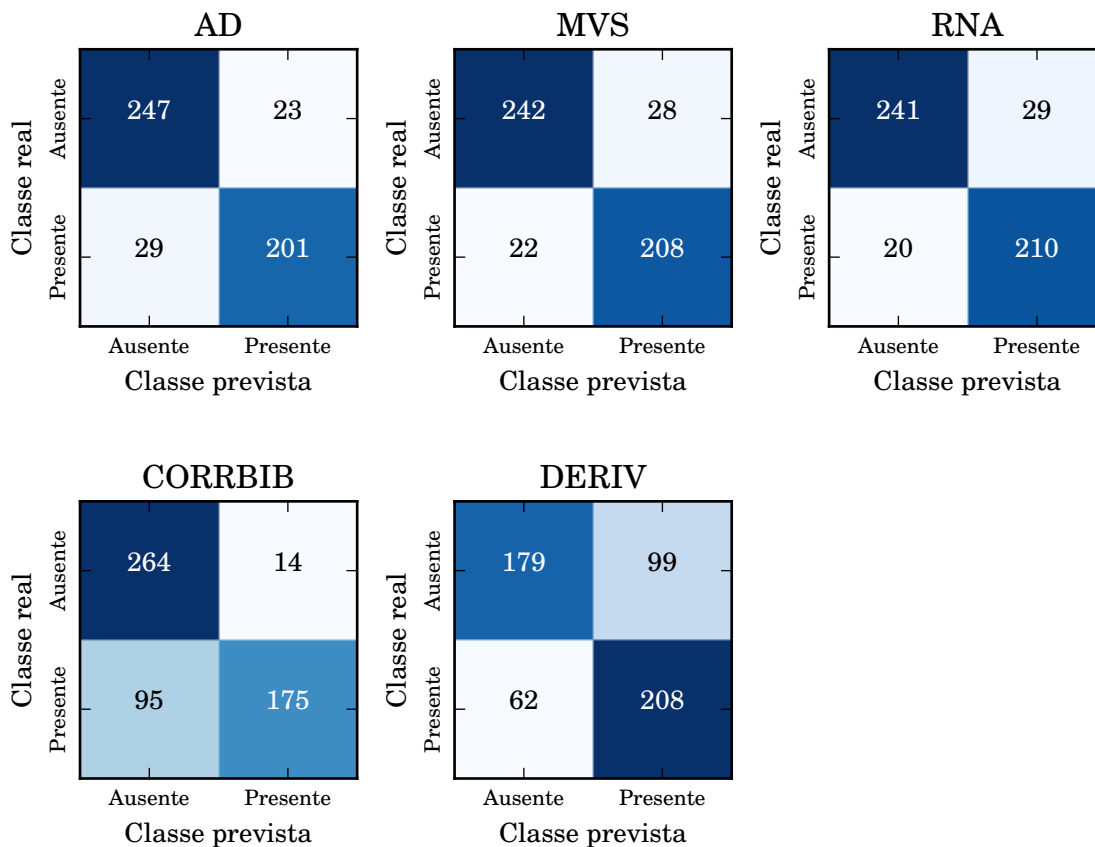
Figura 43 – Curvas ROC dos modelos testados em SNR-BAIXA



Fonte: autoria própria

Legenda: Os modelos de IA se mostraram superiores aos métodos tradicionais em toda a curva ROC. Até uma probabilidade máxima de alarmes falsos de 0,1, o algoritmo CorrBib e Deriv mostram-se equivalentes. A partir deste valor, CorrBib passa a mostra rum desempenho ligeiramente melhor que Deriv.

Figura 44 – Matriz de confusão dos modelos testados em SNR-BAIXA



Fonte: autoria própria

se mostraram superiores (Tabela 22). Uma visão geral de desempenho dos modelos pode ser observada através da curva ROC (Figura 45).

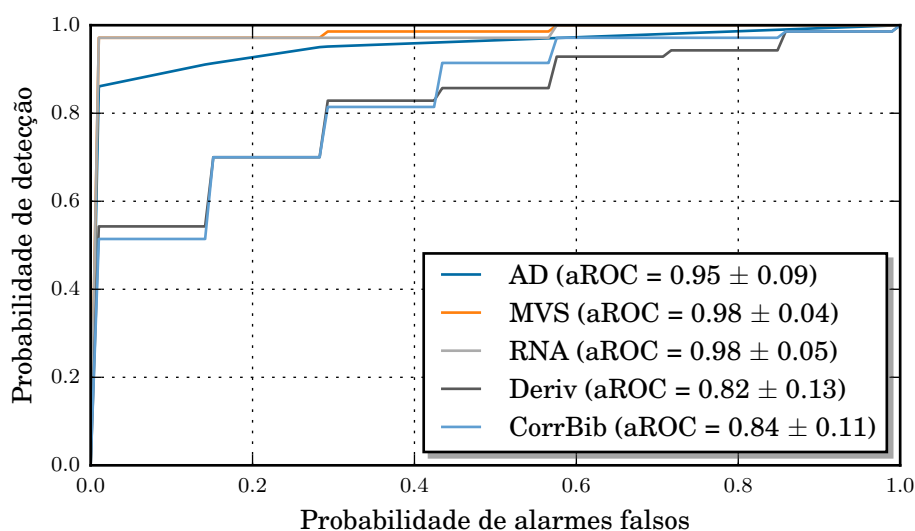
Tabela 22 – Métricas de desempenho dos métodos tradicionais de detecção em SNR-MEDIA

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.96 ± 0.09	0.92 ± 0.09	0.90 ± 0.13	0.06 ± 0.09
MVS	0.99 ± 0.04	0.96 ± 0.09	0.96 ± 0.09	0.04 ± 0.09
RNA	0.98 ± 0.05	0.96 ± 0.09	0.96 ± 0.09	0.04 ± 0.09
Deriv	0.83 ± 0.13	0.73 ± 0.13	0.59 ± 0.20	0.13 ± 0.10
CorrBib	0.84 ± 0.11	0.72 ± 0.11	0.91 ± 0.11	0.47 ± 0.16

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

Figura 45 – Curvas ROC dos modelos testados em SNR-MEDIA



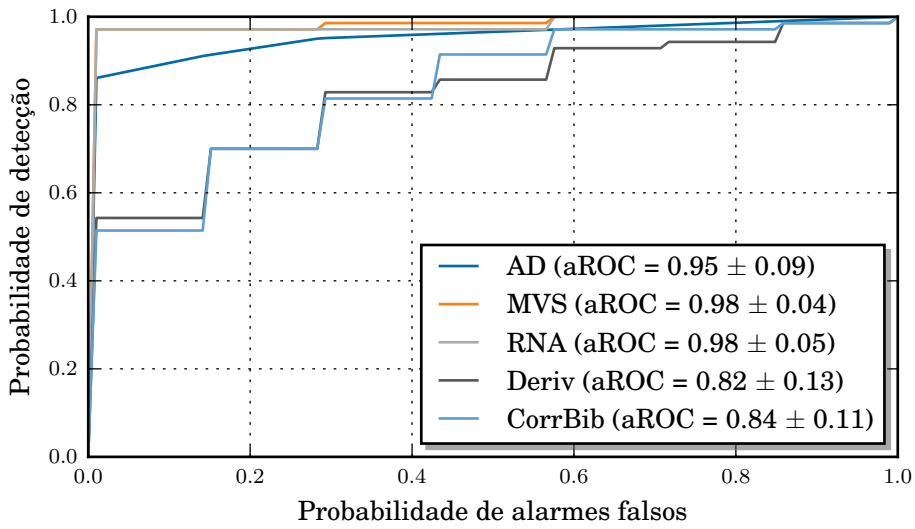
Fonte: autoria própria

Legenda: Os métodos tradicionais obtiveram desempenho semelhantes, enquanto os modelos de IA se mostraram superiores aos métodos tradicionais em toda a curva ROC.

Finalmente, os resultados dos testes com espectros do grupo SNR-ALTA são mostrados na Tabela 23. Observa-se que apesar dos modelos de IA terem superado os métodos tradicionais, a diferença entre o melhor método tradicional (CorrBib) e o melhor modelo de IA (MVS e RNA) passa a ser pouco significativa ($p > 0,05$). O desempenho de todos os métodos testados pode ser melhor avaliado através da curva ROC (Figura 46).

A Figura 47 sumariza os valores de aROC de todos os métodos testados nos diferentes grupos de espectros de ^{210}Pb .

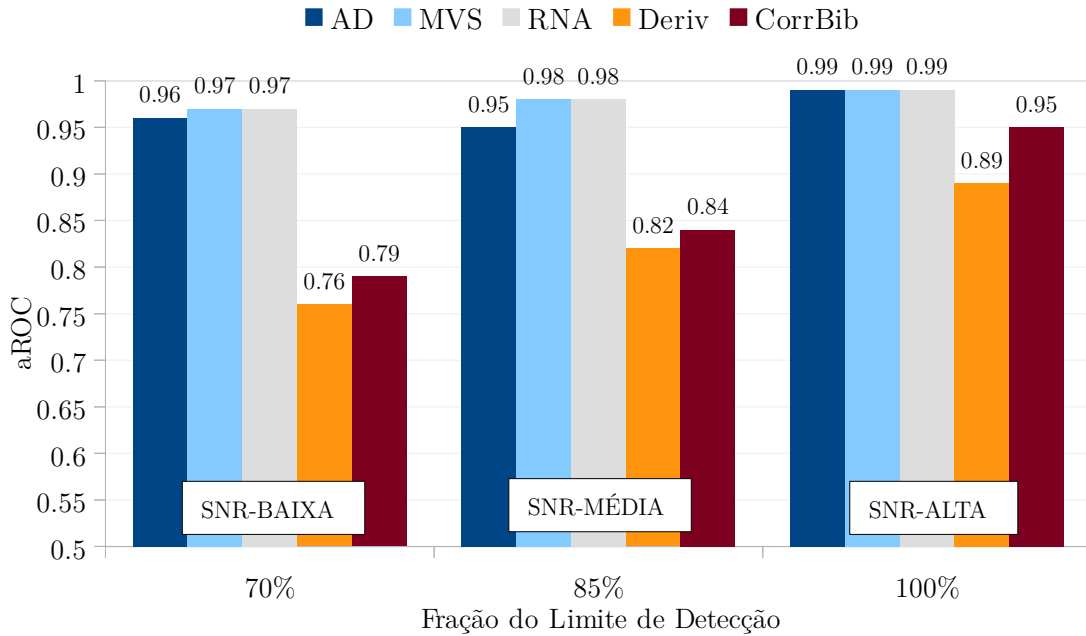
Figura 46 – Curvas ROC dos modelos testados em SNR-ALTA



Fonte: autoria própria

Legenda: Os modelos de IA se mostraram superiores aos métodos tradicionais em toda a curva ROC.

Figura 47 – Sumário dos valores de aROC para os diferentes métodos testados em espectros de 210Pb



Fonte: autoria própria

Legenda: Os modelos de IA mostraram desempenho melhor do que os métodos tradicionais. Porém, em SNR-ALTA, a diferença entre os modelos de IA e os métodos tradicionais passa a ser desprezível.

Tabela 23 – Métricas de desempenho dos métodos tradicionais de detecção em SNR-ALTA

Método	aROC	Acurácia	P_D	P_{AF}
AD	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.02	0.01 ± 0.02
MVS	1.00 ± 0.01	0.99 ± 0.01	0.99 ± 0.02	0.01 ± 0.02
RNA	1.00 ± 0.01	0.99 ± 0.01	0.99 ± 0.02	0.00 ± 0.01
Deriv	0.89 ± 0.04	0.78 ± 0.04	0.75 ± 0.07	0.19 ± 0.06
CorrBib	0.95 ± 0.03	0.74 ± 0.05	0.99 ± 0.01	0.51 ± 0.09

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

5.2.3 Análise dos resultados

Nos três grupos de espectros, as classes “ausente” e “presente” contêm uma fração de amostras cujas contagens totais se encontram em uma região dúbia de classificação, que é onde ocorre a intersecção entre as distribuições das amostras “ausente” e “presente”. A Figura 48, por exemplo, mostra a distribuição de soma de contagens dos espectros SNR-BAIXA. A área de intersecção está preenchida em azul e laranja, representando a probabilidade de falso-negativos (β) e falso-positivos (α) respectivamente. O cálculo de α e β dependerá do valor de L_C , que é o ponto de cruzamento entre as duas curvas. Calculado este valor iterativamente, encontrou-se $L_C = 42$.

A acurácia de um algoritmo que utilize a soma dos canais de um pico para realizar o teste de Currie (como é o caso de CorrBib) pode ser estimada através dos valores de L_C , α e β . Seja $\Phi(x, \mu)$ a função distribuição cumulativa de Poisson com média μ . Desta forma,

$$\alpha = 1 - \Phi(L_C, \mu_B) = 0.08 \quad (5.1)$$

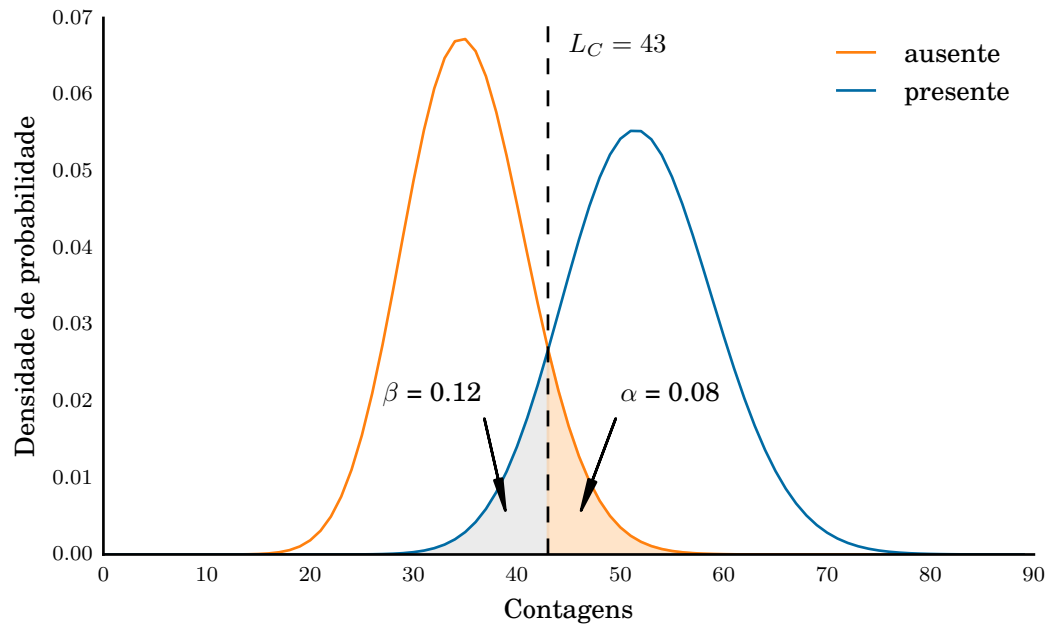
$$\beta = \Phi(L_C, \mu_A) = 0.12, \quad (5.2)$$

onde μ_B representa a média das contagens brutas do branco (classe “ausente”) e μ_A a média das contagens brutas das amostras. Sendo a área sob a curva de probabilidades de uma distribuição igual a 1, temos que

$$\frac{\alpha + \beta}{2 - \alpha - \beta} = 0,90. \quad (5.3)$$

Isto significa que se fossem bem conhecidos o valor médio de diversas contagens de branco e amostra, a classificação de picos baseando-se puramente em sua área e um limiar igual a L_C teria probabilidade de 0,90 de um espectro ser classificado corretamente.

