

Universidade do Estado do Rio de Janeiro

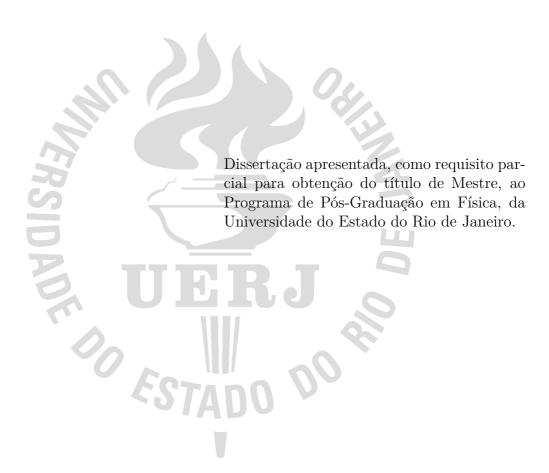
Centro de Tecnologia e Ciências Instituto de Física Armando Dias Tavares

Dalmo da Silva Dalto

Discriminação de sinais de detectores de partículas utilizando técnicas de aprendizado de máquina

Dalmo da Silva Dalto

Discriminação de sinais de detectores de partículas utilizando técnicas de aprendizado de máquina



Orientador: Helio Nogima

CATALOGAÇÃO NA FONTE UERJ/ REDE SIRIUS / BIBLIOTECA CTC/D

D152d Dalto, Dalmo da Silva.

Assinatura	Data
Autorizo, apenas para fins acadêmicos e científicos, a reproduç desta dissertação, desde que citada a fonte.	ão total ou parcial
Bibliotecária: Teresa da Silva CRB7/5209	
C	DU 539.1.074
1. Partículas (Física nuclear) - Detectores – Teses. 2. Processame Inteligência artificial - Teses. 3. Aprendizado do computador – Teses Helio (Orient.). II. Universidade do Estado do Rio de Janeiro. Institut Armando Dias Tavares. III. Título.	. I. Nogima,
Orientador: Helio Nogima. Dissertação (mestrado) - Universidade do Estado do Rio de Janeir Física Armando Dias Tavares.	ro, Instituto de
Discriminação de sinais de detectores de partículas utilizando técr aprendizado de máquina / Dalmo da Silva Dalto. – 2023. 145 f.: il.	nicas de

Dalmo da Silva Dalto

Discriminação de sinais de detectores de partículas utilizando técnicas de aprendizado de máquina

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Física, da Universidade do Estado do Rio de Janeiro.

Aprovada em 18 de Agosto de 2023. Banca Examinadora:

> Helio Nogima (Orientador) Instituto de Física Armando Dias Tavares – UERJ

> Prof. Dra. Eliza Melo da Costa Instituto de Física Armando Dias Tavares – UERJ

> Prof. Dr. Dilson de Jesus Damião Instituto de Física Armando Dias Tavares – UERJ

Prof. Dr. Gilvan Augusto Alves Centro Brasileiro de Pesquisas Físicas

Prof. Dr. Anderson Campos Fauth Universidade Estadual de Campinas

DEDICATÓRIA

Eu dedico essa dissertação aos meus falecidos professores Paulo Brigagão e Jorge Trovão, o legado continua. Dedico principalmente a minha avó Jacira Maria da Silva. Dedico também, a todos meus familiares que ajudaram na minha formação.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Gostaria de expressar minha profunda gratidão a várias pessoas que tornaram possível a conclusão desta dissertação e enriqueceram minha jornada acadêmica. Em primeiro lugar, minha sincera gratidão ao Dr. Helio Nogima, cujo apoio foi fundamental para o êxito deste trabalho. Sua orientação e paciência em compartilhar seus vastos conhecimentos foram inestimáveis, permitindo-me desenvolver meu senso crítico e abraçar novos horizontes.

Agradeço ao professor Paulo Orenstein pelo curso de machine learning lecionado no IMPA, que foi de extrema importância para entender o princípio e os fundamentos de aprendizado de máquina. Agradeço ao meu amigo Leonardo por ter me convidado a fazer o curso de verão do IMPA. Também quero agradecer a Andre Sznajder e Antônio Vilela Pereira, pois o curso de aprendizado de máquina lecionado da pós-graduação, me preparou tecnicamente para a construção dos modelos de aprendizado de máquina.

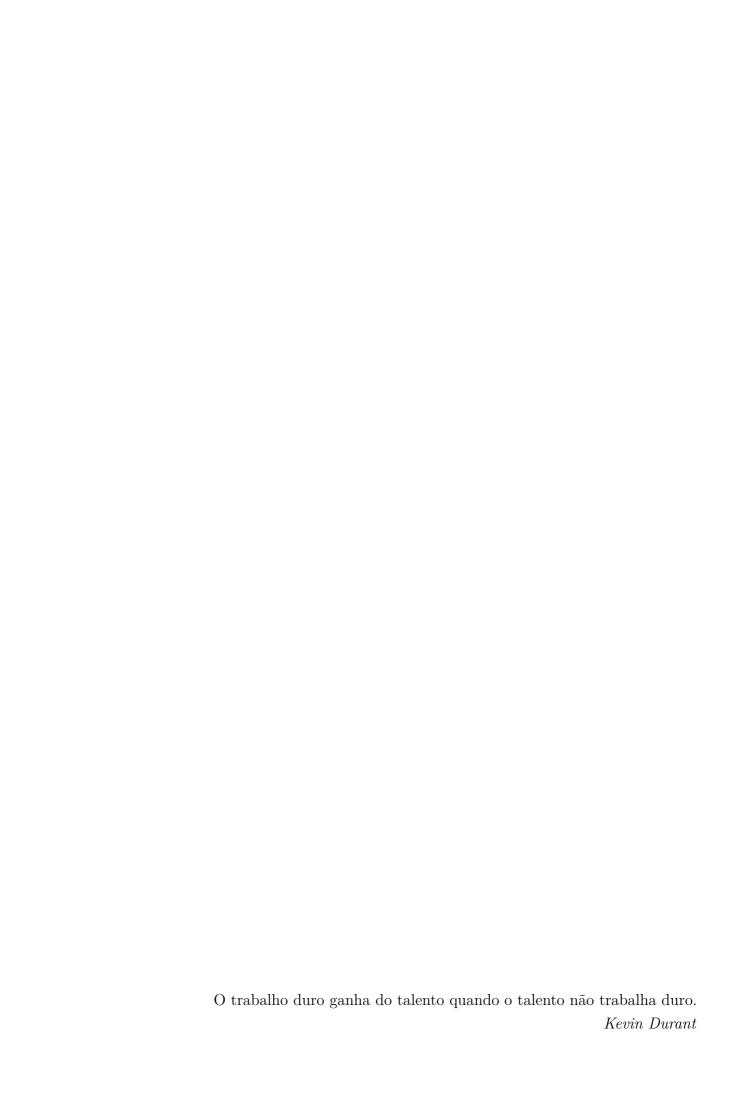
Agradeço também aos grandes amigos fiz durante a faculdade e que continuaram ao meu lado durante a pós-graduação, o nosso famoso time 7: Em especial, quero mencionar Felipe Fernandes Garcia, que desde o início da graduação em física, sempre esteve presente para me auxiliar e me motivou a seguir no mestrado, sem os seus concelhos não teria entrado para o mestrado. A João Paulo Sampaio, que tem sido um apoio desde os primeiros dias na faculdade, agradeço sua constante ajuda e espero que daqui a 4 anos, todos nós três alcancemos o título de doutor.

Não posso deixar de mencionar meus queridos professores, que foram verdadeiras inspirações para minha escolha na área da física. Ao saudoso professor Paulo Brigagão, cuja memória continua a me guiar, e ao professor Nelson Lage, que me incentivou a seguir esse caminho na física, minha eterna gratidão.

Minha família também desempenhou um papel fundamental nesta jornada, sempre me apoiando incondicionalmente. À minha avó Jacira Maria da Silva, meu irmão Igor da Silva Dalto, minha mãe Ana Paula da Silva e minhas tias Iracema da Silva, Severina da Silva e Célia da Silva meu profundo agradecimento por todo amor e incentivo que me concederam.

E a Jorge Trovão, um amigo e pai especial que me ensinou a ser resiliente e preparou-me para enfrentar os desafios da vida, meu sincero agradecimento.

Por fim, sou grato a todos que, de alguma forma, participaram positivamente em minha vida. Cada contribuição, por menor que tenha sido, ajudou a moldar o caminho que percorri até aqui. Que essa jornada de aprendizado e crescimento continue, e que a gratidão que sinto hoje se converta em ações positivas para o mundo ao meu redor.



RESUMO

DALTO, D. S. Discriminação de sinais de detectores de partículas utilizando técnicas de aprendizado de máquina. 2023. 145 f. Dissertação (Mestrado em Física) – Instituto de Física Armando Dias Tavares, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

Um dos problemas básicos na instrumentação de detectores de partículas é a otimização da relação sinal/ruído no estágio inicial de coleta do pulso elétrico. Na eletrônica de processamento, além da adoção de filtros e formatação, a discriminação de sinais é usualmente obtida impondo um valor limiar de amplitude. Normalmente esta solução é suficiente, dado que os ruídos possuem menor amplitude comparados aos sinais na maior parte dos detectores. Entretanto, a depender da aplicação e do tipo de detector empregado este método filtra apenas parte dos ruídos mantendo aqueles de alta amplitude, além de causar o efeito indesejável de eliminar sinais de baixa amplitude. Em casos como esses métodos de discriminação independente do tamanho dos sinais elétricos devem ser considerados. O uso de técnicas de aprendizado de máquina é uma alternativa interessante, dada a crescente disponibilidade de dispositivos programáveis do tipo FPGA (Field Programmable Gate Arrays) de alta velocidade para aplicações de inteligência artificial. O método leva em conta as diferenças nas formas de onda entre sinais e ruídos que podem ser usadas na discriminação dos pulsos elétricos gerados pelos detectores, independente do seu tamanho. Nesta dissertação apresentamos um estudo da discriminação de sinais utilizando diferentes algoritmos de aprendizado de máquina aplicados a um detector cintilador. Foram utilizados dados simulados de sinais e ruídos e dados experimentais para o estudo da discriminação de sinais por um algoritmo não-supervisionado HCA (Hierarchical Clustering Analysis) e os supervisionados: Floresta Aleatória, Redes Neurais Profundas e Redes Neurais Convolucionais. Uma comparação do desempenho dos diferentes algoritmos supervisionados com o método de discriminação por amplitude de pulso é apresentada.

Palavras-chave: Aprendizado de máquina. Detectores de partículas. Formas de onda. Discriminação de sinais.

ABSTRACT

DALTO, D. S. Waveform discrimination of particle detectors using machine learning techiques. 2023. 145 f. Dissertação (Mestrado em Física) – Instituto de Física Armando Dias Tavares, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

One of the fundamental problems in particle detector instrumentation is optimizing the signal-to-noise ratio in the early stage of electrical pulse collection. Besides adopting filters and formatting, signal discrimination is usually achieved in electronic signal processing by imposing an amplitude threshold value. This solution is generally enough since the noise has a lower amplitude than the signals in most detectors. However, depending on the application and the type of detector used, this method filters only a part of the noise, retaining those with high amplitude, and it also has the undesirable effect of eliminating low-amplitude signals. In cases like these, methods of discrimination independent of the size of the electrical signals should be considered. The use of machine learning techniques is an interesting alternative, given the increasing availability of high-speed programmable devices like Field Programmable Gate Arrays (FPGAs) for artificial intelligence applications. The method considers the differences in waveforms between signals and noise, which can be used in discriminating the electrical pulses generated by the detectors, regardless of their size. In this dissertation, we present a study of signal discrimination using different machine learning algorithms applied to a scintillator detector. Simulated data of signals and noise and experimental data were used to study signal discrimination using an unsupervised algorithm called Hierarchical Clustering Analysis (HCA) and supervised algorithms such as Random Forest, Deep Neural Networks, and Convolutional Neural Networks. A comparison of the performance of the different supervised algorithms with the pulse amplitude discrimination method is presented.

Keywords: Machine learning. Waveforms. Particle detectors. Signal discrimination.

LISTA DE FIGURAS

Figura 1 - Perda de energia da interação de múons com o metal	25	2
Figura 2 $$ - Perda de energia para elétrons, múons, píons, prótons, dê	uterons e	
partículas α no ar	25	2
Figura 3 - Perda de energia por interações de fótons	24	4
Figura 4 - Efeito Compton	2	5
Figura 5 - Contador proporcional	2'	7
Figura 6 - Efeito avalanche	2'	7
Figura 7 - Dependência de íons coletados em função do campo	28	8
Figura 8 - Linhas do campo elétrico de uma MWPC	29	9
Figura 9 - Câmara multifilar proporcional	29	9
Figura 10 - Câmara de arrasto	30	C
Figura 11 - Esquema de um detector cintilador	3	1
Figura 12 - Esquematização do mecanismo de cintilação de materiais org	gânicos 32	2
Figura 13 - Banda de estrutura de cristais inorgânicos $\ \ldots \ \ldots \ \ldots$	3	3
Figura 14 - Esquema Básico de uma fotomultiplicadora	3	3
Figura 15 - Eficiência quântica de diversos materiais do fotocátodo	34	4
Figura 16 - Esquema de um circuito de divisor de tensão	30	6
Figura 17 - Detector semicondutor	3'	7
Figura 18 - Detector APD	38	8
Figura 19 - Detector SiPM	38	8
Figura 20 - Relação de interpretabilidade X flexibilidade	40	(
Figura 21 - Relação entre viés e variância de acordo com a complexidade	e do modelo 4:	2
Figura 22 - Esquematização em blocos de aprendizado supervisionado $$.	43	3
Figura 23 - Esquematização da FA	43	3
Figura 24 - Unidade linear de limiar	4	4
Figura 25 - Rede neural de uma camada oculta	4	5
Figura 26 - Rede neural de 2 camadas oculta $\ \ldots \ \ldots \ \ldots \ \ldots \ \ldots$	40	6
Figura 27 - Camadas convolucionais com campos receptivos retangulares $% \left(1\right) =\left(1\right) \left(1$	3 48	8
Figura 28 - Conexões entre camadas convolucionais com ${\it zero}\ padding$	48	8
Figura 29 - Strides de camadas convolucionais $\ \ \ldots \ \ \ldots \ \ \ldots \ \ \ldots$	49	9
Figura 30 - Filtros de CNNs $\ \ldots \ \ldots \ \ldots \ \ldots \ \ldots \ \ldots$	49	S
Figura 31 - Empilhamento de mapas de característica em camadas convo	olucionais . 50	C
Figura 32 - Camada Pooling	5	1
Figura 33 - Arquitetura de uma CNN $\ \ldots \ \ldots \ \ldots \ \ldots \ \ldots$	52	2
Figura 34 - Método gradiente descendente	5	3
Figura 35 - Representação da Matriz de Confusão	54	4

Figura 36 -	- Curva da ROC	55
Figura 37 -	Exemplo de uma clusterização	56
_	Passos de uma clusterização Hierárquica	57
_	dendrograma de clusterização aglomerativa	58
_	Seleção das componentes do PCA	59
Figura 41 -	Proporção de variância explicada em função das componentes c	60
_		64
Figura 43 -	· Ajustes de funções em um sinal de cintilação experimental (trigger de	cc
T3: 4.4	,	66
_	v 3	67
_	Sistema experimental com trigger de coleta de sinais de cintilação	69 7 0
<u> </u>	Formas de onda experimentais de sinais de cintilação	70
<u> </u>	Sistema experimental com <i>trigger</i> de coleta de ruídos	71
_	1	71
Figura 49 -	· Taxa de variância explicada em função das dimensões para FOs de sinais de múons cósmicos e ruídos de corrente escura para 1600V	74
Figura 50 -	Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente escura para 1600V	75
Figura 51 -	· Clusterização de FOs de sinais de múons cósmicos e ruídos de corrente	
J		75
Figura 52 -	Formas de onda da clusterização para FOs de sinais de múons cósmicos	
Figura 53 -	e ruído de corrente escura de 1600V	76
	múons cósmicos e ruídos de corrente escura	77
Figura 54 -	· Taxa de variância explicada em função das dimensões de FOs de sinais	
	de múons cósmicos e de fundo de 1500V	80
Figura 55 -	Dendrograma de FOs de sinais de múons cósmicos e de fundo de 1500V	80
Figura 56 -	· Clusterização de FOs de sinais de múons cósmicos e de fundo de 1500V	81
Figura 57 -	Formas de onda da clusterização de FOs de sinais de múons cósmicos	
	e de fundo de 1500V	82
Figura 58 -	FOs de sinais de múons cósmicos e de ruídos que apresentam pulso cintilação de 1500V	83
Figura 59 -	· Clusterização de FOs de sinais de múons cósmicos e de fundo para cada	
F): CO	S	83
9	· Clusterização de FOs de sinais de múons cósmicos e de fundo para 1850V	
_	FOs contidas na região de protuberância	84
_	Rede DNN com função ativação de saída softmax	86
_	•	87
rıgura 64 -	· Arquitetura da DNN para treinamento de FOs simuladas	88

Figura 65 - Arquitetura da DNN de FOs de sinais de múons cósmicos e ruído de corrente escura	8
Figura 66 - Treinamento da DNN para FOs simuladas	9
Figura 67 - Treinamento da DNN para FOs de sinais de múons cósmicos e ruído	
de corrente escura	0
Figura 68 - Arquitetura DNN de FOs de sinais de múons cósmicos e de fundo 92	1
Figura 69 - Treinamento da DNN para FOs de sinais de múons cósmicos e de fundo 99	
Figura 70 - Arquitetura da CNN para FOs simuladas	
Figura 71 - Arquitetura da CNN para FOs de sinais de múons cósmicos e ruído de	
corrente escura	4
Figura 72 - Treinamento da CNN para FOs simuladas	
Figura 73 - Treinamento da CNN para FOs do tipo sinais de múons cósmicos e	
ruído de corrente escura	റ
Figura 74 - Arquitetura da CNN para FOs de sinais de múons cósmicos e de fundo 9'	
Figura 75 - Treinamento da CNN de FOs de sinais de múons cósmicos e de fundo . 98	
Figura 76 - Desempenho da FA treinada por FOs simuladas de sinais de múons	_
cósmicos e ruído de corrente escura	n
Figura 77 - Desempenho da DNN treinada por FOs simuladas de sinais de múons	,
cósmicos e ruído de corrente escura	n
Figura 78 - Desempenho da CNN treinada por FOs simuladas de sinais de múons	J
cósmicos e ruído de corrente escura	1
Figura 79 - Desempenho da FA treinada por FOs de sinais de múons cósmicos e	_
ruídos de corrente escura	1
Figura 80 - Desempenho da DNN treinada por FOs de sinais de múons cósmicos e	T
ruídos de corrente escura	2
Figura 81 - Desempenho da CNN treinada por FOs de sinais de múons cósmicos e	_
ruídos de corrente escura	า
Figura 82 - Desempenho dos modelos treinados por amostra simuladas de FOs de	_
sinais de múons cósmicos e ruídos de corrente escura	3
Figura 83 - Desempenho dos modelos treinado por dados experimentais de FOs de	J
sinais de múons cósmicos e ruídos de corrente escura	9
Figura 84 - Desempenho da FA treinada por FOs de sinais de múons cósmicos e de)
fundo	1
	Ŧ
Figura 85 - Desempenho da DNN treinada por FOs de sinais de múons cósmicos e	_
de fundo	Э
Figura 86 - Desempenho da CNN treinada por FOs de sinais de múons cósmicos e	_
de fundo	Ó
Figura 87 - Desempenho dos modelos treinados com FOs sinais de múons cósmicos	_
e de fundo	1

Figura 88 - Discriminação por altura de pulso de FOs de sinais de múons cósmicos	
e ruídos de corrente escura	. 108
Figura 89 - Discriminação por altura de pulso de FOs de sinais de múons cósmicos	
e de fundo	. 109
Figura 90 - Árvore de decisões	. 115
Figura 91 - Árvore de decisão mais complexa	. 116
Figura 92 - Algoritmo $Bagging$. 117
Figura 93 - Método de reamostragem bootstrap	. 119
Figura 94 - Método de reamostragem Validação cruzada K -fold	. 120
Figura 95 - Variância explicada em função das dimensões de FOs de sinais de múons	
cósmicos e ruídos de corrente escura de 1750V)	. 121
Figura 96 - Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente	
escura de 1750 V	. 122
Figura 97 - Clusterização de FOs de sinais de múons cósmicos e ruídos de corrente	
escura de 1750 V	. 122
Figura 98 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e ruídos de corrente escura de 1750V	. 123
Figura 99 - Variância explicada em função das dimensões de FOs de sinais de múons	
cósmicos e ruídos de corrente escura de 1850V	. 123
Figura 100 - Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente	
escura de 1850V \dots	. 124
Figura 101 - Clusterização de FOs de sinais de múons cósmicos e ruídos de corrente	
escura de 1850 V	. 124
Figura 102 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e ruídos de corrente escura de 1850 V	. 125
Figura 103 - Variância explicada em função das dimensões para 1950 V 	. 125
Figura 104 - Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente	
escura de 1950 V	. 126
Figura 105 - Clusterização de FOs de sinais de múons cósmicos e ruídos de corrente	
escura de 1950V)	. 126
Figura 106 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e ruídos de corrente escura de 1950 V	. 127
Figura 107 - Variância explicada em função das dimensões de FOs de sinais de múons	
cósmicos e de fundo de 1550 V	. 128
Figura 108 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de 1550	V 128
Figura 109 - Clusterização de FOs de sinais de múons cósmicos e de fundo de 1550	V 129
Figura 110 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e de fundo de 1550 V	. 129

Figura 111 - Variância explicada em função das dimensões de FOs de sinais de múons	
cósmicos e de fundo de 1600 V	130
Figura 112 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de $1600\mathrm{V}$	130
Figura 113 - Clusterização de FOs de sinais de mú ons cósmicos e de fundo de $1600\mathrm{V}$	131
Figura 114 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e de fundo de 1600 V	131
Figura 115 - Variância explicada em função das dimensões de FOs de sinais de múons	
cósmicos e de fundo de 1750V) $\dots \dots \dots \dots \dots \dots$	132
Figura 116 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de $1750\mathrm{V}$	132
Figura 117 - Clusterização das formas de onda de FOs de sinais de múons cósmicos	
e de fundo de 1750 V	133
Figura 118 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e de fundo de 1750 V	134
Figura 119 - Variância explicada em função das dimensões de FOs de sinais de múons	
cósmicos e de fundo de 1850V) $\dots \dots \dots \dots \dots \dots$	134
Figura 120 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de $1850\mathrm{V}$	135
Figura 121 - Clusterização de FOs de sinais de mú ons cósmicos e de fundo de $1850\mathrm{V}$	135
Figura 122 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e de fundo de 1850 V	136
Figura 123 - Variância explicada em função das dimensões de FOs de sinais de múons	
cósmicos e de fundo de 1950 V	136
Figura 124 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de 1950 V	137
Figura 125 - Clusterização de FOs de sinais de mú ons cósmicos e de fundo de $1950\mathrm{V}$	137
Figura 126 - Formas de onda da clusterização de FOs de sinais de múons cósmicos	
e de fundo de 1950 V	138
Figura 127 - Função densidade de probabilidade	140
Figura 128 - Eficiência da função densidade de probabilidade $P(\epsilon;8,10)$	142
Figura 129 - Parada antecipada	143
Figura 130 - $Dropout$	144

LISTA DE TABELAS

Tabela 1	- Materiais do fotocátodo
Tabela 2	- Proporção de classe em cada <i>cluster</i> para FOs de sinais de múons cós-
	micos e ruídos de corrente escura para 1600 e 1750 V 76
Tabela 3	- Proporção de classe em cada $cluster$ para FOs de sinais de múons cós-
	micos e ruídos de corrente escura para 1850 e 1950V
Tabela 4	- Proporção de classe em cada cluster para FOs de sinais de múons cós-
	micos e de fundo para 1500, 1550 e 1600 V 79
Tabela 5	- Proporção de classe em cada ${\it cluster}$ para FOs de sinais de múons cós-
	micos e fundo para 1750, 1850 e 1950 V $\dots \dots $
Tabela 6	- Estimativa da acurácia por tratamento Bayesiano de modelos treinados
	por FOs simuladas de sinais de múons cósmicos e ruído de corrente escura 100
Tabela 7	- Estimativa da acurácia por meio de um tratamento Bayesiano de mode-
	los treinados por FOs de sinais de múons cósmicos e ruídos de corrente
	escura
Tabela 8	- Estimativa da acurácia por meio de tratamento Bayesiano de modelos
	treinado por FOs de sinais de múons cósmicos e de fundo 105

LISTA DE ABREVIATURAS E SIGLAS

APD fotodiodo de avalanche PMT Fotomultiplicadora

FO Forma de onda

MLP Perceptron multicamada

FA Floresta Aleatória

LTU Unidade linear de limiar MLP Perceptron multicamada DNN Rede neural profunda

CNN Rede neural convolucionais

NN Rede neural FP Falso positivo

TN Verdadeiro negativo TP Verdadeiro positivo

FN Falso negativo

ROC Características operacionais do receptor

AUC Área abaixo da curva

TPR Taxa de verdadeiros positivosTNR Taxa de verdadeiros negativosNIM Nuclear Instrument Module

LFNP Laboratório de Física Nuclear e Partículas

OOB out of bag

AM Aprendizado de máquina

PCA Análise de Componentes Principais

SUMÁRIO

	INTRODUÇÃO
1	DETECTORES DE PARTÍCULAS 20
1.1	Interação da radiação com a matéria
1.1.1	Interação de partículas carregadas
1.2	Bremsstrahlung
1.3	Interação de fótons
1.3.1	Efeito fotoelétrico
1.3.2	<u>Efeito Compton</u>
1.3.3	Produção de Pares
1.4	Detectores de ionização
1.4.1	<u>Câmara proporcional multifilar</u>
1.4.2	<u>Câmara de arrasto</u>
1.5	Detector cintilador
1.5.1	Materiais orgânicos
1.5.2	Materiais inorgânicos
1.6	Fotomultiplicadora
1.6.1	<u>Fotocátodo</u>
1.6.2	Seção eletro-multiplicadora e eletrodos focalizadores
1.6.3	<u>Divisor de tensão</u>
1.7	Detectores semicondutores
1.7.1	<u>Detector APD</u>
1.7.2	<u>Detector SiPMs</u>
2	APRENDIZADO DE MÁQUINA 39
2.1	Seleção do algoritmo de aprendizado de máquina 40
2.1.1	Relação entre interpretabilidade e flexibilidade
2.1.2	Relação entre viés e variância
2.2	Aprendizado supervisionado
2.2.1	Floresta Aleatória (Random Forest)
2.3	Aprendizado profundo
2.3.1	<u>Perceptron</u>
2.3.2	Rede neural de uma camada interna
2.3.3	Rede neural de múltiplas camadas
2.4	Rede neural convolucional
2.4.1	Camada convolucional
2.4.2	Filtros e empilhamento de múltiplos mapas de características 47
2.4.3	Camada de pooling

2.4.4	Arquiteturas CNN	. 51
2.5	Retropropagação	. 52
2.6	Medições de Desempenho	. 53
2.6.1	<u>Curva ROC</u>	. 54
2.7	Aprendizado não supervisionado	. 55
2.7.1	Medição da distância do agrupamento	. 55
2.8	Clusterização hierárquica	. 56
2.9	Redução de dimensionalidade	. 57
2.9.1	Análise de Componentes Principais (PCA)	. 58
3	METODOLOGIA	. 61
4	FORMAÇÃO DOS SINAIS E RUÍDOS E SUA SIMULAÇÃO	. 63
4.1	Formação do sinal elétrico	. 63
4.2	Ruídos da fotomultiplicadora	. 64
4.2.1	Emissão termoiônica	. 64
4.2.2	Fuga de corrente	. 65
4.2.3	Afterpulse	. 65
4.3	Simulação das formas de onda	. 65
4.3.1	Sinais de cintilação	. 65
4.3.2	Ruídos de corrente escura	. 66
5	SISTEMA EXPERIMENTAL	. 69
5.1	Coleta de FOs de sinais de cintilação	. 69
5.2	Coleta de FOs de ruídos	. 70
5.3	Processamento das FOs e linguagem de programação	. 70
6	ANÁLISE DE CLUSTERIZAÇÃO	. 73
6.1	FOs de sinais de múons cósmicos e ruído de corrente escura	. 73
6.2	FOs de sinais de múons cósmicos e de fundo	. 78
7	CONSTRUÇÃO DOS MODELOS SUPERVISIONADOS	. 85
7.1	FA	. 85
7.2	DNN e CNN	. 85
7.2.1	DNN - FOs de sinais de múons cósmicos e ruído de corrente escura	. 87
7.2.2	DNN - FOs de sinais de múons cósmicos e de fundo	. 87
7.2.3	CNN - FOs de sinais de múons cósmicos e ruído de corrente escura	. 91
7.2.4	$\underline{\mathrm{CNN}}$ - FOs de sinais de múons cósmicos e de fundo	. 95
8	ANÁLISE DOS RESULTADOS	. 99
8.1	FOs de sinais de múons cósmicos e ruídos de corrente escura	. 99
8.2	FOs de sinais de múons cósmicos e de fundo	. 103
8.3	Discriminação por altura de pulso	. 107
	CONCLUSÃO	. 110
	REFERÊNCIAS	. 112

APÊNDICE	${\bf A}$ – Algoritmos de aprendizado de máquina
APÊNDICE	${f B}$ – Métodos de reamostragem
APÊNDICE	${f C}$ – Clusterização das formas de onda
APÊNDICE	D – Tratamento Bayesiano para estimar as incertezas das
eficiências .	
APÊNDICE	${f E}$ – Técnicas de regularização

INTRODUÇÃO

Detectores de partículas são dispositivos que registram interações de partículas com a matéria. Eles surgiram junto com o descobrimento e os estudos da radioatividade natural no final do século XIV e início do século XX. Desde então uma alta multiplicidade de tecnologias de dispositivos detectores foram desenvolvidos. Partindo de um uso puramente científico, os detectores de partículas logo encontraram aplicações em outros campos, principalmente na medicina. Porém, o berço do desenvolvimento das novas tecnologias tem acontecido na maior parte no campo da física experimental de partículas, onde as aplicações exigem alto desempenho desses dispositivos induzindo inovações no seu desenho.

