



COPPE/UFRJ

ANÁLISE NÃO-LINEAR DE COMPONENTES INDEPENDENTES PARA
UMA FILTRAGEM ONLINE BASEADA EM CALORIMETRIA DE ALTA
ENERGIA E COM FINA SEGMENTAÇÃO

Eduardo Furtado de Simas Filho.

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

Orientadores: José Manoel de Seixas
Luiz Pereira Calôba

Rio de Janeiro
Dezembro de 2010

ANÁLISE NÃO-LINEAR DE COMPONENTES INDEPENDENTES PARA
UMA FILTRAGEM ONLINE BASEADA EM CALORIMETRIA DE ALTA
ENERGIA E COM FINA SEGMENTAÇÃO

Eduardo Furtado de Simas Filho.

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

Examinada por:

Prof. José Manoel de Seixas, D.Sc.

Prof. Luiz Pereira Calôba, Dr.Ing.

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Leandro Salazar de Paula, D.Sc.

Prof. Marley Maria Bernardes Rebuzzi Vellasco, Ph.D.

Prof. Guilherme de Alencar Barreto, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

DEZEMBRO DE 2010

Aos meus amores Cléa e Letícia.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro concedido em parte do tempo no qual este trabalho foi desenvolvido.

Aos meus orientadores, os professores José Manoel de Seixas e Luiz Pereira Calôba, pelo apoio e motivação, sem os quais este trabalho não existiria.

Aos colegas, professores e funcionários do Laboratório de Processamento de Sinais (LPS/COPPE/UFRJ) pela inestimável ajuda na condução do trabalho, em especial aos amigos Natanael Moura, Rodrigo Torres, Danilo Enoque e Werner Freund.

Ao Instituto Federal de Educação, Ciência e Tecnologia da Bahia, pelo apoio concedido para a finalização deste trabalho.

À minha família e aos meus amigos (do Roger) por suportarem minha ausência durante este período e por me incentivarem sempre a continuar, especialmente ao amigo Frederico Wegelin, pela companhia e apoio durante o tempo em que morei no Rio de Janeiro.

Em especial, dedico esse trabalho a minha esposa Cléa e minha filha Letícia pelo amor e apoio em todos os momentos.

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

ANÁLISE NÃO-LINEAR DE COMPONENTES INDEPENDENTES PARA
UMA FILTRAGEM ONLINE BASEADA EM CALORIMETRIA DE ALTA
ENERGIA E COM FINA SEGMENTAÇÃO

Eduardo Furtado de Simas Filho.

Dezembro/2010

Orientadores: José Manoel de Seixas
Luiz Pereira Calôba

Programa: Engenharia Elétrica

O ATLAS é o maior detector do acelerador de partículas LHC. Nas colisões, uma enorme quantidade de informação é produzida, porém, apenas uma pequena parcela é importante para a caracterização dos fenômenos físicos de interesse, o que exige um eficiente sistema para detecção (filtragem) *online* de eventos. Os elétrons são partículas extremamente importantes para o LHC, e aparecem mascarados por um intenso ruído de fundo composto de jatos hadrônicos, pois esses últimos podem apresentar perfil de deposição de energia nos calorímetros semelhante ao de elétrons (os calorímetros são medidores de energia com fina segmentação e no ATLAS são divididos em sete camadas, totalizando mais de 100.000 sensores). Neste trabalho, é proposta a utilização do modelo não-linear da análise de componentes independentes no processo de extração de características, visando otimizar o desempenho do sistema neural de filtragem *online* de elétrons no ATLAS (*Neural Ringer*). Para explorar toda a segmentação e granularidade disponíveis, a extração de características foi efetuada separadamente para cada camada do calorímetro. Diversos algoritmos foram utilizados na estimação dos componentes independentes. Através da abordagem proposta, foi possível alcançar alta eficiência de discriminação, gerando dados mais limpos para a análise *offline*.

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

INDEPENDENT COMPONENT ANALYSIS FOR ONLINE FILTERING
BASED ON HIGH-ENERGY AND HIGHLY SEGMENTED CALORIMETRY

Eduardo Furtado de Simas Filho.

December/2010

Advisors: José Manoel de Seixas

Luiz Pereira Calôba

Department: Electrical Engineering

ATLAS is the largest detector of the Large Hadron Collider (LHC). A large amount of information is produced in the collisions, but only a small fraction is important for characterizing interesting physics, demanding an efficient online event detection (trigger) system. Electrons are extremely important for the LHC and are immerse in a huge background noise of hadronic jets, as these last signatures present calorimeter energy deposition profile similar to electrons one (calorimeters are highly segmented energy measurement systems and in ATLAS are split into seven layers and more than 100,000 sensing elements). In this work, the nonlinear independent component analysis (NLICA) model is proposed to extract features aiming at the optimization of the ATLAS electron online neural discriminator (Neural Ringer). In order to cope with the full segmentation and granularity available, feature extraction was performed at layer-level. Different algorithms were used for the independent components estimation. Through the proposed approach, higher discrimination efficiency was achieved, producing cleaner data for offline analysis.