Figura 48 – Distribuição das contagens brutas na região de interesse dos espectros SNR-BAIXA



Fonte: autoria própria

Legenda: Enquanto boa parte das amostras são classificadas com alto grau de confiança, uma porção destas se encontram em uma zona confusa de classificação onde as curvas de distribuição das classes ausente e presente se sobrepõem. Este ponto central de intersecção é chamado de limite crítico pela definição de Currie (1968). Utilizando o teste de Currie para avaliar se um efeito físico está presente em função da contagem obtida, estima-se que, neste caso específico dos espectros SNR-BAIXA, haverá uma probabilidade de alarmes falsos de aproximadamente 0,08 (região laranja) e uma probabilidade de não-deteção (região azul) de aproximadamente 0,12.

Este valor coincide com a acurácia dos modelos de IA, mas está bem além do desempenho obtido pelos algoritmos tradicionais.

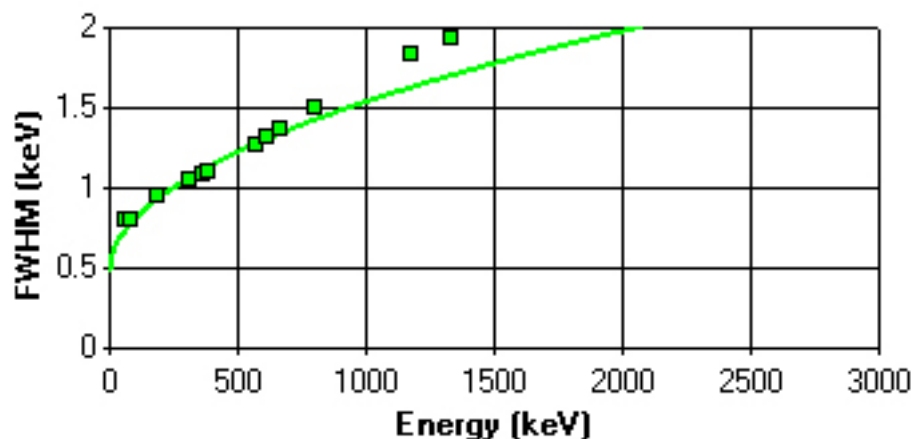
Segundo Blaauw (2016), a probabilidade de detecção de um pico na espectrometria gama depende intrinsecamente do funcionamento do algoritmo de busca de picos e de como ele relaciona área, formato e contínuo. Logo, a diferença entre o desempenho teórico e empírico do algoritmo CorrBib pode ter pelo menos duas causas consideráveis.

Primeiramente, o baixo SNR dos espectros faz com que o algoritmo tenha dificuldade em definir com precisão onde o pico começa e termina através do ajuste não-linear. O cálculo de área do pico depende amplamente desta etapa e pode levar a conclusões erradas pelo teste de limite crítico. Entretanto, os modelos de IA obtiveram acurácia, P_D e P_{AF} tal como previsto pela fórmula de Curie.

Em segundo lugar, a função “FWHM vs. energia” que é estimada pelo Genie 2000 não é compatível com os valores de FWHM observados no espectrômetro utilizado. Alguns autores já observaram que em certos sistemas de espectrometria, a família de

curvas utilizada por este programa não é ideal (GILMORE, 2008)(YIXIANG, 2013). Na Figura 49, a região de baixa energia (menor que 100 keV) passa a sofrer um desvio significativo devido ao mau ajuste dos pontos, o que pode penalizar o desempenho do algoritmo CorrBib, já que este depende de um sistema bem calibrado para estimar a região onde picos deveriam estar presentes.

Figura 49 – Curva de calibração de FWHM em função da energia pelo Genie 2000



Fonte: autoria própria

Legenda: Apesar de alguns sistemas apresentarem um comportamento quase linear de resolução (FWHM) em função da energia no espectro, o programa Genie 2000 da Canberra não permite especificar outro tipo de polinômio senão o programado em seu código. Esta função pode prejudicar algoritmos de detecção e cálculo de área do programa quando se deseja utilizar valores da biblioteca como parâmetro de comparação, como é o caso do algoritmo Library Correlation NID.

Devido aos resultados bastante semelhantes entre o teste de hipótese de Curie e os modelos de IA, pode-se concluir que os modelos possivelmente utilizaram a soma dos canais de interesse ao pico para realizar a discriminação de um espectro entre “ausente” e “presente”. Se este for o caso, isto indicaria que os resultados satisfatórios só puderam ocorrer porque os modelos de IA aprenderam qual a largura ideal do pico e deram mais ênfase aos parâmetros correspondentes aos canais corretos.

5.3 Teste com amostras de urânio

A Tabela 24 mostra o desempenho de cada um dos métodos testados na identificação das amostras que continham $300 \mu\text{g L}^{-1}$ ou mais de urânio. Como é possível observar, o modelo de IA com melhor desempenho foi a RNA. O método Deriv, que teve desempenho pior que CorrBib durante os testes prévios em condições mais controladas, mostrou-se superior nos testes com amostras de urânio.

Inesperadamente, o modelo linear de MVS se mostrou inadequado para a classificação de espectros de urânio, apesar de ter mostrado bom desempenho nos testes com

Tabela 24 – Desempenho dos modelos de IA e dos métodos tradicionais na identificação de amostras com concentração igual ou maior que 300 ug L⁻¹ de urânio

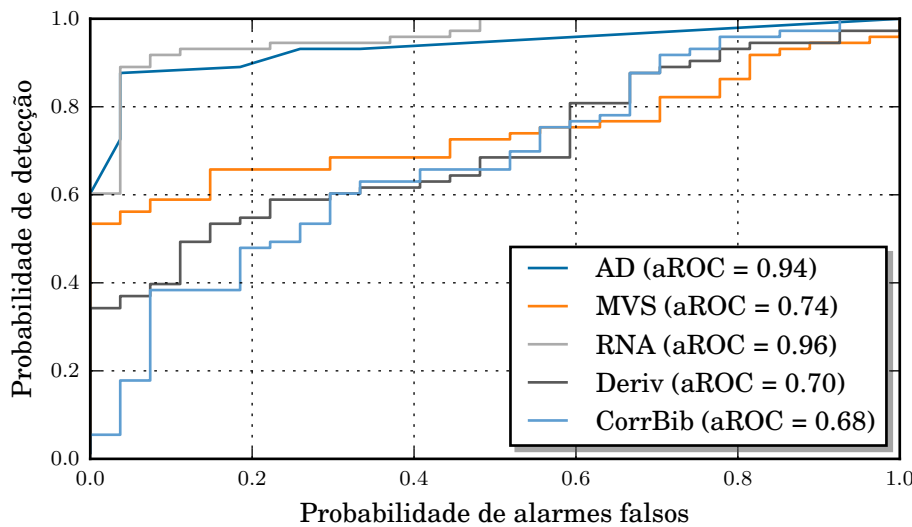
Método	aROC	Acurácia	P_D	P_{AF}
AD	0.94 ± 0.03	0.84 ± 0.03	0.86 ± 0.05	0.22 ± 0.14
MVS	0.75 ± 0.04	0.68 ± 0.05	0.67 ± 0.09	0.30 ± 0.15
RNA	0.96 ± 0.06	0.91 ± 0.07	0.89 ± 0.13	0.04 ± 0.32
Deriv	0.70 ± 0.08	0.70 ± 0.03	0.59 ± 0.04	0.20 ± 0.06
CorrBib	0.68 ± 0.06	0.65 ± 0.03	0.50 ± 0.05	0.25 ± 0.06

Fonte: autoria própria

Legenda: aROC = área sob curva ROC; acurácia = índice de acertos; P_D = probabilidade de detecção; P_{AF} = probabilidade de alarmes falsos

²¹⁰Pb e com espectros artificiais. Este fato pode ser melhor observado através da curva ROC dos métodos testados (Figura 50).

Figura 50 – Curvas ROC dos modelos testados em espectros de nitrato de uranila



Fonte: autoria própria

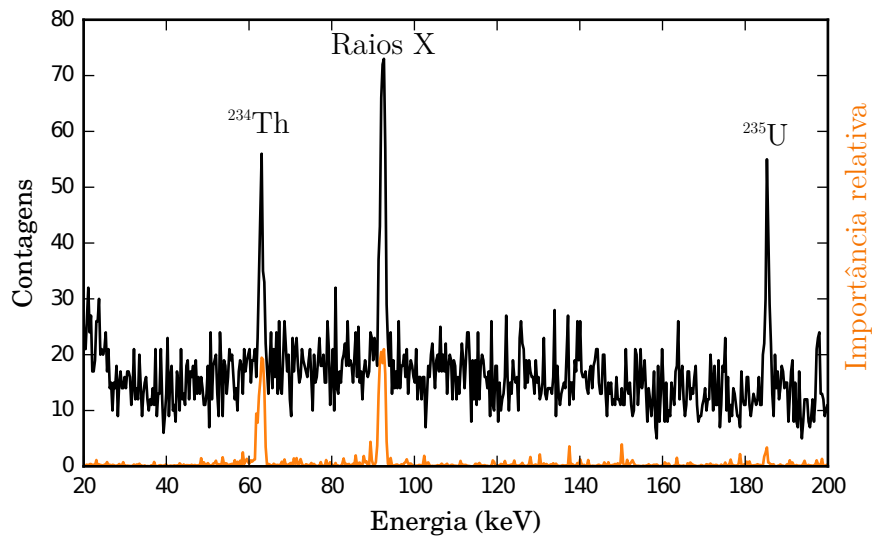
Legenda: Os modelos de IA se mostraram superiores aos métodos tradicionais em toda a curva ROC.

É possível observar através da curva ROC que o modelo de MVS falhou em demonstrar um desempenho satisfatório em toda a faixa de limiares possíveis. Este resultado pode ser atribuído ao fato do modelo de MVS utilizado possuir um núcleo linear. Este núcleo, embora apropriado para tarefas das mais variadas naturezas, não foi suficiente para a tarefa de classificação dos espectros de urânio, visto que, diferente dos cenários testados anteriormente, há mais de um pico e interferências complexas a serem consideradas.

A Figura 51 mostra uma região restrita de um dos espectros de urânio utilizados. Mais especificamente, as contagens dos canais desta região de interesse serviram como

dados de entradas para os modelos de IA (600 valores de entrada). Entretanto, após a remoção do contínuo de fundo, cálculo da transformada *wavelet* e normalização dos parâmetros, o modelo de Floresta Aleatória utilizada para a seleção dos parâmetros mais relevantes estimou que apenas os picos de ^{234}Th e a área formada por diversos picos pequenos de raios X, provenientes do fenômeno de auto-fluorescência da amostra, eram relevantes para a classificação dos espectros.

Figura 51 – Região de interesse do espectro de urânio



Fonte: autoria própria

Legenda: O eixo primário (direita) mostra o número de contagens em relação à energia das emissões. O eixo secundário (esquerda) mostra a importância relativa estimada pelos modelos de IA durante a fase de seleção dos parâmetros mais relevantes a serem utilizados. Nota-se que apenas os picos do ^{234}Th e do multipletto de raios x oriundos da do fenômeno de auto-fluorescência do urânio.

A decisão da FA de descartar o pico de ^{235}U , apesar de ser um pico com número de contagens bastante alta, pode ter sido feita devido à interferência do pico de ^{226}Ra , que possui praticamente a mesma energia. O ^{226}Ra um dos produtos de decaimento do ^{238}U , que se encontra naturalmente no ambiente.

6 CONCLUSÃO

Tanto nos experimentos com espectros simples (gerados artificialmente e provenientes da fonte mono-energética de ^{210}Pb), quanto nos experimentos com espectros mais complexos (provenientes de amostras de nitrato de uranila), observou-se que os modelos de IA possuem excelente capacidade de diferenciação de espectros gama, sendo capazes de superar os métodos tradicionais na tarefa de detecção de radionuclídeos quando o espectro possui baixa relação sinal-ruído. Os modelos de IA tiveram área sob a curva ROC superior a 90% em todos os experimentos propostos. O cenário de pior desempenho foi a classificação das amostras de nitrato de uranila, cujo modelo de máquinas de vetores de suporte com núcleo linear teve um desempenho equivalente ao dos métodos tradicionais na classificação das amostras.

Foi possível constatar que a vantagem dos modelos de inteligência artificial em relação aos métodos tradicionais diminui conforme a relação sinal-ruído dos espectros aumenta. A partir dos testes com ^{210}Pb , concluiu-se que quando a relação sinal-ruído é próxima ao limite de detecção (com intervalo de confiança de 95%), não é possível observar vantagem estatisticamente significativa dos modelos de IA sobre os métodos tradicionais.

Em vista do aparente potencial de interpretação de espectros pelos modelos de aprendizado de máquina, soa promissor a concepção de um sistema autônomo, capaz de tomar providências rápidas com menor dependência da intervenção humana. A aplicação deste tipo de tecnologia implicaria por exemplo, em maiores chances de identificação em tempo hábil de incidentes envolvendo radionuclídeos ao ambiente antes que estes atinjam concentrações catastróficas. As ferramentas para este sistema podem ter custo reduzido, dada a disponibilidade de códigos bem implementados, livres, sem custo ao usuário e mantido pela própria comunidade científica.

Este trabalho, por fim, mostrou que os avanços no campo da inteligência artificial permitem que o aprendizado de máquina aprimore métodos analíticos nucleares, especialmente na área de espectrometria gama. Apesar desta técnica ser muito bem fundamentada e resistir à prova do tempo, os algoritmos de análise ainda podem proporcionar espaço para melhorias com os recursos computacionais modernos.

O próximo passo na elaboração deste trabalho será a implementação de um modelo de aprendizado de máquina que seja capaz não somente de identificar a presença de analitos específicos em um espectro, mas também que possa realizar uma estimativa quantitativa da concentração de atividade deste na amostra.

REFERÊNCIAS

- ALEXANDRA, C. M. *Depleted Uranium: Properties, Uses, and Health Consequences*. [S.l.]: CRC Press, 2006. ISBN 978-1-4200-0456-4. Citado 2 vezes nas páginas 73 e 74.
- ALLEN, T.; ELLIS, R.; PETRIDIS, M. *Applications and Innovations in Intelligent Systems XVI: Proceedings of AI-2008, The Twenty-Eighth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*. [S.l.]: Springer Science & Business Media, 2010. ISBN 978-1-84882-215-3. Citado 2 vezes nas páginas 47 e 68.
- ALTMAN, D. G.; BLAND, J. M. Statistics Notes: Diagnostic tests 1: Sensitivity and specificity. *BMJ*, v. 308, n. 6943, p. 1552–1552, jun. 1994. ISSN 0959-8138, 1468-5833. Citado na página 44.
- ALTSHULER, B.; PASTERNAK, B. Statistical Measures of the Lower Limit of Detection of a Radioactivity Counter. *Health Physics*, v. 9, n. 3, p. 293–298, 1963. Citado na página 37.
- AMERICAN NATIONAL STANDARDS INSTITUTE. *Performance Criteria for Radiobioassay*. New York, 1996. Citado na página 37.
- ANDRELLO, A. C.; APPOLONI, C. R.; GUIMARÃES, M. F. Uso do Césio-137 para avaliar taxas de erosão em cultura de soja, café e pastagem. *Revista brasileira de Ciência do Solo*, v. 27, n. 2, p. 223–229, 2003. Citado na página 71.
- ARCAS, B. A. Y.; FAIRHALL, A. L.; BIALEK, W. What can a single neuron compute? In: *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*. [S.l.]: MIT Press, 2001. v. 13, p. 75. ISBN 0-262-12241-3. Citado na página 46.
- BABU, M. N. S. et al. Concentration of uranium levels in groundwater. *International Journal of Environmental Science & Technology*, v. 5, n. 2, p. 263–266, 2008. Citado na página 72.
- BENGIO, Y.; GRANDVALET, Y. No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research*, v. 5, n. Sep, p. 1089–1105, 2004. Citado na página 42.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006. (Information science and statistics). ISBN 0-387-31073-8 978-0-387-31073-2. Citado 8 vezes nas páginas 40, 46, 47, 49, 54, 57, 64 e 65.
- BLAAUW, M. Detection limits should be a thing of the past in gamma-ray spectrometry in general as well as in neutron activation analysis. *Journal of Radioanalytical and Nuclear Chemistry*, v. 309, n. 1, p. 39–43, jun. 2016. ISSN 0236-5731, 1588-2780. Citado 2 vezes nas páginas 110 e 149.