Em todos os modernos detectores de partículas, o processo básico para a detecção é sempre o mesmo: o depósito de energia da partícula incidente no meio é convertido direta ou indiretamente em uma corrente elétrica. Por se tratar de um processo rápido, isso corresponde a um pulso elétrico que identifica a presença da partícula, que nesse trabalho é o sinal da partícula. Na prática, outros processos podem ocorrer gerando pulsos elétricos não produzidos pela partícula incidente, esses são denominados ruídos ou **fundo**, a depender do contexto. Os pulsos elétricos são tratados por uma eletrônica de leitura que tem como uma das suas funções a discriminação dos sinais das partículas, ou seja, a separação de sinais de ruídos. Além da adoção de filtros e da formatação do sinal, a discriminação normalmente é feita impondo um valor de limiar de amplitude do sinal elétrico. Geralmente, essa abordagem é suficiente, uma vez que os ruídos tendem a ter menor amplitude em comparação com os sinais, por força do projeto dos detectores. No entanto, dependendo da aplicação e do tipo do detector empregado esse método filtra apenas parte dos ruídos, mantendo aqueles de alta amplitude, podendo também causar o efeito indesejável de eliminar sinais de baixa amplitude. Para contornar esses problemas, métodos mais sofisticados devem ser empregados.

Da mesma forma como se verifica no cotidiano, o uso da inteligência artificial (IA) vem se tornando cada vez mais comum nas diversas áreas do conhecimento e em vários extratos das tecnologias. Isso inclui o processamento de sinais elétricos. Em medicina, por exemplo, modelos de aprendizado de máquina (AM) são implementados em eletrocardiogramas para classificação de batimentos normais e anormais (Muhammad et. al., 2020) e identificação de aneurismas (Tianqi et. al., 2023.); em produtos de consumo como carros e celulares, os padrões dos sinais dos sensores identificam a movimentação, obstáculos e inúmeras outras aplicações. O estudo de aplicações de IA em pulsos elétricos de detectores de radiação não é novo (Clements; Miao, 1995), mas a crescente disponibilidade de dispositivos FPGAs (do inglês, field programmable gate array) de alta velocidade (Furletov, 2022) tornam cada vez mais interessantes o aproveitamento da informação analógica

do pulso elétrico dos detectores. A ideia é processar o sinal elétrico no momento da sua captura e discriminar os pulsos de interesse na malha eletrônica de tratamento de sinais sem necessariamente impor um limiar de discriminação em amplitude.

Este trabalho tem como objetivo estudar a discriminação de sinais de detectores de partículas com o uso de IA. Para isso foi utilizado como objeto de teste um detector cintilador plástico acoplado a um tubo fotomultiplicador (PMT, do inglês). Esta escolha foi feita por ser um detector popular e cujo pulso elétrico, tanto de sinal como de ruído, pode ser fisicamente descrito de forma relativamente simples. Além disso, esse tipo de detector é conhecido por produzir alta taxa de ruídos de baixa amplitude, devido à corrente escura da PMT, o que facilita a avaliação do método. Foram construídos modelos supervisionados e não supervisionado de aprendizado de máquina para estudar o desempenho da discriminação de sinais com base na forma de onda (FO) dos pulsos elétricos. Numa primeira etapa foram usadas formas de onda simuladas, e posteriormente dados experimentais. Embora tenha sido aplicada para o caso específico do detector cintilador, o método estudado pode ser estendido para outros tipos de detectores.

A dissertação está dividida da seguinte forma: no Capítulo 1 são discutidos os processos físicos de interação da radiação com a matéria e alguns dos tipos de detectores mais utilizados em física de partículas; o Capítulo 2 resume o princípio do aprendizado de máquina, descrevendo o funcionamento dos algoritmos que foram utilizados. Também são apresentadas as métricas de desempenho que foram usadas para avaliar os modelos; o Capítulo 3 faz uma descrição da metodologia de análise que segue nos próximos capítulos; a formação de sinais e ruídos de detectores cintilador são abordados no Capítulo 4. Preferiu-se fazer essa discussão neste capítulo pela possibilidade de envolvê-la com a análise e facilitar a leitura. Logo em seguida, o Capítulo 5 expõe o sistema experimental e a coleta de dados; a análise de clusterização não supervisionada é discutida no (Capítulo 6); a construção de cada modelo supervisionado é explicada no Capítulo 7; o Capítulo 8 apresenta os resultados de cada classificador e a eficiência das previsões. Também são expostas as eficiências que seriam obtidas pela discriminação por altura de pulso. Por fim, a conclusão da dissertação é feita na Seção 8.3.

1 DETECTORES DE PARTÍCULAS

Neste capítulo será feita uma breve discussão sobre os processos físicos fundamentais utilizados na detecção de partículas e uma rápida descrição do funcionamento de alguns dos principais detectores utilizados em física de altas energias.

1.1 Interação da radiação com a matéria

O funcionamento dos detectores se baseia na interação da partícula com o material que compõem o detector. Nos detectores eletrônicos, que é a única categoria que consideramos neste trabalho, esse processo deve resultar na produção de um sinal elétrico que identifica a passagem da partícula. O estudo da interação de partícula com a matéria usualmente é dividido em três categorias segundo o tipo da partícula interagente: partículas carregas pesadas, radiação gama, elétrons e pósitrons. Neste capítulo serão descritos os principais processos de perda de energia importantes para a detecção de partículas. A geração de fótons por efeito Cherenkov (Leo, 1987) e por radiação de transição (Grupen et. al., 2009) embora sejam importantes para a detecção de partículas, não são relevantes para a perda de energia e portanto não serão discutidas aqui.

1.1.1 Interação de partículas carregadas

Quando uma partícula carregada leve atravessa um material ocorre uma perda de energia cinética por meio de excitação e ionização. Interações de excitações como a da R1 resultam em emissões de fótons de baixa energia que são úteis para detecção de partículas. Sabendo que a energia cinética máxima transferida para um elétron depende da massa em repouso (m_0) e do momento da partícula incidente (Equação 1), onde γ é o fator de Lorentz (= E/m_0c^2), β é a razão entre a velocidade da partícula (v) e a velocidade da luz no vácuo (c) e m_e é a massa do elétron, o trabalho realizado pelo elétron é dado pela Equação 2.

$$e^- + \text{átomo} \longrightarrow \text{átomo}^* + e^- \longrightarrow \text{átomo} + \gamma + e^-$$
 (R1)

$$p = \gamma m_0 \beta c \tag{1}$$

$$T_{max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e / m_0 + (m_e / m_0)^2}$$
 (2)

A perda de energia se dá exclusivamente por colisões inelásticas, das quais frações muito pequenas de energia são transferidas. Entretanto, o número de colisões por unidade de comprimento é muito grande, produzindo uma perda de energia suficiente para uma medição. A perda de energia (dE) por unidade de comprimento (dX) pode ser tratada através da fórmula de Bethe-Bloch (Equação 3), onde z é a carga da partícula incidente, Z e A número atômico e massa atômica do material absorvedor, I é a energia média de excitação (uma característica do material absorvedor), r_e é o raio clássico do elétron e N_A é o número de Avogadro. O parâmetro δ descreve o quanto o campo elétrico transversal de uma partícula relativística é rastreável pela densidade de carga do átomo de elétrons.

A Equação 3 é válida apenas para perda de energia por ionização ou excitação de partículas carregadas pesadas. A Figura 1 mostra tal perda de energia por ionização e excitação de múons no ferro, e a Figura 2 exibe perda de energia de elétrons, múons, píons, prótons, dêuterons e partículas α no ar.

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A} \frac{1}{\beta^2} \left(ln \left(\frac{2m_e c^2 \gamma^2 \beta^2}{I} \right) - \frac{\delta}{2} \right)$$
 (3)

Para partículas carregadas leves a equação de Bethe-Bloch pode ser aproximada para a Equação 4, onde δ^* apresenta um valor diferente do δ da Equação 3, uma forma mais completa da equação para a perda de energia do elétron é mostrada na Equação 5, onde leva-se em conta colisões elétron-elétron e efeito de trajetória.

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A} \frac{1}{\beta^2} \left(\ln \left(\frac{\gamma m_e c^2}{2I} \right) - \beta^2 - \frac{\delta^*}{2} \right)$$
 (4)

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\ln \left(\frac{\gamma m_e c^2 \beta \sqrt{\gamma - 1}}{\sqrt{2}I} \right) + \frac{1}{2} \left(1 - \beta^2 \right) - \frac{2\gamma - 1}{2\gamma^2} \ln 2 + \frac{1}{16} \left(\frac{\gamma - 1}{\gamma} \right)^2 \right]$$
(5)

O tratamento de perda de energia para os pósitrons, onde a massa é a mesma do elétron, mas a carga é diferente, a equação de Bethe-Bloch pode ser aproximada para Equação 6.

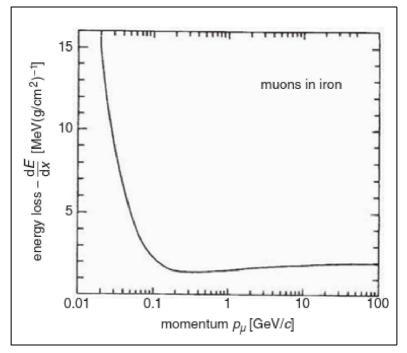


Figura 1 - Perda de energia da interação de múons com o metal.

Legenda: Perda de energia por ionização e excitação por múons no metal, e sua dependência com o momento do múon.

Fonte: Grupen et. al., 2009, p. 676.

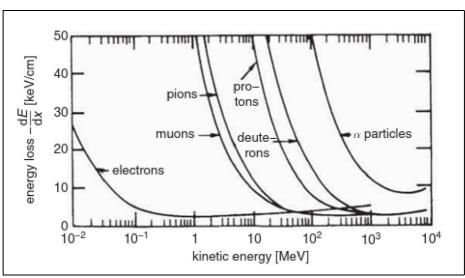


Figura 2 - Perda de energia para elétrons, múons, p
íons, prótons, dêuterons e partículas α no ar
.

Fonte: Grupen et. al., 2009, p. 676.

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\ln \left(\frac{\gamma m_e c^2 \beta \sqrt{\gamma - 1}}{\sqrt{2}I} \right) - \frac{\beta^2}{24} \left(23 + \frac{14}{\gamma + 1} + \frac{10}{(\gamma + 1)^2} + \frac{4}{(\gamma + 1)^3} \right) \right]$$
(6)

1.2 Bremsstrahlung

Quando partículas carregadas interagem com o campo elétrico do núcleo e são desaceleradas, elas emitem uma fração da energia cinética na forma de fótons, a radiação emitida é chamada de Bremsstrahlung. A perda de energia por Bremsstrahlung para altas energias pode ser descrita pela Equação 7, caso a energia perdida for por incidência de elétrons a Equação 8 deve ser considerada.

$$-\frac{dE}{dx} \approx 4\alpha N_A \frac{Z^2}{A} z^2 \left(\frac{1}{4\pi\epsilon_0} \frac{e^2}{mc^2}\right)^2 E ln \frac{183}{Z^{1/3}}$$

$$\tag{7}$$

$$-\frac{dE}{dx} \approx 4\alpha N_A \frac{Z^2}{A} r_e^2 E ln \frac{183}{Z^{1/3}} \tag{8}$$

A perda de energia por Bremsstrahlung depende exclusivamente das características do meio, para os elétrons ($z=1, m=m_e$) por exemplo, a Equação 7 e Equação 8 podem ser reescritas pela Equação 9, onde a partícula incidente atravessa uma camada (x) e X_0 é o comprimento da radiação. Ao integrar a respectiva equação, chega-se a Equação 10.

$$-\frac{dE}{dx} = \frac{E}{X_0} \tag{9}$$

$$-E = E_0 e^{-x/X_0} (10)$$

A energia total perdida por unidade de comprimento na interação e^{\pm} com a matéria pode ser dada pela soma de contribuições de efeitos de ionização e Bremsstrahlung (Equação 11).

$$\left(\frac{dE}{dx}\right)_{total} = \left(\frac{dE}{dx}\right)_{ioniza\tilde{\zeta}ao} + \left(\frac{dE}{dx}\right)_{Bremsstrahlung}$$
(11)

1.3 Interação de fótons

Fótons são detectados indiretamente por meio de conversão de energia. Eles interagem de três de formas bem distintas, podendo ser completamente absorvidos (efeito fotoelétrico, produção de pares) ou espalhados por um ângulo relativamente grande (efeito Compton).

No efeito fotoelétrico, um fóton é absorvido e sua energia é suficiente para emitir um elétron. Na produção de pares, o fóton penetra no átomo e atinge uma região próxima do núcleo, onde sua energia é convertida em um par elétron-pósitron, esse processo exige que a energia do fóton seja igual ou maior que a energia de repouso dos dois produtos. O efeito Compton é o espalhamento de um fóton por um elétron livre que resulta na diminuição da energia (aumento do comprimento de onda) do fóton espalhado. A Figura 3 exibe o alcance de operação de cada interação de fóton.

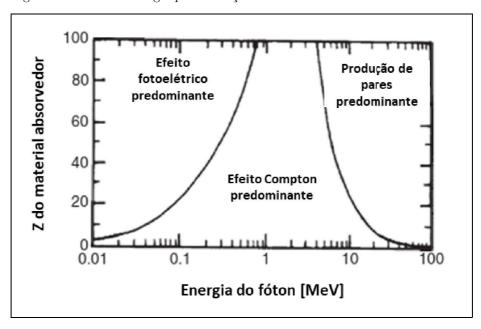


Figura 3 - Perda de energia por interações de fótons.

Legenda: Regiões que o efeito fotoelétrico, efeito Compton e produção de pares dominam, em função da energia do fóton e da carga alvo de número Z [14, 50, 53]

Fonte: Grupen et. al., 2009, p. 676. Adaptado pelo autor.

1.3.1 Efeito fotoelétrico

O efeito fotoelétrico consiste na absorção de um fóton por um material, geralmente metálico, que adquire energia suficiente para a emissão de um elétron. A energia de

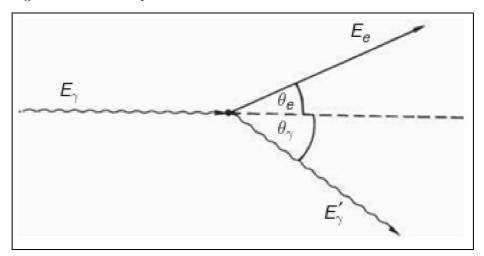
liberação é dada pela Equação 12, onde E é a energia com o qual o elétron é emitido, ν é a frequência da luz incidente e ϕ é a energia necessária para emitir o elétron (função trabalho).

$$E = h\nu - \phi \tag{12}$$

1.3.2 Efeito Compton

É o espalhamento de um fóton de energia E_{γ} em um elétron como ilustra a Figura 4. A probabilidade diferencial do espalhamento Compton $(\phi_c(E_{\gamma}, E'_{\gamma})dE'_{\gamma})$ para $m_ec^2 < E'_{\gamma} < E_{\gamma}$ é dada pela fórmula de Klein–Nishina (Equação 13), onde θ_{γ} é o espalhamento do fóton no referencial do laboratório, E_{γ} é a energia do fóton incidente, e a seção de choque total por elétron é dada pela Equação 14, onde $\epsilon = E_{\gamma}/m_ec^2$. A taxa de energia de fóton de espalhamento e fóton incidente (Equação 15) está relacionada com a seção de choque (σ_{cs}) e de seção de espalhamento (Equação 16).

Figura 4 - Efeito Compton.



Legenda: Definição das variáveis cinemáticas no espalhamento Compton Fonte: Grupen et. al., 2009, p. 676.

$$\phi_c(E_\gamma, E_\gamma') dE_\gamma' = \pi r_e^2 \frac{N_A Z}{A} \frac{m_e c^2}{E_\gamma} \frac{dE_\gamma'}{E_\gamma'} \left[1 + \left(\frac{E_\gamma'}{E\gamma} \right)^2 - \frac{E_\gamma'}{E\gamma} \operatorname{sen}^2 \theta_\gamma \right]$$
(13)

$$\sigma_c^e = 2\pi r_e^2 \left(\frac{1+\epsilon}{\epsilon^2}\right) \left[\frac{2(1+\epsilon)}{1+2\epsilon} - \frac{1}{\epsilon} \ln\left(1+2\epsilon\right) + \frac{1}{2\epsilon} \ln(1+2\epsilon) - \frac{1+3\epsilon}{(1+2\epsilon)^2}\right] \left\{ \text{cm}^2/\text{elétron} \right\}$$
(14)

$$\frac{E_{\gamma}'}{E_{\gamma}} = \frac{1}{1 + \epsilon(1 - \cos\theta_{\gamma})}\tag{15}$$

$$\sigma_{cs} = \frac{E_{\gamma}'}{E_{\gamma}} \sigma_c^e \tag{16}$$

1.3.3 Produção de Pares

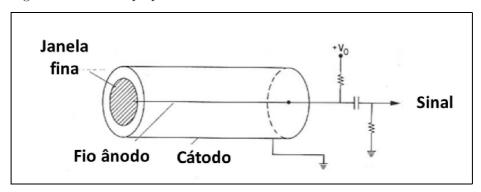
O processo de produção de pares envolve a interação de um fóton com um campo elétrico de um núcleo atômico, e sua transformação em um par elétron-pósitron. Para que ocorra, o fóton precisa ter uma energia mínima (Equação 17), onde $m_{núcleo}$ é a massa do núcleo. Desde que $m_{núcleo} \gg m_e$, a energia necessária tem que ser maior ou igual a 1,022 MeV.

$$E_{\gamma} \ge 2m_e c^2 + \frac{2m_e^2 c^2}{m_{n\'ucleo}} \tag{17}$$

1.4 Detectores de ionização

A Figura 5 mostra a estrutura básica de um detector a gás cilíndrico, a partícula incidente ioniza os átomos do gás, gerando íons e elétrons primários, que por sua vez, são acelerados pelo campo elétrico, obtendo energia suficiente para ionizar mais átomos do meio, tal processo desencadeia um efeito conhecido como avalanche de *Townsend* (Figura 6), que multiplica a carga elétrica a ser coletada pelo fio ânodo (Leo, 1987). Os modos de operação de uma câmara a fio são dados em função do campo elétrico aplicado ou da diferença de potencial aplicado (Figura 7).

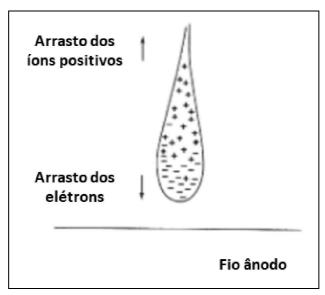
Figura 5 - Contador proporcional.



Legenda: Exemplo de um simples detector de ionização a gás.

Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

Figura 6 - Efeito avalanche.



Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

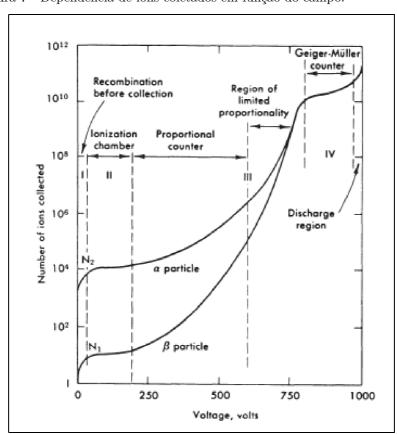


Figura 7 - Dependência de íons coletados em função do campo.

Legenda: Número de íons coletados em função da voltagem aplicada, em um único fio de um detector a gás. As regiões I,II,III e IV mostram a voltagem necessária para se operar com cada tipo de detector a gás. Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

1.4.1 Câmara proporcional multifilar

A câmara proporcional multifilar (MWPC, do inglês) possui o mesmo princípio básico de funcionamento, porém, é formado por conjunto de fios ânodos separados e espaçados igualmente entre um plano de dois cátodos. Devido à diferença de potencial ao qual o detector é submetido, geram-se linhas de campos elétricos equipotenciais em torno de cada fio (Figura 8). Desta forma, a avalanche de elétrons produzida se propaga na direção do fio que gera o campo ao qual esta sujeita. A posição da partícula em uma coordenada espacial pode ser medida através da presença de sinal no fio correspondente como indicado na Figura 9. Outra coordenada espacial pode ser obtida, por exemplo, com uma MWPC adicional sobreposta, com a direção dos fios perpendicular ao outro.

0.4 0.5 0.6 0.7 0.80 0.85 0.90 0.5

Figura 8 - Linhas do campo elétrico de uma MWPC.

Legenda: Linhas de campos equipotenciais dos fios de uma MWPC.

Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

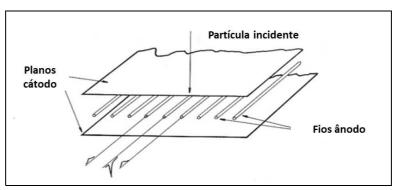


Figura 9 - Câmara multifilar proporcional.

Legenda: Ilustração de uma câmara multifilar proporcional.

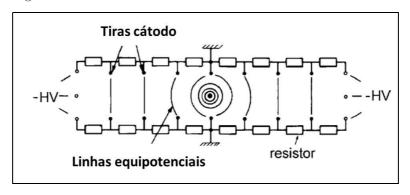
Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

1.4.2 Câmara de arrasto

As câmaras de arrasto são um aperfeiçoamento das MWPCs, a distância (x) entre o ponto de passagem da partícula e o ânodo, pode ser mensurada a partir da velocidade de arrasto do meio (u), do tempo de chegada da partícula (t_0) e o tempo de chegada dos elétrons no fio ânodo (t_1) , onde o tempo t_0 é fornecido por um detector auxiliar, normalmente um detector cintilador, que será descrito mais à frente. As câmaras de arrasto apresentam uma quantidade de fios ânodos menor se comparados a uma MWPC, utilizando uma distribuição de fios cátodos com gradiente de tensão para uniformizar o campo elétrico na região sensível de forma a permitir o uso da Equação 18, a Figura 10 ilustra um exemplo de um campo gerado na câmara de arrasto.

$$x = \int_{t_0}^{t_1} u \, dt \tag{18}$$

Figura 10 - Câmara de arrasto.



Legenda: Ilustração do campo formado em uma área do detector de arrasto.

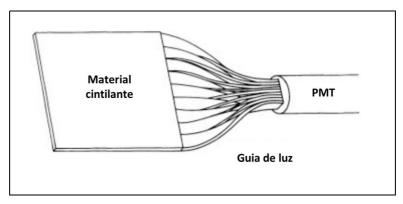
Fonte: Grupen et. al., 2009, p. 676. Adaptado pelo autor.

1.5 Detector cintilador

Os detectores cintiladores são compostos por um material cintilador acoplado a um fotossensor (Figura 11). Quando uma partícula atravessa esse material, os elétrons são excitados e quando retornam ao seu estado fundamental emitem fótons. O dispositivo fotossensível absorve a energia dos fótons, emitindo elétrons via efeito fotoelétrico. Estes então passam por um processo de multiplicação, gerando um pulso elétrico mensurável no estágio final. Os materiais cintiladores são divididos em duas categorias: orgânicos e inorgânicos. As próximas subseções irão detalhar o mecanismo de cintilação de cada tipo

de material.

Figura 11 - Esquema de um detector cintilador.



Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

1.5.1 Materiais orgânicos

Cintiladores orgânicos são baseados em cadeias aromáticas de benzeno, uma das características dessa estrutura é o rápido tempo de decaimento, da ordem de nanosegundos. A luz cintilante desses materiais surge através de transições de níveis de energia realizadas pelos elétrons livres na camada de valência das moléculas. Esses elétrons não são associados a nenhum átomo específico e ocupam as ligações π .

Na Figura 12 é mostrado um diagrama de níveis de energia contendo estados singletos, tripletos e o estado fundamental denotado por S_0 . Acima desse nível estão os estados excitados $(S^*, S^{**}, ...)$ e T_0 o menor estado dos tripletos e seus estados excitados $(T^*, T^{**}, ...)$. As excitações singleto geralmente decaem imediatamente (≤ 10 ps) para o estado excitado (S^*) sem a emissão de radiação. Esse processo é conhecido como conversão interna. No estado S^* existe uma alta probabilidade de produzir um decaimento ao estado fundamental S_0 , emitindo luz fluorescente (setas onduladas).

Para os estados tripletos, o mesmo processo de conversão interna ocorre com o decaimento para estado de tripleto fundamental (T_0) . As transições T_0 a S_0 são possíveis, mas são altamente proibidas pelas regras de seleção multipolo. Em vez disso, o decaimento de T_0 ocorre principalmente por interação com outra molécula.

1.5.2 Materiais inorgânicos

Os cintiladores inorgânicos são, em sua maioria, compostos de cristais de halogenetos alcalinos. Possuem tempo de decaimento maior se comparado aos materiais orgânicos,

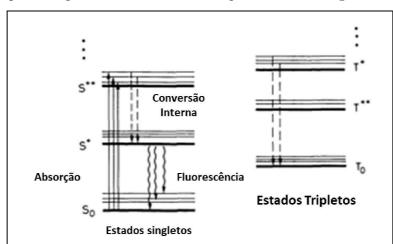


Figura 12 - Esquematização do mecanismo de cintilação de materiais orgânicos.

Legenda: Esquematização de possíveis transições em um material cintilador com a passagem de uma partícula carregada. O decaimento para estado fundamental (S_0) com a emissão de radiação (luz fluorescente) é representado por setas onduladas e a conversão interna por setas tracejadas. Os estados singletos são denotados como S e os estados tripletos por T.

Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

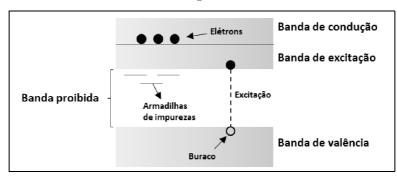
e uma de suas vantagens é o seu alto poder de frenamento. Tal atributo é devido a sua alta densidade, o que permite construir detectores compactos com capacidade de calorimetria. Além disso, possuem alta emissão de luz, resultando em uma melhor resolução de energia. Essas características são adequadas para detecção de radiação gama e elétrons de altas energias.

O mecanismo de cintilação nos materiais inorgânicos pode ser entendido a partir da estrutura de banda de cristais, como ilustrado pela Figura 13. Caso o material seja dopado, níveis de energia podem ser criados na região da banda proibida (armadilhas de impurezas); dependendo da dopagem, a impureza pode ser aceitadora ou doadora de elétrons. Quando a impureza é doadora, a migração de buracos pode ocorrer para as armadilhas de impurezas, fazendo com que elétrons preencham os espaços abertos pelos buracos, emitindo radiação (cintilação) nesse processo. Na impureza aceitadora, os níveis criados funcionam como armadilha de elétrons, capturando e aprisionando os portadores de cargas. Assim que ocorre a interação, a energia é cedida aos elétrons aprisionados, liberando energia na forma de radiação.