Sumário

Lista de Figuras	xiii
Lista de Tabelas	xxi
Símbolos e Abreviaturas	xxv
1 Introdução	1
1.1 Contexto	1
1.2 Motivação	2
1.3 Trabalhos Anteriores	4
1.4 Objetivos	5
1.5 Metodologia	6
1.6 Conteúdo do Trabalho	7
2 Física de Partículas e o Detector ATLAS	8
2.1 Panorama Geral da Física de Partículas Elementares	8
2.2 O Acelerador LHC	13
2.3 Características Gerais do Detector ATLAS	15
2.4 Principais Objetos de Interesse no ATLAS	19
2.5 O Sistema de Calorimetria do ATLAS	21
2.5.1 Breve Introdução à Calorimetria	21
2.5.2 Características dos Calorímetros do ATLAS	23
2.5.3 Desempenho Esperado dos Calorímetros	27
3 Filtragem Online no ATLAS	30
3.1 Introdução aos Sistemas de Filtragem em HEP	30
3.1.1 Aplicações de Redes Neurais Artificiais	32

3.2	O Sistema de Filtragem <i>Online</i> do ATLAS	36
3.2.1	Primeiro Nível de Filtragem	39
3.2.2	Filtragem de Alto Nível	41
3.2.3	Plataforma de <i>Software</i> do Sistema de Filtragem	45
4	Deteção de Elétrons a partir de Informações de Calorimetria no ATLAS	47
4.1	Filtragem de Elétrons no L1	48
4.2	Filtragem de Elétrons no L2 - Algoritmo T2Calo	50
4.3	Neural Ringer - Alternativa para Filtragem de Elétrons no L2	53
4.3.1	Extração de Características - Anelamento	53
4.3.2	Normalização	56
4.3.3	Teste de Hipóteses - Classificador Neural	58
4.3.4	Tempo de Execução	59
4.4	Extensões ao Neural Ringer	61
4.4.1	Importância do Pré-Processamento para Classificadores Neurais	62
4.4.2	Pré-Processamento Linear	63
4.4.3	Pré-Processamento Não-linear	64
5	Análise de Componentes Independentes	65
5.1	Modelo Linear da ICA	65
5.2	ICA Não-Linear	68
5.2.1	Unicidade da Solução em NLICA	71
5.2.2	Modelos com Restrições Estruturais	71
5.2.3	Algoritmos sem Restrições Estruturais	73
5.2.4	ICA Local	78
5.3	Aplicações de ICA e NLICA para Extração de Características	79
5.4	Aplicações em Física de Altas Energias e Áreas Correlatas	82
5.5	Utilizando Informação das Classes na Estimação dos Componentes Independentes	86
5.5.1	Componentes Principais de Discriminação	87
5.5.2	Utilizando os Rótulos de Classe como Sinais de Entrada para os Algoritmos de ICA	88

5.5.3	Proposta de Algoritmo para Estimação de Componentes Independentes e Discriminantes	89
6	Metodologia Proposta e Bases de Dados Utilizadas	93
6.1	Parâmetros de Avaliação do Desempenho	94
6.2	Extração de Características	96
6.2.1	Algoritmos de Extração de Características	98
6.3	Classificação	106
6.3.1	Motivação para o Uso de Classificadores Segmentados	106
6.3.2	Combinação de Múltiplos Classificadores	108
6.4	Especificações de Treinamento	110
6.5	Bases de Dados	110
6.5.1	Dados Simulados	111
6.5.2	Dados Experimentais	116
7	Resultados - Dados Simulados	128
7.1	Conjunto E10	128
7.1.1	Resultados com os Discriminadores Existentes	129
7.1.2	Resultados com Pré-Processamento por NLICA	137
7.1.3	Estudo da Relevância por Camada	156
7.1.4	Comparação com Discriminadores Lineares	162
7.1.5	Comentários e Discussão	163
7.2	Conjunto E15i	165
7.2.1	Resultados com os Discriminadores existentes	166
7.2.2	Resultados com Pré-Processamento por NLICA	167
7.2.3	Estudo da Relevância por Camada	180
7.2.4	Comentários e Discussão	182
8	Resultados - Dados Experimentais	184
8.1	Eficiência dos Sistemas de Classificação Propostos na Rejeição de Raios Cósmicos	184
8.2	Colisões do LHC	187
8.2.1	Resultados - Anéis	189
8.2.2	Resultados - ICA e NLICA	195

8.2.3	Estudo da relevância por camada	200
8.2.4	Comentários e Discussão	202
8.3	Estimativa do custo computacional dos algoritmos propostos	204
8.3.1	Abordagens segmentadas	206
9	Conclusões	209
9.1	Trabalhos Futuros	212
Referências Bibliográficas		213
A	Aspectos Teóricos das Técnicas de Extração de Características	239
A.1	Mapas Auto-Organizáveis	239
A.1.1	Quantização Vetorial por Aprendizado	241
A.1.2	Classificação a Partir do Mapa de Características	243
A.2	Análise de Componentes Principais	244
A.3	Técnicas de Pré-Processamento - Compactação	244
A.3.1	Análise de Componentes Principais	244
A.3.2	Redução de Dimensão	246
A.4	Análise de Componentes Independentes	247
A.4.1	Princípios de Estimação dos Componentes Independentes	247
A.4.2	Pré-Processamento dos Sinais para ICA	250
A.4.3	Principais Algoritmos para ICA	251
A.4.4	Extensões ao Modelo Básico de ICA	254
A.5	ICA Não-Linear	256
A.5.1	Algoritmo Taleb-Jutten para o Modelo PNL	257
B	Conceitos Fundamentais em Classificação de Sinais	259
B.1	Teste de Hipóteses	259
B.2	Critério de Bayes	260
B.3	Filtros Casados	261
B.4	Discriminante Linear de Fisher	263
B.5	Classificadores Neurais	263
C	Algoritmos Genéticos	266
C.1	Algoritmo Genético como Método de Otimização	266