- BLAAUW, M. et al. The 1995 IAEA intercomparison of γ -ray spectrum analysis software. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 387, n. 3, p. 416–432, 1997. Citado na página 32.
- BODANSKY, D. *Nuclear Energy: Principles, Practices, and Prospects*. [S.l.]: Springer Science & Business Media, 2007. ISBN 978-0-387-26931-3. Citado na página 23.
- BOUCKAERT, R. R. Choosing between two learning algorithms based on calibrated tests. In: *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. [S.l.: s.n.], 2003. p. 51–58. Citado na página 84.
- BOYD, S. P.; VANDENBERGHE, L. *Convex Optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004. ISBN 978-0-521-83378-3. Citado na página 57.
- BRADLEY, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, v. 30, n. 7, p. 1145–1159, jul. 1997. ISSN 0031-3203. Citado na página 44.
- BREIMAN, L. et al. *Classification and Regression Trees*. [S.l.]: CRC press, 1984. ISBN 0-412-04841-8. Citado na página 64.
- BRENNECKA, G. A. et al. Natural variations in uranium isotope ratios of uranium ore concentrates: Understanding the $^{238}\text{U}/^{235}\text{U}$ fractionation mechanism. *Earth and Planetary Science Letters*, v. 291, n. 1–4, p. 228–233, mar. 2010. ISSN 0012-821X. Citado na página 71.
- BRODZINSKI, R. L. et al. An ultralow background germanium gamma-ray spectrometer. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 239, n. 2, p. 207–213, 1985. Citado na página 31.
- BROMOVÁ, P.; ŠKODA, P.; VÁŽNÝ, J. Classification of Spectra of Emission Line Stars Using Machine Learning Techniques. *International Journal of Automation and Computing*, v. 11, n. 3, p. 265–273, jun. 2014. ISSN 1476-8186, 1751-8520. Citado na página 134.
- BUCHTELA, K. Radiochemical Methods - Gamma-Ray Spectrometry. In: *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*. [S.l.]: Elsevier, 2014. ISBN 978-0-12-409547-2. Citado na página 25.
- BURBIDGE, R.; BUXTON, B. An introduction to support vector machines for data mining. *Keynote papers, young OR12*, p. 3–15, 2001. Citado na página 54.
- CAMARGO, I. M. C.; MAZZILLI, B. Estimativa de risco devido à ingestão de isótopos de urânio em fontes de águas minerais. *Revista de Saúde Pública*, v. 32, p. 317–320, 1998. Citado na página 72.
- CANADIAN STANDARDS ASSOCIATION. *Environmental Monitoring Programs at Class I Nuclear Facilities and Uranium Mines and Mills*. Toronto, 2010. Citado na página 37.
- CANBERRA. *Germanium Detectors*. Meriden, 2003. 99 p. Citado na página 27.

- CANBERRA. *Genie 2000 Spectroscopy Software: Customization Tools*. Meriden, 2006. Citado 8 vezes nas páginas 29, 32, 33, 76, 83, 147, 148 e 151.
- CANNIZZARO, F. et al. Results of the measurements carried out in order to verify the validity of the Poisson-exponential distribution in radioactive decay events. *The International Journal of Applied Radiation and Isotopes*, v. 29, n. 11, p. 649–652, 1978. Citado na página 39.
- CLARKE, R. et al. The properties of high-dimensional data spaces: Implications for exploring gene and protein expression data. *Nature Reviews Cancer*, v. 8, n. 1, p. 37–49, jan. 2008. ISSN 1474-175X. Citado na página 59.
- CLULOW, F. V. et al. Radium-226 in water, sediments, and fish from lakes near the city of Elliot Lake, Ontario, Canada. *Environmental Pollution*, v. 99, n. 1, p. 13–28, jan. 1998. ISSN 0269-7491. Citado na página 36.
- CONSELHO NACIONAL DO MEIO AMBIENTE. *RESOLUÇÃO CONAMA Nº 396, de 3 de Abril de 2008*. [S.l.], 2008. 64–68 p. Citado na página 73.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, set. 1995. ISSN 0885-6125, 1573-0565. Citado na página 59.
- CRAVEN, P. G.; FAIRHURST, S. A.; SUTCLIFFE, L. H. A simple approach to derivative spectroscopy. *Spectrochimica Acta Part A: Molecular Spectroscopy*, v. 44, n. 5, p. 539–545, 1988. Citado na página 33.
- CRESSIE, N. Kriging nonstationary data. *Journal of the American Statistical Association*, v. 81, n. 395, p. 625–634, 1986. Citado na página 71.
- CROWTHER, J. A. Discovery of Radioactivity. *Nature*, v. 161, n. 4094, p. 609–609, abr. 1948. ISSN 0028-0836. Citado na página 23.
- CURRIE, L. A. Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Analytical chemistry*, v. 40, n. 3, p. 586–593, 1968. Citado 3 vezes nas páginas 36, 37 e 39.
- CURRIE, L. A. Detection and quantification limits: Origins and historical overview. *Analytica Chimica Acta*, v. 391, n. 2, p. 127–134, 1999. Citado na página 36.
- DEBERTIN, K.; HELMER, R. G. *Gamma- and x-Ray Spectrometry with Semiconductor Detectors*. Amsterdam: North-Holland, 1988. ISBN 978-0-444-87107-7. Citado na página 29.
- DENIS, F. *Apprentissage automatique: des modèles formels aux applications*. Tese (Dissertação de Doutorado) — Université de Lille, Villeneuve-d'Ascq, jan. 2000. Citado na página 41.
- DIETTERICH, T. G. *Statistical Tests for Comparing Supervised Classification Learning Algorithms*. Corvallis, 1997. Citado na página 84.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: *International Workshop on Multiple Classifier Systems*. [S.l.]: Springer, 2000. p. 1–15. Citado na página 43.

- DONE, L.; IOAN, M.-R. Minimum Detectable Activity in gamma spectrometry and its use in low level activity measurements. *Applied Radiation and Isotopes*, v. 114, p. 28–32, ago. 2016. ISSN 09698043. Citado na página 38.
- DRAGOVIC, S.; ONJIA, A. Classification of soil samples according to geographic origin using gamma-ray spectrometry and pattern recognition methods. *Applied Radiation and Isotopes*, v. 65, n. 2, p. 218–224, fev. 2007. ISSN 09698043. Citado na página 71.
- EAST, L. V.; PHILLIPS, R. L.; STRONG, A. R. A fresh approach to NaI scintillation detector spectrum analysis. *Nuclear Instruments and Methods in Physics Research*, v. 193, n. 1-2, p. 147–155, 1982. Citado na página 131.
- EBAID, Y. Y.; EL-MONGY, S. A.; ALLAM, K. A. ^{235}U - γ emission contribution to the 186 keV energy transition of ^{226}Ra in environmental samples activity calculations. *International Congress Series*, v. 1276, p. 409–411, fev. 2005. ISSN 0531-5131. Citado na página 69.
- EILERS, P. H.; BOELEN, H. F. Baseline correction with asymmetric least squares smoothing. *Leiden University Medical Centre Report*, v. 1, p. 1, 2005. Citado na página 131.
- EL-SAHARTY, A. A. Radioactive survey of coastal water and sediments across Alexandria and Rashid coasts. *The Egyptian Journal of Aquatic Research*, v. 39, n. 1, p. 21–30, 2013. ISSN 1687-4285. Citado na página 36.
- EMSLEY, J. *Nature's Building Blocks: An A-Z Guide to the Elements*. Reprinted with corrections. Oxford: Oxford Univ. Press, 2003. ISBN 0-19-850340-7 0-19-850340-7 978-0-19-850340-8. Citado na página 72.
- ESPOSITO, F. et al. A comparative analysis of methods for pruning decision trees. *IEEE transactions on pattern analysis and machine intelligence*, v. 19, n. 5, p. 476–491, 1997. Citado na página 66.
- FAGUNDES, R. C.; MORAES, M. A. P. Environmental monitoring programme: Radiometric measurements in Centro Experimental Aramar and and region (CTMSP-Brazil). 2009. Citado na página 72.
- FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, jun. 2006. ISSN 0167-8655. Citado na página 44.
- FERREE, M. A.; SHANNON, R. D. Evaluation of a second derivative UV/visible spectroscopy technique for nitrate and total nitrogen analysis of wastewater samples. *Water Research*, v. 35, n. 1, p. 327–332, 2001. Citado na página 33.
- FISCHER, D. *History of the International Atomic Energy Agency: The First Forty Years*. Vienna: The Agency, 1997. ISBN 978-92-0-102397-1. Citado na página 24.
- FORD, J. Radiation, people and the environment. 2004. Citado 2 vezes nas páginas 23 e 24.
- FRANCOEUR, D. Machines à Vecteurs de Support - Une Introduction. *Cahiers de Mathématique de l'Université de Sherbrooke*, v. 1, p. 7–25, nov. 2010. Citado 6 vezes nas páginas 53, 56, 57, 59, 61 e 62.

- FUKUDA, H.; KITAMURA, S. I. Application of Projection Neural Network to Monitoring for Environmental Gamma Radiation. 1994. Citado na página 68.
- GARCÍA-TALAVERA, M. Evaluation of the suitability of various γ lines for the γ spectrometric determination of ^{238}U in environmental samples. *Applied radiation and isotopes*, v. 59, n. 2, p. 165–173, 2003. Citado 2 vezes nas páginas 75 e 76.
- GEER, L.-E. D. Currie detection limits in gamma-ray spectroscopy. *Applied radiation and isotopes*, v. 61, n. 2, p. 151–160, 2004. Citado 2 vezes nas páginas 32 e 136.
- GEMAN, S.; BIENENSTOCK, E.; DOURSAT, R. Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, v. 4, n. 1, p. 1–58, jan. 1992. ISSN 0899-7667, 1530-888X. Citado na página 51.
- GEORGIEV, A.; GAST, W. Digital pulse processing in high resolution, high throughput, gamma-ray spectroscopy. *IEEE Transactions on nuclear science*, v. 40, n. 4, p. 770–779, 1993. Citado na página 31.
- GILMORE, G. *Practical Gamma-Ray Spectrometry*. 2nd ed. ed. Chichester, England ; Hoboken, NJ: Wiley, 2008. ISBN 978-0-470-86196-7. Citado 11 vezes nas páginas 27, 29, 30, 33, 34, 35, 36, 38, 76, 111 e 136.
- GREDILLA, A. et al. Non-destructive Spectroscopy combined with chemometrics as a tool for Green Chemical Analysis of environmental samples: A review. *TrAC Trends in Analytical Chemistry*, v. 76, p. 30–39, fev. 2016. ISSN 0165-9936. Citado na página 74.
- HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, v. 143, n. 1, p. 29–36, 1982. Citado na página 44.
- HASSANAT, A. B. et al. Solving the problem of the K parameter in the kNN classifier using an ensemble learning approach. *International Journal of Computer Science and Information Security*, v. 12, n. 8, 2014. Citado na página 42.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. ed. New York, NY: Springer, 2009. (Springer series in statistics). ISBN 978-0-387-84857-0 978-0-387-84858-7. Citado 5 vezes nas páginas 41, 53, 55, 57 e 66.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. [S.l.]: MacMillan Publishing Company, 1994. Citado 2 vezes nas páginas 42 e 52.
- HAYNES, W. M. (Ed.). *Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data*. 95. rev. ed. ed. Boca Raton, Fla.: CRC Press, 2014. ISBN 978-1-4822-0867-2 978-1-4822-0867-2. Citado na página 75.
- HELMER, R.; LEE, M. Analytical functions for fitting peaks from Ge semiconductor detectors. *Nuclear Instruments and Methods*, v. 178, n. 2-3, p. 499–512, dez. 1980. ISSN 0029554X. Citado na página 86.
- HELMER, R.; LEUN, C. V. D. Recommended standards for γ -ray energy calibration (1999). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 450, n. 1, p. 35–70, ago. 2000. ISSN 01689002. Citado na página 88.

- HODGKIN, A. L.; HUXLEY, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, v. 117, n. 4, p. 500–544, 1952. Citado na página 46.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural networks*, v. 2, n. 5, p. 359–366, 1989. Citado na página 46.
- HSU, C. W.; CHANG, C. C.; LIN, C. J. A practical guide to support vector classification. 2003. Citado na página 141.
- HUANG, G. B. Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Transactions on Neural Networks*, v. 14, n. 2, p. 274–281, mar. 2003. ISSN 1045-9227. Citado na página 51.
- HUANG, J.; LING, C. X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, v. 17, n. 3, p. 299–310, 2005. Citado na página 44.
- HUANG, T. C.; PARRISH, W. A combined derivative method for peak search analysis. In: *Advances in X-Ray Analysis*. [S.l.]: Springer, 1984. p. 45–52. ISBN 1-4612-9713-3. Citado na página 33.
- INTERNATIONAL ATOMIC ENERGY AGENCY (Ed.). *Measurement of Radionuclides in Food and the Environment: A Guidebook*. Vienna: [s.n.], 1989. (Technical reports series / International Atomic Energy Agency, 295). OCLC: 246710493. ISBN 978-92-0-125189-3. Citado 2 vezes nas páginas 24 e 74.
- INTERNATIONAL ATOMIC ENERGY AGENCY (Ed.). *Analytical Techniques in Uranium Exploration and Ore Processing*. Vienna: International Atomic Energy Agency, 1992. (Technical reports series, no. 341). ISBN 92-0-103792-9. Citado 2 vezes nas páginas 73 e 74.
- INTERNATIONAL ATOMIC ENERGY AGENCY. *Our Work: Roles and Responsibilities*. 2015. <<https://www.iaea.org/technicalcooperation/programme/Prog-Man/Roles/index.html>>. Citado na página 24.
- INTERNATIONAL STANDARDS ORGANIZATION. *Determination of the Characteristic Limits (Decision Threshold, Detection Limit and Limits of the Confidence Interval) for Measurements of Ionizing Radiation - Fundamentals and Application*. [S.l.], 2010. 60 p. Citado 2 vezes nas páginas 37 e 38.
- IZHIKEVICH, E. M.; EDELMAN, G. M. Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences*, v. 105, n. 9, p. 3593–3598, abr. 2008. ISSN 0027-8424, 1091-6490. Citado na página 52.
- JÄDERSTRÖM, H. et al. Experimental verification of mathematical efficiency modeling and true coincidence summing correction of the SAGe Well detector. *Journal of Radioanalytical and Nuclear Chemistry*, v. 307, n. 3, p. 1961–1966, out. 2015. ISSN 0236-5731, 1588-2780. Citado na página 27.
- JAROSZ, Q. *Recreated Neuron in Inkscape and Hand-Tuned to Reduce Filesize*. 2009. Citado na página 46.