1.6 Fotomultiplicadora

Fotomultiplicadoras (PMT, *Photomultiplier Tube*) são dispositivos largamente utilizados na detecção de radiação. Basicamente uma fotomultiplicadora é composta de

Figura 13 - Banda de estrutura de cristais inorgânicos

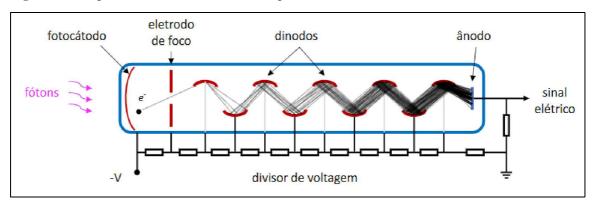


Legenda: Nessa estrutura o elétron é excitado, deixando o buraco livre na banda de valência, formando um par elétron-buraco excitado. Este par pode migrar através do cristal e ser capturado pelas armadilhas de impurezas. Sendo a impureza doadora, a migração de buracos pode ocorrer para as armadilhas de impurezas, vagando espaços para elétrons ocuparem e emitindo radiação nesse processo (cintilação).

Fonte: O autor, 2023.

um invólucro de vidro contendo uma série de componentes para a captação de fótons, conversão em corrente elétrica e intensificação do sinal. Para auxiliar na transmissão de luz, o material geralmente é feito de vidro ou quartzo, dependendo da largura espectral necessária.

Figura 14 - Esquema Básico de uma fotomultiplicadora.



Fonte: O autor, 2023.

A Figura 14 mostra uma ilustração básica de uma fotomultiplicadora, onde um fóton chega ao fotocátodo ejetando um elétron (fotoelétron) via efeito fotoelétrico. O fotoelétron é direcionado ao estágio de multiplicação através de um eletrodo de foco. Os estágios de multiplicação são compostos pelos dínodos que estão sob um gradiente de potencial elétrico que ao choque, outros elétrons são arrancados multiplicando o seu número a cada dínodo. O processo se repete até a chegada no ânodo, produzindo um sinal elétrico amplificado.

1.6.1 Fotocátodo

A eficiência do conversor fotoelétrico varia fortemente com a frequência incidente da luz e a estrutura do material. De modo geral o espectro de resposta é expresso pela eficiência quântica (Equação 19), onde λ é o comprimento de onda incidente. A Figura 15 mostra a eficiência quântica do fotocátodo quando feito de diferentes materiais. Observa-se que cada material possui um espectro de eficiência melhor para determinados comprimentos de onda incidente, Tabela 1.

$$\eta(\lambda) = \frac{\text{número de fotoelétrons liberados}}{\text{número de fótons incidente no fotocátodo}(\lambda)}$$
(19)

Quantum Efficiency % Q.E. %

28

24

20

10.0

5.0

Modified / S-20

Extended / S-20

S-20

Extended / S-20

S-10

8

Bialkali (K-Cs)

S-11

S-10

S-1

Figura 15 - Eficiência quântica de diversos materiais do fotocátodo.

Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

1.6.2 Seção eletro-multiplicadora e eletrodos focalizadores

Os eletrodos focalizadores são responsáveis pelo direcionamento do fotoelétron ao primeiro dínodo, para aumentar a eficiência de produção de sinal elétrico. A seção eletromultiplicadora é composta por um arranjo de várias superfícies denominadas dínodos, que

Tabela 1 - Materiais do fotocátodo.

Tipo do Cátodo	Composição	λ no pico de resposta (nm)	Eficiência quântica por pico
S1(C)	Ag-O-Cs	800	0,36
S4	SbCs	400	16
S11(A)	SbCs	440	17
Super (A)	SbCs	440	22
S13 (U)	SbCs	440	17
S20 (T)	SbNa- KCs	420	20
S20R	SbNa- KCs	550	8
TU	SbNa- KCs	420	20
Bialkali	SbRb-Cs	420	26
Bialkali D	Sb-K-Cs	400	26
Bialkali DU	Sb-K-Cs	400	26
SB	Cs-Te	235	10

Fonte: Leo, 1987, p. 377. Adaptado pelo autor.

são postas sob crescente diferença de potencial.

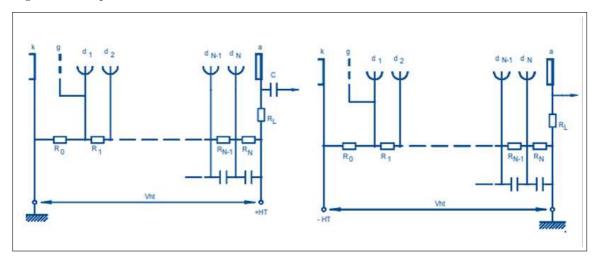
O fotoelétron é direcionado ao primeiro dínodo, onde colide transferindo energia cinética, emitindo elétrons secundários. Eles são direcionados ao próximo dínodo gerando mais elétrons, o mesmo efeito é causado nos próximos dínodos, produzindo um sinal elétrico amplificado ao serem coletados pelo ânodo. Tipicamente, esse processo gera em torno de 10^4 a 10^6 elétrons para cada fotoelétron emitido pelo fotocátodo. Esta multiplicidade é o chamado de ganho G.

1.6.3 Divisor de tensão

Para gerar a distribuição de potencial elétrico entre fotocátodo, eletrodos focalizadores, dínodos e ânodo, é necessário um circuito divisor de tensão. Existem dois tipos básicos de divisor de tensão: de polaridade positiva com o cátodo aterrado e o ânodo em alto potencial positivo (Figura 16, à esquerda); de polaridade negativa, com o ânodo aterrado e o cátodo em alto potencial negativo (Figura 16, à direita). A escolha depende do tipo de aplicação.

Para contagem de pulsos, a polaridade positiva é preferencialmente utilizada. O capacitor C isola o circuito de medição da alta tensão. Ela é conveniente porque não é necessário isolar a região em torno do fotocátodo. Para aplicações que envolvem detecção de fluxo contínuo ou muitos pulsos breves, onde o uso de um capacitor de acoplamento não seria prático, a polaridade negativa deve ser usada. Isso requer precauções especiais para minimizar seu efeito na corrente escura e para proteger o tubo contra o efeito destrutivo

Figura 16 - Esquema de um circuito de divisor de tensão.



Legenda: Esquema de um circuito divisor de tensão, onde a esquerda temos uma fonte de tensão de polaridade positiva, e a direita polaridade negativa.

Fonte: Photonis, 2007, p. 313.

de gradientes de tensão no vidro.

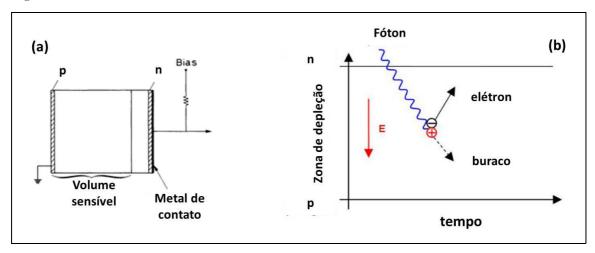
1.7 Detectores semicondutores

A Figura 17 (a) mostra uma configuração básica de um detector semicondutor, composto por uma junção pn. A composição gera um campo elétrico na interface, chamada de zona de depleção, tal volume é sensível à radiação. A partir de uma polarização reversa aplicada nas extremidades do semicondutor, os elétrons são atraídos pela polaridade positiva na região n e o mesmo ocorre para os buracos da região p na polaridade negativa. A polarização aumenta a espessura da zona de depleção, fazendo com que a coleta de carga e a sensibilidade à radiação seja mais eficiente. A Figura 17 (b) exemplifica um fóton sendo absorvido pela zona de depleção, gerando um par elétron-buraco.

1.7.1 Detector APD

A Figura 18 mostra a estrutura básica de um detector fotodiodo de avalanche (APD), ela é composta por junções que geram uma região de absorção (A), onde o campo elétrico é menos intenso; e uma região de multiplicação (M), onde o campo elétrico é mais intenso. As partículas interagem com a região de absorção, gerando os pares elétronsburacos. Os elétrons produzidos são atraídos para a região de multiplicação, desencadeando um efeito de avalanche, e ao final desse processo, a carga é coletada pelo circuíto

Figura 17 - Detector semicondutor.



Legenda: (a) Detector semicondutor. (b) Processo de criação de um par elétron-buraco, pela absorção de um fóton em um material semicondutor.

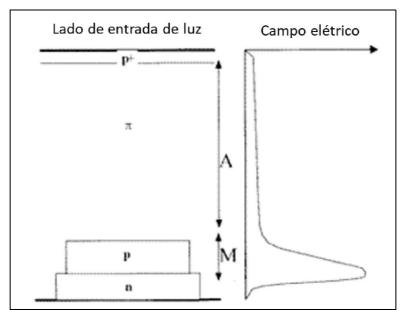
Fonte: (a) Leo, 1987, p. 377 (b) Kim, 2010, p. 364. Adaptado pelo autor.

elétrico.

1.7.2 Detector SiPMs

Detectores SiPMs (fotomultiplicadoras de silício) são arranjos de micro-células APDs em que cada célula é associada a um APD diferente. Na Figura 19 (a) é mostrada a esquematização desse detector, exemplificando a interação de 3 fótons, onde cada fóton interage com uma micro-célula diferente, iniciando um processo de avalanche na zona de depleção, o circuito associado a cada célula se encontra na Figura 19 (b). O detector opera em modo não proporcional, ou seja, independe do número de pares íonsburacos gerados pela partícula interagente e cada célula contribui para a carga final do arranjo.

Figura 18 - Detector APD



Legenda: Estrutura básica de um detector APD. A região A é onde ocorre a absorção dos fótons, produzindo pares elétrons-buracos. Na região M o campo elétrico é suficiente para gerar um efeito avalanche que multiplica a carga original.

Fonte: Optoelectronics, 2003, p. 8. Adaptado pelo autor.

(a)

Fóton

O.Elementos

P D. Saída

Figura 19 - Detector SiPM

Legenda: (a) Esquematização de um detector SiPM. (b) Circuito elétrico associado. Fonte: Kim, 2010, p. 364. Adaptado pelo autor.

2 APRENDIZADO DE MÁQUINA

Aprendizado de máquina (AM) é um método de análise de dados que permite a automatização da construção de modelos analíticos. A partir de dados amostrais é possível construir previsores que identificam certos padrões e tomam decisões com o mínimo de interversão humana.

Existem três tipos de aprendizado de máquina: o aprendizado supervisionado, onde os algoritmos são treinados com supervisão humana; o aprendizado não supervisionado, no qual o algoritmo aprende sem interversão humana; o aprendizado por reforço, cujo método observa o ambiente e toma a estratégia com o maior número de recompensas (Géron, 2019).

De forma geral, um problema em aprendizado de máquina pode ser representado pela Equação 20, onde f(X) é a função que define o comportamento do observável (função alvo) Y e ϵ é um erro aleatório associado a Y. Basicamente, assume-se uma relação entre uma amostra (Equação 21) e um observável Y.

$$Y = f(X) + \epsilon \tag{20}$$

Para explicar os algoritmos supervisionados, será adotada a notação convencional da referência (James et. al., 2017). A partir de uma matriz X com elementos x_{ij} (Equação 21), onde as linhas representam os dados e colunas as características (features), a construção de um estimador $\hat{f}(X)$ que tenha como resultado da previsão \hat{Y} pode ser realizada.

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$
(21)

$$\widehat{Y} = \widehat{f}(X) \tag{22}$$

A construção de modelos de aprendizado de máquina é feita a partir da minimização do risco empírico (Equação 23) de uma amostra de treino (S) dada por pares $[(x_1, y_1), ..., (x_n, y_n)]$. Assume-se que, para um determinado conjunto de hipóteses (F), onde cada observação é vinculada a um rótulo y, ocorre uma perda de desempenho (L).

$$ERM(S) = \arg\min_{f \in F} \left[\frac{1}{n} \sum_{i=1}^{n} L(y_i - \hat{f}(x_i)) \right]$$
(23)

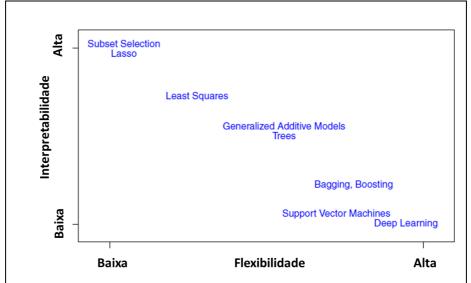
2.1 Seleção do algoritmo de aprendizado de máquina

Devido a extensa gama de algoritmos de aprendizado de máquina, torna-se um desafio escolher o método que melhor se ajusta aos dados. Nas próximas subseções serão discutidas as relações que impactam na predição dos modelos.

2.1.1 Relação entre interpretabilidade e flexibilidade

Algoritmos flexíveis (mais graus de liberdade) são aqueles cujos métodos apresentam alta capacidade de generalizar os dados de uma forma muito ampla. No entanto, apresentam baixa interpretabilidade de predição, os métodos mais interpretativos são menos flexíveis. Escolher um algoritmo de aprendizado que melhor se ajuste aos dados, pode exigir a perda da interpretabilidade dos resultados em troca de um melhor desempenho, a relação em questão de alguns algoritmos é mostrado na Figura 20.

Figura 20 - Relação de interpretabilidade X flexibilidade



Legenda: Relação de interpretabilidade X flexibilidade de alguns algoritmos de aprendizado de máquina.

Fonte: James et. al., 2017, p. 419. Adaptado pelo autor.

2.1.2 Relação entre viés e variância

Esta relação pode ser extraída a partir de uma métrica utilizada em aprendizado de máquina, o erro médio quadrático do estimador (MSE, do inglês, Equação 24), tal métrica quantifica o desempenho de um estimador, onde y_i são os observáveis (alvos) e $\hat{f}(x_i)$ é a previsão do estimador para os valores de entrada (x_i) .

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2$$
 (24)

Através do valor esperado (E) de um teste (x_*, y_*) MSE é possível decompor a relação entre viés e variância, como mostrada na Equação 25 (James et. al., 2017), onde $\text{Vi\'es}(\widehat{f}(x_*) = E[\widehat{f}(x_*)] - f(x_*))$ e. A equação em questão, exibe a soma de 3 quantidades fundamentais: a variância de $\widehat{f}(x_*)$, viés quadrático de $\widehat{f}(x_*)$ e a variância do erro irredutível.

A variância de um estimador nos diz o quanto um modelo se altera ao mudar os dados de treinamento. Considerar um modelo mais flexível pode aumentar a variância do previsor, causando sobreajuste aos dados de treinamento (Figura 21). O aumento do viés ocorre por erro de escolha, como utilizar um previsor menos flexível que não consegue captar as nuances dos dados, nesse caso o teste se subajusta ao treinamento (Figura 21). O erro irredutível refere-se aos pontos fora da curva, exceções.

A otimização do modelo é alcançada quando o método de aprendizagem atinge baixa variância e baixo viés. Na prática, modelos de alta flexibilidade estão mais sujeitos a sobreajuste, com a redução da variância o desempenho melhora. Os modelos mais interpretativos (baixa flexibilidade) estão suscetíveis a um alto viés e baixa variância, ao reduzir o viés, o modelo é otimizado. A Figura 21 mostra a relação de erro de predição e flexibilidade do modelo.

$$E[(y_* - \hat{f}(x_*))^2] = E[(y_* - E[f(x_*)] + E[f(x_*)] - \hat{f}(x_*))^2] = Var(\hat{f}(x_*)) + Vi\acute{e}s(\hat{f}(x_*))^2 + Var(\epsilon)$$
(25)

2.2 Aprendizado supervisionado

Na construção dos modelos supervisionados, os dados são inseridos no algoritmo, onde cada observável é rotulado com a sua respectiva classe para o processamento do treinamento. Ao final desta etapa, o algoritmo torna-se um modelo (classificador ou

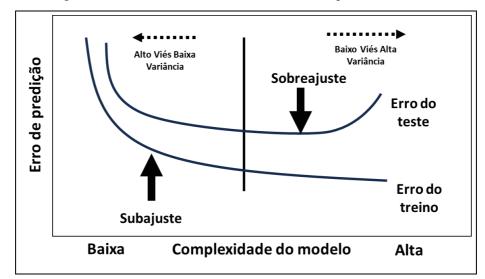


Figura 21 - Relação entre viés e variância de acordo com a complexidade do modelo

Legenda: Modelos de alta complexidade possuem um viés baixo e variância alta, por isso, estão sujeitos a sobreajuste, enquanto os modelos de baixa complexidade são suscetíveis a subajuste, por apresentaram um alto viés e baixa variância.

Fonte: O autor, 2023.

previsor) e está pronto para classificar as amostras de teste (Figura 22).

Nesta dissertação foram utilizados os seguintes algoritmos de aprendizado de máquina supervisionado:

- Floresta Aleatória (do inglês, Random Forest),
- Rede neural profunda (DNN),
- Rede neural convolucional (CNN).

2.2.1 Floresta Aleatória (Random Forest)

A Floresta Aleatória (FA)¹ (RF) é um método de aprendizado de ensemble, composto por um conjunto de árvores de decisões (Seção A.1). O algoritmo treinada por meio de bootstrap (Seção B.1) e aleatoriamente algumas árvores são "podadas" (profundidade reduzida). O algoritmo agrega a previsão de um conjunto de modelos (Hard Voting), elegendo os dados por meio de voto majoritário (moda estatística). Árvores de decisões (Seção A.1) e Bagging e Pasting (Seção A.2) são os dois algoritmos utilizados na construção da Floresta Aleatória. A Seção B.1 detalha o método de reamostragem bootstrap.

¹ Em algumas figuras a Floresta Aleatória é mencionada no termo em inglês Random Forest

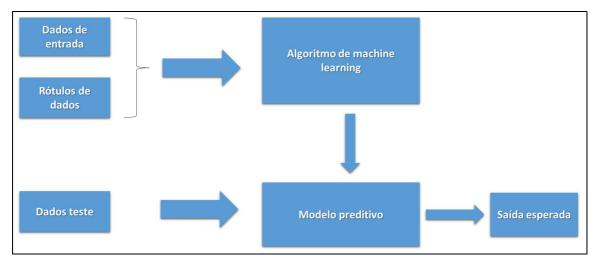


Figura 22 - Esquematização em blocos de aprendizado supervisionado

Legenda: No aprendizado supervisionado os dados de entradas são rotulados para o processo de treinamento. Após esta etapa, o modelo preditivo foi construído, ao inserir os dados de teste, uma saída esperada (previsão) é produzida.

Fonte: O autor, 2023.

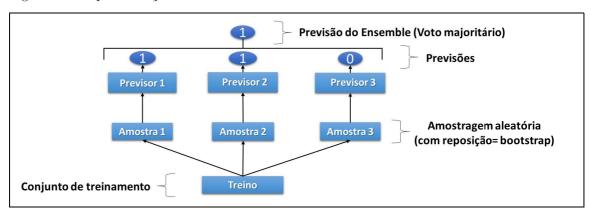


Figura 23 - Esquematização da FA

Legenda: Representação em blocos do algoritmo FA, onde três previsores (árvores de decisões) são treinados por sub-amostras de *bootstrap* que classificam um observável por meio de voto majoritário (moda estatística).

Fonte: O autor, 2023.

2.3 Aprendizado profundo

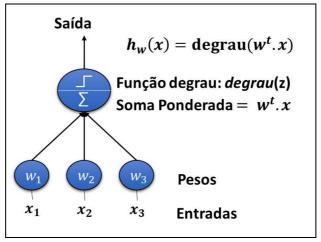
Conhecido também como deep learning em inglês, é um aprendizado de máquina capaz de identificar padrões em dados de alta dimensão (número de características). A estrutura desses métodos de aprendizagem se assemelham a uma rede de neural, elas pode ser composta por inúmeras camadas de processamento que aprendem padrões de imagens e de texto, por exemplo. Nas subsecções a seguir, é explicado o funcionamento de um perceptron (neurônio) e de uma rede neural de uma camada.

2.3.1 Perceptron

O perceptron ou unidade linear de limiar (LTU, do inglês) é baseado em um neurônio onde suas entradas e saídas são números. Ou seja, vai além de valores binários e cada conexão tem um peso associado. A LTU calcula a soma ponderada de suas entradas (x_i) , cada qual com o peso w_i e que submetidos a uma função ativação que produz uma saída z (Equação 26), como mostrado na Figura 24.

$$z = w_1 x_1 + w_2 x_2 + \dots + w_n x_n = w^t x \tag{26}$$

Figura 24 - Unidade linear de limiar



Fonte: O autor, 2023.

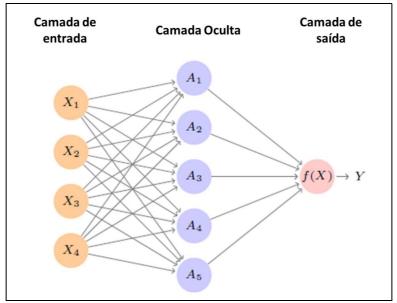
2.3.2 Rede neural de uma camada interna

Uma rede neural (NN, do inglês) simples é composta por 3 camadas: uma camada de entrada (com p característica, Equação 27), uma camada oculta e uma camada de saída, como ilustrado na Figura 25. Cada conexão possui um peso associado e cada neurônio realiza a soma ponderada de suas respectivas conexões, produzindo uma saída a partir de uma função ativação, a resposta da rede é dada pela camada de saída (Equação 28), onde g é a função ativação, b o viés da camada e \widehat{W}_i a matriz peso da camada.

$$X = (X_1, X_2, ..., X_p) (27)$$

$$\hat{f} = \widehat{W}_2 g(\widehat{W}_1 x + \widehat{b}_1) + \widehat{b}_2 = \widehat{W}_2 \widehat{A}^{(1)}(x) + \widehat{b}_2$$
(28)

Figura 25 - Rede neural de uma camada oculta



Fonte: James et. al., 2017, p. 419. Adaptado pelo autor.

2.3.3 Rede neural de múltiplas camadas

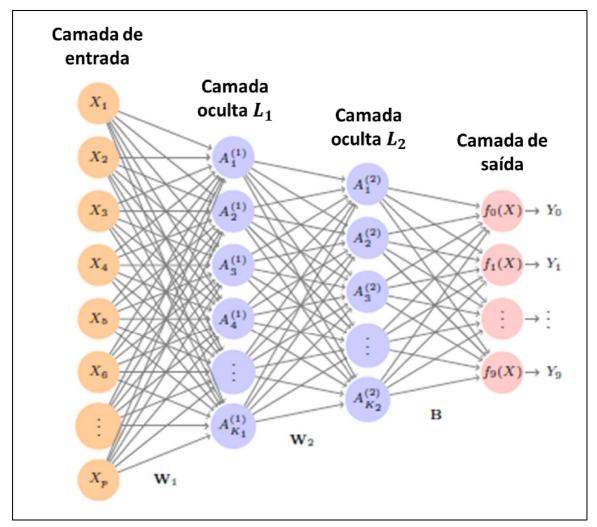
Com apenas uma camada interna e um grande número de neurônios, a DNN (do inglês, deep neural network) possui a capacidade de ser ajustar a maioria dos dados. No entanto, foi observado que as tarefas de aprendizagem apresentam boas soluções quando

se aumenta o número de camadas ocultas. A Figura 26 exibe as seguintes saídas para cada camada:

- Primeira camada: $\widehat{A}^{(1)}(x) = g_1(\widehat{W}_1x + \widehat{b}_1),$
- Segunda camada: $\widehat{A}^{(2)}(x) = g_2(\widehat{W}_2\widehat{A}^{(1)}(x) + \widehat{b}_2),$
- Camada final: $\widehat{A}^{(f)}(x) = g_{final}(\widehat{W}_f \widehat{A}^{(2)}(x) + \widehat{b}_f),$

A rede em questão é composta pela seguinte arquitetura: p=784, camada L_1 ($K_1=256$ neurônios) e camada L_2 ($K_2=128$ neurônios), totalizando 235.146 parâmetros.

Figura 26 - Rede neural de 2 camadas oculta



Legenda: Rede neural de 2 camadas ocultas contendo a seguinte composição: a camada de entrada com p neurônios, uma camada oculta L_1 com K_1 neurônios, uma segunda camada oculta L_2 com K_2 neurônios e uma camada de saída com multiclassifição composta 9 neurônios.

Fonte: James et. al., 2017, p. 419. Adaptado pelo autor.

2.4 Rede neural convolucional

As construções das chamadas redes neurais convolucionais (CNN, do inglês) se baseiam nos estudos do córtex visual do cérebro. A alta quantidade de dados a disposição atualmente e técnicas de aprendizagem de máquina de redes profunda, resultaram em CNNs com um alto desempenho em tarefas complexas.

2.4.1 Camada convolucional

Diferentemente de uma DNN, onde neurônios estão conectados unidimensionalmente, as CNNs possuem conexões através de um campo receptivo. A Figura 27 ilustra a conexão entre as camadas convolucionais, onde os neurônios são conectados por campos receptivos retangulares.

A arquitetura da CNN permite que a rede se concentre em características de baixo nível na primeira camada, reunindo-as em características de nível superior na próxima camada convolucional. Tal construção permite que a CNN seja eficiente em reconhecimento de imagens.

A Figura 28 ilustra as conexões através do campo receptivo, onde os neurônios se distribuem bidimensionalmente (linha i e coluna j) e são conectados por um campo receptivo de altura f_h e largura f_w na camada superior. Normalmente adicionam-se zeros ao redor das camadas de entrada para que se tenha a mesma largura e altura da camada anterior, a técnica é chamada de zero padding. Além disso, é possível conectar uma grande camada de entrada a uma muito menor, dando espaçamentos aos campos receptivos (Figura 29). A distância entre dois campos receptivos é ajustável por parâmetros, cujas distâncias são chamadas de strides.

2.4.2 Filtros e empilhamento de múltiplos mapas de características

Na CNN os pesos são distribuídos nos chamados filtros ou kernels de convoluções em que cada neurônio de um mesmo conjunto compartilha de um filtro semelhante. A Figura 30 mostra dois conjuntos de possíveis filtros. Enquanto o primeiro filtro está focado nas linhas brancas verticais, o segundo filtro irá distribuir os pesos nas linhas brancas horizontais. Ou seja, o padrão em que o filtro estiver focado será destacado quando for detectado. Assim, uma camada cheia de neurônios, com a utilização do mesmo filtro fornece um mapa de características que destacam as áreas mais semelhantes ao filtro.

Resumidamente, o treinamento de uma CNN é capaz de encontrar filtros mais úteis e aprender a combiná-los em padrões mais complexos.

Camada convolucional 2

Camada convolucional 1

Camada de entrada

Figura 27 - Camadas convolucionais com campos receptivos retangulares

Legenda: Ilustração das conexões através de campos receptivos.

Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

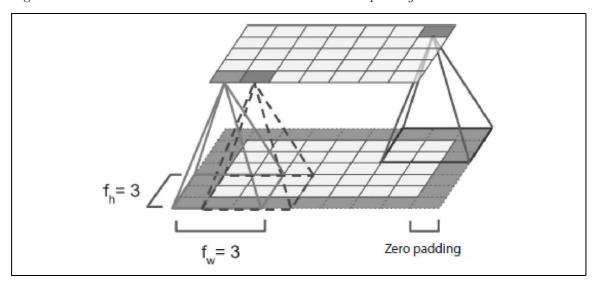
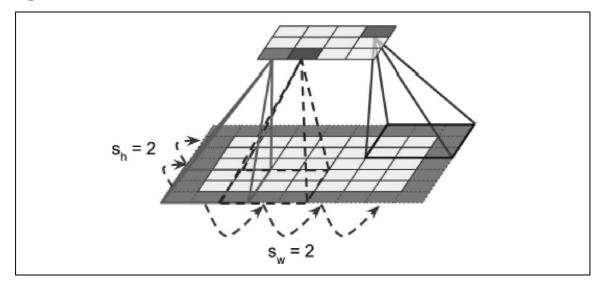


Figura 28 - Conexões entre camadas convolucionais com zero padding

Legenda: Conexões entre camada convolucionais com $zero\ padding$, composta por um campo receptivo de 3 unidades de largura e 3 unidades de altura.