C.2	Estrutura de um Algoritmo Genético	267
C.2.1	Conceitos Principais	267
C.2.2	Escalonamento de Aptidão	269
C.2.3	Implementação de um Algoritmo Genético	271
C.2.4	O Algoritmo Genético Utilizado	271
D	Produção Científica	274
D.1	Publicações em periódicos	274
D.2	Capítulo de Livro	276
D.3	Artigos em Conferências	277
D.4	Resumos em Conferências	281

Listas de Figuras

2.1	Diagrama do Modelo Padrão, mostrando as partículas elementares incluindo o, ainda não confirmado, bóson de Higgs.	10
2.2	Mapa de localização dos detectores do LHC.	14
2.3	Diagrama esquemático do ATLAS.	16
2.4	Eixo de coordenadas do ATLAS.	16
2.5	Cortes (a) transversal e (b) axial do ATLAS.	18
2.6	Disposição em camadas dos calorímetros do ATLAS.	24
2.7	Granularidade e profundidade das camadas do calorímetro eletromagnético.	26
2.8	Energia perdida por elétrons antes do calorímetro.	27
2.9	Energia normalizada média depositada nas seções eletromagnética e hadrônica em função de η para elétrons.	28
2.10	Erro relativo do calorímetro na medição da energia de elétrons para diferentes valores de (a) energia e (b) η .	29
3.1	Diagrama em blocos do sistema de filtragem genérico para um experimento de HEP.	31
3.2	Diagrama em blocos do sistema de filtragem e aquisição de dados do ATLAS.	37
3.3	Diagrama em blocos do primeiro nível de filtragem.	41
3.4	Diagrama em blocos do segundo nível de filtragem.	44
4.1	Janela deslizante analisada pelo L1 no calorímetro eletromagnético.	48
4.2	Processo de identificação de elétrons no L2.	50
4.3	Distribuição de elétrons e jatos quanto à razão $\frac{E_{HAD}}{E_{EM}}$ para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.	52

4.4	Distribuição de elétrons e jatos quanto à razão de forma ($R_{shape} = \frac{E_{3\times 7}}{E_{7\times 7}}$) para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.	52
4.5	Fluxo de processamento do <i>Neural Ringer</i>	53
4.6	Diagrama do processo de construção dos anéis.	55
4.7	Sinais típicos medidos na camada EM2 (acima) e os respectivos sinais em anéis (abaixo), respectivamente para um elétron (esquerda) e um jato (direita).	56
4.8	Sinais em anéis para (a) elétron típico, (b) jato típico e (c) jato com perfil semelhante ao de elétrons.	57
4.9	Tempos de processamento considerando todas as etapas (esquerda) e somente as etapas específicas do <i>Neural Ringer</i> (direita).	61
4.10	Fluxo de processamento das extensões ao <i>Neural Ringer</i>	61
5.1	Diagrama do <i>cocktail party problem</i>	68
5.2	Sinais (a) fonte, (b) observados e (c) recuperados através da ICA. . .	69
5.3	Diagrama do modelo de mistura PNL.	72
5.4	Diagrama do modelo PNL-L.	73
5.5	Diagrama do modelo da Mono não-linearidade.	74
5.6	NLICA a partir de SOM.	74
5.7	Diagrama do algoritmo MISEP.	77
5.8	Diagrama do modelo da ICA local.	78
5.9	Modelos neurais para estimar (a) a primeira e (b) a k-ésima PCD. .	88
5.10	Diagramas de (a) treinamento e (b) operação dos algoritmos de ICA/NLICA utilizando informação das classes na entrada.	89
5.11	Modelo Pós Não-linear modificado.	90
5.12	Procedimento de treinamento para o modelo pós não-linear modificado.	90
6.1	Histogramas das saídas de dois classificadores distintos, Exemplo 1 à esquerda e Exemplo 2 à direita.	95
6.2	Curvas ROC (esquerda) e SP (direita) para dois classificadores distintos.	95
6.3	Processo de extração de características no modo não-segmentado. .	97
6.4	Processo de extração de características no modo segmentado.	97