- JORDAAN, E. M. Robust outlier detection using SVM regression. In: . [S.l.]: IEEE, 2004. p. 2017–2022 vol.3. ISBN 978-0-7803-8359-3. Citado na página 53.
- KAISER, H. Zum Problem der Nachweisgrenze. *Fresenius' Zeitschrift für Analytische Chemie*, v. 209, n. 1, p. 1–18, 1965. ISSN 0016-1152, 1618-2650. Citado na página 37.
- KANEVSKI, M. et al. Artificial Neural Networks and Spatial Estimation of Chernobyl Fallout. *Geoinformatics*, v. 7, n. 1, p. 5–11, 1996. Citado 2 vezes nas páginas 46 e 71.
- KELLER, P. E. et al. *Applications of Neural Networks in Environment, Energy and Health*. [S.l.]: World Scientific, 1996. ISBN 978-981-4547-54-3. Citado na página 77.
- KELLER, P. E.; KOUZES, R. T. Gamma spectral analysis via neural networks. In: *Nuclear Science Symposium and Medical Imaging Conference, 1994., 1994 IEEE Conference Record*. [S.l.]: IEEE, 1994. v. 1, p. 341–345. ISBN 0-7803-2544-3. Citado 2 vezes nas páginas 68 e 69.
- KEYSER, R. M.; TWOMEY, T. R.; UPP, D. L. A Comparison of an HPGe-based and NaI-based Radionuclide Identifier (RIID) for Radioactive Materials. In: *ESARDA Annual Meeting*. [S.l.: s.n.], 2005. Citado na página 30.
- KHANDAKER, M. U. High purity germanium detector in gamma-ray spectrometry. *International Journal of Fundamental Physical Sciences*, v. 1, n. 2, p. 42–46, jun. 2011. ISSN 2231-8186. Citado na página 30.
- KINGMA, D.; BA, J. Adam - A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, dez. 2014. Citado na página 81.
- KNOLL, G. F. *Radiation Detection and Measurement*. [S.l.]: John Wiley & Sons, 2010. ISBN 978-0-470-13148-0. Citado 4 vezes nas páginas 25, 26, 27 e 30.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *IJCAI*. [S.l.: s.n.], 1995. v. 14, p. 1137–1145. Citado na página 41.
- KOSKELO, M. J.; AARNIO, P. A.; ROUTTI, J. T. SAMPO80 - An accurate gamma spectrum analysis method for minicomputers. *Nuclear Instruments and Methods in Physics Research*, v. 190, n. 1, p. 89–99, 1981. Citado 4 vezes nas páginas 33, 83, 95 e 147.
- LAMEGO, F. F. et al. Impactos de Mineralização e Sustentabilidade no Semi-árido. Estudo de caso: Unidade de Concentração de Urânio-URA (Caetité). *Comunicação Técnica elaborada para o XV Simpósio Brasileiro de Recursos Hídricos, Curitiba. Porto Alegre: ABRH*, p. 18, 2003. Citado na página 72.
- LANGLEY, P. *Elements of Machine Learning*. [S.l.]: Morgan Kaufmann, 1996. ISBN 1-55860-301-8. Citado na página 39.
- L'ANNUNZIATA, M. F. (Ed.). *Handbook of Radioactivity Analysis*. 2nd ed. ed. San Diego: Academic Press, 2003. ISBN 0-12-436603-1. Citado 6 vezes nas páginas 25, 26, 27, 32, 33 e 36.
- LE, Q. V. Building high-level features using large scale unsupervised learning. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 2013. p. 8595–8598. Citado na página 51.

- LECUN, Y.; KANTER, I.; SOLLA, S. A. Second order properties of error surfaces: Learning time and generalization. *Advances in neural information processing systems*, v. 3, p. 918–924, 1991. Citado na página 53.
- LECUN, Y. A. et al. Efficient BackProp. In: MONTAVON, G.; ORR, G. B.; MÜLLER, K.-R. (Ed.). *Neural Networks: Tricks of the Trade*. Heidelberg: Springer, 2012, (Lecture Notes in Computer Science, 7700). p. 9–48. ISBN 978-3-642-35288-1 978-3-642-35289-8. Citado na página 50.
- LEE, K. et al. Clarification of the calculation of minimum detectable activity in low-level radioactivity measurements. *Applied Radiation and Isotopes*, v. 109, p. 449–451, mar. 2016. ISSN 09698043. Citado na página 38.
- LOBO, J. M.; JIMÉNEZ-VALVERDE, A.; REAL, R. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, v. 17, n. 2, p. 145–151, mar. 2008. ISSN 1466-822X, 1466-8238. Citado na página 45.
- LOH, W.-Y. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 1, n. 1, p. 14–23, jan. 2011. ISSN 19424787. Citado na página 64.
- LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado 5 vezes nas páginas 41, 54, 56, 57 e 62.
- LOUPPE, G. et al. Understanding variable importances in forests of randomized trees. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2013. p. 431–439. Citado na página 67.
- MANGINI, A. et al. Evidence for a higher natural uranium content in world rivers. *Nature*, v. 278, n. 5702, p. 337–339, mar. 1979. ISSN 0028-0836. Citado na página 72.
- MARISCOTTI, M. A. A method for automatic identification of peaks in the presence of background and its application to spectrum analysis. *Nuclear Instruments and Methods*, v. 50, n. 2, p. 309–320, maio 1967. ISSN 0029-554X. Citado 3 vezes nas páginas 33, 83 e 147.
- MCCROAN, K. D.; KELLER, J. M. Data Calculation, Analysis, and Reporting. In: KAHN, B. (Ed.). *Radioanalytical Chemistry*. [S.l.]: Springer New York, 2006. p. 189–219. ISBN 978-0-387-34122-4 978-0-387-34123-1. Citado na página 37.
- MICHIE, D.; SPIEGELHALTER, D. J.; TAYLOR, C. C. (Ed.). *Machine Learning, Neural and Statistical Classification*. New York: Ellis Horwood, 1994. (Ellis Horwood series in artificial intelligence). ISBN 978-0-13-106360-0. Citado na página 141.
- MISDAQ, M. A.; BOURZIK, W. Determination of committed effective doses from annual intakes of ²³⁸U and ²³²Th from the ingestion of cereals, fruits and vegetables by using CR-39 and LR-115 II SSNTD. *Journal of Radioanalytical and Nuclear chemistry*, v. 254, n. 3, p. 551–555, 2002. Citado na página 73.
- MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning*. [S.l.]: MIT press, 2012. ISBN 0-262-01825-X. Citado 10 vezes nas páginas 40, 53, 54, 55, 57, 59, 63, 64, 65 e 67.

- MONARD, M. C.; BARANAUSKAS, J. A. Indução de Regras e Árvores de Decisão. In: REZENDE, S. O. (Ed.). *Sistemas inteligentes: fundamentos e aplicações*. [S.l.]: Editora Manole Ltda, 2003. ISBN 978-85-204-1683-9. Citado 3 vezes nas páginas 63, 65 e 66.
- MURRAY, A. et al. Analysis for naturally occurring radionuclides at environmental concentrations by gamma spectrometry. *Journal of Radioanalytical and Nuclear Chemistry*, v. 115, n. 2, p. 263–288, 1987. Citado na página 30.
- MURTHY, S. K.; SALZBERG, S. Decision Tree Induction: How Effective Is the Greedy Heuristic? In: *KDD-95*. Montreal: [s.n.], 1995. p. 222–227. Citado na página 65.
- MUSTOE-PLAYFAIR, H. *Scintillation Counter as a Spectrometer*. 2011. Citado na página 26.
- NELSON, W. R.; HIRAYAMA, H.; ROGERS, D. W. *EGS4 Code System*. [S.l.], 1985. Citado na página 68.
- O’HAVER, T. C. et al. Derivative spectroscopy and its applications in analysis. In: *Analytical Proceedings*. [S.l.]: Royal Society of Chemistry, 1982. v. 19, p. 22–46. Citado na página 33.
- OLIVEIRA, M. C. C. D. *Estudo Da Exposição Ambiental a Metais e Radionuclídeos Em Crianças Do Entorno Da Represa Guarapiranga*. Tese (Dissertação de Mestrado) — Universidade de São Paulo, São Paulo, 1999. Citado 2 vezes nas páginas 71 e 72.
- ORTEC. *GammaVision-32 - A66-B32 Software User’s Manual*. 6. ed. Oak Ridge: [s.n.], 2009. Citado na página 33.
- ORTEC. *Gamma Spectroscopy | ORTEC Scientific Equipment*. 2014. <<http://www.ortec-online.com/Solutions/gamma-spectroscopy.aspx>>. Citado na página 25.
- ORTEC. *Hand-Held Radioisotope Identifiers - HPGe Advantage*. 2014. <<http://www.ortec-online.com/Products-Solutions/Hand-Held-Radioisotope-Identifiers-HPGe-Proposition.aspx>>. Citado na página 30.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 81.
- PEREZ-ANDUJAR, A.; PIBIDA, L. Performance of CdTe, HPGe and NaI(Tl) detectors for radioactivity measurements. *Applied Radiation and Isotopes*, v. 60, n. 1, p. 41–47, jan. 2004. ISSN 0969-8043. Citado 2 vezes nas páginas 25 e 30.
- PÉREZ, D. V. et al. Concentração total de urânio e tório em alguns solos brasileiros. *Pesquisa Agropecuária Brasileira*, v. 33, n. 8, p. 1417–1423, ago. 1998. ISSN 1678-3921. Citado na página 71.
- PHILLIPS, G. W.; MARLOW, K. W. Automatic analysis of gamma-ray spectra from germanium detectors. *Nuclear Instruments and Methods*, v. 137, n. 3, p. 525–536, 1976. Citado 2 vezes nas páginas 33 e 133.
- POLLÁN, R. R. *Improving Multilayer Perceptron Classifiers AUC Performance*. Tese (Dissertação de Mestrado) — Universidade do Porto, Porto, out. 2011. Citado na página 43.

- POWERS, D. M. Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. dez. 2011. ISSN 2229-3981. Citado na página 43.
- PRADO, G. R. *Estudo de Contaminação Ambiental Por Urânio No Município de Caetité (BA), Utilizando Dentes Humanos Como Bioindicadores*. Tese (Dissertação de Mestrado) — Universidade Estadual de Santa Cruz, Ilhéus, 2007. Citado na página 71.
- QUINLAN, J. R. Induction of Decision Trees. *Machine Learning*, v. 1, n. 1, p. 81–106, mar. 1986. ISSN 0885-6125, 1573-0565. Citado na página 63.
- REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-Validation. In: LIU, L.; ÖZSU, M. T. (Ed.). *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. p. 532–538. ISBN 978-0-387-39940-9. Citado na página 41.
- RÉVAY, Z. et al. Recent developments in HYPERMET PC. *Journal of Radioanalytical and Nuclear Chemistry*, v. 248, n. 2, p. 401–405, 2001. Citado na página 33.
- ROZMARIĆ, M.; IVSIĆ, A. G.; GRAHEK, Z. Determination of uranium and thorium in complex samples using chromatographic separation, ICP-MS and spectrophotometric detection. *Talanta*, v. 80, n. 1, p. 352–362, nov. 2009. ISSN 1873-3573. Citado na página 73.
- RUELLAN, H. et al. A new spectra processing code applied to the analysis of ²³⁵U and ²³⁸U in the 60 to 200 keV energy range. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 369, n. 2, p. 651–656, fev. 1996. ISSN 0168-9002. Citado na página 70.
- RUSS, W. R. Library correlation nuclide identification algorithm. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 579, n. 1, p. 288–291, ago. 2007. ISSN 01689002. Citado 4 vezes nas páginas 84, 95, 149 e 150.
- RUSSELL, I. Introduction to Neural Networks. *The UMAP Journal*, v. 14, p. 1, 1991. Citado na página 46.
- RYAN, C. G. et al. SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, v. 34, n. 3, p. 396–402, 1988. Citado na página 32.
- SANAGI, M. M. et al. Comparison of Signal-to-Noise, Blank Determination, and Linear Regression Methods for the Estimation of Detection and Quantification Limits for Volatile Organic Compounds by Gas Chromatography. *Journal of AOAC International*, v. 92, n. 6, p. 1833–1838, nov. 2009. Citado na página 36.
- SCHÖLKOPF, B.; SMOLA, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, Mass: MIT Press, 2002. (Adaptive computation and machine learning). ISBN 978-0-262-19475-4. Citado na página 53.
- SCHÖN, R.; WINKLER, G.; KUTSCHERA, W. A critical review of experimental data for the half-lives of the uranium isotopes ²³⁸U and ²³⁵U. *Applied Radiation and Isotopes*, v. 60, n. 2–4, p. 263–273, fev. 2004. ISSN 0969-8043. Citado na página 71.

SCHROEDER, D. J. *Astronomical Optics*. 2nd ed. ed. San Diego: Academic Press, 2000. ISBN 978-0-12-629810-9. Citado na página 35.

SEMKOW, T. M. et al. Low-background gamma spectrometry for environmental radioactivity. *Applied radiation and isotopes*, v. 57, n. 2, p. 213–223, 2002. Citado na página 31.

SETHY, N. et al. Natural Radionuclide (U and ²²⁶Ra) in Water, Sediment, Fish and Plant Species in the Aquatic Environment around Uranium Mining and Ore processing Site at Jaduguda, India. *Journal of Ecosystem & Ecography*, v. 01, n. 01, 2011. ISSN 21577625. Citado na página 36.

SHARMA, S. et al. Anomaly detection in gamma ray spectra: A machine learning perspective. In: *Computational Intelligence for Security and Defence Applications (CISDA), 2012 IEEE Symposium On*. [S.l.]: IEEE, 2012. p. 1–8. ISBN 1-4673-1416-1. Citado na página 71.

SHULTIS, J. K.; FAW, R. E. *Fundamentals of Nuclear Science and Engineering Second Edition*. [S.l.]: CRC Press, 2007. Google-Books-ID: BR6AxeMlzbIC. ISBN 978-1-4200-5136-0. Citado na página 30.

SINGH, M. et al. Spatial mapping of uranium in groundwater and risk assessment around an atomic power station in India. *Environ Eng Manag J*, 2013. Citado na página 72.

SITEK, A.; CELLER, A. M. Limitations of Poisson statistics in describing radioactive decay. *Physica Medica*, v. 31, n. 8, p. 1105–1107, dez. 2015. ISSN 1120-1797. Citado na página 35.

SZELES, M. S. M. F. *Determinação de Urânio e Suas Razões Isotópicas Em Amostras Ambientais*. Tese (Dissertação de Mestrado) — Instituto de Pesquisas Energéticas e Nucleares, São Paulo, 1990. Citado na página 23.

UNITED NATIONS (Ed.). *Sources and Effects of Ionizing Radiation: United Nations Scientific Committee on the Effects of Atomic Radiation: UNSCEAR 2008 Report to the General Assembly, with Scientific Annexes*. New York: United Nations, 2010. OCLC: ocn680422162. ISBN 978-92-1-142274-0. Citado na página 23.

VAPNIK, V. N. *Statistical Learning Theory*. New York: Wiley, 1998. (Adaptive and learning systems for signal processing, communications, and control). ISBN 978-0-471-03003-4. Citado na página 53.

VAPNIK, V. N.; CHERVONENKIS, A. Y. A class of algorithms for pattern recognition learning. *Avtomat. i Telemekh*, v. 25, n. 6, p. 937–945, 1964. Citado na página 53.

VARLEY, A. et al. Development of a neural network approach to characterise ²²⁶Ra contamination at legacy sites using gamma-ray spectra taken from boreholes. *Journal of environmental radioactivity*, v. 140, p. 130–140, 2015. Citado na página 69.

VENKATRAM, A. On the use of kriging in the spatial analysis of acid precipitation data. *Atmospheric Environment (1967)*, v. 22, n. 9, p. 1963–1975, 1988. Citado na página 71.