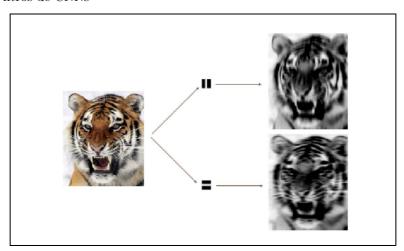
Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

Figura 29 - Strides de camadas convolucionais



Legenda: Diagrama de uma camada de entrada 5x7 (preenchida com zeros) que é conectada a uma camada 3x4 através de campos receptivos 3x3 com *strides* de 2 unidades. Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

Figura 30 - Filtros de CNNs



Legenda: Filtros de convolução utilizados na imagem de um tigre, onde os filtros de linhas verticais destacam as listras verticais, desfocando as horizontais. O inverso ocorre para os filtros horizontais.

Fonte: James et. al., 2017, p. 419. Adaptado pelo autor.

CNNs são representadas com mais precisão em 3D, onde cada camada é empilhada por vários mapas de características (Figura 31). A saída de um neurônio é dado pela Equação 29, onde $z_{i,j,k}$ é a saída do neurônio localizado na linha i, coluna j e no mapa de característica k; s_h e s_w são strides horizontais e verticais, respectivamente; f_{nt} é o número do mapa de característica da camada anterior; b_k é termo de polarização para o mapa de característica k.

Mapa de Características 1

Mapa 2

Filtros

Mapa 2

Camada convolucional 1

Camada convolucional 1

Camada convolucional 1

Canais

Vermelho Verde Azul

Figura 31 - Empilhamento de mapas de característica em camadas convolucionais

Legenda: Representação de empilhamentos de camadas convolucional, com seus mapas de características.

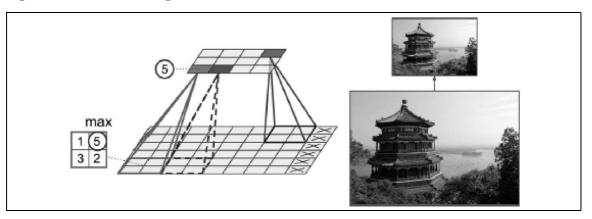
Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

$$z_{i,j,k} = b_k + \sum_{u=0}^{f_h - 1} \sum_{v=0}^{f_w - 1} \sum_{k' = 0}^{f_{n'-1}} x_{i',j',k',\cdot} w_{u,\nu,k',k} \begin{cases} i' = i \times s_h + u \\ j' = j \times s_w + u \end{cases}$$
(29)

2.4.3 Camada de pooling

A camada de *pooling* tem como objetivo reduzir a carga computacional, preservando recursos de memória, essa técnica permite também reduzir o risco de sobreajuste, os parâmetros da camada são semelhantes a de uma camada convolucional. No entanto, os neurônios da camada de *pooling* não possuem peso. A partir de uma função de agregação as entradas são unidas. A Figura 32 demonstra o mecanismo de *pooling*, onde apenas o valor máximo de entrada em cada *kernel* chega à próxima camada, enquanto as outras entradas são descartadas.

Figura 32 - Camada Pooling



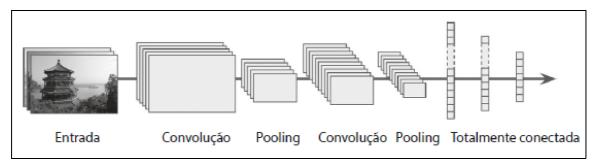
Legenda: Camada max pooling (kernel pooling 2x2, stride de 2, sem padding)

Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

2.4.4 Arquiteturas CNN

As arquiteturas das CNNs empilham as camadas convolucionais com mapas de característica de profundidade, seguidas por camada *pooling* e assim por diante até que a informação é achatada na saída. A técnica permite que a entrada esteja totalmente conectada com a saída da rede. A arquitetura em questão, permite que a rede seja capaz de reconhecer padrões independente do local onde ocorra, diferentemente de uma DNN que reconhece o padrão apenas em um local específico. Além disso, também é possível conectar uma DNN com a camada de achatamento (*flatten*, do inglês) que comprime a informação e possibilita tal conexão.

Figura 33 - Arquitetura de uma CNN



Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

2.5 Retropropagação

As redes neurais inicializam os pesos (θ) aleatoriamente, e a cada iteração (θ^i) os pesos dos neurônios são atualizados de forma que a perda seja minimizada. O método que otimiza a perda de uma rede neural é chamado de retropropagação. A técnica reajusta os pesos a cada iteração por meio do método de gradiente descendente (Equação 30), a Equação 31 mostra como o peso é atualizado, onde ν é a taxa de aprendizagem da rede.

A retropropagação funciona da seguinte forma: a instância de treinamento flui pela rede e calcula a saída de cada camada consecutiva; logo em seguida, mede-se o erro de saída da rede $(y_j - \hat{y}_j)$ e contabiliza-se quanto cada neurônio da última camada oculta contribuiu para o erro da saída. Assim, o método passa a medir a contribuição do erro das camadas anteriores até chegar a camada de entrada. Esta passagem reversa mede eficientemente o gradiente de erro em todos os pesos de conexão na rede, propagando para trás o gradiente de erro.

A Figura 34 mostra como funciona o gradiente descendente em 1D para o parâmetro θ , onde a cada interação θ^i há uma aproximação da perda mínima. Na última interação (θ^7) , atinge-se a perda mínima $(R(\theta^7))$, otimizando o desempenho do modelo.

$$\nabla R(\theta^i) = \left. \frac{\partial R(\theta)}{\partial \theta} \right|_{\theta = \theta^i} \tag{30}$$

$$w_{i,j}^{(pr\'oximodegrau)} = w_{i,j} + \nu(y_j - \hat{y}_j)x_i$$
(31)

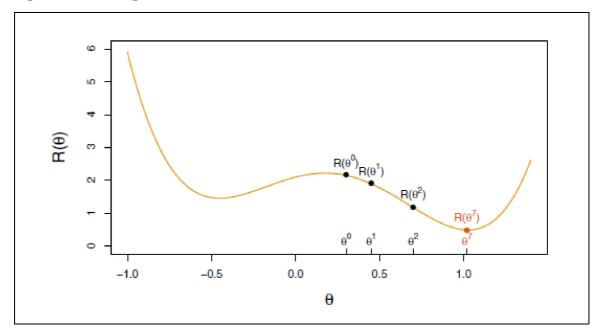


Figura 34 - Método gradiente descendente

Legenda: Processo de aprendizado de uma rede através do método de gradiente descendente para θ , onde θ^i é a interação.

Fonte: James et. al., 2017, p. 419. Adaptado pelo autor.

2.6 Medições de Desempenho

Uma forma de avaliar o desempenho dos modelos de aprendizado de máquina é através da matriz confusão (Figura 35). A métrica fornece a quantidade de erros e acertos de cada classe, possuindo as seguintes previsões: verdadeiro positivo (TP, do inglês), a classe positiva que de fato é classificada corretamente; o falso positivo (FP), a classe negativa que está sendo predita como positiva; o falso negativo (FN), a classe positiva que é classificada como negativa; por último, o verdadeiro negativo (TN, do inglês) que é a classe negativa classificada corretamente.

Utilizando a matriz confusão são construídas métricas mais concisas como: precisão (Equação 32), revogação (Equação 33) que também é conhecida como taxa de verdadeiros positivos (TPR), acurácia (Equação 34) e taxa de verdadeiros negativos (TNR), do inglês, Equação 35).

$$\operatorname{precisão} = \frac{TP}{TP + FP} \tag{32}$$

$$TPR = \frac{TP}{TP + FN} \tag{33}$$

$$acurácia = \frac{TP + TN}{TP + FN + TN + FP}$$
(34)

$$TNR = \frac{TN}{FP + TN} \tag{35}$$

Figura 35 - Representação da Matriz de Confusão



Legenda: Matriz de confusão, onde TP e TN são as classes classificadas corretamente, FP e FN são as predições incorretas.

Fonte: O autor, 2023.

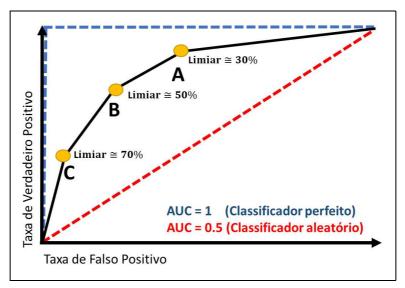
2.6.1 Curva ROC

A curva das características operacionais do receptor (ROC, do inglês) e a área abaixo da curva (AUC, do inglês) são métricas utilizadas para medir e comparar desempenho dos modelos de classificação binária. É um gráfico simples que permite estudar a variação da taxa de verdadeiro positivo (Equação 33) e taxa de falso positivo (Equação 36), para diferentes pontos de corte de probabilidade estimada (limiares).

A Figura 36 mostra uma curva ROC, onde três valores de limiar são destacados, se o valor é igual a 0%, todos os valores são preditos como positivo, quando é 100%, todos os observáveis são classificados como negativo. O classificador ideal apresenta a AUC igual ou o mais próximo de 1, caso a área seja menor ou igual a 0,5, o classificador é aleatório.

$$FPR = \frac{FP}{FP + TN} \tag{36}$$

Figura 36 - Curva da ROC



Legenda: Representação de uma curva ROC, mostrando o valor de limiar em três pontos, onde a linha tracejada vermelha representaria um classificador aleatório (AUC=0,5) e a linha azul seria um classificador perfeito (AUC=1).

Fonte: O autor, 2023.

2.7 Aprendizado não supervisionado

No aprendizado não supervisionado o algoritmo aprende sobre os dados sem interversão humana, ou seja, não exige rotulação dos observáveis. Dentre os métodos de aprendizagem mais utilizados estão os de agrupamento, a chamada clusterização, onde são agrupados os dados que apresentam características semelhantes, os *clusters*. A Figura 37 exemplifica a técnica de clusterização, onde uma amostra é segmentada em quatro *clusters*.

2.7.1 Medição da distância do agrupamento

A clusterização é realizada por meio de métricas que computam a distância entre dois observáveis, a métrica em questão é chamada de dissimilaridade ou matriz distância. Existem diversos tipos de métricas de distância, porém, as mais usuais são a distância euclidiana (Equação 37) e a distância de Manhattan (Equação 38). A partir dessas métricas a dissimilaridade entre dois observáveis (x_i e y_i) pode ser mensurada, outras distâncias são apresentadas na referência (Kassambara, 2017).

$$d_{euc}(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$
(37)

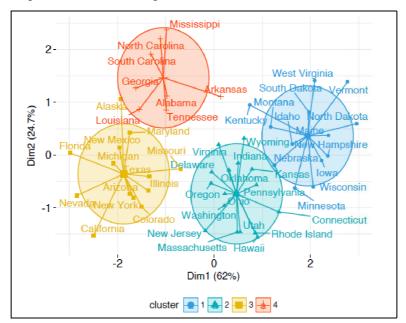


Figura 37 - Exemplo de uma clusterização

Legenda: Clusterização de uma amostra, onde foi utilizada um redução de dimensionalidade chamada de PCA, o eixo X representa a componente c_1 do PCA e o Y a componente c_2 . Os valores de porcentagem significam a proporção de variabilidade armazenada em cada componente.

Fonte: Kassambara, 2017, p. 187.

$$d_{man}(x,y) = \sum_{i=1}^{n} |(x_i - y_i)|$$
(38)

2.8 Clusterização hierárquica

A clusterização hierárquica (HCA, do inglês) é uma forma alternativa para segmentar uma amostra em *clusters*, ela pode ser realizada por clusterização aglomerativa ou clusterização divisiva (Kassambara, 2017). Na clusterização aglomerativa (AGNES, do inglês) cada objeto é considerado um único *cluster* e a cada passo, de acordo com a similaridade são juntados em um *cluster* maior, ao final, são aglomerados em um único *cluster* (Figura 38). A clusterização divisiva (DIANA, do inglês, Figura 38) realiza um processo inverso de clusterização, onde um único *cluster* é dividido em *clusters* menores e a cada passo se segmenta em um conjunto de observações menores.

Basicamente, a clusterização hierárquica exige os seguintes passos:

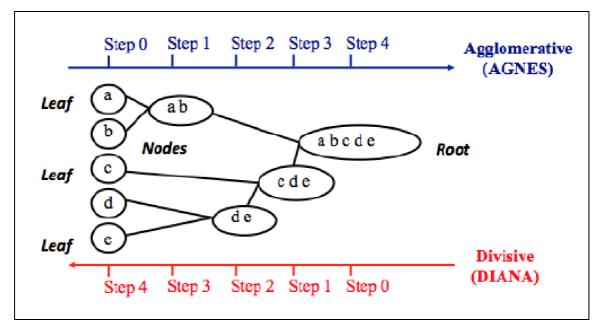
- Preparo dos dados.
- Associar os pares da distância de agrupamento (Equação 37) a uma matriz de dis-

similaridade.

- Uma função ligação para agrupar os objetos na árvore de clusterização hierárquica.
- Definir um corte na árvore hierárquica para a formação dos *clusters*.

A clusterização aglomerativa permite analisar uma amostra através de um dendrograma², onde um limiar definido interrompe as próximas junções de *clusters*. A Figura 39 mostra um dendrograma com um limiar, separando a amostra em três *clusters*.

Figura 38 - Passos de uma clusterização Hierárquica



Legenda: Construção da clusterização aglomerativa (AGNES) e análise divisiva (DIANA), a aglomerativa cria pequenos aglomerados (folhas) que são juntados em *clusters* maiores, formando um único *cluster* (raiz) ao final. A DIANA apresenta uma construção inversa.

Fonte: Kassambara, 2017, p. 187.

2.9 Redução de dimensionalidade

Diversos problemas de aprendizado de máquina envolvem inúmeras características, isso acaba tornando o treinamento mais lento e muito mais difícil de encontrar uma solução ideal, esse problema pode ser resolvido por redução de dimensionalidade. Existem diversos algoritmos que reduzem a dimensionalidade da amostra, o mais popular dentre eles é o PCA (*Principal Components Analysis*).

² Dendrograma é uma representação esquemática ou diagrama que lembra a estrutura de uma árvore.

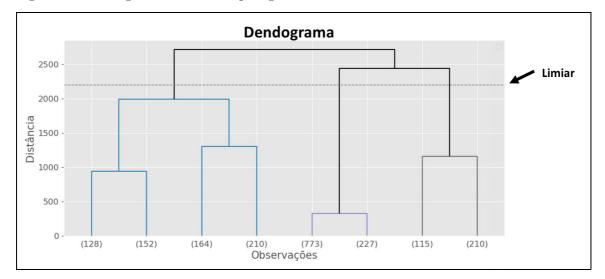


Figura 39 - dendrograma de clusterização aglomerativa

Legenda: Dendrograma de uma clusterização aglomerativa, onde foi definido um limiar de 2200, resultando em 3 clusters para uma possível análise.

Fonte: O autor, 2023.

2.9.1 Análise de Componentes Principais (PCA)

O PCA maximiza a variância empírica dos dados e a distribui nas chamadas componentes, através de uma técnica padrão de fatoração de matriz chamada de decomposição em valores singulares (SVD, do inglês) (James et. al., 2017). A partir dessa técnica é possível decompor uma matriz X de treinamento em um produto escalar de três matrizes $U\Sigma V^T$, onde a matriz V contém as componentes (Equação 39) e U é uma matriz mxm real ou complexa. Σ é uma matriz retangular diagonal mxn, com números reais não-negativos na diagonal.

A Figura 40 exemplifica visualmente como o algoritmo funciona em 2D, onde são traçadas as seguintes componentes, a que apresenta a direção que melhor maximiza a variância da amostra (c_1) ; a componente ortogonal a c_1 que maximiza a segunda melhor direção (c_2) . A partir dessa técnica, torna-se possível analisar uma amostra de muitas dimensões em 2D, projetando as componentes c_1 e c_2 no hiperplano.

$$V = \begin{pmatrix} | & | & | & | \\ c_1 & c_2 & \dots & c_n \\ | & | & | & | \end{pmatrix}$$

$$(39)$$

A classe PCA da biblioteca *Scikit-Learn* (Pedregosa *et. al.*, 2011) implementa o PCA com decomposição SVD. Uma informação útil desse algoritmo é a taxa de **variân-**

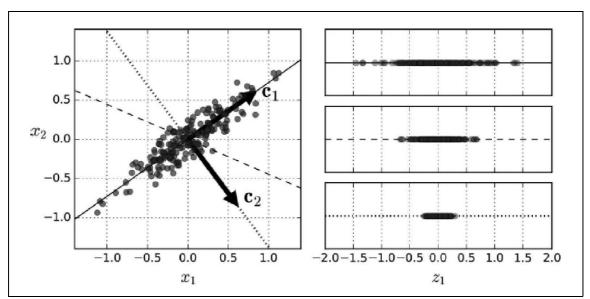


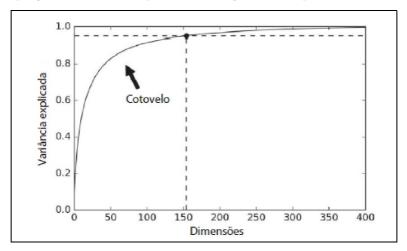
Figura 40 - Seleção das componentes do PCA

Legenda: Selecionando as componentes ortogonais entre si que maximizam a variância empírica da amostra.

Fonte: Géron, 2019, p. 549.

cia explicada, que indica a proporção de variância dos conjuntos de dados distribuída ao longo do eixo de cada componente, permitindo analisar o quanto de informação é armazenada em cada componente (Figura 41).

Figura 41 - Proporção de variância explicada em função das componentes \boldsymbol{c}



Legenda: Proporção de variância explicada em função do número de dimensões, onde o **cotovelo** indica até que ponto o aumento de componentes não seria substancial.

Fonte: Géron, 2019, p. 549.

3 METODOLOGIA

Os modelos de aprendizado de máquina construídos foram treinados e testados com dados de simulação e por formas de onda coletadas por meio de um sistema de experimental contendo 3 detectores.

Os dados experimentais passaram por um processo de redução de taxa de amostragem e centralização da forma de onda. Os eventos são então formatados para a análise de clusterização, treinamento e avaliação dos modelos. A formatação em questão é exemplificada na matriz da Equação 40, onde cada linha contém uma forma de onda com seus respectivos valores de voltagem. As soluções dos modelos mostraram-se suficiente ao se considerar formas de onda com 250 características (valores de voltagens em mV).

$$X = \begin{pmatrix} V_{11} & V_{12} & \dots & V_{1p} \\ V_{21} & V_{22} & \dots & V_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ V_{n1} & V_{n2} & \dots & V_{np} \end{pmatrix}$$

$$(40)$$

As amostras primeiramente são analisadas por clusterização não supervisionada. Ela permite uma avaliação global da variação dos dados da amostra. A análise possibilita, por exemplo, eliminar pulsos com saturações de escala.

Após a análise de clusterização, as amostras são separadas para a construção e avaliação dos modelos supervisionados, sendo separadas da seguinte forma: 80% para o treinamento, 10% para a validação e 10% para o teste. A amostra de validação é utilizada para encontrar os hiperparâmetros que melhor se ajustam aos dados. Quando o treinamento termina, o modelo tem acesso ao teste e suas previsões são analisadas.

Inicialmente, o algoritmo analisado foi a Floresta Aleatória, e em seguida, previsores mais flexíveis de aprendizado profundo (DNN e CNN). Na primeira etapa, são construídos modelos que classificam sinais de cintilação e ruídos de corrente escura. Para uma primeira aproximação, foram geradas FOs simuladas para treinamento, tendo em vista que essas são mais controláveis e facilitam a avaliação dos algoritmos em estudo. Após essa etapa, os modelos foram treinados com formas de onda de dados experimentais. Todas as amostras de teste foram compostas por dados experimentais. Para uma análise com pulsos de diferentes amplitudes, foram coletadas amostras com diferentes voltagens no detector. As eficiências dos classificadores foram comparadas ao método de discriminação por altura de pulso. Na segunda etapa, foram construídos modelos que classificam sinais de cintilação (múons cósmicos) e ruídos de diversas fontes (fundo).

No Capítulo 4 é mostrado como ocorre a formação do pulso elétrico produzido pelo

detector cintilador plástico e a origem das fontes de ruídos do respectivo dispositivo. Também é explicado como foi realizada a modulação dos pulsos experimentais para a geração de formas de onda simuladas. O Capítulo 5 exibe o sistema experimental e como ele é ajustado para a coleta de cada tipo de FO. O Capítulo 6 discute a análise de clusterização para cada amostra de FO. O Capítulo 7 mostra como os modelos supervisionados foram construídos. No Capítulo 8, os resultados dos modelos são analisados e comparados ao método de discriminação por altura de pulso.

4 FORMAÇÃO DOS SINAIS E RUÍDOS E SUA SIMULAÇÃO

4.1 Formação do sinal elétrico

O sinal dos detectores cintiladores possuem uma forma bem definida, contendo a convolução de duas constante de decaimento: a constante de cintilação associada a emissão dos fótons (τ_s) ; e a resposta do circuito elétrico, associado à fotomultiplicadora (τ) .

O ânodo é o eletrodo responsável pela coleta dos elétrons na seção multiplicadora e pela transmissão do pulso elétrico. O sinal de saída do ânodo é um pulso proporcional ao número de elétrons emitido pelo fotocátodo. A saída do sinal da fotomultiplicadora deve se comportar como um gerador de corrente ideal.

A Figura 42 mostra o circuito equivalente ao de uma fotomultiplicadora (Equação 41), onde sua resistência e sua capacitância interna são representadas por R e C, respectivamente. Dado que a chegada dos fótons na PMT tem uma constante de tempo associada (τ_s) , a corrente do ânodo é descrita pela Equação 42; onde G é o fator de ganho da fotomultiplicadora, N número de fotoelétrons emitidos pelo cátodo, e carga do elétron e τ_s a constante de decaimento do cintilador. A combinação da Equação 42 com a Equação 41 forma um equação diferencial, cuja a solução é dada pela Equação 43, onde $\tau = RC$. Esta equação descreve a forma de onda esperada de um sinal de cintilação em duas situações: a do caso mais genérico e aquela em que a constante de tempo de cintilação coincide com a do circuito elétrico.

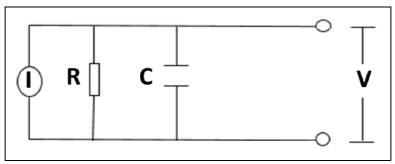
$$I(t) = \frac{V}{R} + C\frac{dV}{dt} \tag{41}$$

$$I(t) = \frac{GNe}{\tau_s} exp(\frac{-t}{\tau_s}) \tag{42}$$

$$V(t) = \begin{cases} -\frac{GNeR}{\tau - \tau_s} \left[e^{\left(-\frac{t}{\tau_s}\right)} - e^{\left(-\frac{t}{\tau}\right)} \right] & \tau \neq \tau_s \\ \left(\frac{GNeR}{\tau_s^2} \right) t e^{\left(-\frac{t}{\tau_s}\right)} & \tau = \tau_s \end{cases}$$

$$(43)$$

Figura 42 - Circuito equivalente de uma fotomultiplicadora.



Fonte: O autor, 2023.

4.2 Ruídos da fotomultiplicadora

Mesmo quando não há interações de partículas com o detector, ainda há fluxos de pequenas correntes que podem ser detectados, os chamados ruídos que surgem de diversas fontes. A seguir discutimos brevemente algumas das principais fontes de ruídos.

4.2.1 Emissão termoiônica

A emissão de elétrons do cátodo e dos dínodos devido ao calor é uma fonte de ruído térmico, estes elétrons são capturados pelo sistema multiplicador, produzindo sinais que se assemelham ao dos fotoelétrons. Essa contribuição tem uma dependência exponencial com a temperatura, dada pela equação de Richardson Equação 44 (Leo, 1987).

$$I = AT^2 exp^{\left(\frac{-e\phi}{kT}\right)} \tag{44}$$

A é uma constante de proporcionalidade. T é a temperatura. ϕ é a função trabalho. k é a constante de Boltzmann

(45)

Esse fenômeno faz com que os sinais das PMTs produzam pulsos continuamente, independente da incidência de fótons no fotocátodo, produzindo a maior contribuição à chamada corrente escura.

4.2.2 Fuga de corrente

A fuga da corrente elétrica por fios, condutores e circuito externos devido ao baixo isolamento elétrico é também uma fonte de ruído. Isto pode ter origem, por exemplo, na conexão com o divisor de tensão com a presença de condução parasita entre os terminais, gerando correntes indesejadas, com a diminuição da umidade esse efeito tende a reduzir.

4.2.3 Afterpulse

Outro gerador de ruídos indesejáveis são os afterpulse, sendo sinais que ocorrem posteriormente ao pulso principal, possuindo relação com ele. Dado que a fotomultiplicadora não possui um vácuo perfeito, gases estão presentes no seu interior, a corrente elétrica produzida no processo de multiplicação podem ionizar seus átomos liberando íons que são acelerados em direção ao fotocátodo ou aos dínodos. Isso faz com que seja produzido um sinal que podem ocorrer em dezenas a centenas de nanosegundos após o pulso principal. Além disso, processos de recombinação de elétrons da corrente com o meio podem também gerar fótons que chegam ao fotocátodo, produzindo novos sinais.

4.3 Simulação das formas de onda

O uso de formas de onda simuladas permitiu avaliar os diferentes algoritmos de aprendizado de máquina em estudo. O modelamento das FOs de sinais e de ruídos foi obtido por ajuste de função em dados experimentais de sinais de cintilação (trigger de múons cósmicos) e de ruído de corrente escura, respectivamente. A variação dos parâmetros destas funções permitiu a geração de formas de ondas simuladas. Nas próximas subseções será detalhado como as FOs de sinais e ruídos simuladas são geradas.

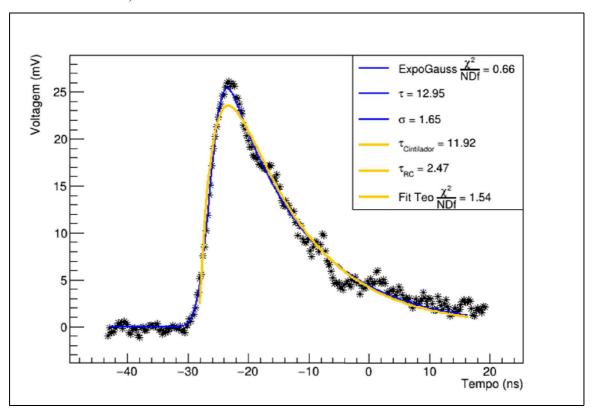
4.3.1 Sinais de cintilação

A Equação 46 (Ratcliff, 1979) é a função expo-gaussina utilizada para simular o sinal de cintilação onde: g(t) é a flutuação aleatória da linha de base; τ a constante de decaimento; A é a amplitude da componente gaussiana; μ e σ é a média e desvio padrão da gaussiana, respectivamente. Para construir uma simulação realísticas dos sinais, a função foi ajustada a uma amostra de dados experimentais de cintilação, de onde esses parâmetros foram extraídos. A Figura 43 mostra os seguintes ajustes em uma FO experimental: a função simulação de cintilação (Equação 46) em azul e a função teórica do sinal de

cintilação (Equação 43) em amarelo.

$$f(t; A, \mu, \sigma, \tau) = \frac{A\sigma}{\tau} \sqrt{\frac{\pi}{2}} exp\left(\frac{1}{2} \left(\frac{\sigma}{\tau}\right)^2 - \frac{t - \mu}{\tau}\right) erfc\left(\frac{1}{\sqrt{2}} \left(\frac{\sigma}{\tau} - \frac{t - \mu}{\sigma}\right)\right) + g(t)$$
(46)

Figura 43 - Ajustes de funções em um sinal de cintilação experimental (trigger de múons cósmicos)



Legenda: Forma de onda de cintilação com os ajustes de função da Equação 46 (azul) e Equação 43 (amarelo).

Fonte: O autor, 2023.

4.3.2 Ruídos de corrente escura

A função utilizada para gerar FOs de ruídos de corrente escura foi a lognormal (Equação 47), onde s é o parâmetro de escala e $m=e^{mediana}$. Os parâmetros da função são estimados através de ajustes nos dados experimentais, a Figura 44 mostra os seguintes ajustes na FO experimental de corrente escura: a função de ajuste de ruído de corrente escura (Equação 47) na cor verde e a função teórica com os termos de constante de decaimento do cintilador igualada a zero $\tau_s=0$ (Equação 43) em amarelo.