6.5	Diagrama do discriminador baseado em ICA Local.	101
6.6	Diagramas dos processos de treinamento dos mapas auto-organizáveis utilizando informação a respeito das classes através de LVQ (es- querda) e rótulos de classe como entradas (direita).	102
6.7	Diagrama do discriminador baseado no modelo PNL da NLICA. . . .	105
6.8	Decisão utilizando classificador global.	106
6.9	Decisão utilizando classificadores segmentados.	107
6.10	Modelo de rede combinadora para os classificadores segmentados. . .	109
6.11	Distribuição em energia, η e ϕ das assinaturas de elétrons (esquerda) e jatos (direita) aprovadas pelos cortes de primeiro nível tipo E10 e E15i.	114
6.12	Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E10.	115
6.13	Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E15.	115
6.14	Correlação (acima) e informação mútua (abaixo) entre os 100 anéis. .	117
6.15	Visualização de um evento do LHC candidato ao decaimento de um bóson Z em dois elétrons (elétron e anti-elétron).	118
6.16	Histogramas em energia η e ϕ dos eventos de raios cósmicos.	119
6.17	Exemplo de evento fantasma em $(\eta; \phi) = (-0,3; 1,6)$	120
6.18	Taxas de aceitação de assinaturas eletromagnéticas noL1 em função da luminosidade.	121
6.19	Eficiência dos cortes do <i>offline</i> em função de η e da energia.	123
6.20	Histogramas em energia η e ϕ dos eventos de colisões do LHC. . . .	125
6.21	Sinais experimentais médios de “elétrons” e “jatos”.	126
7.1	Eficiência (esquerda) e falso alarme (direita) em energia, η e ϕ para o <i>Neural Ringer</i> e o T2Calo.	131
7.2	Contagem (%) das intersecções acerto×erro dos discriminadores <i>Neu- ral Ringer</i> e T2Calo para elétrons (esquerda) e jatos (direita). . . .	132
7.3	Histogramas da contagem (em %) em energia (esquerda) e η (direita) para assinaturas de elétrons e jatos de acordo com as decisões dos discriminadores T2Calo e <i>Neural Ringer</i>	133

7.4	Comparação (<i>Neural Ringer</i> ×T2Calo) dos eventos médios para elétrons (acima) e jatos (abaixo).	134
7.5	Curvas ROC para os discriminadores <i>Neural Ringer</i> e ICA+MLP. . .	135
7.6	Dispersão das saídas dos discriminadores para elétrons (esquerda) e jatos (direita).	136
7.7	Contagem (%) das intersecções acerto×erro dos discriminadores neurais operando diretamente sobre os anéis (<i>Ringer</i>) e com pré-processamento por ICA para elétrons (esquerda) e jatos (direita). . .	136
7.8	Eficiência (esquerda) e falso alarme (direita) em energia e η para o discriminador neural operando diretamente sobre os anéis (<i>Ringer</i>) e com pré-processamento por ICA.	137
7.9	Comparativo de desempenho entre o discriminador neural operando diretamente sobre os anéis (<i>Ringer</i>) e com pré-processamento por ICA para elétrons (esquerda) e jatos (direita).	138
7.10	Probabilidade de agrupamento nos <i>clusters</i>	139
7.11	Distribuições em energia e η dos eventos nos <i>clusters</i> 1 (acima) e 2 (abaixo).	140
7.12	Eventos médios de elétrons e jatos calculados em cada <i>cluster</i> , comparados aos perfis médios para todo o conjunto E10.	141
7.13	Curvas ROC obtidas para os classificadores neurais nos diferentes <i>clusters</i>	142
7.14	Variação do máximo SP com o número de neurônios ocultos.	142
7.15	Diagrama do discriminador baseado em ICA Local.	143
7.16	Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores <i>Neural Ringer</i> ICA+MLP e Local ICA+MLP.	144
7.17	Comparativo de eficiência entre os discriminadores <i>Neural Ringer</i> ICA+MLP e Local ICA+MLP.	146
7.18	Erro de reconstrução médio e SP calculados variando-se o número de neurônios do mapa (para mapas não-segmentados treinados de modo não-supervisionado).	147

7.19	Erro de reconstrução médio e SP calculados variando-se o número de neurônios do mapa (para mapas segmentados treinados de modo supervisionado)	148
7.20	Probabilidade de ativação dos neurônios antes (esquerda) e depois da LVQ (direita) num mapa segmentado treinado para a camada E2.	149
7.21	Curvas ROC obtidas para os diferentes modos de treinamento do SOM após a classificação por uma rede MLP.	150
7.22	Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores <i>Neural Ringer ICA+MLP</i> e <i>SOM+MLP</i>	151
7.23	Visualização 2-D de eventos de elétron (esquerda) e jato (direita) típicos através do SOM.	153
7.24	Visualização 3-D de eventos de elétron (esquerda) e jato (direita) típicos através do SOM.	153
7.25	Curvas ROC para os discriminadores baseados no modelo PNL.	156
7.26	Desempenho de classificação obtido quando as camadas são substituídas pelo seu valor médio.	158
7.27	Desempenho de classificação obtido re-treinando o classificador neural retirando-se parte da informação.	158
7.28	Desempenho de classificação obtido por classificadores treinados com a informação de apenas uma das camadas em termos das curvas ROC (esquerda) e do máximo SP (direita), onde a linha tracejada vertical indica o desempenho obtido pelo discriminador neural operando sobre todas as camadas.	159
7.29	Gráficos de correlação entre a saída do classificador treinado para a camada E1 e para as demais camadas, para assinaturas de elétrons (acima) e jatos (abaixo).	160
7.30	Curvas ROC obtidas para os diversos métodos de combinação de classificadores.	161
7.31	Variação do máximo SP com o número de neurônios ocultos para a rede combinadora (esquerda) e máximo SP obtido através da rede combinadora reirando-se, de modo segmentado, parte da informação. .	161
7.32	Eficiência (acima) e falso alarme (abaixo) em energia e η	162