- VIGNERON, V. et al. Statistical modelling of neural networks in γ -spectrometry. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 369, n. 2, p. 642–647, fev. 1996. ISSN 0168-9002. Citado 2 vezes nas páginas 69 e 70.
- WERNICK, M. N. et al. Machine learning in medical imaging. *Signal Processing Magazine, IEEE*, v. 27, n. 4, p. 25–38, 2010. Citado na página 39.
- WORLD HEALTH ORGANIZATION (Ed.). *Guidelines for Drinking-Water Quality*. 4th ed. ed. Geneva: World Health Organization, 2011. ISBN 978-92-4-154815-1 92-4-154815-0. Citado na página 72.
- WUNSCH, S.; FINK, J.; JONDRAL, F. K. Improved Detection by Peak Shape Recognition Using Artificial Neural Networks. In: *Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd*. [S.l.]: IEEE, 2015. p. 1–5. Citado na página 67.
- YIXIANG, W. Y. W. Implicit FWHM calibration for gamma-ray spectra. *Nuclear Science and Techniques*, v. 24, p. 020403, 2013. Citado na página 111.
- YÜCEL, H.; CETINER, M. A.; DEMIREL, H. Use of the 1001keV peak of ^{234m}Pa daughter of ^{238}U in measurement of uranium concentration by HPGe gamma-ray spectrometry. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 413, n. 1, p. 74–82, 1998. Citado na página 75.
- YÜCEL, H.; DIKMEN, H. Uranium enrichment measurements using the intensity ratios of self-fluorescence X-rays to ^{92}keV gamma ray in $\text{UXK}\alpha$ spectral region. *Talanta*, v. 78, n. 2, p. 410–417, abr. 2009. ISSN 0039-9140. Citado na página 69.
- ZAGNI, F. et al. The concept of minimum detectable activity of radionuclide activity meters and their suitability for routine quality control of radiopharmaceuticals. An experimental study. *Applied Radiation and Isotopes*, v. 113, p. 22–27, jul. 2016. ISSN 09698043. Citado na página 38.
- ZAHN, G. S.; GENEZINI, F. A.; MORALLES, M. Evaluation of peak-fitting software for gamma spectrum analysis. 2009. Citado na página 76.
- ZGUROVSKY, M. Z.; ZAYCHENKO, Y. P. Neural Networks with Feedback and Self-organization. In: *The Fundamentals of Computational Intelligence: System Approach*. Cham: Springer International Publishing, 2016. v. 652, p. 39–79. ISBN 978-3-319-35160-5 978-3-319-35162-9. Citado na página 68.
- ZIEGLER, L. *Quality Control for Environmental Measurements Using Gamma-Ray Spectrometry*. Washington, 2002. 158 p. Citado na página 25.
- ZUPAN, J.; GASTEIGER, J. *Neural Networks for Chemists: An Introduction*. New York, NY, USA: John Wiley & Sons, Inc., 1993. ISBN 978-3-527-28592-1. Citado na página 47.

Apêndices

APÊNDICE A – PRÉ-PROCESSAMENTO

A.1 Correção do contínuo

Foram testados diversos modos de correção do contínuo de fundo dos espectros utilizando dados artificiais:

Divisão pela mediana. Uma janela de interesse de 100 canais percorre o espectro calculando-se a soma total das contagens dos canais. O fator de normalização é igual à mediana de cada janela. O espectro é dividido por este fator de normalização.

Erosão de picos. Este algoritmo foi proposto por East, Phillips e Strong (1982) e se baseia na suavização iterativa do espectro, onde em cada iteração é calculada a média de uma pequena janela do espectro. Na iteração seguinte, compara-se o valor de contagem de cada canal com a média da janela da iteração anterior, substituindo-se o valor de contagens daquele canal pelo menor valor comparado. Maiores detalhes sobre este algoritmo podem ser encontrados em Anexos seção A.2.

Mínimos quadrados assimétricos. Um ajuste não-linear de todo o espectro é realizado através de um método insensível a *outliers* chamado de “mínimos quadrados assimétricos” ou MQA (EILERS; BOELEN, 2005). Em seguida, o ajuste é subtraído do espectro.

A Figura 52 mostra o resultado de cada um dos métodos de correção do contínuo, aplicados a um espectro de nitrato de uranila. O desempenho de classificação de espectros artificiais, submetidos a diferentes métodos de correção de contínuo, são mostrados na Tabela 25.

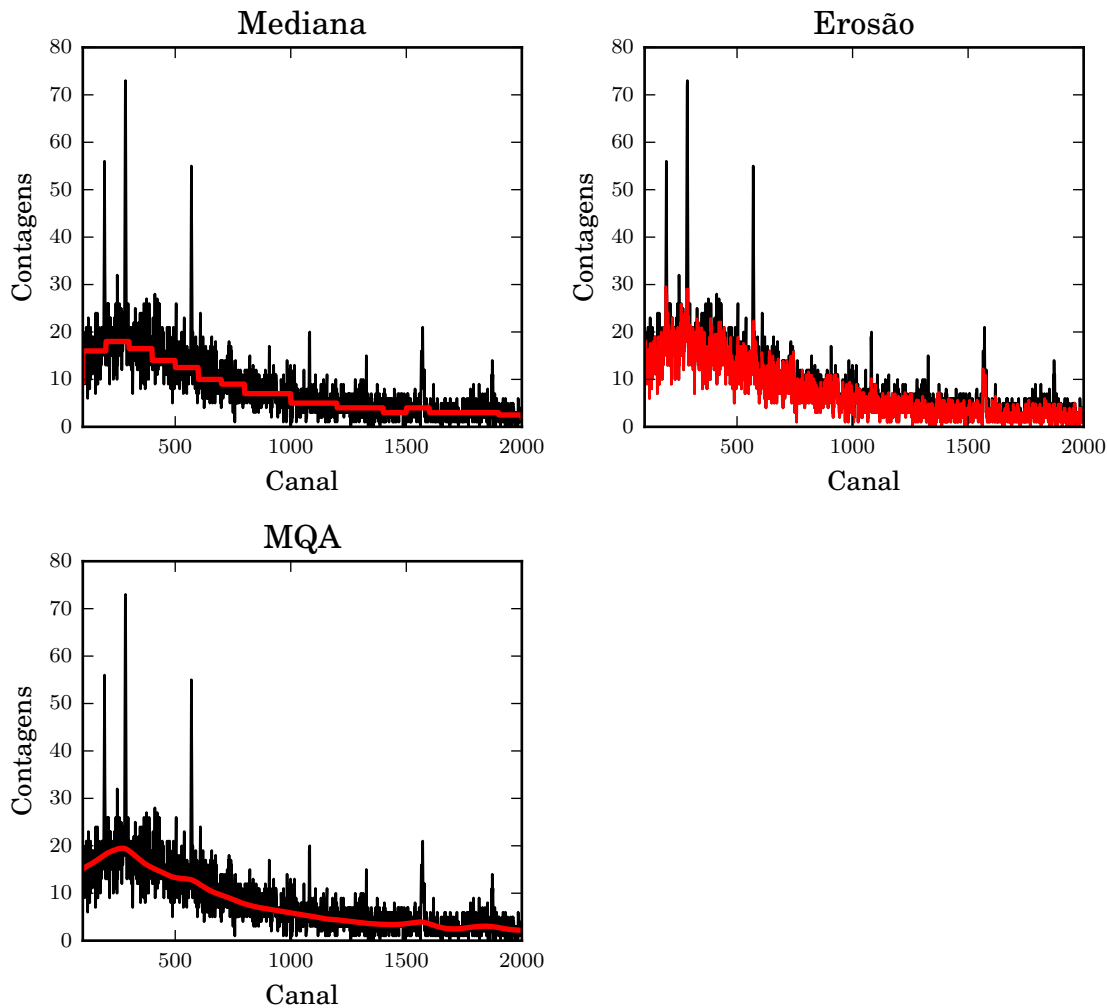
Tabela 25 – Desempenho de uma MVS com diferentes métodos de correção do contínuo de fundo dos espectros

Método	Acurácia	P_D	P_{AF}	aROC
Sem correção	0.70 ± 0.04	0.70 ± 0.05	0.28 ± 0.08	0.80 ± 0.04
Mediana	0.75 ± 0.05	0.76 ± 0.06	0.27 ± 0.09	0.83 ± 0.04
Erosão	0.62 ± 0.04	0.54 ± 0.06	0.30 ± 0.06	0.65 ± 0.04
MQA	0.75 ± 0.05	0.74 ± 0.06	0.24 ± 0.09	0.83 ± 0.04

Fonte: autoria própria

De acordo com a Tabela 25, os espectros que não foram corrigidos tiveram um desempenho menor do que aqueles corrigidos através do método da mediana e pelos MQA. No entanto, o método da remoção do contínuo por erosão mostrou um desempenho pior do que a utilização dos espectros sem correção.

Figura 52 – Resultados de diferentes métodos de correção do contínuo aplicados a um espectro de nitrato de uranila



Fonte: autoria própria

Legenda: a linha preta representa o espectro original e a linha vermelha a estimativa do contínuo pelos diferentes métodos propostos.

O desempenho pelo método da mediana e MQA foram bastante parecidos, mas deve-se considerar que os espectros utilizados possuem um contínuo, embora variável, retilíneo do primeiro ao último canal. Normalmente, é possível observar curvas e declives em parte dos espectros reais, os quais podem afetar a correção pela mediana. Portanto, apesar deste trabalho ter obtido bons resultados através da remoção do contínuo pelo cálculo da mediana, o método de remoção por MQA, apesar de computacionalmente custoso, pode apresentar comportamento mais robusto em alguns casos.

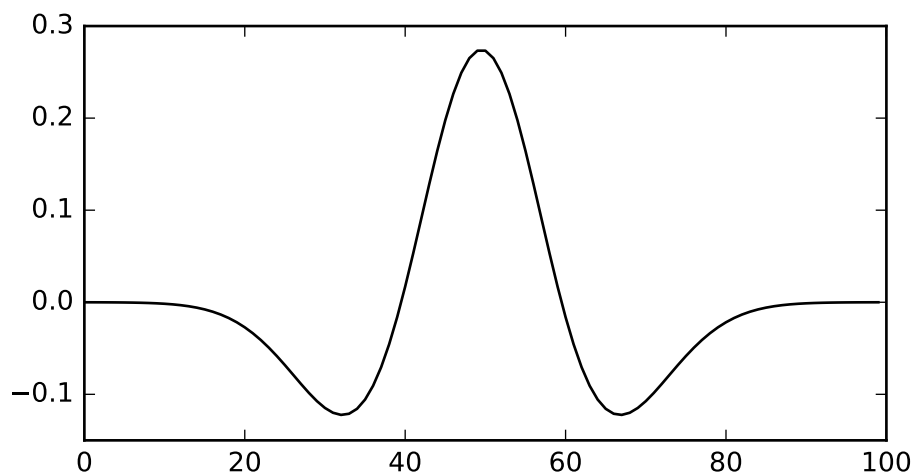
A.2 Extração de novos parâmetros

Resolveu-se testar diversos métodos de extração de novos parâmetros de entrada a partir dos dados dos espectros originais, avaliando a melhoria que cada método causou na medida de acurácia dos testes. Os métodos testados foram:

Primeira e segunda derivada. os espectros foram derivados uma e duas vezes e o espectro resultante foi repassado para os modelos de IA como parâmetro de entrada.

Convolução com operador de Ricker. foi realizada uma convolução do espectro com uma máscara que se assemelha a uma curva gaussiana (Figura 53) de largura igual a $FWHM(E)/2,36$, onde E é a energia do canal por onde a máscara é aplicada. A premissa desta técnica é suavizar o espectro de forma que os picos sejam salientados, enquanto o ruído de baixa amplitude do contínuo é transformado para próximo de zero.

Figura 53 – Operador de Ricker



Fonte: autoria própria

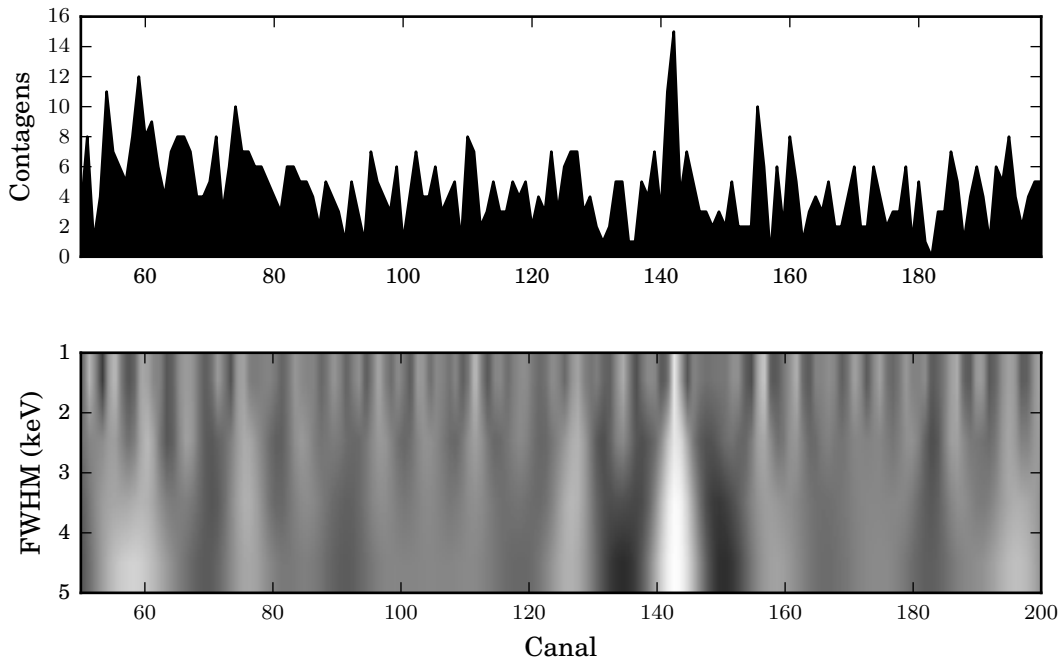
Esta máscara também é conhecida como “operador de Ricker” ou “operador chapéu mexicano” e equivale à segunda derivada da função gaussiana. Este é um operador morfológico bastante conhecido na área do processamento de sinais.

Várias máscaras possuem as características necessárias (ortogonais e simétricas) para se eliminar o contínuo de fundo. Entretanto, melhores resultados são obtidos quando o formato da máscara se assemelha ao do pico que se deseja isolar, preferencialmente com larguras compatíveis (PHILLIPS; MARLOW, 1976).

Transformada *wavelet*. O procedimento se baseou em cinco convoluções sucessivas do espectro com uma **wavelet** de 4 larguras diferentes. O vetor gerado por essa operação na região de interesse do espectro original possui tamanho igual a n° de canais $\times n^{\circ}$ comprimentos da máscara. Este processo também é conhecido como “transformada

wavelet” (ou TW). Efetivamente, cada área do espectro de 100 canais passa a gerar uma matriz com 500 novos parâmetros para serem passados aos modelos (Figura 54).

Figura 54 – Exemplo de transformada wavelet em um espectro arbitrário



Fonte: autoria própria

Este método de decomposição de sinais foi aplicado com sucesso em diversos problemas relacionados ao reconhecimento de padrões em sinais, como na classificação de estrelas a partir de suas linhas de emissões (BROMOVÁ; ŠKODA; VÁŽNÝ, 2014).