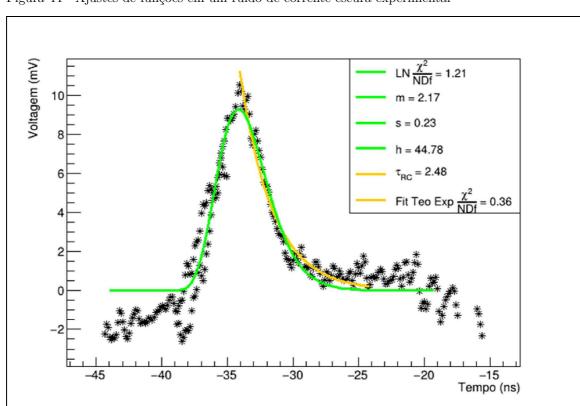


Figura 44 - Ajustes de funções em um ruído de corrente escura experimental $\,$

Legenda: Forma de onda de ruído de corrente escura, com os ajustes de função da Equação 47 (verde) e Equação 43 ($\tau \neq \tau_s$ e $\tau_s = 0$, amarelo).

Fonte: O autor, 2023.

$$R(t) = \frac{1}{t\sqrt{2\pi s^2}} exp\left(-\frac{(\ln x - m)^2}{2s^2}\right) + g(t)$$
(47)

5 SISTEMA EXPERIMENTAL

A coleta dos dados foi realizada através de um sistema composto por três detectores cintiladores alinhados verticalmente e conectados a uma eletrônica padrão NIM (U.S. NIM COMMITTEE, 1990) e a um osciloscópio digital. Os cintiladores são do tipo plástico, modelo BC404 da Saint Gobain com tamanho de $40x40x2~\rm cm^2$. Cada um é conectado a uma fotomultiplicadora XP2020 da Photonis através de um guia de luz. O osciloscópio é um Keysight DSOS404A, que possui largura de banda de 4 GHz, com taxa amostragem máxima de 20 GSa/s, desempenhando a função de digitalizador dos pulsos elétricos. Com este sistema foram coletadas tanto FOs de sinais quanto de ruídos.

5.1 Coleta de FOs de sinais de cintilação

A Figura 45 mostra o esquema do aparato experimental usado para coletar sinais de cintilação. Os pulsos dos três detectores cintiladores são lidos por canais independentes da eletrônica NIM, que estabelece um sinal de coincidência lógica de nível 2 entre os detectores 1 e 3, produzindo um trigger para o digitalizador. O pulso do detector 2 é digitalizado a cada trigger. A Figura 46 mostra algumas das FOs produzidas com esta configuração, onde notam-se subidas praticamente lineares do pulso e descidas não-lineares correspondente à constante de tempo do cintilador e do circuito de acoplamento da fotomultiplicadora discutidos na seção anterior. As FOs foram coletadas com as seguintes voltagens aplicadas no detector 2: 1500, 1550, 1600, 1750, 1850 e 1950 V.

Partícula carregada

Detector 1

Detector 2

Detector 3

Linear
Fan-out
LT-Discriminators

Partícula carregada

Logic
Unit
4GHz
Digitizer

Figura 45 - Sistema experimental com trigger de coleta de sinais de cintilação

Legenda: Esquema do sistema experimental com trigger para coleta de sinais de cintilação, onde é estabelecido uma coincidência de nível 2 entre o detector 1 e 3, permitindo uma coleta de sinal de cintilação do detector 2.

Fonte: O autor, 2023.

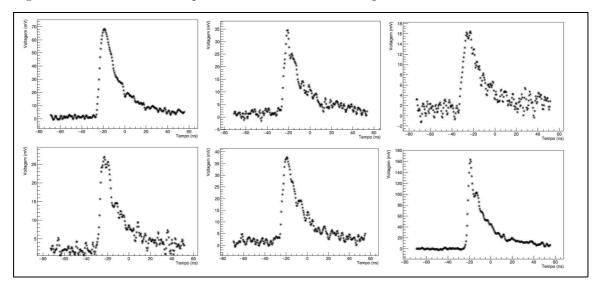


Figura 46 - Formas de onda experimentais de sinais de cintilação

Legenda: Diferentes FOs de sinais de cintilação (trigger de múons cósmicos). O decaimento do pulso elétrico está associado à emissão dos fótons.

Fonte: O autor, 2023.

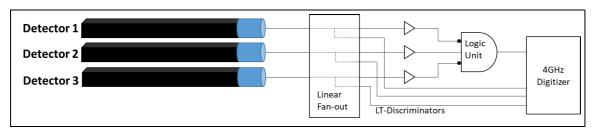
5.2 Coleta de FOs de ruídos

Uma diferente configuração de trigger foi usada para coletar FOs de pulsos de fundo, ou ruídos (Figura 47). Foi fixado um veto à existência de pulsos em quaisquer dos detectores 1 ou 3, permitindo coletar formas de onda quando há pulso apenas no detector 2. A Figura 48 mostra algumas formas de onda coletadas. Na maior parte dos pulsos, não é possível observar o decaimento exponencial anteriormente observado, tornando pouco provável que tenham sido produzidos por fótons de cintilação de uma partícula carregada que atravessa toda a espessura do detector. Pelo menos, a maioria desses pulsos são devidos à corrente escura. Para construir uma amostra com esses pulsos, foram feitos testes de ajustes à Equação 47 e selecionados aqueles com boa concordância. Para a coleta das formas de onda com esta configuração foram aplicados os mesmos valores de tensão: 1500, 1550, 1600, 1750, 1850 e 1950 V.

5.3 Processamento das FOs e linguagem de programação

As formas de onda inicialmente são ajustadas para a aprendizagem dos modelos, esses ajustes consistem em: redução da taxa de amostragem, centralizações dos pulsos e formatação matricial das formas de onda (Equação 40). O pré-processamento ocorre através de macros em linguagem de programação C++, com o uso do pacote ROOT (Brun; Rademakers, 1997).

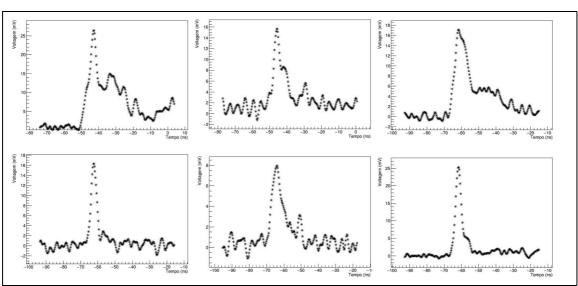
Figura 47 - Sistema experimental com trigger de coleta de ruídos



Legenda: Sistema experimental para coleta de FOs de ruídos, onde é ajustado um veto de coincidência lógica nível 1 entre o detector 1 e 3, permitindo-se coletas de FOs do detector 2 quando não transmissão de pulso pelo outros detectores.

Fonte: O autor, 2023.

Figura 48 - Formas de onda experimentais de trigger de ruído



Legenda: Formas de onda de ruídos de diversas fontes. Em geral os pulsos são mais estreitos do que os sinais cintilação.

Os modelos de aprendizado são construídos por programação em Python, utilizando as seguintes bibliotecas: TensorFlow Keras (Abadi et. al., 2015) para modelos de deep learning (DNNs e CNNs) e Scikit-Learn para clusterização e FA (Pedregosa et. al., 2011).

6 ANÁLISE DE CLUSTERIZAÇÃO

A utilização da clusterização não supervisionada permitiu construir um treinamento mais conciso dos modelos supervisionados. A redução de dimensionalidade junto à clusterização hierárquica aglomerativa (Pedregosa et. al., 2011) possibilitaram analisar as formas de onda coletadas em um hiperplano de duas dimensões. Tal análise mostrou ser útil, pois saturações de escala do osciloscópio puderam ser detectadas e retiradas da amostra.

Esta análise é separada em uma avaliação da discriminação entre sinais de múons cósmicos e de ruídos de corrente escura, e outra de discriminação de sinais de múons cósmicos e de fundo. Para isto, foram utilizadas amostras de dados experimentais, contendo 3 mil FOs de sinais de múons cósmicos e 3 mil de ruído de corrente escura; e 10 mil formas de onda para a amostra de sinais de múons cósmicos e 10 mil de fundo. As amostras foram coletadas com o sistema experimental descrito na seção anterior, sendo uma para cada voltagem aplicada ao detector.

6.1 FOs de sinais de múons cósmicos e ruído de corrente escura

O PCA fornece informações armazenadas em cada componente ao se aplicar a redução de dimensionalidade. A Figura 49 mostra a variância explicada em função do número de dimensões (componentes) para os dados de 1600 V, indicando uma preservação de mais de 95% da informação ao se considerar quatro componentes. No entanto, para cada análise utilizamos sempre as duas primeiras componentes (c_1 e c_2), visto que serão sempre as componentes com mais informação.

O dendrograma (Seção 2.8) da respectiva coleta é apresentado na Figura 50, contendo uma distância relevante para três clusters, por isso é escolhido um limiar de 1000 para o algoritmo. A partir disso, a clusterização pode ser analisada no hiperplano de duas dimensões, como mostrado na Figura 51. A clusterização em questão, mostra que o cluster 1 concentra os ruídos de corrente escura, enquanto o cluster 0 e cluster 2 contém sinais de múons cósmicos, esses dois últimos se diferem pela amplitude. É possível verificar as FOs contidas em cada cluster, a Figura 52 exibe algumas delas, com as cores de cada respectivo cluster. Visando uma identificação de proporção das classes em cada cluster, um rótulo paralelo é definido para as formas de onda, mas o algoritmo não tem acesso a eles. Através desse rótulo, a proporção de classe (sinal e ruído) em cada cluster pode ser observada. A Tabela 2 mostra as proporções de classe para a coleta de 1600 V, onde identifica-se uma excelente separação entre sinal e ruído. O mesmo procedimento é feito para as voltagens de 1750, 1850 e 1950 V. Os dendrogramas, a variância explicada das

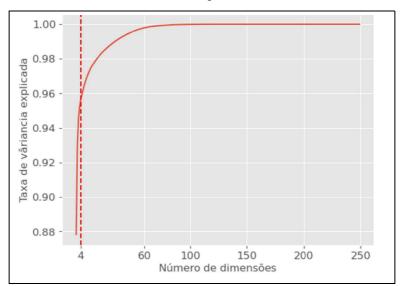


Figura 49 - Taxa de variância explicada em função das dimensões para FOs de sinais de múons cósmicos e ruídos de corrente escura para 1600V

Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

Fonte: O autor, 2023.

componentes, clusterizações e algumas formas de onda de cada respectivo *cluster*, estão contidas na Seção C.1.

A Figura 53 apresenta a clusterização das demais voltagens, as proporções de classe de cada cluster estão na Tabela 2 nos valores de 1600 e 1750 V e Tabela 3 nos valores de 1850 e 1950 V. A análise permitiu eliminar saturações de escalas que são identificadas nos clusters mais distantes e em menor quantidade, principalmente para voltagens de 1850V e 1950V, que obviamente produzem pulsos de maior amplitude. Além disso, para essas mesmas tensões, nota-se uma pequena proporção de sinais que se misturam aos ruídos em um mesmo cluster (nos valores de 1850 e 1950 V, Tabela 3). Ao serem analisados, é verificado que esses sinais são de fato ruídos. Esse efeito é causado pela voltagem elevada de alimentação no detector, produzindo pulsos de corrente escura em maior quantidade que acabam casualmente dentro da janela de coincidência do trigger de sinal. Para um treinamento mais conciso dos modelos supervisionados, esses eventos são retirados da amostra.

De forma geral, foi verificado que as formas de onda de 1600 V e 1750 V apresentam uma variabilidade semelhante, enquanto as FOs de 1850V se comparam com as de 1950 V. A partir disso, as FOs de 1750 V e 1850 V são separadas para o treinamento e teste dos modelos supervisionados (80% treino, 10% teste e 10% validação), as demais amostras de voltagens são inteiramente utilizadas para testes dos modelos (1600 e 1950 V).

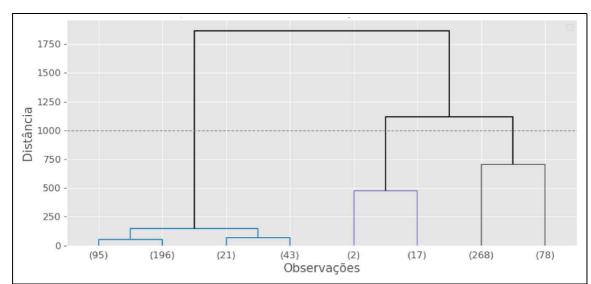


Figura 50 - Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente escura para $1600\mathrm{V}$

Legenda: Dendrograma onde a separação é definida por um limiar de 1000, segmentando a amostra em 3 $\it clusters$.

Fonte: O autor, 2023.

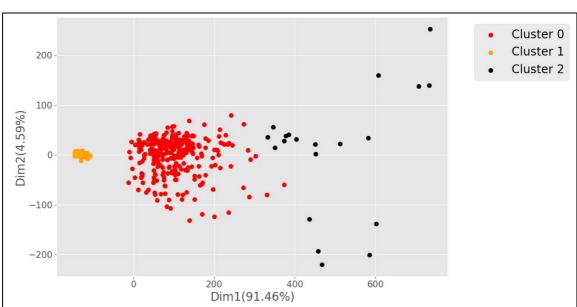


Figura 51 - Clusterização de FOs de sinais de múons cósmicos e ruídos de corrente escura para $1600\mathrm{V}$

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 contém sinais de múons cósmicos, o cluster 1 aglomera ruídos e o cluster 2 agrupa sinais de múons cósmicos de maior amplitude.

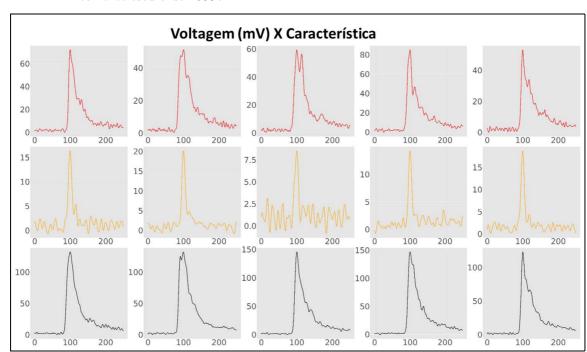


Figura 52 - Formas de onda da clusterização para FOs de sinais de múons cósmicos e ruído de corrente escura de $1600\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada cluster da Figura 51 com a suas respectivas cores.

Fonte: O autor, 2023.

Tabela 2 - Proporção de classe em cada cluster para FOs de sinais de múons cósmicos e ruídos de corrente escura para 1600 e 1750 V

Cluster	Clusterizaç	ção de 1600V	Clusterização de 1750V		
Ciusiei	Sinal (%)	Ruído (%)	Sinal (%)	Ruído (%)	
Cluster 0	94,8	0,0	8,7	0,0	
Cluster 1	0,0	100,0	0,5	100,0	
Cluster 2	5,2	0,0	90,8	0,0	

Fonte: O autor, 2023.

Tabela 3 - Proporção de classe em cada cluster para FOs de sinais de múons cósmicos e ruídos de corrente escura para 1850 e 1950V

Cluster	Clusterizac	ção de 1850V	Clusterização de 1950V		
Ciaster	Sinal (%)	Ruído (%)	Sinal (%)	Ruído (%)	
Cluster 0	96,6	0,0	89,9	0,0	
Cluster 1	2,0	0,0	8,5	100,0	
Cluster 2	1,4	100,0	1,6	0,0	

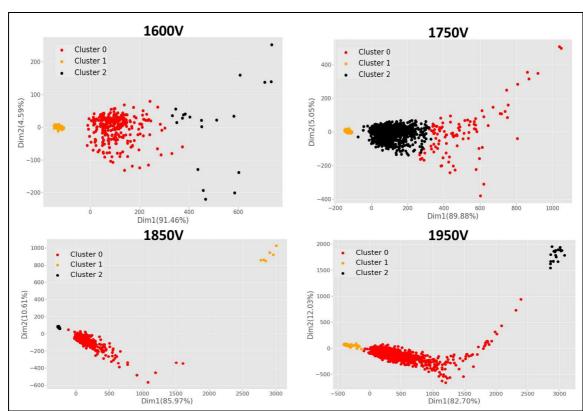


Figura 53 - Clusterização de cada voltagem de operação para FOs de sinais de múons cósmicos e ruídos de corrente escura

Legenda: Projeção das componentes do PCA, com clusterização por meio de HCA para cada voltagem aplicada.

6.2 FOs de sinais de múons cósmicos e de fundo

A análise desta seção contém ruídos de diversas fontes, considerados fundo, onde o mesmo procedimento de análise é realizado inicialmente para uma amostra de 1500 V. A redução de dimensionalidade é aplicada (Figura 54) nas FOs, e as componentes c_1 e c_2 são utilizadas para clusterização e projeção do hiperplano. O dendrograma da clusterização é construído e um limiar de 3000 é ajustado no algoritmo, dividindo a amostra em quatro clusters (Figura 55). A projeção do hiperplano clusterizado é mostrada na Figura 56. Algumas FOs contidas em cada cluster são mostradas na Figura 57, com suas respectivas cores. Através do rótulo paralelo é possível analisar a proporção de classe em cada cluster (Tabela 4). O mesmo procedimento é realizado para as outras amostras de voltagens e estão contidas na Seção C.2.

A análise de clusterização de 1500 V mostra uma concentração relevante de fundo e de sinais em um mesmo *cluster*. Através da verificação individual das formas de onda, observa-se a presença de pulsos de cintilação na amostra de fundo. A Figura 58 mostra na primeira linha algumas FOs de ruído, enquanto na segunda linha são de sinais de múons cósmicos, todas de um mesmo *cluster*. Dada a expectativa de alta eficiência de veto de múons cósmicos pelo *trigger* usado, a quantidade observada de pulsos de cintilação é inesperada. Estudos adicionais precisam ser realizados para compreender a origem dessas FOs.

As clusterizações de todas as voltagens são mostradas na Figura 59, onde as FOs mais distantes são saturações de escala dos sinais de múons cósmicos. A formação de uma protuberância que se destaca de um aglomerado pode ser identificada nas clusterizações, elas são mais intensas e distantes com o aumento da tensão. A Figura 60 por exemplo, mostra a clusterização de 1850 V, onde a protuberância é destacada no cluster 2. As FOs identificadas nessa região são saturações de escala de pulsos de cintilação na amostra de fundo (Figura 61). A frequência desses eventos aumentam para os maiores valores de tensão. Esse efeito ocorre devido à menor escala de voltagem na coleta de ruídos (maior amplificação e, portanto, menor escala dinâmica), enquanto os sinais foram coletados em escala maior.

A Tabela 4 nos valores de 1550 e 1600 V e a Tabela 5 no valor de 1750 V mostram uma razoável separação de sinais e ruídos, enquanto na Tabela 5 nota-se que as voltagens de 1850 e 1950 V contém *clusters* com proporções acima ou em torno de 50% entre sinais e ruídos. Isto provavelmente ocorre porque os pulsos de cintilação inclusos na amostra de ruídos tendem a se assemelhar aos pulsos de múons cósmicos, em proporção muito maior que os pulsos de corrente escura verificados na seção anterior. As FOs de 1500 e 1550 V apresentam proporções de classe menores, porém, esse efeito de aglomeração é causado apenas por FOs de pulsos de cintilação na amostra de fundo.

Tendo em vista um treinamento mais conciso, as saturações de escala tanto de sinais

Tabela 4 - Proporção de classe em cada cluster para FOs de sinais de múons cósmicos e de fundo para $1500,\,1550$ e $1600\mathrm{V}$

Cluster	Clusterização de 1500V		Clusterização de 1550V		Clusterização de 1600V	
Ciusiei	Sinal (%)	Ruído (%)	Sinal (%)	Ruído (%)	Sinal (%)	Ruído (%)
Cluster 0	96,7	20,6	0,7	0,7	98,1	4,5
Cluster 1	0,5	0,8	99,3	9,0	0,3	95,3
Cluster 2	2,8	4,9	0,0	90,3	1,6	0,2
Cluster 3	0,0	73,7	0,0	0,0	0,0	0,0

Fonte: O autor, 2023.

Tabela 5 - Proporção de classe em cada cluster para FOs de sinais de múons cósmicos e fundo para 1750, 1850 e 1950 V

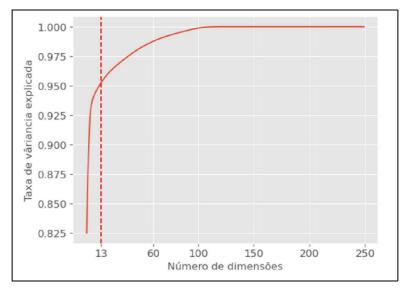
Cluster	Clusterização de 1750V		Clusterização de 1850V		Clusterização de 1950V	
Ciustei	Sinal (%)	Ruído (%)	Sinal (%)	Ruído (%)	Sinal (%)	Ruído (%)
Cluster 0	15,3	0,25	3,1	0,0	2,7	0,0
Cluster 1	1,2	97,5	48,9	98,2	71,5	100,0
Cluster 2	83,5	2,0	48,0	1,8	25,8	0,0

Fonte: O autor, 2023.

quanto de ruídos são retiradas das amostras, os ruídos contidos na amostra de sinais para as tensões de 1850 e 1950 V, também são retirados. Já os ruídos que apresentam pulsos de cintilação são mantidos para verificar se os modelos supervisionados os distinguem dos sinais.

A partir dessa análise, as formas de onda de 1500 e 1750 V foram selecionadas para treinamento, visto que, as FOs de 1500 V apresentam uma proporção considerável de fundo misturados aos sinais de múons cósmicos, se comparados com as FOs de 1550, 1850 e 1950 V. As FOs de 1600 V se comparam as de 1750 V, onde pulsos de ruídos que não apresentam cintilação são menos frequentes. As amostras das formas de onda que são separadas para treinamento, se dividem em: 80% para treino, 10% para validação e 10% para teste, as FOs das demais voltagens de coleta são completamente utilizadas para teste (1550, 1600, 1850 e 1950 V).

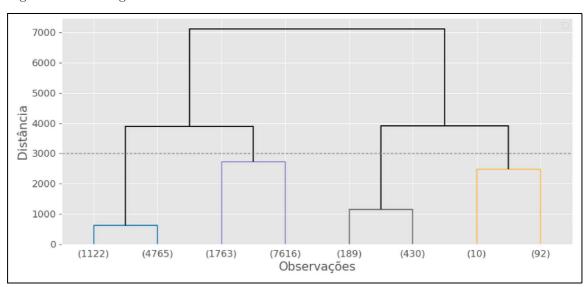
Figura 54 - Taxa de variância explicada em função das dimensões de FOs de sinais de múons cósmicos e de fundo de $1500\mathrm{V}$



Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

Fonte: O autor, 2023.

Figura 55 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de 1500V



Legenda: Dendrograma cuja separação é definida por um limiar de 3000. Assim, a amostra se segmenta em 4 clusters.

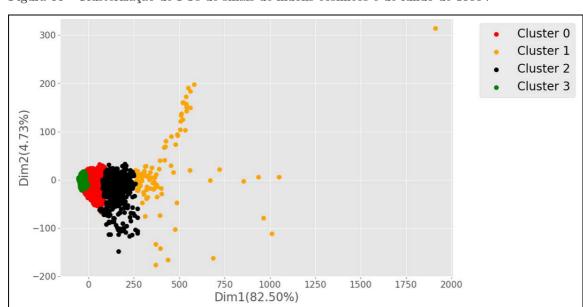
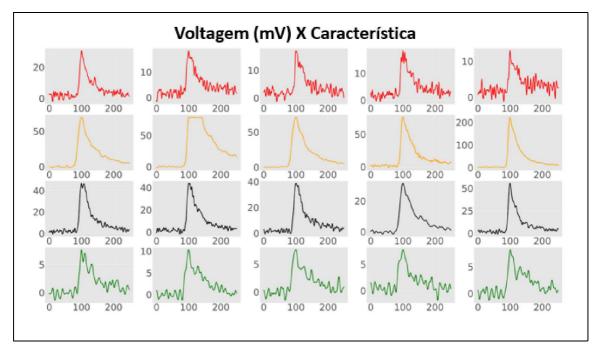


Figura 56 - Clusterização de FOs de sinais de múons cósmicos e de fundo de $1500\mathrm{V}$

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 agrupa os sinais de múons cósmicos e um concentração considerável de pulsos de cintilação na amostra de fundo, o cluster 1 contém pulsos com escalas saturadas tanto de fundo quanto sinais, o cluster 2 aglomera pulsos de cintilação com alta amplitude e por último, o cluster 3 com concentração da maior parte dos ruídos

Figura 57 - Formas de onda da clusterização de FOs de sinais de mú
ons cósmicos e de fundo de $1500\mathrm{V}$



Legenda: Algumas formas de onda contidas em cada cluster da Figura $56~{\rm com}$ a suas respectivas cores.

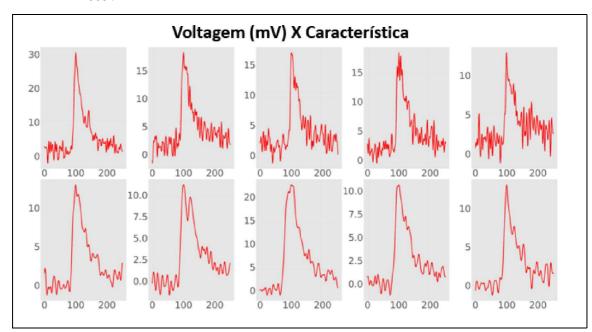


Figura 58 - FOs de sinais de múons cósmicos e de ruídos que apresentam pulso cintilação de $1500\mathrm{V}$

Legenda: Formas de onda contidas na clusterização da Figura 56, com sua respectiva cor. A primeira linha apresenta as FOs de pulsos de cintilação na amostra de fundo, enquanto a segunda linha exibem os sinais de múons cósmicos.

Fonte: O autor, 2023.

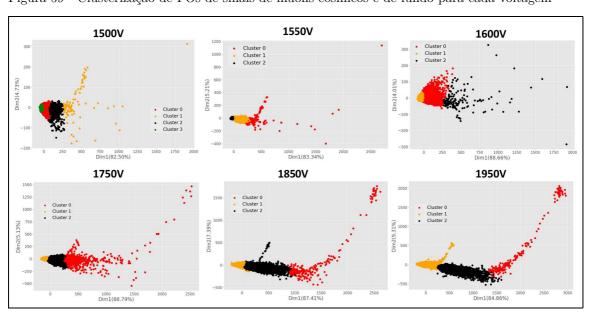


Figura 59 - Clusterização de FOs de sinais de múons cósmicos e de fundo para cada voltagem

Legenda: Projeção das componentes do PCA, com a clusterização HCA para cada voltagem aplicada.

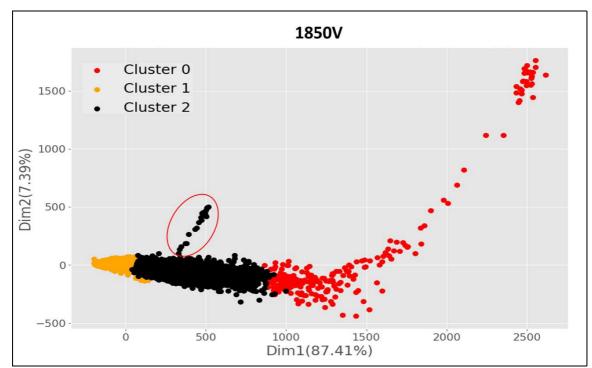


Figura 60 - Clusterização de FOs de sinais de múons cósmicos e de fundo para 1850V

Legenda: Projeção das componentes PCA, com clusterização HCA para formas de onda de 1850 V, onde a protuberância destacada é uma região que contém pulsos de saturações de escala de ruídos.

Fonte: O autor, 2023.

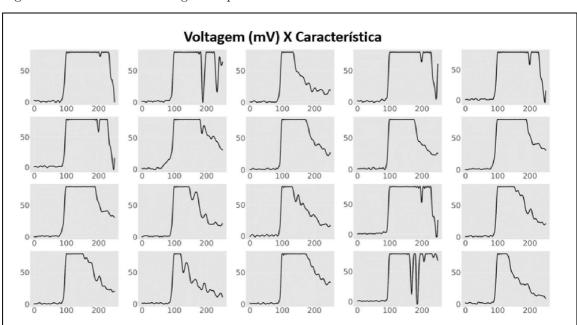


Figura 61 - FOs contidas na região de protuberância

7 CONSTRUÇÃO DOS MODELOS SUPERVISIONADOS

Para a construção dos modelos, foram aplicados os critérios de seleção definidos na clusterização. Primeiramente, foi construído e testado o modelo de Floresta Aleatória, onde os hiperparâmetros do algoritmo foram encontrados via validação cruzada (Seção B.2). Em seguida, foram construídas as arquiteturas das redes neurais (DNN e CNN) e realizadas as técnicas de regularização utilizadas para o treinamento de cada modelo de aprendizado profundo (Seção E). As subsecções apresentam a construção de cada classificador (FOs de sinais de múons cósmicos e corrente escura - FOs de sinais de múons cósmicos e de fundo).