7.33 Curvas ROC para discriminadores lineares (<i>Fisher</i>) e não-lineares (MLP)	163
7.34 Histogramas do terceiro anel da camada E2 (esquerda) e de um dos componentes independentes (ICA linear) estimados para a mesma camada (direita).	164
7.35 Gráficos de dispersão entre componentes independentes estimados por ICA linear (esquerda) e pelo modelo PNL (direita).	164
7.36 Comparativo da eficiência em energia (esquerda) e η (direita) para o T2Calo e o discriminador baseado no pré-processamento em ICA com compactação por PCD.	167
7.37 Divisão dos eventos de elétrons e jatos nos dois agrupamentos.	168
7.38 Distribuições em energia e η dos eventos nos <i>clusters</i> 1 (acima) e 2 (abaixo).	169
7.39 Eventos médios de elétrons e jatos nos <i>clusters</i> 1 e 2 comparados com os eventos médios considerando todo o conjunto de assinaturas E15i.	170
7.40 Variação do máximo SP com o número de neurônios ocultos.	171
7.41 Comparativo de eficiência a partir das Curvas ROC.	171
7.42 Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores <i>Neural Ringer</i> e Local ICA+MLP.	172
7.43 Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores <i>Neural Ringer</i> e Local ICA+MLP.	173
7.44 Erro quadrático médio (esquerda) e SP (direita) calculados variando-se as dimensões dos mapas, acima para o modo não-segmentado e abaixo para o segmentado.	174
7.45 Curvas ROC para as diversas abordagens do SOM (conjunto E15i).	175
7.46 Eficiência (acima) e falso alarme (abaixo) em energia e η	176
7.47 Distribuições em η e energia dos eventos de elétrons (esquerda) e jatos (direita) classificados corretamente pelo discriminador SOM+MLP e incorretamente pelo <i>Neural Ringer</i>	177
7.48 Eventos médios (e desvio padrão) de elétrons (esquerda) e jatos (direita) classificados corretamente pelo discriminador SOM+MLP e incorretamente pelo <i>Neural Ringer</i>	178

7.49	Curvas ROC obtidas variando-se o número de neurônios utilizados para estimar as funções não-lineares do modelo PNL (conjunto E15i).	179
7.50	Eficiência (acima) e falso alarme (abaixo) em energia e η	180
7.51	Máximo SP obtido após treinamento das redes neurais retirando (de modo segmentado) parte da informação.	181
7.52	Variação do desempenho da rede neural combinadora variando-se o número de neurônios da camada oculta.	182
7.53	Máximo SP (esquerda) e curvas ROC (direita) obtidas para a rede combinadora retirando-se a informação (de modo segmentado) de uma ou mais camadas do calorímetro, a linha tracejada vertical indica o desempenho do <i>Neural Ringer</i>	182
8.1	Saída dos classificadores neurais (acima) e comparativo de rejeição (abaixo) para eventos de raios cósmicos.	186
8.2	Perfil médio dos eventos de raios cósmicos aprovados conjuntamente pelo discriminador baseado em ICA Local e pelo <i>Neural Ringer</i> . . O perfil médio dos elétrons do conjunto E15 é mostrado para comparação.	187
8.3	Comparativo de desempenho em energia e η de diferentes discriminadores na rejeição de eventos de raios cósmicos.	188
8.4	Saídas do <i>Neural Ringer</i> (treinado para os sinais simulados E10) para os sinais experimentais.	189
8.5	Máximo SP (média e desvio padrão) obtidos variando-se o número de neurônios ocultos.	190
8.6	Distribuição das saídas para os discriminadores neural (esquerda e acima) e de Fisher (direita e acima) e as curvas ROC (abaixo).	191
8.7	SP.	192
8.8	Comparativo de desempenho entre o <i>Neural Ringer</i> e os dois cortes do T2Calo.	194
8.9	Perfil médio dos eventos classificados incorretamente pelo <i>Neural Ringer</i> . Os sinais médios de “elétrons” e “jatos” são mostrados para comparação.	195
8.10	Distribuições em energia e η dos eventos classificados incorretamente pelo <i>Neural Ringer</i>	196