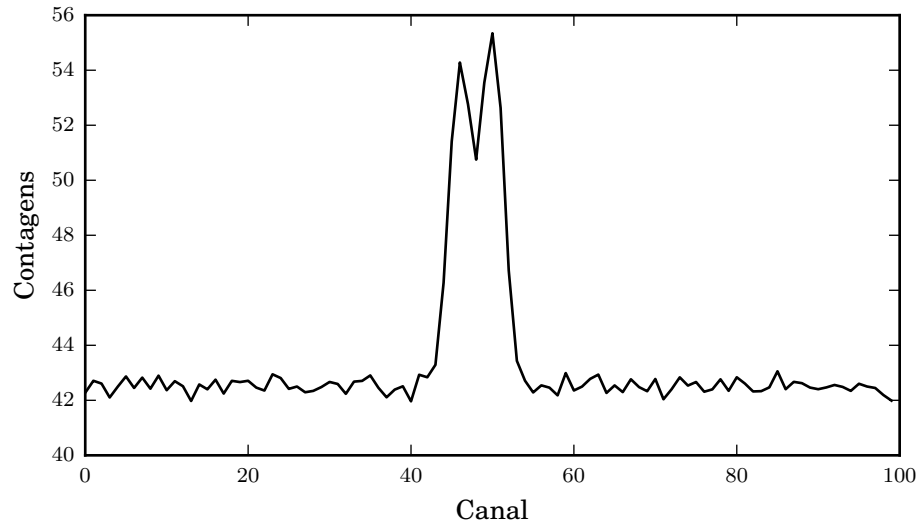
Os diferentes métodos de extração foram testados em espectros artificialmente gerados cujo pico principal havia SNR entre 2 e 4 dB e um pico espúrio em 50% das amostras em um canal bastante próximo do pico principal, de modo que estes se encontrassem parcialmente sobrepostos quando ambos estivessem presentes no mesmo espectro (Figura 55). O resultado dos testes é mostrado na Tabela 26. O método TW obteve o melhor desempenho de forma geral. A acurácia e aROC se correlacionaram ao longo dos testes.

A.3 Centralização

As RNAs e MVSs são sensíveis a dados cuja média não se aproxima de zero. Para esses modelos, os dados de entrada foram centralizados através da expressão

$$x'_j = \frac{x_j^{(i)} - \bar{x}_j}{\sigma_j}, \quad (\text{A.1})$$

Figura 55 – Média dos espectros artificiais utilizados no teste de extração de parâmetros de entrada



Fonte: autoria própria

Legenda: Os espectros artificialmente gerados de 100 canais possuem um pico principal em metade das amostras, localizado no canal 50. Para dificultar a tarefa de classificação, um pico espúrio também foi adicionado aleatoriamente em metade das amostras, no canal 47 de tal forma que este se sobrepusesse parcialmente ao pico principal.

Tabela 26 – Desempenho dos modelos de IA com métodos diversos de extração de parâmetros de entrada

Modelo	Método	Acurácia	P_D	P_{AF}	aROC
AD	Deriv1x	0.73 ± 0.06	0.72 ± 0.09	0.26 ± 0.09	0.79 ± 0.06
	Deriv2x	0.73 ± 0.06	0.72 ± 0.09	0.26 ± 0.09	0.79 ± 0.06
	Convol	0.78 ± 0.06	0.78 ± 0.08	0.22 ± 0.07	0.86 ± 0.05
	TW	0.86 ± 0.03	0.85 ± 0.05	0.12 ± 0.05	0.93 ± 0.02
MVS	Deriv1x	0.82 ± 0.03	0.80 ± 0.04	0.16 ± 0.04	0.89 ± 0.03
	Deriv2x	0.82 ± 0.03	0.80 ± 0.04	0.16 ± 0.04	0.89 ± 0.03
	Convol	0.89 ± 0.02	0.88 ± 0.04	0.09 ± 0.04	0.96 ± 0.02
	TW	0.90 ± 0.03	0.89 ± 0.04	0.08 ± 0.04	0.96 ± 0.02
RNA	Deriv1x	0.81 ± 0.04	0.80 ± 0.06	0.17 ± 0.05	0.89 ± 0.03
	Deriv2x	0.81 ± 0.04	0.80 ± 0.06	0.17 ± 0.05	0.89 ± 0.03
	Convol	0.89 ± 0.03	0.88 ± 0.04	0.11 ± 0.04	0.95 ± 0.02
	TW	0.90 ± 0.03	0.89 ± 0.04	0.11 ± 0.04	0.95 ± 0.02

Fonte: autoria própria

onde $x_j^{(i)}$ é o j -ésimo parâmetro normalizado da i -ésima amostra do conjunto de amostras, \bar{x}_j é a média do j -ésimo parâmetro de todas as amostras do conjunto de treinamento, e σ_j o respectivo desvio padrão.

A.4 Seleção de parâmetros

A utilização de muitos parâmetros de entrada pode promover resultados instáveis ou sobreajustados quando passados aos modelos de IA. Árvores de decisão, por exemplo, procuram dar prioridade no particionamento nos nós aos parâmetros que estejam mais correlacionados com o resultado esperado. Quanto mais a árvore se aprofunda e se aproxima de suas folhas, maior a probabilidade de se encontrar parâmetros menos relevantes, que não contribuirão significativamente para o resultado da AD e pode, inclusive, torná-la instável. Este sobreajuste pode impactar na capacidade de generalização e prejudicar a acurácia do modelo.

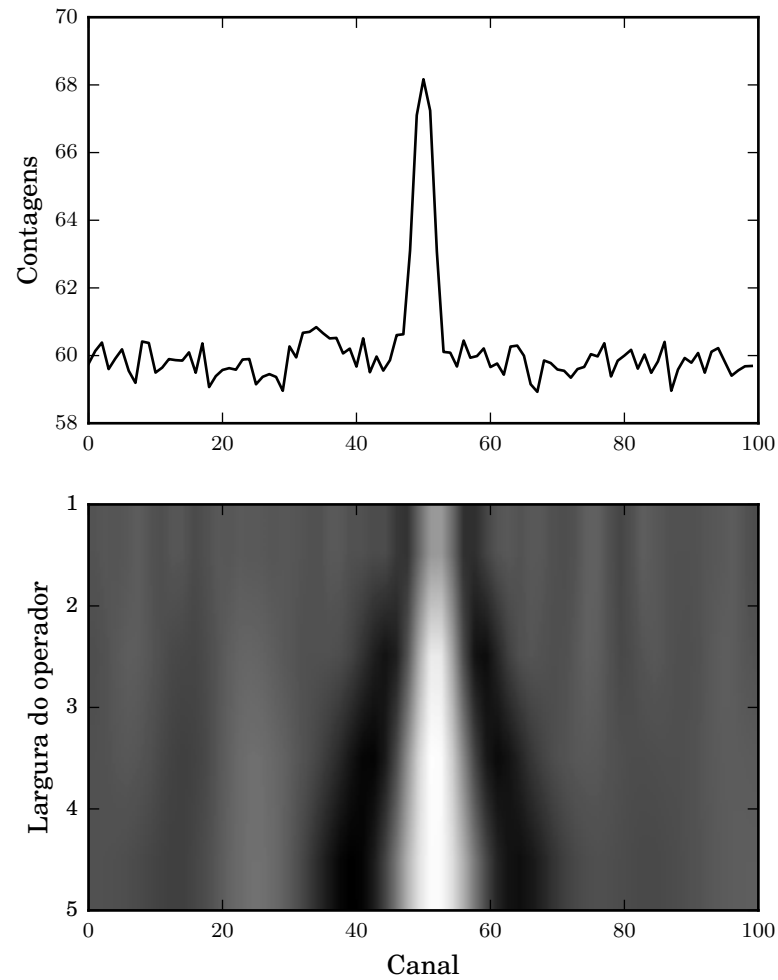
Desta forma, é importante remover parâmetros que sejam ineficientes antes de se avaliar os modelos de IA através de validação cruzada. A Figura 56 mostra os parâmetros que são passados aos modelos antes do pré-processamento. A utilização da TW durante esta fase gera 400 novos parâmetros de entrada quando o espectro possui 100 canais.

A escolha do número correto de parâmetros de entrada é um dos maiores problemas quando se implementa um bom classificador. Um número excessivo de parâmetros irrelevantes ou correlacionados, se presentes, podem afetar o desempenho do modelo. A Figura 57 mostra como o resultado dos modelos de IA são afetados de acordo com o número de parâmetros de entrada de cada espectro. Nota-se que o desempenho da AD degrada a partir de 40 parâmetros de entrada, enquanto a MVS e a RNA mantém um bom desempenho independente do valor.

Desta forma, para garantir que os modelos não tenham o desempenho comprometido devido a um excesso de parâmetros de entrada, assim como para minimizar o esforço computacional, optou-se por utilizar apenas os parâmetros de entrada mais relevantes. Utilizando uma mistura de ADs, estimou-se a importância relativa atribuída a cada parâmetro de entrada pelo modelo (Figura 58).

Na Figura 58, percebe-se que apenas os elementos mais críticos para a delimitação de um pico foram escolhidos. De acordo com Geer (2004) a melhor região para o cálculo de área dos picos (com a menor variância possível) envolve o centróide $\pm 1, 25 \cdot \text{FWHM}$, mais 4 canais de contínuo de fundo à esquerda e direita do pico (GILMORE, 2008). Isto equivale a aproximadamente 21 canais quando $\text{FWHM} = 1$. A utilização de canais adjacentes aos picos é necessária para obter-se uma estimativa da intensidade do contínuo e subtraí-la do pico para o cálculo da área. Resolveu-se fazer com que os modelos de aprendizado utilizassem 21 parâmetros de entrada selecionados pela FA.

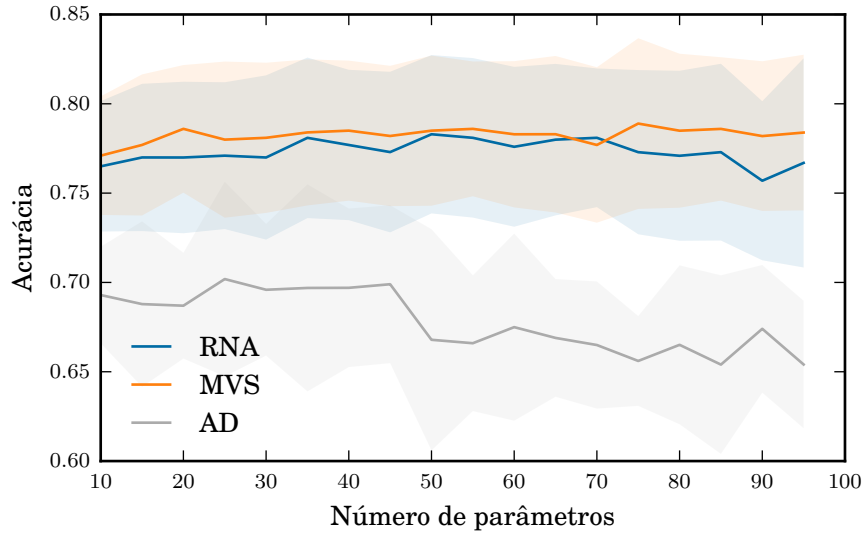
Figura 56 – Exemplo de parâmetros de entrada dos modelos antes do pré-processamento



Fonte: autoria própria

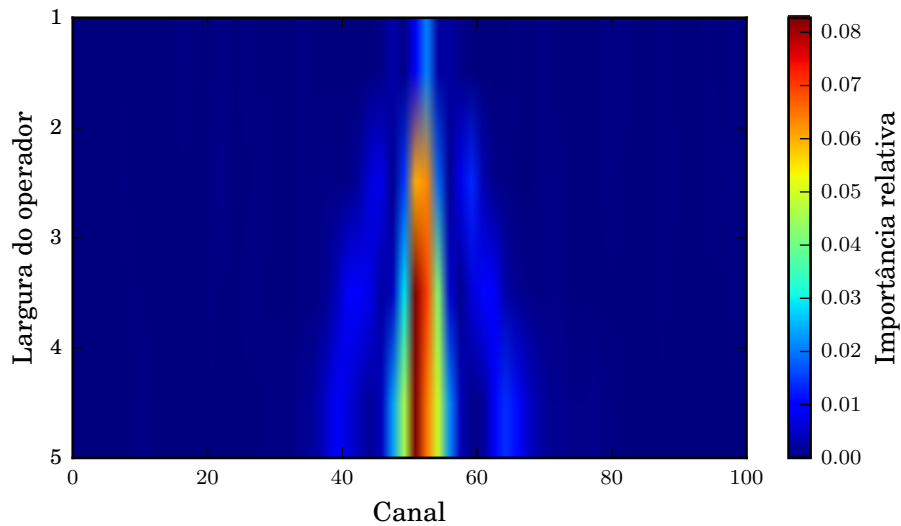
Legenda: A transformada wavelet é tecnicamente a convolução do espectro em sua forma pura com um operador morfológico de tamanhos variados. No caso deste trabalho, o operador morfológico utilizado foi o operador de Ricker, também conhecido como chapéu mexicano. A transformada, portanto, mapeia o espectro de uma dimensão a uma matriz de duas dimensões com um número de elementos iguais ao número de canais do espectro vezes a quantidade de larguras do operador de Ricker, que variou de 1 a 4 FWHM.

Figura 57 – Acurácia do modelo em função do número de parâmetros de entrada



Fonte: autoria própria

Figura 58 – Relevância dos parâmetros de entrada



Fonte: autoria própria

Legenda: As ADs permitem avaliar o ganho de informação em relação a cada parâmetro passado para o modelo. No entanto, ADs são extremamente sensíveis a variações nos dados de treinamento. Através de uma FA (Floresta Aleatória), é possível obter uma média de valores que é mais estável e provavelmente menos enviesada do que o resultado de uma única árvore. Nota-se, pela avaliação da FA mostrada na figura, que o modelo aprendeu quais canais pertencem ao pico e suas devidas importâncias.

APÊNDICE B – OTIMIZAÇÃO DOS HIPERPARÂMETROS

B.1 Árvores de decisão

Primeiramente, foi analisado como a complexidade das ADs afetam o desempenho do modelo. Diversos níveis de complexidade foram conseguidos permutando-se todos os valores possíveis de profundidade máxima e número mínimo de amostras necessárias (em porcentagem do total de amostras) para que um nó da árvore se divida e dê origem a outro nó.

A Figura 59 mostra o desempenho das ADs com diferentes números de camadas e números mínimos de amostras para particionamento de nós. A acurácia das ADs variou entre 0.70 e 0.76, sendo que a AD com melhor desempenho foi aquela com profundidade de 7 camadas e mínimo de 12% das amostras disponíveis para particionamento. Percebe-se pela Figura 59 que o desempenho degrada à medida que a AD se torna mais liberal sobre a criação de novos nós, exigindo que o nó possua um menor número de amostras candidatas para que haja particionamento.

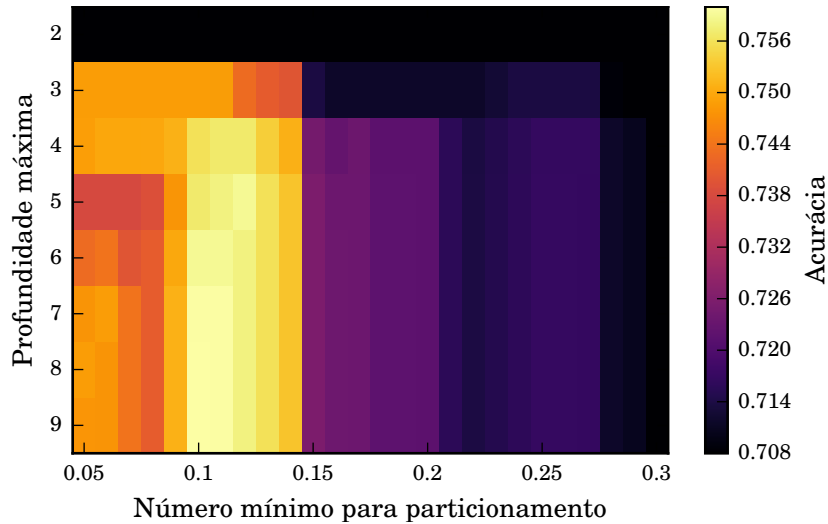
Em seguida, foram testados os critério de avaliações de pureza do nó para se decidir qual é o mais apropriado para a tarefa de classificação. Escolhidas a profundidade e número mínimo para particionamento, as ADs foram configuradas para diferentes critérios de impureza: índice Gini (AD-GINI) e entropia (AD-ENTROPIA). A escolha do critério não causou melhora significativa nos desempenhos (Figura 60).