7.1 **FA**

Os hiperparâmetros da FA são encontrados por meio de um método da biblioteca do scikit-learn, o RandomizedSearchCV (Pedregosa et. al., 2011). Ele configura uma grade de valores de hiperparâmetros e seleciona combinações aleatórias para treinar o modelo, através de uma técnica de reamostragem chamada de validação cruzada (Seção B.2).

A FA possui um artifício chamado de avaliação OOB (do inglês, Out of Bag), que permite aproveitar a amostra que se separa para a validação. A reamostragem bootstrap utilizada no algoritmo (Figura 93), infere que a probabilidade de não incluir um observável i a um determinado previsor é de $((1-\frac{1}{n})^n \to e^{-1} \approx 0,37)$. Portanto, em cada subconjunto usa-se em média 63% para treinamento, permitindo utilizar os 37% como dados de validação.

Após os ajustes dos hiperparâmetros por meio de validação cruzada através do RandomizedSearchCV, os modelos são avaliados por métricas de desempenho com as amostras de teste. A avaliação OOB para sinais de múons cósmicos e de ruído de corrente escura, tanto para formas de onda simuladas quanto FOs experimentais, resultou nas respectivas taxas de acerto 99,87% e 99,84%.

A avaliação OOB para FOs de sinais de múons cósmicos e de fundo atingiu uma taxa de acerto de 97,33%, os hiperparâmetros dessa FA também são encontrados através do RandomizedSearchCV.

7.2 DNN e CNN

Inicialmente foi construída uma rede neural com um mínimo de parâmetros, com baixo número de neurônios e apenas uma camada oculta. No caso em que apresenta um

subajuste, há a indicação de que a rede necessita de maior complexidade, e são adicionados mais neurônios e camadas ocultas. Caso a rede apresente um sobreajuste ao treinamento, as técnicas regularização são aplicadas (Seção E), todos os treinamentos utilizam parada antecipada (Seção E.1).

A função ativação utilizada nas camadas de saída foi a *softmax* (Equação 48). A saída da rede tem a probabilidade do dado de entrada ser classificado como sinal ou como ruído, a Figura 62 exemplifica a resposta da rede.

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \tag{48}$$

Camada de entrada ocultas

Camada de saída

Posinal provido

Figura 62 - Rede DNN com função ativação de saída softmax

Fonte: O autor, 2023.

A função perda definida para os modelos de aprendizado profundo foi a **entropia cruzada** (Equação 49), onde n é o tamanho da saída. O otimizador usado para treinamento das redes foi o ADAM (estimação momentânea adaptativa) (Abadi *et. al.*, 2015), que minimiza a função perda por retropropagação (Seção 2.5).

$$L = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]$$
(49)

Os modelos de CNN construídos são compostos por camadas convolucionais seguidas por camadas *pooling* e a camada de achatamento é conectada a uma DNN, como ilustrado na Figura 63.

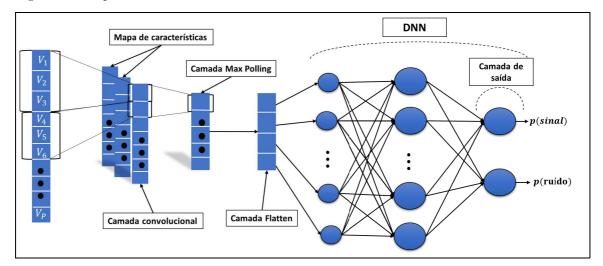


Figura 63 - Arquitetura de uma CNN de 1D conectada a uma DNN

Legenda: CNN de 1D conectada a uma DNN, onde os dados de entrada são formas de onda. A camada de saída é ativada por softmax (Equação 48)

Fonte: O autor, 2023.

7.2.1 DNN - FOs de sinais de múons cósmicos e ruído de corrente escura

A arquitetura da rede treinada por FOs simuladas e FOs experimentais é mostrada na Figura 64 e Figura 65, respectivamente, onde o shape mostra a quantidade de neurônios em cada camada. A rede treinada por FOs simuladas contém duas camadas ocultas, ativadas por função ReLU (Equação 50). O sobreajuste é reduzido por meio de dropout (Seção E.2) e regularização L_1 (Seção E.3). A rede treinada por FOs experimentais é composta por três camadas ocultas, sendo aplicadas as mesmas técnicas de regularização e funções ativações.

$$ReLU(x) = \begin{cases} x & \text{se } x > 0\\ 0 & \text{se } 0 \le x \end{cases}$$
 (50)

As duas DNNs apresentaram uma convergência para as curvas de aprendizado, certificando que o aprendizado foi bem sucedido, como mostrado na Figura 66 (treinamento por dados simulado) e Figura 67 (treinamento por dados experimentais).

7.2.2 DNN - FOs de sinais de múons cósmicos e de fundo

A Figura 68 mostra a arquitetura da rede utilizada para classificar as FOs de sinais de múons cósmicos e de fundo, totalizando 31.477 parâmetros. A DNN contém

Figura 64 - Arquitetura da DNN para treinamento de FOs simuladas

Layer (type)	Output	Shape	Param #
dense_18 (Dense)	(None,	80)	20080
dropout_12 (Dropout)	(None,	80)	0
dense_19 (Dense)	(None,	40)	3240
dropout_13 (Dropout)	(None,	40)	0
dense_20 (Dense)	(None,	2)	82

Legenda: Rede de 3 camadas regularizada por meio de dropout (Seção E.2), totalizando 23402 parâmetros.

Fonte: O autor, 2023.

Figura 65 - Arquitetura da DNN de FOs de sinais de múons cósmicos e ruído de corrente escura

Layer (type)	Output	Shape	Param #
dense_19 (Dense)	(None,	100)	25100
dropout_14 (Dropout)	(None,	100)	0
dense_20 (Dense)	(None,	50)	5050
dropout_15 (Dropout)	(None,	50)	0
dense_21 (Dense)	(None,	25)	1275
dropout_16 (Dropout)	(None,	25)	0
dense_22 (Dense)	(None,	2)	52
otal params: 31,477			
rainable params: 31,477			
on-trainable params: 0			

Legenda: Rede de 4 camadas regularizada por meio de dropout (Seção E.2), totalizando 31477 parâmetros.

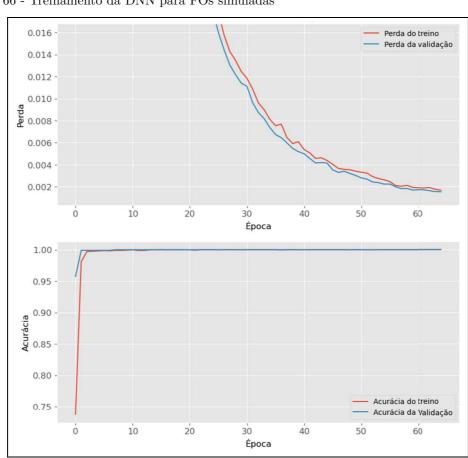


Figura 66 - Treinamento da DNN para FOs simuladas

Legenda: Treinamento de um DNN, onde as curvas de aprendizado apresentam convergência. Fonte: O autor, 2023.

Perda do treino Perda da validação 0.06 0.05 0.04 0.03 0.02 0.01 Ó 20 40 60 80 100 120 Época 1.00 0.95 0.90 Acurácia 08.0 08.0 0.75 0.70 Acurácia do treino Acurácia da Validação 20 40 80 100 120 Época

Figura 67 - Treinamento da DNN para FOs de sinais de múons cósmicos e ruído de corrente escura

Legenda: Treinamento de um DNN, onde as curvas de aprendizado apresentam convergência. Fonte: O autor, 2023.

duas camadas ocultas (ativadas por função ReLU, Equação 50), sendo otimizada por regularização L_2 (Seção E.3).

Figura 68 - Arquitetura DNN de FOs de sinais de múons cósmicos e de fundo

Layer (type)	Output	Shape	Param #
dense_4 (Dense)	(None,	100)	25100
dense_5 (Dense)	(None,	50)	5050
dense_6 (Dense)	(None,	25)	1275
dense_7 (Dense)	(None,	2)	52
otal params: 31,477 rainable params: 31,477 lon-trainable params: 0			

Legenda: Rede de 4 camadas regularizada por meio de *dropout* (Seção E.2), totalizando 31477 parâmetros.

Fonte: O autor, 2023.

O treinamento da DNN é exibido na Figura 69, onde a rede apresenta uma convergência das curvas de aprendizado, mostrando que o modelo consegue generalizar os dados com uma alta acurácia (Equação 34).

7.2.3 CNN - FOs de sinais de múons cósmicos e ruído de corrente escura

A arquitetura da rede treinada por FOs simuladas e FOs experimentais é mostrada na Figura 70 e Figura 71, respectivamente, onde o shape mostra o número de neurônios e o tamanho dos filtros de cada camada. Elas são compostas por 3 camadas convolucionais, cada uma seguida por camada pooling, e a camada de achatamento é conectada a uma DNN. As camada internas da CNN utilizadas para treinamento de FOs simuladas, são ativadas por função ReLU (Equação 50) e regularizadas por L2 (Seção E.3). A rede treinada por FOs experimentais apresentam a mesma composição, exceto pela função ativação da camada interna da DNN (tanh(x), Equação 51), e não necessitou da técnica de regularização L_2 . As redes neurais convolucionais possuem uma arquitetura muito mais complexa do que as DNNs, porém, a sua construção gera menos parâmetros.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{51}$$

O treinamento de ambas CNNs apresentaram uma convergência entre as curvas de aprendizado (Figura 72 e Figura 73), mostrando êxito na construção do modelo, e

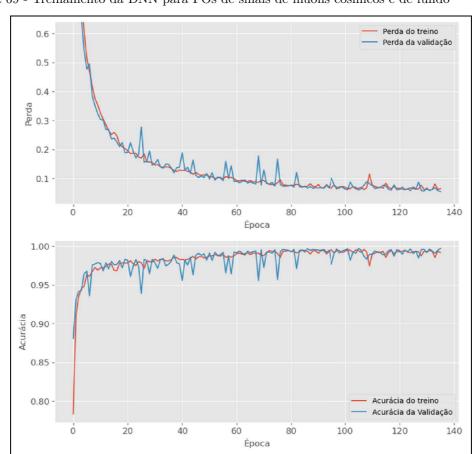


Figura 69 - Treinamento da DNN para FOs de sinais de múons cósmicos e de fundo

Legenda: O treinamento da rede apresentam curvas de validação com certas oscilações, porém, o modelo converge na última época.

Figura 70 - Arquitetura da CNN para FOs simuladas

Layer (type)	Output Shape	Param #
conv1d_3 (Conv1D)	(None, 250, 5)	20
<pre>max_pooling1d_3 (MaxPooling 1D)</pre>	(None, 246, 5)	0
conv1d_4 (Conv1D)	(None, 246, 2)	32
<pre>max_pooling1d_4 (MaxPooling 1D)</pre>	(None, 121, 2)	0
conv1d_5 (Conv1D)	(None, 121, 1)	7
<pre>max_pooling1d_5 (MaxPooling 1D)</pre>	(None, 59, 1)	0
flatten_1 (Flatten)	(None, 59)	0
dense_3 (Dense)	(None, 20)	1200
dense_4 (Dense)	(None, 10)	210
dense_5 (Dense)	(None, 2)	22
Total params: 1,491 Trainable params: 1,491 Non-trainable params: 0		

Legenda: Rede convolucional de 3 camadas convolucionais, 3 camadas pooling, 1 camada de achatamento conectada a uma DNN de 3 camadas, totalizando 1491 parâmetros.

Figura 71 - Arquitetura da CNN para FOs de sinais de múons cósmicos e ruído de corrente escura

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 125, 18)	72
	, , ,	
<pre>max_pooling1d (MaxPooling1D)</pre>	(None, 60, 18)	0
conv1d_1 (Conv1D)	(None, 60, 9)	495
max_pooling1d_1 (MaxPooling 1D)	(None, 18, 9)	0
conv1d_2 (Conv1D)	(None, 18, 9)	252
max_pooling1d_2 (MaxPooling 1D)	(None, 4, 9)	0
<pre>global_max_pooling1d (Globa lMaxPooling1D)</pre>	(None, 9)	0
flatten (Flatten)	(None, 9)	0
dense (Dense)	(None, 7)	70
dense_1 (Dense)	(None, 3)	24
dense_2 (Dense)	(None, 2)	8
Total params: 921 Trainable params: 921		
Non-trainable params: 0		
•		

Legenda: Rede convolucional de 3 camadas convolucionais, 4 camadas pooling, os dados são achatados na camada de achatamento, que por sua vez, são conectados a uma DNN de 3 camadas, totalizando 921 parâmetros.

apresentando resultados de alta acurácia.

Perda do treinamento 0.8 Perda da validação 0.7 0.6 0.5 g ا الا 0.4 0.3 0.2 0.1 10 70 20 30 40 50 Época 1 000 0.975 0.950 0.925 0.900 0.875 0.850 0.825 Acurácia do treino Acurácia da Validação 0.800 Ó 10 20 30 40 50 70 Época

Figura 72 - Treinamento da CNN para FOs simuladas

Legenda: Treinamento da CNN com formas de onda simuladas, mostrando uma convergência das curvas de aprendizado.

Fonte: O autor, 2023.

7.2.4 CNN - FOs de sinais de múons cósmicos e de fundo

A arquitetura da rede CNN utilizada para a classificação contém camadas convolucionais seguidas por camadas *pooling*, os dados são achatados e conectados a uma DNN de 3 camadas, totalizando 7600 parâmetros (Figura 74). Nessa rede, as camadas internas são ativadas por função tangente hiperbólica e são regularizadas por L_2 (Seção E.3).

O treinamento da rede apresentou uma convergência entre as curvas de aprendizado, que é mostrada na Figura 75. O aumento dos *strides* (distância do campo receptivo, Seção 2.4.1) foi necessário para reduzir o sobreajuste, pois diminuiu a complexidade da rede.

0.5 Perda do treino Perda da validação 0.4 Perda 8.0 0.2 0.1 5 15 25 30 ó 10 35 20 Época 1.0 0.9 Acurácia 8.0 8.0 0.6 Acurácia do treino Acurácia da Validação 0.5 5 10 15 25 20 30 35 Época

Figura 73 - Treinamento da CNN para FOs do tipo sinais de múons cósmicos e ruído de corrente escura

Legenda: Treinamento da rede CNN, com as curvas de treino e validação convergindo em ambas métricas.

Figura 74 - Arquitetura da CNN para FOs de sinais de múons cósmicos e de fundo

Layer (type)	Output Shape	Param #
conv1d_3 (Conv1D)	(None, 250, 40)	240
max_pooling1d_3 (MaxPooling 1D)	(None, 63, 40)	0
conv1d_4 (Conv1D)	(None, 13, 20)	4020
<pre>max_pooling1d_4 (MaxPooling 1D)</pre>	(None, 4, 20)	0
conv1d_5 (Conv1D)	(None, 1, 13)	1313
<pre>max_pooling1d_5 (MaxPooling 1D)</pre>	(None, 1, 13)	0
flatten_1 (Flatten)	(None, 13)	0
dense_3 (Dense)	(None, 50)	700
dense_4 (Dense)	(None, 25)	1275
dense_5 (Dense)	(None, 2)	52
Total params: 7,600 Trainable params: 7,600 Non-trainable params: 0		

Legenda: Rede convolucional de 3 camadas convolucionais, 3 camadas pooling, conectadas a uma camada de achatamento, com conexão para uma DNN de 3 camadas, totalizando 7600 parâmetros.

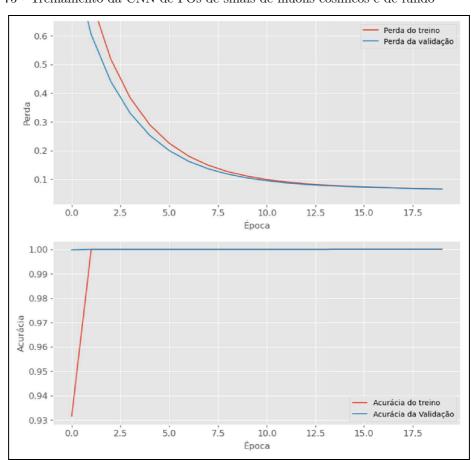


Figura 75 - Treinamento da CNN de FOs de sinais de múons cósmicos e de fundo

Legenda: Treinamento da rede CNN, onde cada época representa um interação (época), apresentando uma convergência entre as curvas de aprendizado, com uma alta acurácia. Fonte: O autor, 2023.

8 ANÁLISE DOS RESULTADOS

As incertezas das métricas de cada modelo foram estimadas por um tratamento Bayesiano (Seção D), tendo em vista que a matriz confusão fornece a quantidade de classes classificadas corretamente e incorretamente, a acurácia, o TPR e o TNR podem ser tratadas como eficiências. A partir de uma derivação da função densidade de probabilidade, torna-se possível estimar a eficiência média (Equação 52) e sua incerteza (Equação 53), onde k são as previsões classificadas corretamente, n as previsões a serem classificadas e ϵ é a eficiência. Os erros avaliados são meramente estatísticos, ou seja, os erros sistemáticos são desconsiderados. Além da estimativa dessas três métricas (acurácia, TPR e TNR), os modelos são avaliados por curva ROC (Figura 36).

$$\bar{\epsilon} = \frac{k+1}{n+2} \tag{52}$$

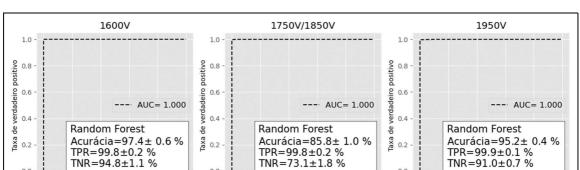
$$V(\epsilon) = \frac{(k+1)(k+2)}{(n+2)(n+3)} - \frac{(k+1)^2}{(n+2)^2}$$
(53)

8.1 FOs de sinais de múons cósmicos e ruídos de corrente escura

Os resultados das métricas de desempenho de cada voltagem para cada tipo de modelo (FA, DNN e CNN) treinados por formas de onda simuladas, são expostos na Figura 76, Figura 77 e Figura 78, respectivamente. A FA apresentou um bom desempenho na classificação dos sinais; no entanto, demonstrou dificuldade em prever os ruídos, especialmente nas formas de onda coletadas com as tensões de 1750 V e 1850 V. A DNN apresenta um desempenho melhor, porém, as previsões das formas de onda de 1600 V apresentaram uma queda de quase 5%. O déficit é causado pela classificação de falsos positivos, onde as FOs de fundo são classificadas como sendo de sinais. A CNN obteve resultados promissores, atingindo acurácia em torno de 99%.

Os resultados dos modelos treinados com formas de onda experimentais mostram acurácia em torno de 99%, com previsões eficientes mesmo com a variação de voltagem. Os resultados de cada classificador (FA, DNN e CNN) são apresentados na Figura 79, Figura 80 e Figura 81, respectivamente.

A Tabela 6 e Tabela 7 exibe a estimativa da acurácia para FOs de teste de 1750 e 1850 V (10% da amostra total) de cada modelo treinados por FOs simuladas e experimentais, respectivamente.



0.6

Taxa de falso positivo

1.0

0.0

0.6

Taxa de falso positivo

Figura 76 - Desempenho da FA treinada por FOs simuladas de sinais de múons cósmicos e ruído de corrente escura

Legenda: Desempenho da curva da ROC do modelo FA, para formas de onda coletadas de diferentes voltagens. O box em branco apresenta o resultado da acurácia, TPR(eficiência dos sinais) e TNR(eficiência dos ruídos).

0.0

0.0

1.0

Fonte: O autor, 2023.

0.6

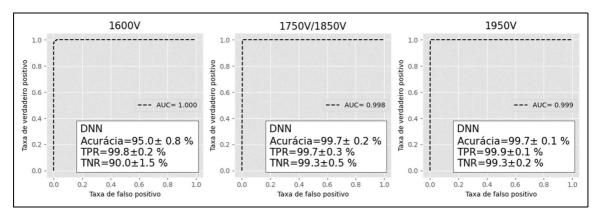
Taxa de falso positivo

0.8

0.0

0.0

Figura 77 - Desempenho da DNN treinada por FOs simuladas de sinais de múons cósmicos e ruído de corrente escura



Legenda: Desempenho da curva da ROC do modelo DNN, para formas de onda coletadas de diferentes voltagens. O box em branco apresenta o resultado da acurácia, TPR (eficiência dos sinais) e TNR(eficiência dos ruídos).

Fonte: O autor, 2023.

Tabela 6 - Estimativa da acurácia por tratamento Bayesiano de modelos treinados por FOs simuladas de sinais de múons cósmicos e ruído de corrente escura

Modelo	Acurácia (%)
FA	$99,8 \pm 0,2$
DNN	$99,9 \pm 0,1$
CNN	$99,56 \pm 0,04$

CNN

Acurácia=98.5± 0.2 % TPR=99.9±0.1 %

0.6

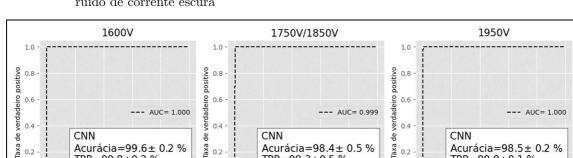
TNR=97.3±0.4 %

Taxa de falso positivo

de

0.0

1.0



0.4

0.2

0.0

0.0

Taxa de

1.0

Figura 78 - Desempenho da CNN treinada por FOs simuladas de sinais de múons cósmicos e ruído de corrente escura

Legenda: Desempenho da curva da ROC do modelo CNN, para formas de onda coletadas de diferentes voltagens. O box em branco apresenta o resultado da acurácia, TPR(eficiência dos sinais) e TNR(eficiência dos ruídos).

CNN

Acurácia=98.4± 0.5 TPR=99.3±0.5 %

0.6

TNR=97.2±1.0 %

Taxa de falso positivo

Fonte: O autor, 2023.

CNN

Acurácia=99.6± 0.2 TPR=99.8±0.2 % TNR=99.3±0.4 %

Taxa de falso positivo

0.6

0.8

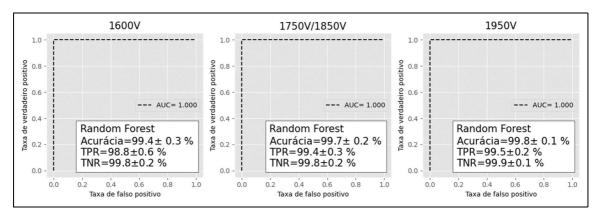
0.4

0.2

0.0

0.0

Figura 79 - Desempenho da FA treinada por FOs de sinais de múons cósmicos e ruídos de corrente escura



Legenda: Desempenho da curva da ROC, para formas de onda coletadas por diferentes tensões do modelo FA. O box em branco apresenta o resultado da acurácia, TPR (eficiência dos sinais) e TNR (eficiência dos ruídos)

Fonte: O autor, 2023.

Tabela 7 - Estimativa da acurácia por meio de um tratamento Bayesiano de modelos treinados por FOs de sinais de múons cósmicos e ruídos de corrente escura

Modelo	Acurácia (%)
FA	$99,7 \pm 0,2$
DNN	$99,91 \pm 0,06$
CNN	$98,0 \pm 0,7$

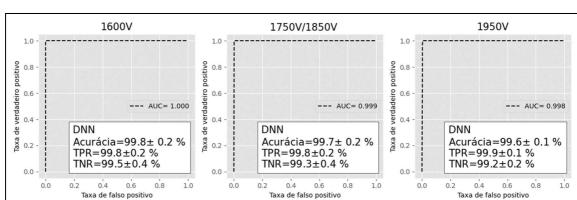


Figura 80 - Desempenho da DNN treinada por FOs de sinais de múons cósmicos e ruídos de corrente escura

Legenda: Desempenho da curva da ROC, para formas de onda coletadas por diferentes tensões do modelo DNN. O box em branco apresenta o resultado da acurácia, TPR (eficiência dos sinais) e TNR (eficiência dos ruídos)

Fonte: O autor, 2023.

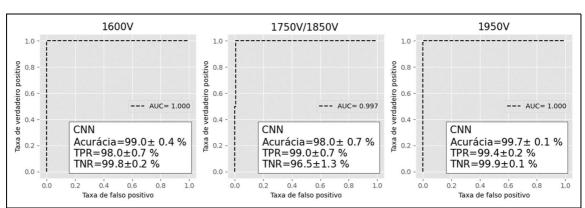
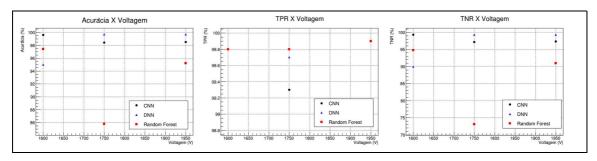


Figura 81 - Desempenho da CNN treinada por FOs de sinais de múons cósmicos e ruídos de corrente escura

Legenda: Desempenho da curva da ROC, para formas de onda coletadas por diferentes tensões do modelo CNN. O box em branco apresenta o resultado da acurácia, TPR (eficiência dos sinais) e TNR (eficiência dos ruídos)

Em todos os tipos de treinamento (simulado e experimental), as curvas ROC atingem valor bem próximo ou igual a um, mostrando que a classificação dos previsores não são aleatórias. A Figura 82 (treinamento por FOs simuladas) e Figura 83 (treinamento por FOs experimentais) mostram os desempenhos (acurácia, TPR e TNR) de cada modelo de uma forma resumida.

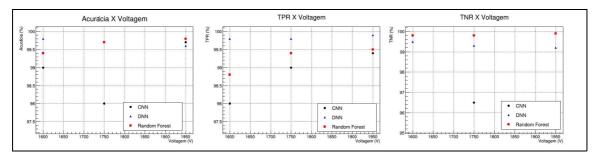
Figura 82 - Desempenho dos modelos treinados por amostra simuladas de FOs de sinais de múons cósmicos e ruídos de corrente escura.



Legenda: Desempenho de cada modelo (acurácia, TPR e TNR) em função da voltagem aplicada para a coleta das formas de onda (sinais de múons cósmicos e ruído de corrente escura). Os classificadores foram treinados com formas de onda simuladas.

Fonte: O autor, 2023.

Figura 83 - Desempenho dos modelos treinado por dados experimentais de FOs de sinais de múons cósmicos e ruídos de corrente escura



Legenda: Desempenho de cada modelo (acurácia, TPR e TNR) em função da voltagem aplicada, para coleta das formas de onda (sinais de múons cósmicos e ruído de corrente escura). O treinamento foi realizado com dados experimentais.

Fonte: O autor, 2023.

8.2 FOs de sinais de múons cósmicos e de fundo

Todos os modelos (FA, DNN e CNN) atingiram métricas de desempenho acima de 90%, como mostram a Figura 84, Figura 85 e a Figura 86, respectivamente. O classificador

CNN atinge taxa de acerto acima de 99%, para todas as tensões. Os outros modelos apresentaram métricas acima de 90%, porém, as previsões de FOs de 1950 V apresentam déficits em torno de 6%, falhando em classificar fundo (TNR). A área da curva ROC para todos os casos atinge uma classificação não aleatória e bem próxima de 1. A Tabela 8 mostra a estimativa de cada modelo, para formas de onda com mesma variabilidade das FOs do treino (10% da amostra de 1500 e 1750 V). A Figura 87 resume o desempenho de cada modelo (FA, DNN e CNN).

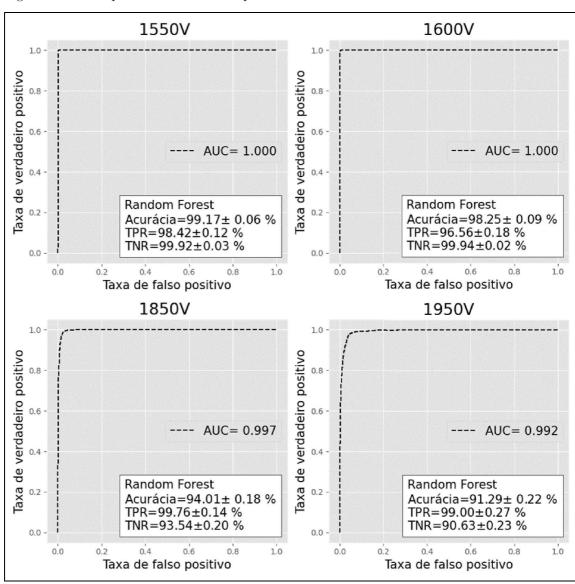


Figura 84 - Desempenho da FA treinada por FOs de sinais de múons cósmicos e de fundo

Legenda: Desempenho da curva da ROC, para FOs de cada coleta de voltagem, para o modelo FA. O box em branco mostra o resultado da acurácia, TPR (eficiência dos sinais) e TNR (eficiência do fundo).