8.11 Distribuição das assinaturas de cada classe nos agrupamento.	197
8.12 Curvas ROC para os discriminadores baseados em ICA/NLICA para os sinais experimentais.	198
8.13 Comparativo de desempenho entre o <i>Neural Ringer</i> e os discriminadores neurais com pré-processamento por ICA (linear) e ICA Local. .	199
8.14 SP máximo×100 obtido por cada classificador especialista (segmentado).	200
8.15 Desempenho obtido,em termos do máximo SP (esquerda) e das curvas ROC (direita), através da rede combinadora.	201
8.16 Comparativo de desempenho entre o <i>Neural Ringer</i> e os discriminadores neurais com pré-processamento por ICA (linear) e ICA Local. .	202
8.17 Comparativo de desempenho entre diferentes discriminadores em termos das curvas ROC.	203
8.18 Parcada do tempo total gasta em cada etapa de processamento do <i>Neural Ringer</i> (esquerda) e parcada do tempo específico das etapas de seleção de região, anelamento e normalização gasto para cada camada do calorímetro.	208
A.1 Diagrama de um mapa auto-organizável	239
A.2 Diagrama da classificação a partir do mapa de características	243
A.3 Compressão e recuperação do sinal x utilizando a transformação por PCA.	246
A.4 Diagrama do algoritmo de Taleb-Jutten para o modelo PNL.	257
B.1 Esquemático do problema de classificação binário.	259
B.2 Diagrama em blocos de um filtro casado.	262
B.3 Exemplo de uma rede neural utilizada para separação dos sinais de entrada em 3 classes.	264
C.1 Exemplo de um cromossomo binário	268
C.2 (a) Recombinação em ponto único e (b) recombinação uniforme.	270
C.3 Fluxo de um Algoritmo Genético	272

Listas de Tabelas

2.1	Intensidade relativa (em comparação com a interação forte) dos diversos tipos de interação.	10
2.2	Exemplos de valores da massa (em energia equivalente) de algumas partículas.	12
2.3	Principais objetos de interesse no ATLAS.	20
2.4	Região de cobertura em η , granularidade e número de canais de leitura das camadas dos calorímetros.	25
3.1	Principais características do sistema de <i>trigger</i> do ATLAS, onde Te e Ts são, respectivamente, as taxas de eventos na entrada e na saída e $Cr = Te/Ts$ é o coeficiente de redução de eventos.	39
3.2	Freqüência esperada para os principais canais de <i>trigger</i> no primeiro nível de filtragem do ATLAS ($L=10^{34}\text{cm}^{-2}s^{-1}$).	42
4.1	Tempo gasto nas diversas etapas necessárias para a execução <i>online</i> do <i>Neural Ringer</i>	60
6.1	Parâmetros de treinamento utilizados.	110
6.2	Composição das bases de dados utilizadas antes e depois do corte de primeiro nível.	113
6.3	Composição das assinaturas de sinais simulados aprovados em diferentes critérios da análise <i>offline</i>	124
7.1	Comparação de desempenho entre o T2Calo, <i>Neural Ringer</i> , e o discriminador neural com pré-processamento por ICA para o conjunto E10. São mostrados o máximo SP, os valores de PD e PF que produzem o máximo SP e PF para $PD=92,5\%$	129

7.2	Comparação de desempenho entre o discriminador baseado na ICA Local e outros classificadores para o conjunto E10. São mostrados o máximo SP, os valores de PD e PF que produzem o máximo SP e PF para PD=92,5%.	143
7.3	Taxas conjuntas (%) de acerto e erro para diferentes discriminadores.	145
7.4	Máximo SP($\times 100$) para diferentes estratégias de treinamento dos mapas (aqui o SP é calculado considerando a operação do SOM diretamente como um classificador).	149
7.5	Comparação de desempenho para diferentes abordagens de treinamento dos mapas auto-organizáveis para o conjunto E10. São mostrados o máximo SP, os valores de PD e PF que produzem o máximo SP e PF para PD=92,5%.	152
7.6	Número de componentes preservados (N) e máximo SP($\times 100$) para diferentes níveis de retenção de energia por PCA.	153
7.7	Número de componentes estimados a partir da compactação por PCA (50% de retenção de energia), PCD e pelo modelo PNL modificado.	154
7.8	Comparação entre discriminadores baseados no pré-processamento através do modelo PNL.	155
7.9	Comparação de desempenho para diferentes discriminadores, corte E10.	155
7.10	Comparação dos resultados obtidos para diferentes discriminadores no conjunto E10 considerando o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=92,5%.	165
7.11	Comparação dos resultados obtidos para oT2Calo, <i>Neural Ringer</i> e o discriminador neural operando em conjunto com a ICA (linear) para o conjunto E15. São mostrados o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.	166

7.12	Comparação dos resultados obtidos com o pré-processamento por ICA local com os de outros discriminadores para o conjunto E15i. Foram considerados o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.	172
7.13	Máximo SP($\times 100$) para diferentes estratégias de treinamento dos mapas (aqui o SP é calculado considerando a operação do SOM diretamente como um classificador).	175
7.14	Comparação dos resultados obtidos através do pré-processamento por SOM com os de outros discriminadores para o conjunto E15i. Foram considerados o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.	176
7.15	Comparação entre discriminadores baseados no pré-processamento através do modelo PNL.	178
7.16	Comparação dos resultados obtidos através do pré-processamento pelo modelo PNL com os de outros discriminadores para o conjunto E15i. Foram considerados o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.	179
7.17	Comparação dos resultados obtidos para diferentes discriminadores no conjunto E15i considerando o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.	183
8.1	Comparação do desempenho de diferentes discriminadores na rejeição de raios cósmicos.	185
8.2	Comparação dos resultados obtidos para o <i>Neural Ringer</i> e o T2Calo considerando o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=96% e PD=98%.	192
8.3	Número de componentes estimados a partir da compactação por PCD.	197