B.2 Máquinas de vetores de suporte

O núcleo utilizado para mapeamento dos parâmetros é efetivamente uma medida de similaridade. Deste modo, conhecendo-se *a priori* a natureza dos dados, aconselha-se escolher o núcleo que se suponha melhor para separar os dados naturalmente. Optou-se por testar o desempenho do núcleo linear, quadrático, o e gaussiano (RBF).

O núcleo da MVS para mapeamento dos valores do espaço de parâmetros ao espaço de características impactou de forma significativa no desempenho dos modelos (Tabela 27 e Figura 61), embora o erro de validação não tenha se correlacionado com a complexidade do modelo. Entre os núcleos não-lineares, houve tanto desempenhos excelentes (MSV-RBF) quanto falhas completas (MSV-QUAD).

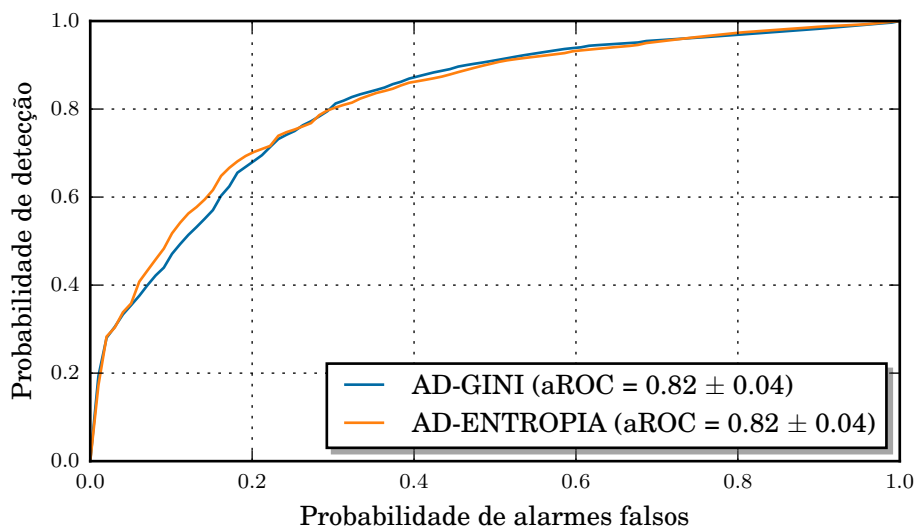
Figura 59 – Acurácia de ADs com diversos valores de profundidade e mínimo de amostras para particionamento



Fonte: autoria própria

Legenda: Os pontos mais claros da figura indicam menor erro de classificação das amostras de validação. O objetivo deste gráfico é fornecer uma idéia geral sobre como o excesso ou falta de complexidade da AD degrada a eficácia do modelo. Para que se obtenha a maior acurácia possível, deve-se chegar a um balanço apropriado entre probabilidade de alarmes falsos e probabilidade de detecção.

Figura 60 – Curvas ROC das ADs com diferentes critérios de avaliação de impureza nodal



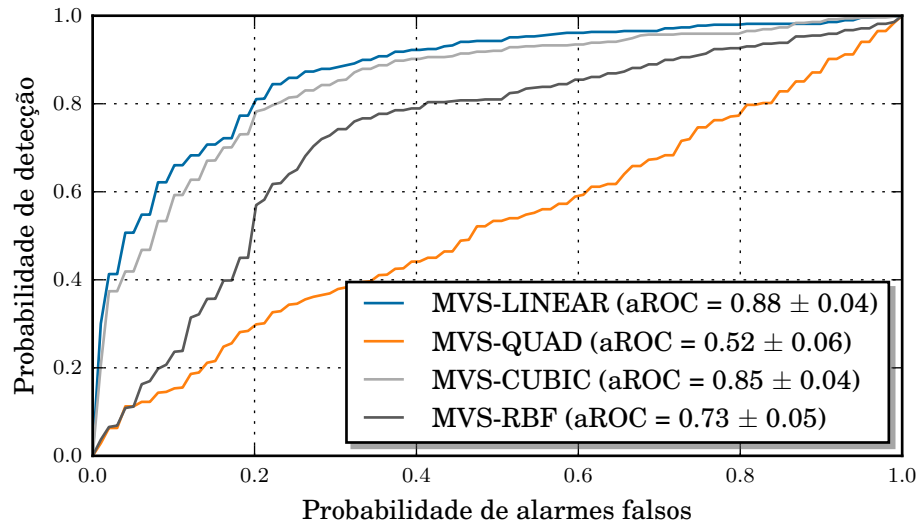
Fonte: autoria própria

Tabela 27 – Desempenho da MVS com diversos tipos de núcleo

Método	Acurácia	P_D	P_{AF}	aROC
MVS-LINEAR	0.80 ± 0.04	0.77 ± 0.06	0.18 ± 0.05	0.88 ± 0.04
MVS-QUAD	0.55 ± 0.03	0.22 ± 0.06	0.14 ± 0.05	0.53 ± 0.06
MVS-CUBIC	0.78 ± 0.04	0.70 ± 0.06	0.15 ± 0.05	0.85 ± 0.04
MVS-RBF	0.54 ± 0.03	0.13 ± 0.04	0.06 ± 0.03	0.73 ± 0.05

Fonte: autoria própria

Figura 61 – Desempenho da MVS de acordo com o tipo de núcleo



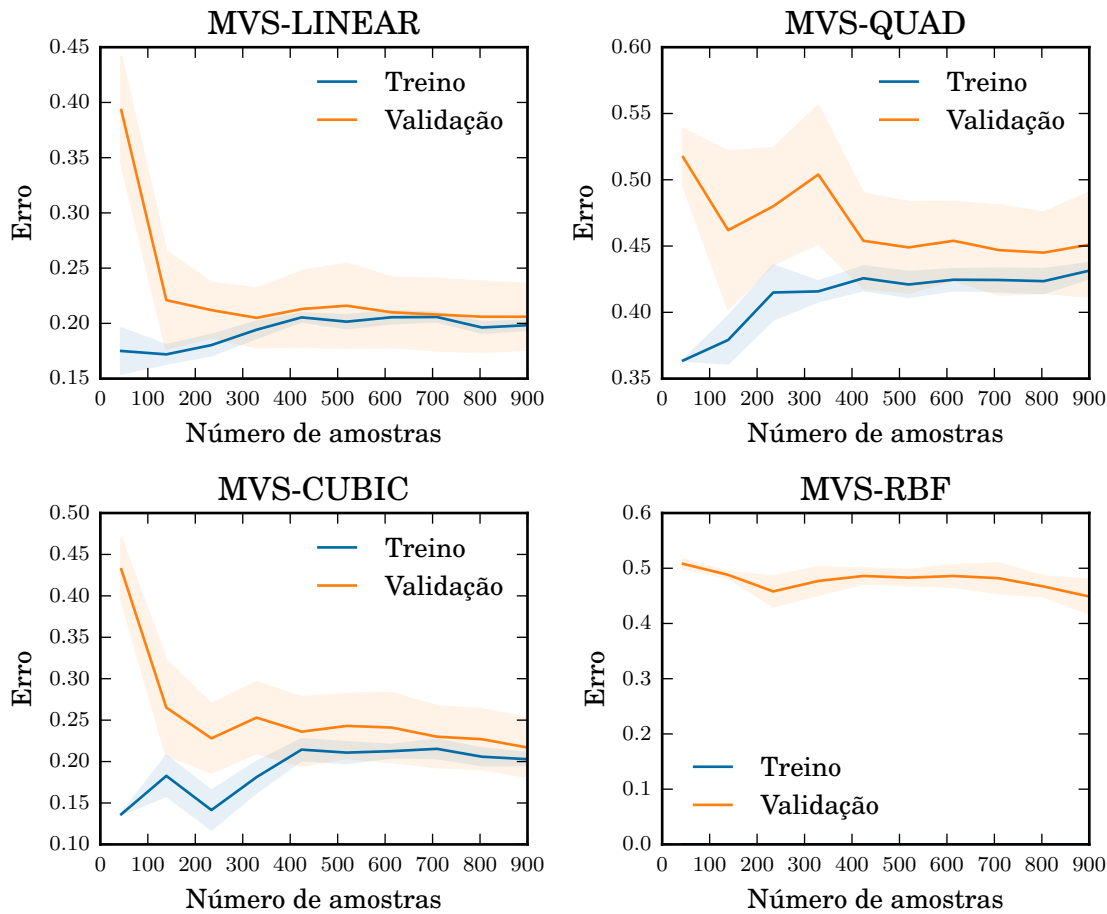
Fonte: autoria própria

Exceto pelo núcleo quadrático (MVS-QUAD), o desempenho geral das MVSs foram satisfatórios, com acurácia variando entre 0,75 e 0,78. Entretanto, a diferença de desempenho não pode ser explicada pelo sobreajuste dos dados, uma vez que a curva de erro da validação e treino tiveram grandezas compatíveis entre si (Figura 62). Isto implicaria em uma piora do desempenho à medida que o modelo utiliza núcleos menos complexos (linear) até mais complexos (cúbico e RBF).

Por outro lado, Michie, Spiegelhalter e Taylor (1994) afirmam que o núcleo linear é uma degeneração do núcleo gaussiano (RBF). Consequentemente, uma MVS com núcleo gaussiano com hiperparâmetros ótimos terá performance superior à MVS com núcleo linear. Uma busca exaustiva pelos melhores hiperparâmetros do núcleo gaussiano foi realizada com este objetivo (Figura 63).

A busca mostrou que a acurácia máxima da MVS é de 0,77 quando $\gamma = 0,00001$ e $C = 1000$, não havendo vantagem significativa sobre o modelo de núcleo linear ($0,78 \pm 0,04$). Isto pode ter ocorrido porque, segundo Hsu, Chang e Lin (2003), se o modelo já possui muitos parâmetros de entrada, não é necessário mapeá-los a um espaço de dimensão maior e o uso de um núcleo mais complexo não trará benefícios. Neste caso, optou-se pela

Figura 62 – Curva de aprendizado da MVS em função do tipo de núcleo



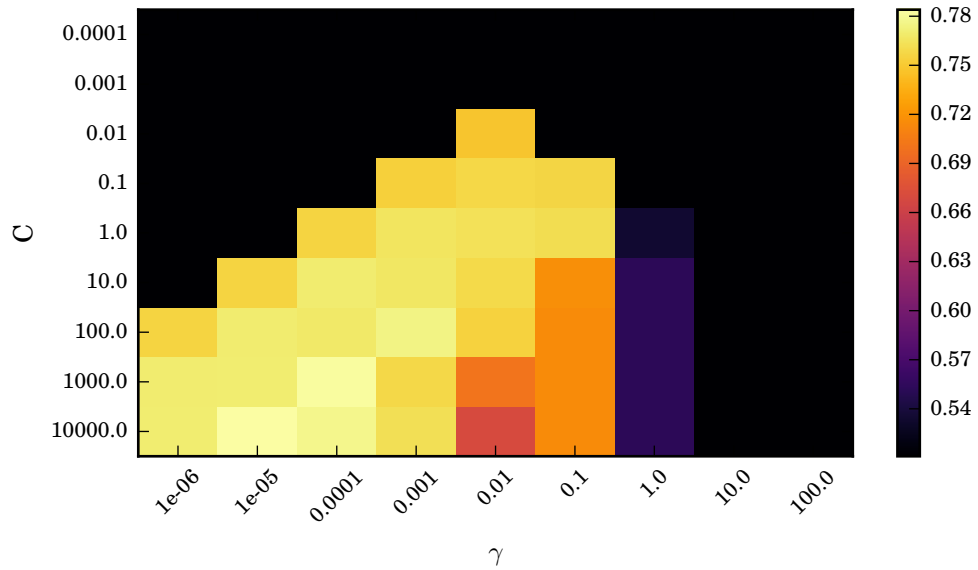
Fonte: autoria própria

Legenda: É normal obter-se curvas de aprendizado e validação com comportamentos distintos quando os modelos estão sobreajustados; a curva de treinamento geralmente indica um baixo índice de erros, enquanto o índice de erros da validação cresce ou permanece alto. A MVS-LINEAR apresenta um comportamento típico de quando a complexidade do modelo é compatível com a complexidade do problema; as curvas de aprendizado e validação convergem ao mesmo tempo para um valor comum. A MVS-QUAD não obteve desempenho dentro dos limites do gráfico.

MVS com núcleo linear devido ao menor custo computacional. A simplicidade do modelo linear oferece ainda a vantagem de não ser necessária a busca pelo hiperparâmetro γ , mas somente por C .

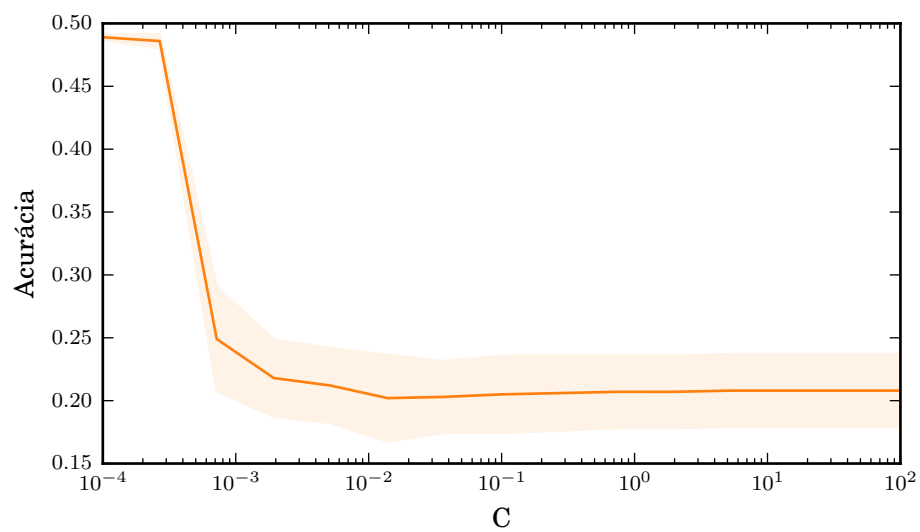
O valor de C na implementação da MVS representa um parâmetro de regularização que controla o balanço entre conseguir uma menor taxa de erros nos dados de treinamento e minimizar o vetor normal dos pesos. Para grandes valores de C , a otimização optará por margens menores adjacentes ao hiperplano. Por outro lado, um valor muito pequeno de C fará com que o modelo busque por margens maiores, mesmo que isto sacrifique uma melhor pontuação na classificação. Os valores utilizáveis de C estão contidos em um intervalo com várias ordens de grandeza. Portanto, a otimização de C ocorreu em intervalos logarítmicos (Figura 64)

Figura 63 – Acurácia da MVS com núcleo gaussiano para diversos hiperparâmetros



Fonte: autoria própria

Figura 64 – Erro na validação da MVS em função do valor de C



Fonte: autoria própria

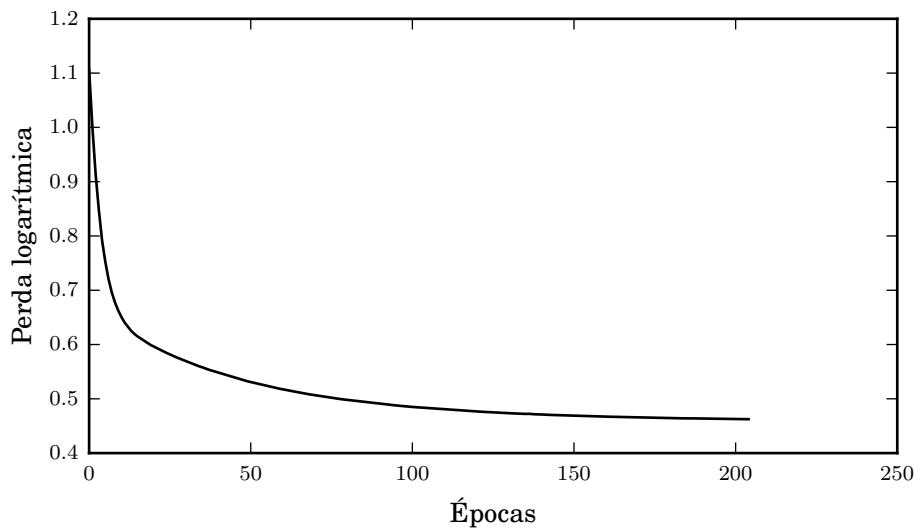
Nota-se que o valor de penalização C da MVS é capaz de afetar severamente o desempenho do modelo, já que valores menores correspondem a uma maior regularização, causando subajuste dos dados (Figura 64). O valor ótimo deste hiperparâmetro foi em torno de $C = 0,2$. Quanto mais o valor de C se afastou de 1, maior foi o tempo de execução do algoritmo durante o treinamento.