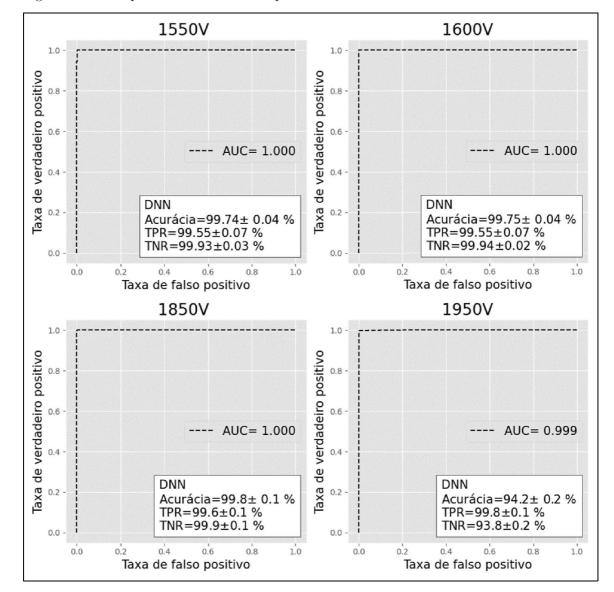


Figura 85 - Desempenho da DNN treinada por FOs de sinais de múons cósmicos e de fundo

Legenda: Desempenho da curva da ROC, para FOs de cada coleta de voltagem, para o modelo DNN. O box em branco mostra o resultado da acurácia, TPR (eficiência dos sinais) e TNR (eficiência do fundo).

Fonte: O autor, 2023.

Tabela 8 - Estimativa da acurácia por meio de tratamento Bayesiano de modelos treinado por FOs de sinais de múons cósmicos e de fundo

Modelo	Acurácia (%)
FA	$97,74 \pm 0,17$
DNN	$99,08 \pm 0,16$
CNN	$99,97 \pm 0,02$

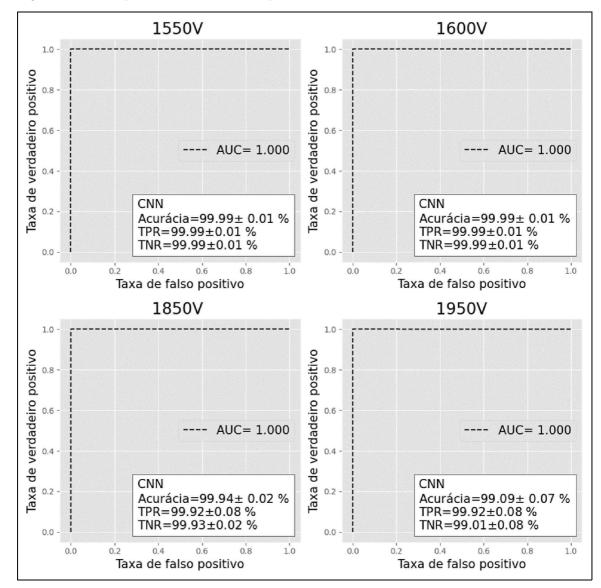


Figura 86 - Desempenho da CNN treinada por FOs de sinais de múons cósmicos e de fundo

Legenda: Desempenho da curva da ROC, para FOs de cada coleta de voltagem, para o modelo CNN. O box em branco mostra o resultado da acurácia, TPR (eficiência dos sinais) e TNR (eficiência do fundo).

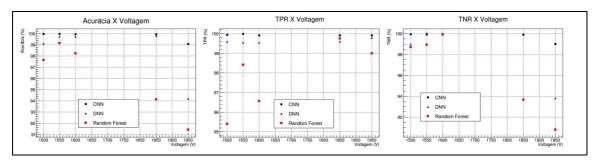


Figura 87 - Desempenho dos modelos treinados com FOs sinais de múons cósmicos e de fundo

Legenda: Desempenho de cada modelo (acurácia, TPR e TNR) em função da voltagem aplicada para coleta das formas de onda (sinais de múons cósmicos e de fundo).

Fonte: O autor, 2023.

8.3 Discriminação por altura de pulso

As formas de onda classificadas pelos modelos foram analisadas por discriminação de altura de pulso. O objetivo é comparar a eficiência na detecção obtida pelo dois métodos.

A Figura 88 mostra a eficiência obtida pela discriminação por altura de pulso, para FOs de sinais de múons cósmicos e de ruído de corrente escura. O limiar definido (traço vertical) é escolhido quando a detecção dos ruídos é nula. As FOs de 1600 V apresentam um queda considerável na eficiência dos sinais (curva vermelha), quase 20% dos sinais são perdidos. Todos modelos supervisionados atingiram uma eficiência acima de 98%. Para as demais voltagens, o método de discriminação de altura de pulso atinge um desempenho semelhante aos dos classificadores.

A discriminação para as FOs de sinais de múons cósmicos e de fundo, onde o limiar definido foi de 15 mV (limiar de operação), obteve-se a eficiência de detecção mostrada na Figura 89. O traço vertical marca o limiar utilizado, a curva em vermelho representam os sinais e a curva em azul os ruídos. Assim como no caso anterior, as FOs são as mesmas classificadas pelos previsores. As FOs de voltagem de 1600 à 1950 V apresentam um boa eficiência para os sinais de múons cósmicos, porém, existe um detecção considerável de ruídos. O aumento do limiar pode ser uma solução para eliminar os ruídos, mas para valores menores de voltagem (1500, 1550, 1600 e 1750 V), há uma queda significativa na eficiência dos sinais. As outras tensões, principalmente a de 1500 V, atingem uma baixa eficiência, praticamente 60% com a coleta considerável de fundo (16,5%), a discriminação de 1550V mostra uma eficiência de 90% dos sinais e detecção de 5,8% do fundo. O modelo de CNN construído, atinge uma eficiência maior que a discriminação por altura de pulso mais 99% para todas as voltagens (Figura 86).

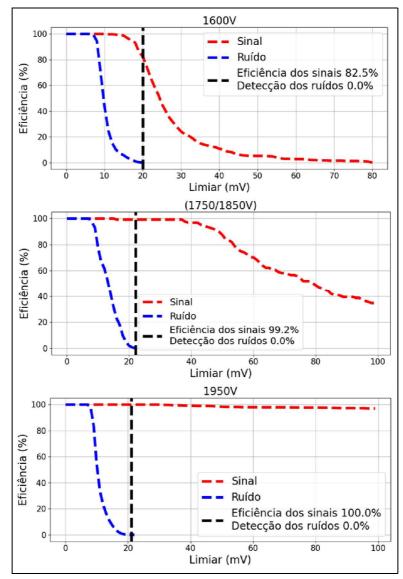


Figura 88 - Discriminação por altura de pulso de FOs de sinais de múons cósmicos e ruídos de corrente escura

Legenda: Discriminação por altura de pulso das FOs de sinais de múons cósmicos e ruídos de corrente escura. As FOs foram utilizadas nos testes dos modelos de aprendizado supervisionado. O traço vertical mostra o limiar de altura de pulso selecionado, o box em branco mostra a eficiência de sinais e a coleta de ruídos que seria obtida, caso utilizasse o respectivo limiar.

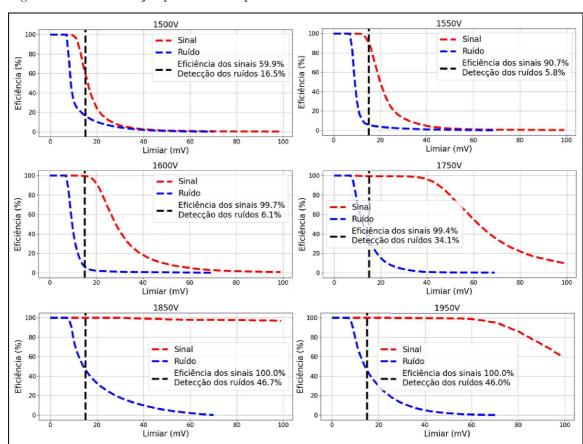


Figura 89 - Discriminação por altura de pulso de FOs de sinais de múons cósmicos e de fundo

Legenda: Discriminação por altura de pulso das FOs de sinais de múons cósmicos e de fundo. As FOs foram utilizadas nos testes dos modelos supervisionados. O traço vertical mostra o limiar de altura de pulso selecionado (15 mV), o box em branco mostra a eficiência de sinais e a coleta de ruídos que seria obtida, caso operasse com o respectivo limiar.

CONCLUSÃO

Foram estudados três algoritmos de aprendizados de máquina para a discriminação de pulsos elétricos de detectores de partícula. O detector estudado é um cintilador plástico conectado à uma fotomultiplicadora, sendo os múons cósmicos que passam por ele a fonte de sinal. Para isso, foram construídos modelos de Floresta Aleatória e de Aprendizado Profundo dos tipos DNN e CNN com base em dados simulados e dados experimentais. Além disso, foi também utilizado métodos de aprendizado não-supervisionado para o filtro e seleção de dados para o treinamento dos modelos supervisionados. Os modelos treinados por formas de onda simuladas tinham como objetivo inicial verificar a capacidade dos algoritmos supervisionados com o uso de amostras mais controláveis. No entanto, os modelos de Aprendizado Profundo apresentaram resultados mais favoráveis para classificar as formas de onda de dados experimentais (Figura 82), com desempenho superior ao da Floresta Aleatória em todas as voltagens. Mesmo assim, a FA ainda atinge acertos acima de 90%, exceto pelas amostras de 1750 V aplicados ao detector.

O desempenho dos modelos treinados com FOs experimentais (sinais de múons cósmicos e de ruído de corrente escura) obtiveram uma excelente performance (Figura 83), onde cada modelo atinge acertos em torno de 99%, para cada voltagem. Os resultados evidenciam uma eficiência superior à discriminação por altura de pulso em 1600 V e um desempenho equivalente para demais tensões de coleta (Figura 88).

Para os classificadores de FOs de sinais de múons cósmicos e de fundo, a CNN apresentou um resultado promissor, com uma taxa de acerto acima de 99% para todos os casos (independe da voltagem de coleta). O classificador DNN se aproxima bem dos resultados da CNN, porém, com um déficit em torno de 5% para classificar o fundo da coleta de 1950 V. Contudo, os classificadores de uma forma geral, apresentaram uma eficiência acima de 90% para todas as métricas. As áreas das curvas ROC foram praticamente igual ou bem próximas de 1, reforçando que os modelos não possuem uma classificação aleatória. O bom desempenho do modelo CNN, pode ser justificado pela arquitetura dos filtros, onde neurônios compartilham o mesmo mapa de característica que aprimora padrões específicos, uma vez que a rede aprende a reconhecer um padrão independentemente do local onde ocorra.

Os resultados dos modelos mostram também, um desempenho superior à discriminação por altura de pulso (Figura 89). Principalmente para as formas de onda coletadas em tensões abaixo de 1600 V, a discriminação por limiar para alguns casos coleta uma quantidade considerável de ruídos. Os modelos de aprendizado de máquina apresentaram uma excelente distinção pela forma de onda independente da tensão de coleta aplicada no detector.

A análise de clusterização, além de possibilitar um treinamento mais conciso para

os modelos supervisionados, permitiu detectar pulsos de cintilação na amostra de fundo de origem não identificada. Estudos adicionais precisam ser realizados para o entendimento desse efeito nas amostras de fundo. No entanto, os modelos apresentaram boa capacidade de distinguir fundo e sinais de múons cósmicos (Figura 87).

Embora o desempenho dos modelos supervisionados tenham apresentado resultados promissores, a praticidade do seu uso no processamento de sinais depende da implementação em dispositivos como ASICs (Application-Specific Integrated Circuits) ou FPGAs (Field Programmable Gate Arrays). No caso da aplicação em detectores de partículas a velocidade de resposta é um importante fator, pois tipicamente geram pulsos elétricos na ordem de 20 ns a 500 ps de largura. Portanto, o dispositivo deve ser capaz de fazer amostragens de sinais a taxas elevadas e fazer a inferência do modelo. Além disso, deve processar a uma velocidade compatível com a taxa de eventos esperada. Nesse caso, o desempenho de classificação do modelo de aprendizado de máquina a ser utilizado pode ter importância menor do que a velocidade de resposta que possa produzir.

REFERÊNCIAS

- ABADI, M. et. al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Disponível em: https://www.tensorflow.org/. Acesso em: 2 dez. 2023.
- BRUN, R.; RADEMAKERS, F. ROOT- An Object Oriented Data Analysis Framework. Nucl. Inst. Meth. in Phys. Res. A, v. 389, n. 1-2, p. 81-86: [s.n.], 1997.
- CLEMENTS, M. A.; MIAO, J. Unsupervised pattern recognition for digital waveform classification from radiation detectors. In: International Conference on Acoustics, Speech, and Signal Processing, Detroit, 1995. [Proceedings ...]. p. 2751-2754 v. 4. DOI: 10.1109/ICASSP.1995.480131.
- FURLETOV, S. Machine learning on FPGA for event selection. Journal of Instrumentation JINST, v. 17, C06009, 2022.
- GÉRON, A. Mãos à Obra: Aprendizado de Máquina com Scikit-Learn TensorFlow: Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes. [S.l.]: Alta Books Editora, 2019. 549 p.
- GRUPEN, C. et. al. Particle Detectors. 2. ed. [S.l.]: Cambridge University Press, 2009. 676 p.
- JAMES, G. et. al. An Introduction to Statistical Learning with Applications in R. 8. ed. Germany, Berlin Heidelberg: Springer-Verlag, 2017. 419 p.
- KASSAMBARA, A. Practical Guide To Cluster Analysis in R: Unsupervised machine learning. [S.l.]: STHDA, 2017. 187 p.
- KIM, K. Y. Advances in Optical and Photonic Devices. InTech, 2010. Disponível em: http://dx.doi.org/10.5772/127. Acesso em: 12 dez. 2023.
- LEO, W. R. Techniques for Nuclear and Particle Physics Experiments. Germany, Berlin Heidelberg: Springer-Verlag, 1987. 377 p.
- MUHAMMAD, W. et. al. Stages-Based ECG Signal Analysis From Traditional Signal Processing to Machine Learning Approaches: A Survey. IEEE Access, Digital Object, 2020. Disponível em: http://dx.doi.org/10.1109/ACCESS.2020.3026968. Acesso em: 12 dez. 2023.
- OPTOELECTRONICS, P. E. Avalanche Photodiodes: A User's Guide. [S.l.], 2003.
- PEDREGOSA, F. et. al. Scikit-learn: Machine Learning in Python. 2011. Disponível em: https://scikit-learn.org/stable/. Acesso em: 2 dez. 2023.
- PHOTONIS. MOD. XP2020Q. Photomultiplier Tubes Catalogue. [S. l.: s. n.], 2007.
- RATCLIFF, R. Group Reaction Time Distributions and an Analysis of Distribution Statistics. Psychological Bulletin, p. 446–461, 1979.
- TIANQI, W. et. al. Machine Learning-Based Pulse Wave Analysis for EarlyDetection of Abdominal Aortic Aneurysms Using In SilicoPulse Waves. symmetry MDPI, 2023. Disponível em: https://doi.org/10.3390/sym13050804. Acesso em: 2 dez. 2023.

ULLRICH, T.; XU, Z. Treatment of Errors in Efficiency Calculations. [S. l.]: Brookhaven National Laboratory, 2008.

U.S. NIM COMMITTEE. Standard NIM Instrumentation System. U.S. Department of Commerce, Springfield, Virginia 22161, 1990.

APÊNDICE A – Algoritmos de aprendizado de máquina

As secções desse apêndice descrevem como os algoritmos que compõem a FA funcionam: o método de árvore de decisões e o método *Bagging*.

A.1 Árvores de decisão

As árvores de decisões são métodos de aprendizagem que envolvem segmentações de um espaço de característica, a partir dos chamados cortes, como mostrado na Figura 90. A qualidade dessas divisões podem ser mensuradas pelo coeficiente de Gini (Equação 54), onde p_{ki} é a média das instâncias da classe k entre as instâncias de treinamento no nó (corte) i. A medida de impureza avalia o corte que contém mais observações da mesma classe.

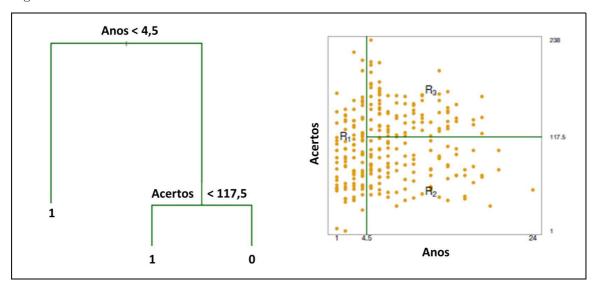
$$G_i = 1 - \sum_{k=1}^n p_{i,k}^2 \tag{54}$$

A profundidade de uma árvore pode ser muito maior, com diversas divisões (Figura 91), uma forma de reduzir a flexibilidade desse algoritmo, é definir um limite para essa profundidade, o processo em questão é chamado de **poda**. Portanto, quando se poda uma árvore de decisão reduz o sobreajuste do modelo, pois a flexibilidade do algoritmo diminui.

A.2 Bagging e Pasting

O termo bagging vem do inglês bootstrap (Seção B.1), esse algoritmo é um método de aprendizado de ensemble que agrega a previsão de um conjunto de modelos. A partir de uma técnica de reamostragem (bootstrap) que segmenta o conjunto de treinamento em amostras ligeiramente diferentes (Figura 92). O procedimento permite criar previsores mais diversificados.

Figura 90 - Árvore de decisões



Legenda: Representação de uma árvore de decisão com 3 divisões.

Fonte: James et. al., 2017, p. 419. Adaptado pelo autor.

Years < 4.5

Putouts < 82
Years < 3.5
5.487
Years < 3.5
5.394
6.189

Walks < 43.5
Walks < 43.5
Walks < 52.5

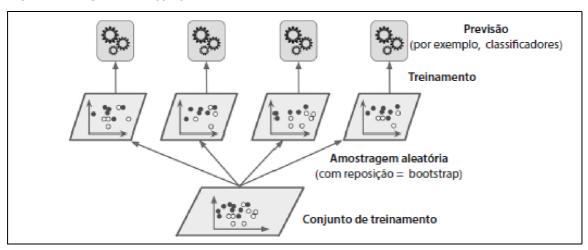
Runs < 47.5
6.015
5.571
6.407
6.549
Years < 6.5
7.289

Figura 91 - Árvore de decisão mais complexa

Legenda: Representação de uma árvore de decisão com diversas divisões.

Fonte: James et. al., 2017, p. 419. Adaptado pelo autor.

Figura 92 - Algoritmo Bagging



Legenda: Algoritmo Bagging, onde é aplicado a técnica de reamostragem bootstrap, construindo previsores diversificados através de sub-amostras.

Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

APÊNDICE B – Métodos de reamostragem

Reamostragem são técnicas utilizadas para gerar variações de amostra de treino. Tal procedimento permite encontrar os hiperparâmetros de um algoritmo que atinge um desempenho otimizado. Além disso, permite até mesmo estimar incertezas para o treinamento dos modelos, os métodos mais comuns são o bootstrap e a validação cruzada.

B.1 Bootstrap

O bootstrap é de longe uma das técnicas mais utilizada de reamostragem. Nesse método, a amostra de treinamento é dividida em sub-amostras, onde as observações são aleatoriamente retiradas e substituídas por dados já contidos na amostra (Figura 93), permitindo gerar variações de uma mesma amostra. Assim, cada sub-amostra gera um estimador diferente, possibilitando estimar uma incerteza para a previsão de um modelo.

B.2 Validação cruzada K-fold

A técnica de reamostragem segmenta a amostra nos chamados folds (Figura 94), para cada diferente fold é realizado um treinamento e validado o modelo gerado, a validação final é a média dos erros dos previsores. A técnica também permite estimar a incerteza para as previsões e até mesmo encontrar hiperparâmetros.

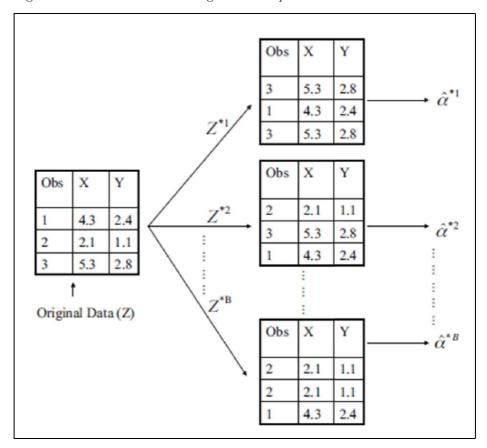


Figura 93 - Método de reamostragem bootstrap

Legenda: Uma representação de reamostragem bootstrap para uma amostra $(Z) \ {\rm com} \ 3 \ {\rm observações}, \ {\rm onde} \ {\rm a} \ {\rm reamostragem} \ {\rm de} \ bootstrap \ {\rm \acute{e}}$ realizada B vezes, cada sub-amostra gera um estimador diferente. Fonte: James $et. \ al., \ 2017, \ {\rm p.} \ 419.$ Adaptado pelo autor.

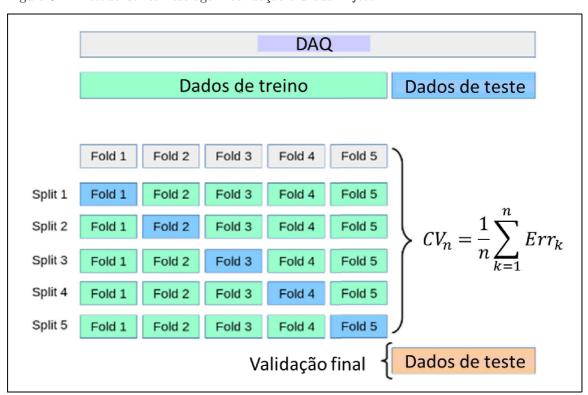


Figura 94 - Método de reamostragem Validação cruzada K-fold

Legenda: O dado coletado é dividido entre treinamento e teste, e é segmentado em folds. Para cada fold de cor azul é feito um treinamento e o desempenho é validado com uma amostra de teste. A técnica permite estimar a incerteza do modelo e até mesmo encontrar os hiperparâmetros do algoritmo que melhor se ajustam aos dados.

Fonte: Pedregosa et. al., 2011, p. 2825–2830. Adaptado pelo autor.

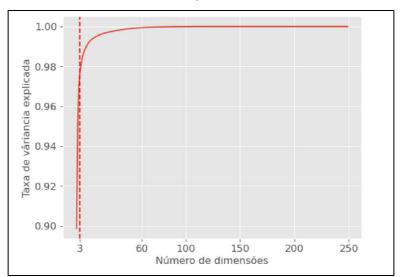
APÊNDICE C – Clusterização das formas de onda

A análise de clusterização foi realizada para cada tensão de coleta, nas secções seguintes são mostrados os resultados para cada amostra de voltagem.

C.1 FOs de sinais de múons cósmicos e ruído de corrente escura

A variância explicada para cada voltagem é mostrada na Figura 95 (1750 V), Figura 99 (1850 V) e Figura 103 (1950 V); o dendrograma gerado para cada coleta na Figura 96, Figura 100 e Figura 104, respectivamente; a clusterização de cada tensão são exibidas nas Figura 97, Figura 101 e Figura 105, e exemplos das formas de onda de cada clusterização na Figura 98, Figura 102 e Figura 106.

Figura 95 - Variância explicada em função das dimensões de FOs de sinais de múons cósmicos e ruídos de corrente escura de 1750V)



Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

Fonte: O autor, 2023.

C.2 FOs de sinais de múons cósmicos e de fundo

Quando se considera todas as possíveis fontes de ruídos, os *clusters* apresentam uma proximidade muito maior entre observações de sinal e de fundo. A taxa de variância explicada para cada voltagem é mostrado na Figura 107 (1550 V), Figura 111 (1600 V),

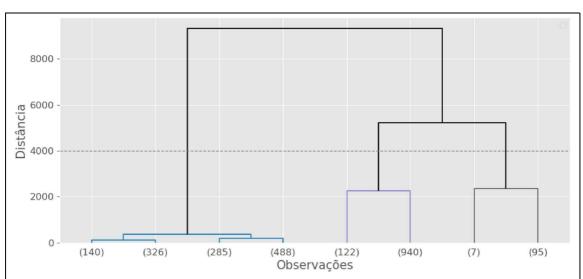


Figura 96 - Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente escura de $1750\mathrm{V}$

Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 4000. Assim, a amostra segmenta em 3 *clusters*.

Fonte: O autor, 2023.

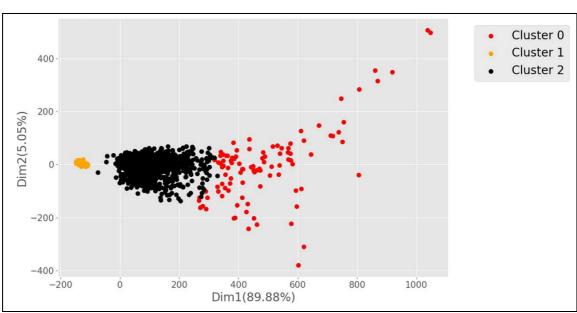


Figura 97 - Clusterização de FOs de sinais de mú
ons cósmicos e ruídos de corrente escura de $1750\mathrm{V}$

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 contém os sinais de múons cósmicos de maior amplitude e algumas saturações de escala de sinais, o cluster 1 aglomera os ruídos e o cluster 2 agrupa os sinais de múons cósmicos.

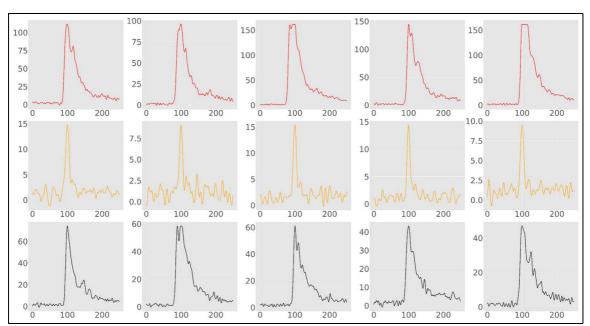


Figura 98 - Formas de onda da clusterização de FOs de sinais de mú
ons cósmicos e ruídos de corrente escura de $1750\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada *cluster* da Figura 97, com a suas respectivas cores.

Fonte: O autor, 2023.

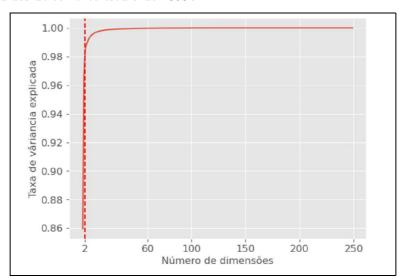


Figura 99 - Variância explicada em função das dimensões de FOs de sinais de múons cósmicos e ruídos de corrente escura de $1850\mathrm{V}$

Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

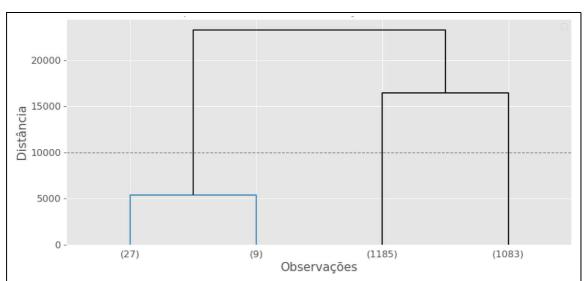


Figura 100 - Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente escura de $1850\mathrm{V}$

Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 10000. Assim, a amostra segmenta em 3 clusters.

Fonte: O autor, 2023.

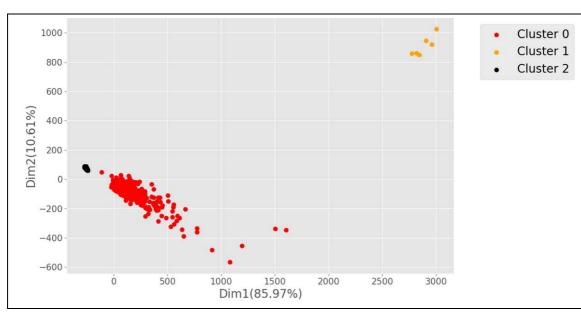


Figura 101 - Clusterização de FOs de sinais de mú
ons cósmicos e ruídos de corrente escura de $1850\mathrm{V}$

Legenda: Projeção das componentes do PCA, onde os dados separados são pela seguinte clusterização: o cluster 0 agrupa a maioria dos sinais, o cluster 1 contém saturação de escala de sinais, enquanto o cluster 2 aglomera os ruídos e alguns sinais.