8.4	Comparação dos resultados obtidos para diferentes discriminadores baseados no pré-processamento por ICA/NLICA considerando o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=93%, PD=96% e PD=98%.	199
8.5	Comparação dos resultados obtidos para diferentes discriminadores considerando o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=93%, PD=96% e PD=98%.	201
8.6	Comparação dos resultados obtidos para diferentes discriminadores considerando o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=93%, PD=96% e PD=98%.	203
8.7	Comparação dos tempos de processamento (em ms) para diferentes discriminadores. O tempo gasto na transformação através da NLICA, quando for o caso, está incluído no tempo do classificador neural. . .	206
8.8	Tempo gasto para seleção da região de interesse e construção dos sinalis em anéis para cada camada do calorímetro, os valores percentuais são referentes apenas ao total destas duas etapas.	207
8.9	Comparativo entre o <i>Neural Ringer</i> e a rede combinadora que utiliza apenas as informações das camadas E1 e E2.	208

Símbolos e Abreviaturas

Símbolos

c	Velocidade da Luz no Vácuo
E	Energia
eV	Elétron-Volt
$\mathcal{E}(.)$	Operador esperança
Ef_i	Eficiência de discriminação da classe i
E_T	Energia transversa
E_T^{miss}	Energia transversa faltante
$H(.)$	Entropia
$I(.)$	Informação mútua
$J(.)$	Negentropia
$kurt(.)$	Curtose
m	Massa
p_T	Momento transverso
$p_x(x)$	Função de densidade de probabilidade
η	Pseudo-rapidez (sistema de coordenadas do ATLAS)
θ	Ângulo polar (sistema de coordenadas do ATLAS)
$\lambda(.)$	Razão de semelhança
ϕ	Ângulo azimutal (sistema de coordenadas do ATLAS)
e^-/j	Elétron/Jato
\mathbf{s}	Vetor das fontes independentes
\mathbf{x}	Vetor dos sinais observados (medidos)
\mathbf{y}	Vetor dos componentes independentes estimados
\mathbf{z}	Vetor dos sinais branqueados

Abreviaturas

No caso de algumas abreviaturas internacionalmente conhecidas, optou-se por mantê-las em inglês.

AG	Algoritmo Genético
ALEPH	Detector do acelerador LEP
ALICE	Detector do LHC
ASIC	<i>Application specific integrated circuit</i>
ATLAS	<i>A Toroidal LHC Aparattus</i>
BOOSTER	Experimento do laboratório Fermilab
cdf	Função de distribuição cumulativa
CERN	Centro Europeu para Pesquisa Nuclear
CHOOZ	Experimento de Física de Alta Energia
CMB	<i>Cosmic Microwave Background</i>
CMS	Detector do LHC
CP	<i>Charge parity</i>
CPLD	<i>Complex Programmable Logic Device</i>
CTP	<i>Central Trigger Processor</i>
D0	Detector do Fermilab
DELPHI	Detector do LEP
DESY	<i>Deutsches Elektronen-Synchrotron</i> <i>Die</i>
DL	Discriminante Linear
DNA	Ácido desoxirribonucleico
DSP	<i>Digital signal processing</i>
E1	Primeira camada eletromagnética do calorímetro do ATLAS
E2	Segunda camada eletromagnética do calorímetro do ATLAS
E3	Terceira camada eletromagnética do calorímetro do ATLAS
EEG	Eletro-encefalograma
EELS	Espectropia por perda de energia de elétrons
EF	Filtro de Eventos (<i>Event Filter</i>)

EL	<i>Emsemble learning</i>
EM	Eletromagnético
FastICA	Algoritmo de extração das componentes independentes
FDL	Discriminante linear de Fisher
FPGA	<i>Field-Programmable Gate Array</i>
Fermilab	<i>Fermi National Accelerator Laboratory</i>
GEANT	Simulador de Monte Carlo para colisões de partículas
GTM	<i>Generative topographic mapping</i>
H0	Primeira camada hadrônica do calorímetro do ATLAS
H1	Segunda camada hadrônica do calorímetro do ATLAS
H2	Terceira camada hadrônica do calorímetro do ATLAS
HAD	Hadrônico
HLT	<i>High Level Trigger</i>
HEGRA	Experimento de astrofísica de alta energia
HEP	<i>High-Energy Physics</i>
HERA	Experimento do laboratório DESY
HERWIG	Simulador de Monte Carlo para colisões de partículas
IATC	Técnica de Cherenkov para imageamento atmosférico
IC	Componentes independentes
ICA	<i>Independent components analysis</i>
INFOMAX	<i>Information maximization algorithm</i>
ISAJET	Simulador de Monte Carlo para colisões de partículas
JADE	<i>Joint Approximate Diagonalization of Eigen-matrices</i>
KEK	Centro de pesquisas Japonês
JEP	Processador de jatos e soma de energia
LEAR	Experimento do CERN (desativado)
LEP	Experimento do CERN (desativado)
LHC	<i>Large Hadron Collider</i>
LHCb	Detector do LHC
LHCf	Detector do LHC
L1	Primeiro nível de filtragem do ATLAS
L2	Segundo nível de filtragem do ATLAS