B.3 Redes neurais artificiais

A quantidade insuficiente ou excessiva de neurônios em uma RNA pode afetar o desempenho do modelo e, portanto, decidiu-se construir uma matriz 20×20 de combinações de neurônios da primeira e segunda camada e escolher o modelo com melhor desempenho.

A RNA com funções de ativação do tipo tangente hiperbólico e topologia de 64 unidades na primeira camada oculta e 12 na segunda, foi testada com várias taxas de regularização entre 0,001 e 100. A RNA foi capaz de convergir e obter acurácia de 0,80 dentro de 200 épocas de treino com regularização igual a 6,0, onde em cada época 90% das amostras do conjunto de treino são iteradas, sendo 10% separadas para validação interna.

Figura 65 – Curva de aprendizado da RNA



Fonte: autoria própria

Anexos

ANEXO A – ALGORITMOS TRADICIONAIS DE DETECÇÃO

A.1 Segunda Diferença Generalizada

O método derivativo de detecção de picos proposto por Mariscotti (1967) se baseia na convolução de um espectro através de uma máscara cuja soma seja zero e sua parte positiva é aproximadamente a largura dos picos que se desejam encontrar. Em seguida, a localização dos picos ocorre aplicando-se um limiar sobre o grau de significância do espectro transformado em função de seu desvio padrão.

A implementação do código utilizado neste trabalho para o método derivativo de busca de picos (DERIV) foi realizada de acordo com os detalhes no manual do *software* Genie 2000 (CANBERRA, 2006) e no artigo de Koskelo, Aarnio e Routti (1981). Seja n_i o número de contagens do canal i de um espectro. A transformação do espectro é feita através da seguinte expressão:

$$ss_i = \frac{dd_i}{sd_i}, \quad (\text{A.1})$$

onde dd_i é a segunda diferença generalizada (forma discreta da segunda derivada) e sd_i é o desvio padrão de dd_i . A Equação A.1 é, efetivamente, uma medida de relação sinal-ruído do espectro.

A fórmula da segunda diferença generalizada dd_i é dada por

$$dd_i = \sum_{j=-k}^k c_j n_{i+j} \quad (\text{A.2})$$

e seu desvio padrão sd_i é dado por

$$sd_i = \sqrt{\sum_{j=-k}^k c_j^2 n_{i+j}}. \quad (\text{A.3})$$

A somatória na Equação A.2 e Equação A.3 é realizada de $-k$ a $+k$, ou seja, abrangendo $2k + 1$ canais em torno de i . No entanto, k é um valor que depende de c_j , que é o vetor de pesos da somatória. O formato de c_j é aproximadamente a segunda derivada

da função gaussiana, ou seja:

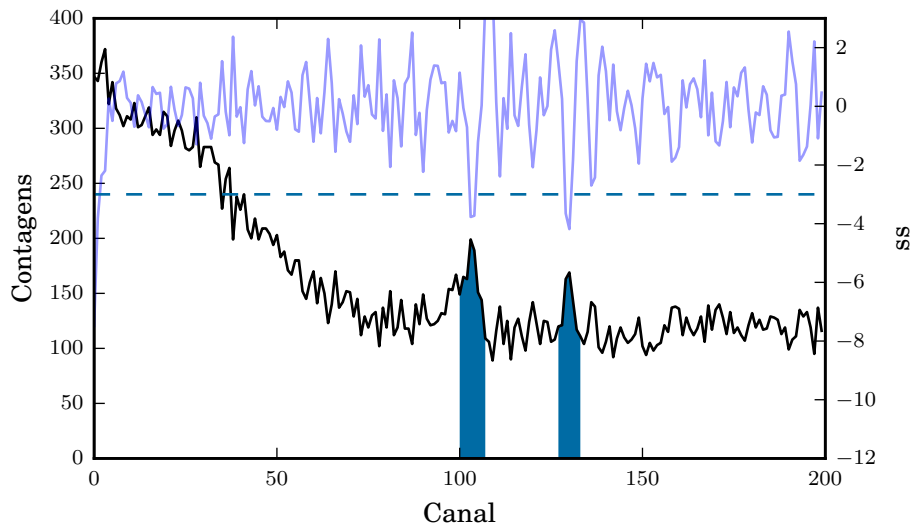
$$c_j = \frac{100(j^2 - \sigma^2)}{\sigma^2} \exp -\frac{j^2}{2\sigma^2}, \quad j = 0, \pm 1, \pm 2, \dots, \pm k, \quad (\text{A.4})$$

onde $\sigma \approx \text{FWHM}/2.355$ dos picos contidos em determinado intervalo. O valor de c_0 é sempre -100 e o conjunto é terminado em k tal que $|c_{k+1}| < 1$. O coeficiente $c_{\pm 1}$ é ajustado de tal forma que a soma de todo o vetor seja zero:

$$c'_1 = c'_{-1} = \frac{c_1}{2} - \frac{\sum_{j=0}^k c_j}{2}. \quad (\text{A.5})$$

Os possíveis picos no espectro são identificados aplicando-se um limiar negativo ao espectro transformado, menor que o qual um pico é dado como encontrado. A Figura 66 mostra um exemplo desta operação em um espectro gama obtido de amostra real. O limiar recomendado por CANBERRA (2006) é $-3, 0$.

Figura 66 – Exemplo de segunda diferença generalizada aplicada a um espectro gama



Fonte: autoria própria

Legenda: Todos os canais cujo valor ss for menor que -3 será rotulado como possível pico. O centróide é calculado pela média dos valores de ss da parte negativa desta região.

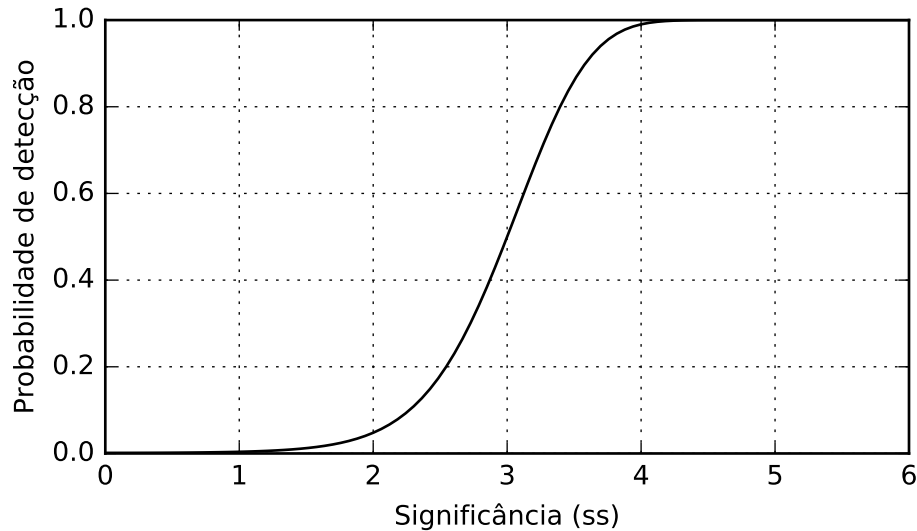
Os programas que utilizam este método de varredura de picos retornam ao usuário o número dos canais onde possíveis picos se encontram e seus respectivos valores de significância, sem uma estimativa das probabilidades de detecção.

Resolveu-se, neste trabalho, relacionar a significância estimada pela segunda diferença generalizada e a probabilidade de detecção do pico através da seguinte expressão:

$$p(ss|L) = \phi(-ss^2|L^2, L), \quad (\text{A.6})$$

onde ϕ representa a função cumulativa normal, ss representa a significância do pico encontrado pelo método derivativo e L representa o limiar escolhido para detecção. A ideia se baseia no conceito de limites de detecção debatido por Blaauw (2016). A Figura 67 mostra um exemplo de valores de probabilidade para $L = -3.0$.

Figura 67 – Probabilidade de detecção de um pico com limiar igual a -3



Fonte: autoria própria

Legenda: A probabilidade de detecção de um pico com $ss = 0$ é de 0.0013, enquanto um pico com $ss = -3$ possui probabilidade de detecção de 0.50 e um pico com $ss = -6$, uma probabilidade de detecção de aproximadamente 1.0

A.2 Identificação de núclídeos por correlação com biblioteca

Este algoritmo, chamado de *library correlation nuclide identification* ou (LibCor-rNID) foi proposto por Russ (2007) e implementado no *software* Genie 2000 a partir da versão 3.1. O algoritmo se baseia na filtragem da linha base do contínuo do espectro e no subsequente teste de correlação entre o formato e amplitude dos picos presentes e as características de picos esperados de acordo com uma biblioteca.

O isolamento do contínuo do espectro é efetuado através de uma técnica chamada de “erosão de picos”, que se assemelha a um filtro passa-baixa. O processo de suavização ocorre iterativamente (total de 8 iterações), comparando cada canal do espectro com canais das extremidades da janela, que varia de tamanho de acordo com o FWHM do canal central (sabido previamente através da calibração de resolução do sistema). A posição das extremidades é dada por

$$i \pm m = i \pm \frac{3}{n} \text{FWHM}, \quad (\text{A.7})$$

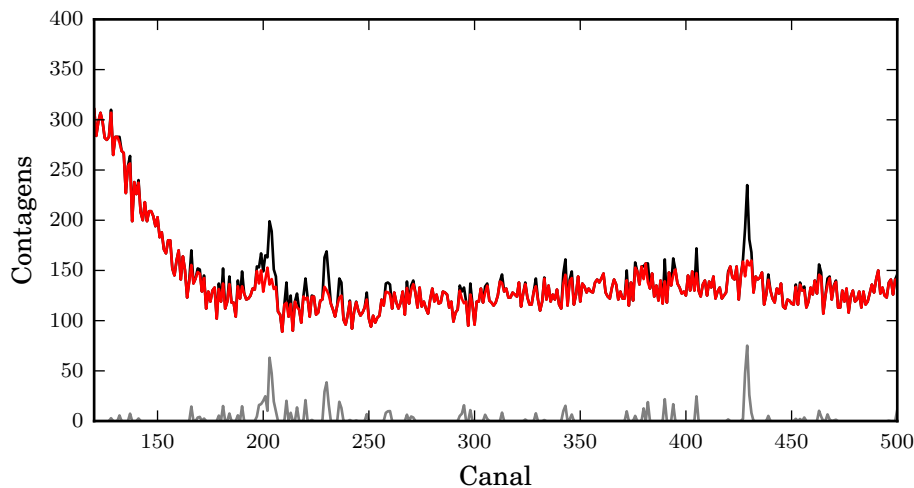
onde i é o número do canal em questão, m é a distância de i até a extremidade da janela, n é o número da iteração atual e FWHM é a resolução esperada para um pico no canal i .

Se o valor do canal central for maior que a média dos canais das extremidades mais uma incerteza (2 desvios padrões), o canal central é reduzido para este valor em um novo espectro temporário. Uma vez que as variáveis são altamente correlacionadas, a incerteza do canal central será dado por

$$\sigma_i = \frac{\sqrt{\sigma_{i-m}^2 + \sigma_{i+m}^2 + 2\sigma_{i-m}\sigma_{i+m}}}{2}. \quad (\text{A.8})$$

Este processo é então repetido 8 vezes com janelas cada vez menores de acordo com a Equação A.7. A Figura 68 mostra um exemplo de espectro cujo contínuo foi subtraído pelo método de erosão.

Figura 68 – Exemplo de espectro cujo contínuo foi subtraído pelo método de erosão



A etapa seguinte consiste em criar espectros para cada nuclídeo de uma biblioteca (contendo todas as informações necessárias). Um processo de filtragem se baseando em valores de tolerância ao ganho não-linear de energia durante aquisição e um limiar de sensibilidade dos picos é utilizado para eliminar nuclídeos da biblioteca que provavelmente não estão presentes. A tolerância do ganho não-linear é uma característica bastante presente em detectores NaI e, portanto, pode ser desconsiderada no presente trabalho, já que o detector HPGe e sua eletrônica acoplada para os experimentos mostraram boa estabilidade.

Segundo Russ (2007), o limiar de sensibilidade é a mínima área espectral de um nuclídeo necessária para identificação, como uma porcentagem do espectro original subtraindo-se o contínuo. De forma mais simples, a significância de um nuclídeo é a máxima porcentagem de contribuição daquele nuclídeo para a área total do espectro sub-

traído. Núclídeos com significância acima do limiar são considerados para identificação. O limiar sugerido por CANBERRA (2006) é de 5%.

Com os núclídeos candidatos, uma otimização desses limiares é realizada através de um processo iterativo onde o χ^2 é calculado em canais dentro de $1,5 \times FWHM$ a partir do centróide de cada pico, variando-se a atividade do pico em intervalos espaçados entre 0% e 100% (0%, 25%, 50%, 75% e 100%) da atividade provável (estimada como 0 e o dobro da mínima atividade estimada anteriormente). A cada novo parâmetro ótimo encontrado, o número de intervalos de busca é reduzido para 60% da lista anterior, até que um valor ótimo seja encontrado após suficientes iterações.

Para o presente estudo, este passo foi simplificado efetuando-se um ajuste não-linear da curva gaussiana sobre o pico de interesse, uma vez que o processo de ajuste inevitavelmente minimizará o valor de χ^2 .

A identificação dos núclídeos é feita posteriormente calculando-se a correlação entre as áreas dos picos no espectro subtraído e as áreas esperadas. As áreas são estimadas, novamente, dentro de $1,5 \times FWHM$ a partir do centróide de cada pico. O coeficiente de correlação é calculado a partir da seguinte expressão:

$$r = \frac{\sum_i (M_i - \bar{M})(N_i - \bar{N})}{\sqrt{\sum_i (M_i - \bar{M})^2} \sqrt{\sum_i (N_i - \bar{N})^2}}, \quad (\text{A.9})$$

onde M_i é o valor de contagem no espectro subtraído no canal i , \bar{M} é a média de valores do espectro subtraído, N_i é o valor estimado de contagem do espectro do núclídeo no canal i e \bar{N} é a média das contagens esperadas no espectro estimado do núclídeo. Valores de r acima do limiar de sensibilidade são dados como identificados.