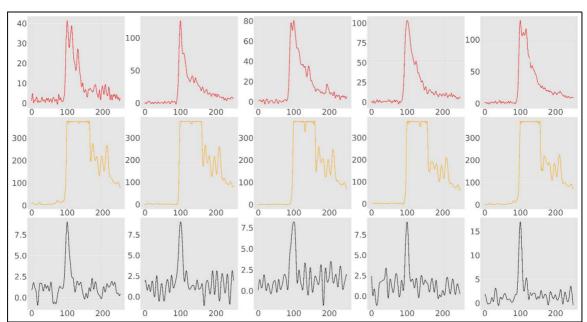


Figura 102 - Formas de onda da clusterização de FOs de sinais de múons cósmicos e ruídos de corrente escura de $1850\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada *cluster* da Figura 101 com a suas respectivas cores.

Fonte: O autor, 2023.

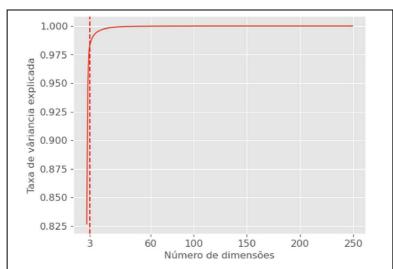


Figura 103 - Variância explicada em função das dimensões para 1950 V

Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

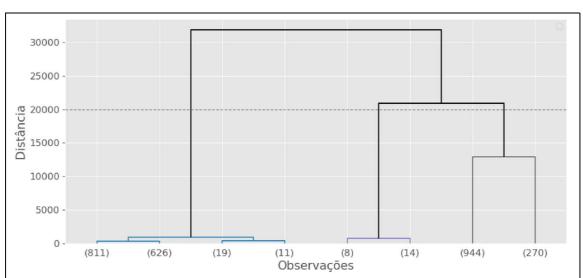


Figura 104 - Dendrograma de FOs de sinais de múons cósmicos e ruídos de corrente escura de $1950\mathrm{V}$

Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 20000. Assim, a amostra segmenta em 3 clusters.

Fonte: O autor, 2023.

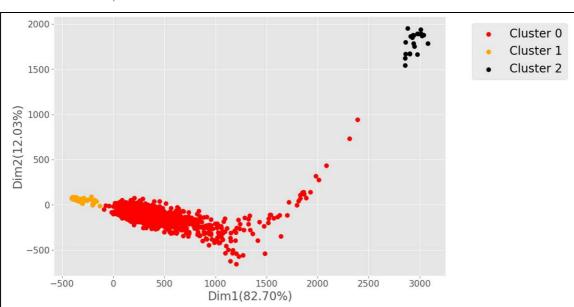


Figura 105 - Clusterização de FOs de sinais de múons cósmicos e ruídos de corrente escura de $1950\mathrm{V})$

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 contém os sinais de múons cósmicos, o cluster 1 agrupa os ruídos e alguns sinais e o cluster 2 aglomera saturação de escala de sinais.

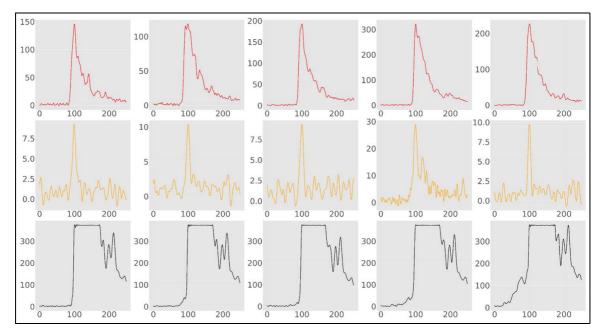


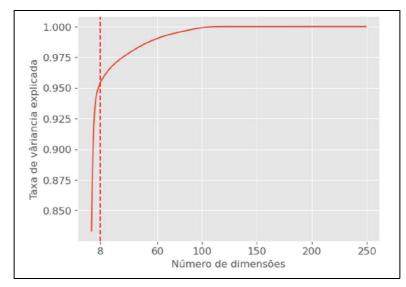
Figura 106 - Formas de onda da clusterização de FOs de sinais de múons cósmicos e ruídos de corrente escura de $1950\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada cluster da Figura 105, com a suas respectivas cores.

Fonte: O autor, 2023.

Figura 115 (1750 V), Figura 119 (1850 V) e Figura 123 (1950 V); o dendrograma gerado para cada coleta está na Figura 108, Figura 112, Figura 116, Figura 120, Figura 124, respectivamente; a clusterização de cada tensão são exibidas nas Figura 109, Figura 113, Figura 117, Figura 121 e Figura 125, e algumas formas de onda de cada clusterização: Figura 110, Figura 114, Figura 118, Figura 122 e Figura 126.

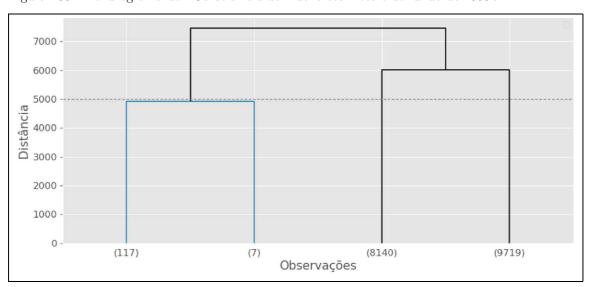
Figura 107 - Variância explicada em função das dimensões de FOs de sinais de múons cósmicos e de fundo de $1550\mathrm{V}$



Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

Fonte: O autor, 2023.

Figura 108 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de $1550\mathrm{V}$



Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 5000. Assim, a amostra se segmenta em 3 clusters.

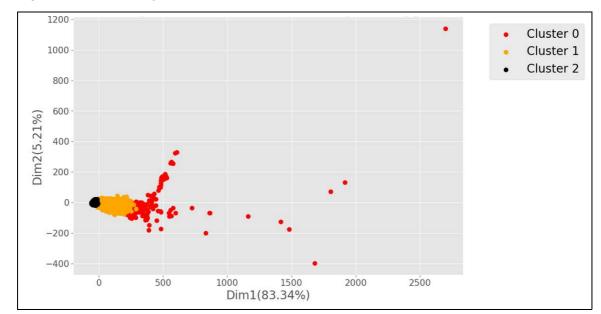


Figura 109 - Clusterização de FOs de sinais de múons cósmicos e de fundo de $1550\mathrm{V}$

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 agrupa os pulsos de escalas saturadas tanto de fundo quanto sinais, o cluster 1 aglomera sinais de múons cósmicos e uma concentração considerável de ruídos que apresentam pulsos de cintilação e o cluster 2 contém FOs de fundo.

Fonte: O autor, 2023.

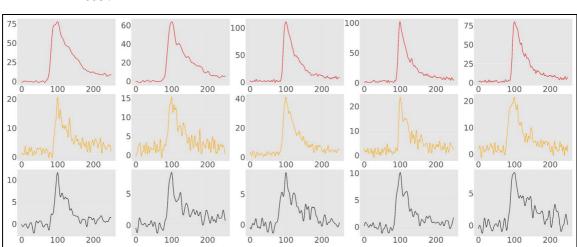
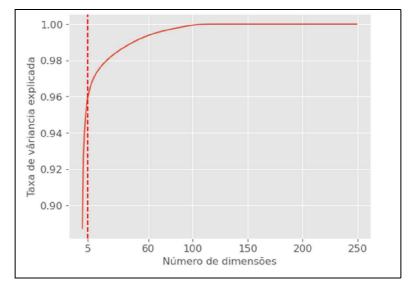


Figura 110 - Formas de onda da clusterização de FOs de sinais de múons cósmicos e de fundo de $1550\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada cluster da Figura 109, com a suas respectivas cores.

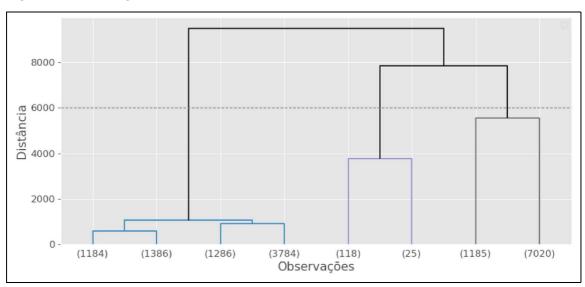
Figura 111 - Variância explicada em função das dimensões de FOs de sinais de múons cósmicos e de fundo de $1600\mathrm{V}$



Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

Fonte: O autor, 2023.

Figura 112 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de $1600\mathrm{V}$



Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 6000. Assim, a amostra se segmenta em 3 clusters.

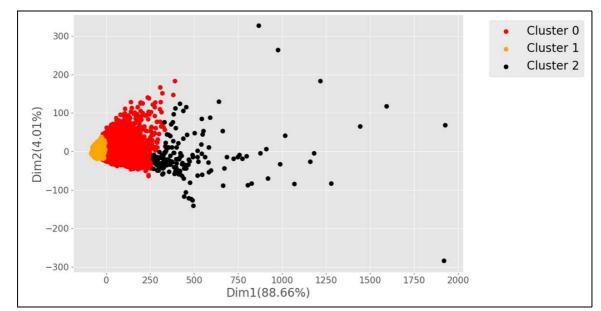


Figura 113 - Clusterização de FOs de sinais de múons cósmicos e de fundo de 1600V

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 agrupa os sinais de múons cósmicos e uma pequena concentração de ruídos que apresentam pulso de cintilação, o cluster 1 concentra a maior proporção de fundo e o cluster 2 contém pulsos com escalas saturadas tanto de fundo quanto sinais, também aglomera alguns sinais de alta amplitude.

Fonte: O autor, 2023.

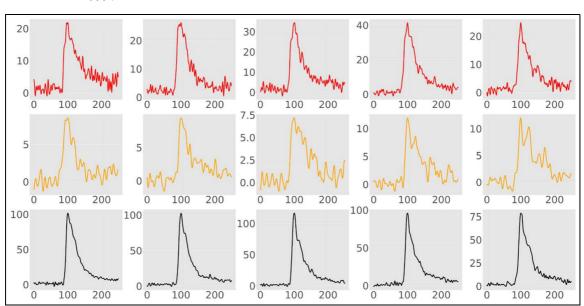
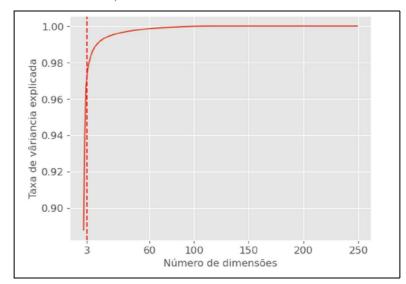


Figura 114 - Formas de onda da clusterização de FOs de sinais de múons cósmicos e de fundo de $1600\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada cluster da Figura 113, com a suas respectivas cores.

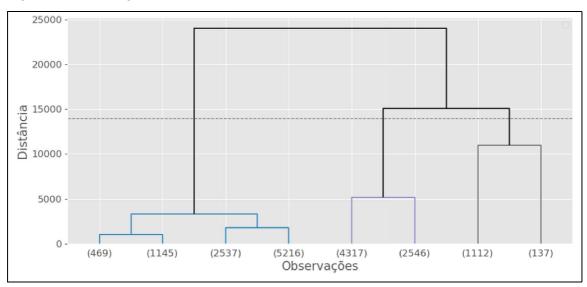
Figura 115 - Variância explicada em função das dimensões de FOs de sinais de múons cósmicos e de fundo de 1750V)



Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

Fonte: O autor, 2023.

Figura 116 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de $1750\mathrm{V}$



Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 14000. Assim, a amostra se segmenta em 3 clusters.

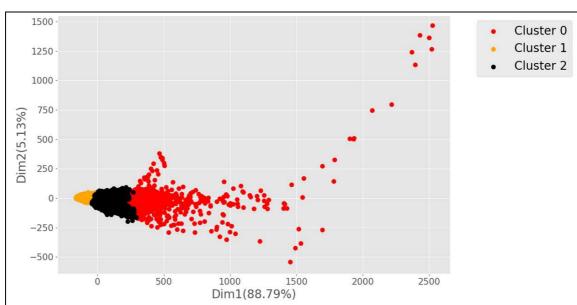


Figura 117 - Clusterização das formas de onda de FOs de sinais de mú
ons cósmicos e de fundo de $1750\mathrm{V}$

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 agrupa pulsos com escalas saturadas de sinais de múons cósmicos e sinais de múons cósmicos de alta amplitude, o cluster 1 contém FOs da maior parte por fundo e o cluster 2 aglomera os sinais de múons cósmicos, também concentra os pulsos de cintilação na amostra de fundo.

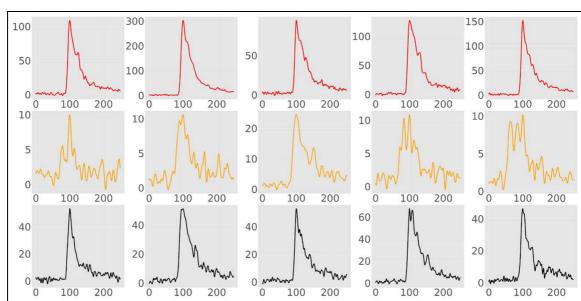


Figura 118 - Formas de onda da clusterização de FOs de sinais de múons cósmicos e de fundo de $1750\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada cluster da Figura 117, com a suas respectivas cores.

Fonte: O autor, 2023.

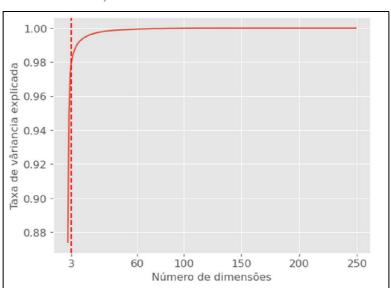


Figura 119 - Variância explicada em função das dimensões de FOs de sinais de múons cósmicos e de fundo de $1850\mathrm{V})$

Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

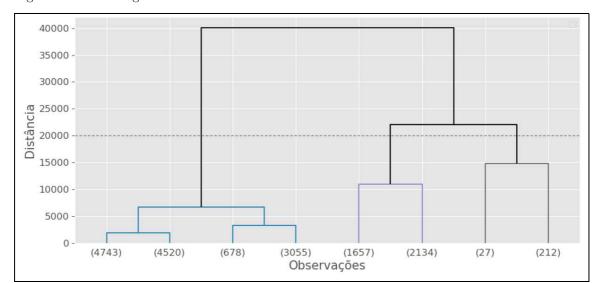


Figura 120 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de 1850V

Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 20000. Assim, a amostra se segmenta em 3 *clusters*.

Fonte: O autor, 2023.

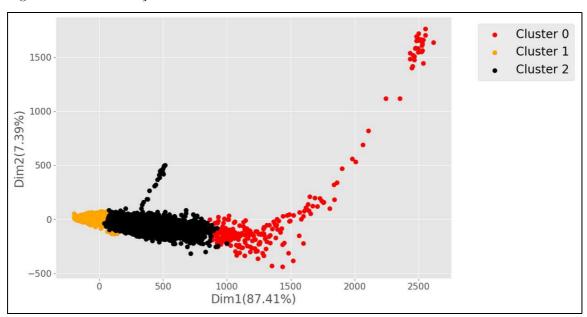


Figura 121 - Clusterização de FOs de sinais de múons cósmicos e de fundo de 1850V

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 agrupa os pulsos com escalas saturadas de sinais de múons cósmicos e sinais de múons cósmicos de alta amplitude, o cluster 1 contém a maior proporção de FOs de fundo e o cluster 2 aglomera os sinais de múons cósmicos e concentra os pulsos de cintilação na amostra de fundo.

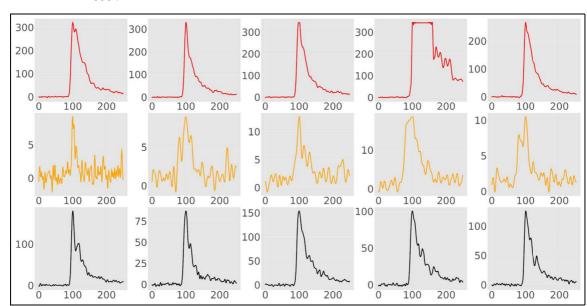


Figura 122 - Formas de onda da clusterização de FOs de sinais de múons cósmicos e de fundo de $1850\mathrm{V}$

Legenda: Algumas formas de onda contidas em cada cluster da Figura 121, com a suas respectivas cores.

Fonte: O autor, 2023.

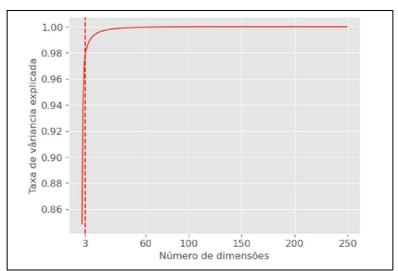


Figura 123 - Variância explicada em função das dimensões de FOs de sinais de múons cósmicos e de fundo de $1950\mathrm{V}$

Legenda: A linha tracejada marca o número de componentes que atinge mais de 95% da informação preservada.

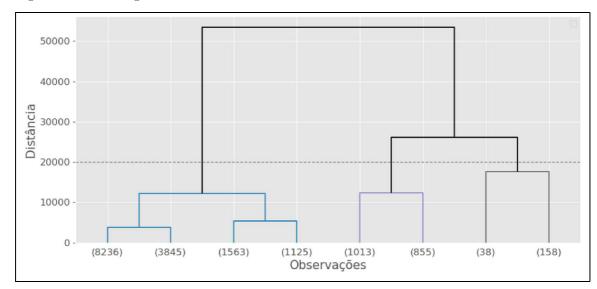


Figura 124 - Dendrograma de FOs de sinais de múons cósmicos e de fundo de 1950V

Legenda: Dendrograma de FOs, onde a separação é definida por um limiar de 20000. Assim, a amostra se segmenta em 3 *clusters*.

Fonte: O autor, 2023.

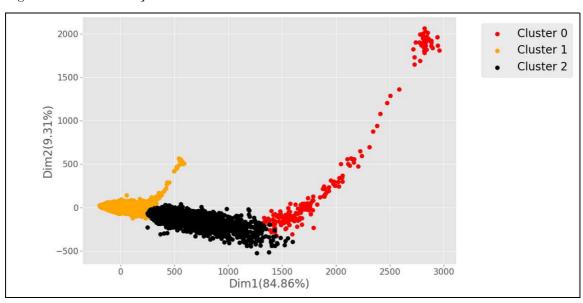
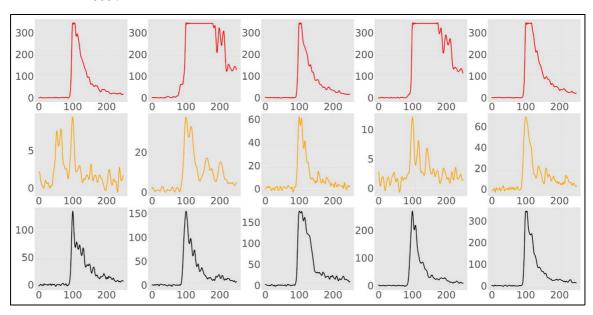


Figura 125 - Clusterização de FOs de sinais de múons cósmicos e de fundo de 1950V

Legenda: Projeção das componentes do PCA, onde os dados são separados pela seguinte clusterização: o cluster 0 agrupa pulsos com escalas saturadas de sinais de múons cósmicos e sinais de múons cósmicos de alta amplitude, o cluster 1 contém a maior proporção de FOs de fundo e o cluster 2 aglomera os sinais de múons cósmicos e concentra os ruídos que apresentam pulso de cintilação.

Figura 126 - Formas de onda da clusterização de FOs de sinais de múons cósmicos e de fundo de $1950\mathrm{V}$



Legenda: Algumas formas de onda contidas em cada cluster da Figura 125, com a suas respectivas cores.

APÊNDICE D – Tratamento Bayesiano para estimar as incertezas das eficiências

O tratamento estatístico utilizado para estimar as incertezas das previsões dos modelos foram baseadas no artigo *Treatment of Errors in Efficiency Calculations* (Ullrich; Xu, 2008). A partir de uma derivação da função densidade de probabilidade é possível estimar os erros estatísticos de uma eficiência.

D.1 Derivação da função densidade de probabilidade

A relação entre as observações classificadas corretamente sobre o total de observáveis, pode ser definida como uma eficiência da previsão.

Do ponto de vista estatístico, as previsões TP e TN são consideradas como k_i , ou seja, eventos que satisfazem a condição de classificação correta pelo modelo e as observações a serem classificadas como n_i . A incerteza da eficiência (ϵ) é estimada a partir de um cálculo baseado no teorema de Bayes.

Por simplicidade, será considerado um único evento, onde k é o total de eventos que são classificados corretamente e n é o total de eventos a serem classificados. O erro a ser avaliado é apenas estatístico, ou seja, ignora-se todos o erros sistemáticos.

Para calcular a função densidade de probabilidade, usa-se o teorema Bayesiano com a abordagem da Equação 55, onde C é a constante de normalização a ser determinada e $P(\epsilon,n)$ é a probabilidade de atribuir um eficiência verdadeira antes de considerar os dados.

$$P(k;\epsilon,n) = \frac{P(k;\epsilon,n)P(\epsilon;n)}{C}$$
(55)

A partir de um dado n e o fato de estar lidando com um processo binomial (Equação 56), se diz que ϵ deve estar incluso no intervalo $0 \le \epsilon \ge 1$, caso contrário, não teria motivo para ser considerado (Equação 57). Portanto é razoável tomar um n independente (Equação 57).

$$P(k;\epsilon,n) = \binom{n}{k} \epsilon^k (1-\epsilon)^{n-k} \tag{56}$$

$$P(\epsilon; n) = \begin{cases} 1 & \text{se } 0 \le \epsilon \ge 1\\ 0 & \text{se } \epsilon < 0 \lor \epsilon > 1 \end{cases}$$
(57)

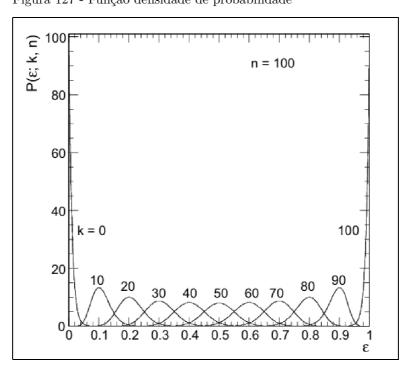
A constante de normalização C pode ser determinada resolvendo a integral da Equação 58, a solução é encontrada a partir de uma definição da função Beta. Assim, o C e a função de densidade de probabilidade são definidos pela Equação 59 e Equação 60, respectivamente. Na Figura 127 é mostrada a função densidade de probabilidade para diferentes valores de k com um n=100.

$$\int_{-\infty}^{\infty} P(\epsilon; k, n) d\epsilon = \frac{1}{C} \binom{n}{k} \int_{0}^{1} \epsilon^{k} (1 - \epsilon)^{n - k} d\epsilon = 1$$
(58)

$$C = \frac{n!}{(n+1)!} = \frac{1}{n+1} \tag{59}$$

$$P(\epsilon; k, n) = \frac{(n+1)!}{k!(n-k)!} \epsilon^k (1-\epsilon)^{n-k}$$
(60)





Legenda: A função densidade de probabilidade $P(\epsilon;n,k)$, para n=100 e k=0,10...100.

Fonte: Ullrich; Xu, 2008, p. 8.

D.2 Características da função densidade de probabilidade

Dado a sua forma analítica é possível calcular os momentos da distribuição, para a média (Equação 61) e variância (Equação 62), observa-se que ambas equações funcionam para os casos limites, em que k=0 levaria a um eficiência não física e para k=n, gerando um incerteza nula.

$$\bar{\epsilon} = \int_0^1 \epsilon P(\epsilon, k, n) d\epsilon = \frac{k+1}{n+2} \tag{61}$$

$$V(\epsilon) = \bar{\epsilon}^2 - \bar{\epsilon}^2 = \frac{(k+1)(k+2)}{(n+2)(n+3)} - \frac{(k+1)^2}{(n+2)^2}$$
(62)

A Figura 128 mostra a distribuição da função densidade de probabilidade para o caso de n=10 e k=8, as linhas verticais sólida e tracejada indicam a média e o valor modal, respectivamente. A região $\bar{\epsilon} \pm \sigma$, onde $\sigma_e = \sqrt{V(\epsilon)}$ (Equação 53), que corresponde a um nível de confiança de 67,3%.

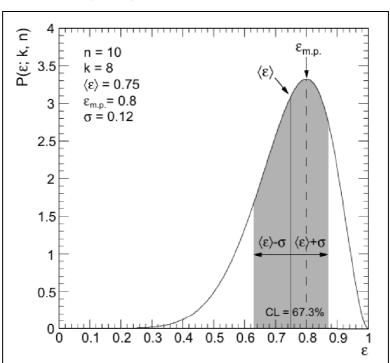


Figura 128 - Eficiência da função densidade de probabilidade $P(\epsilon;8,10)$

Legenda: Eficiência da função densidade de probabilidade $p(\epsilon;8,10), \text{ onde a primeira linha vertical sólida indica}$ o valor médio e a linha tracejada indica o valor mais $\text{provável } (\epsilon_{m.p.})$

Fonte: Ullrich; Xu, 2008, p. 8.

APÊNDICE E – Técnicas de regularização

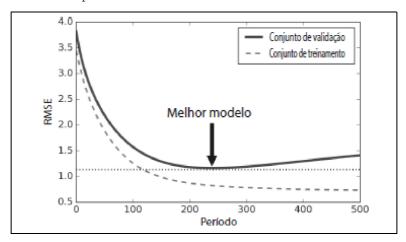
Técnicas de regularização são essenciais para treinamento de rede neurais. Tais procedimentos auxiliam na redução de erros de sobreajuste, dentre as mais utilizadas estão as seguintes:

- Parada antecipada.
- Dropout,
- Regularização L1 e L2.
- Diminuir a complexidade da sua rede.

E.1 Parada antecipada

Uma forma prática de reduzir o sobreajuste é o uso da parada antecipada, onde o treinamento é interrompido antes dos dados de validação começarem a divergir (Figura 129).

Figura 129 - Parada antecipada



Legenda: O treinamento é interrompido antes do modelo começar a divergir.

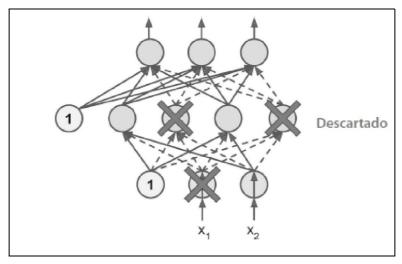
Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

E.2 Dropout

Uma das regularizações que provou ser altamente bem sucedida é o *dropout* (descarte). Nesta técnica as conexões de um neurônio tem uma probabilidade de aleatoria-

mente serem desligados em uma interação (Figura 130), porém podem ser reativados na próxima época do treinamento. A regularização força os neurônios a se coadaptarem com os desligamentos dos neurônios vizinhos, permitindo construir uma rede com unidades mais independentes dessa forma.

Figura 130 - Dropout



Legenda: Regularização por meio de *dropout*, onde neurônios em cada camada tem a probabilidade de serem desligados aleatoriamente, cortando suas conexões.

Fonte: Géron, 2019, p. 549. Adaptado pelo autor.

E.3 Regularização L1 e L2

As regularizações L_1 e L_2 (Equação 63 e Equação 64, respectivamente) possuem uma alta eficiência para redução do sobreajuste. Na L_1 , adiciona-se um custo proporcional à norma do peso, exibindo um modelo esparso (poucos pesos diferentes de zero). A L_2 apenas força o algoritmo a se ajustar aos dados, mas também mantém os pesos reduzidos ao mínimo, adicionando um custo proporcional a w^2 .

$$L_1 = (wx + b - y)^2 + \lambda |w| \tag{63}$$

$$L_2 = (wx + b - y)^2 + \lambda w^2 \tag{64}$$

E.4 Diminuição da complexidade da rede

Uma das possíveis causas do sobreajuste é o uso de uma rede muito complexa para os dados que se deseja ajustar. Uma abordagem prática para reduzir o sobreajuste envolve a diminuição dos parâmetros da rede, seja reduzindo o número de neurônios ou de camadas.