L2PU	Unidade de processamento do L2
L2SV	Supervisor do L2
LVQ	Quantização vetorial por aprendizado
MISEP	Algoritmo para ICA/NLICA
MLP	Perceptron de múltiplas camadas
MSSM	<i>Minimal supersymmetric Standard Model</i>
NLICA	<i>Non-linear independent component analysis</i>
NLPCA	<i>Non-linear principal component analysis</i>
PC	Computador pessoal
PCA	<i>Principal components analysis</i>
PCD	<i>Principal Components for Discrimination</i>
pdf	Função densidade de probabilidade
PMT	Foto-multiplicadora
PNL	<i>Post Nonlinear model</i>
PYTHIA	Simulador de Monte Carlo para colisões de partículas
OGLE	Telescópio instalado no Chile
QCD	Eletrodinâmica quântica
QED	Cromodinâmica quântica
QV/VQ	Quantização vetorial
RBF	Função de base radial
RNA	Rede Neural Artificial
ROB	<i>Buffers</i> de saída do sistema de filtragem do ATLAS
ROC	<i>Receiver operating characteristic</i>
ROD	<i>Drivers</i> de saída do sistema de filtragem do ATLAS
RoI	Região de interesse (<i>Region of Interest</i>)
RoIB	<i>Buffers</i> de RoI
RPC	<i>Resistive Plate Chamber</i>
RPROP	<i>Resilient back-propagation</i>
SFI	<i>Sub-farmer input</i>
SICA	<i>Segmented independent component analysis</i>
SLAC	<i>Stanford Linear Accelerator Center</i>
SPCA	<i>Segmented principal component analysis</i>

SM	Modelo Padrão de interação entre as partículas elementares (<i>Standard Model</i>)
SOBI	Algoritmo de ICA
SOM	Mapa auto-organizável
SP	Figura de mérito de desempenho de classificação
SUSY	Supersimetria
SVM	Máquina de vetor de suporte
T2Calo	Algoritmo de discriminação de elétrons no L2 do ATLAS
TDAQ	<i>Trigger and data aquisition</i>
TGC	<i>Thin Gap Chamber</i>
TOTEM	Experimento do LHC
TT	Torre de <i>trigger</i>

Capítulo 1

Introdução

O processamento estatístico de sinais encontra aplicações nas diversas áreas do conhecimento, desde medicina e saúde pública até as bolsas de valores. Seu uso pode simplificar as tarefas de análise de dados e classificação, pois permite mapear o conjunto de sinais em um espaço onde sua estrutura fundamental está mais acessível.

Este trabalho descreve a aplicação das técnicas de processamento estatístico no sistema de filtragem (detecção) *online* de um detector de partículas elementares de altas energias. O objetivo é extrair características relevantes para guiar o processo de identificação das partículas.

1.1 Contexto

Com os constantes avanços tecnológicos dos sistemas eletrônicos de aquisição de informações, é crescente a necessidade de técnicas eficientes para o processamento *online* de sinais. As grandezas físicas são registradas por elementos sensores, que podem ser únicos (como na medição da velocidade de um motor), ou combinados aos milhares para obter o resultado final (como na captura de vídeo e imagem digitais).

Em aplicações onde a fina granularidade da informação é necessária para descrever adequadamente o processo físico em questão, e sinais com alta dimensão são assim gerados por sistemas de medição compostos por um número elevado de sensores, o custo computacional geralmente é alto. Em alguns casos, a informação disponível pode estar segmentada, pois foi produzida a partir de conjuntos de sensores com características distintas.

Se há a necessidade de uma resposta rápida, pode ser utilizada a combinação de técnicas de compactação de sinais e processamento distribuído. O cenário pode ficar ainda mais complicado quando o volume de dados é alto e o problema a ser resolvido apresenta elevado grau de complexidade.

O ambiente de aplicação deste trabalho é o sistema *online* de filtragem do ATLAS (*A Toroidal LHC Aparatus*) [1], maior detector de propósito geral do acelerador de partículas LHC (*Large Hadron Collider*) [2]. O LHC entrou em operação no final de 2008, logo depois, passou por reparos no sistema de resfriamento de um dos seus supercondutores, voltando a operar no final de 2009.

No LHC, os sinais de interesse são raros, estão imersos em um intenso ruído de fundo e a perda de um desses eventos compromete severamente o desempenho dos detectores. Neste caso, é necessária uma estratégia de filtragem capaz de remover, ou pelo menos atenuar a intensidade do ruído de fundo sem perder os eventos de interesse.

Combinado a isso, a taxa de ocorrência de eventos é extremamente elevada, fazendo com que o intervalo entre eventos consecutivos seja extremamente pequeno. Considerando ainda que os detectores são altamente segmentados e apresentam fina granularidade de células detectoras, a quantidade de informação produzida é enorme (~ 60 TB/s). Neste contexto, a seleção de eventos precisa ser realizada de modo *online* e sob severas restrições no tempo de processamento.

Técnicas que utilizam informações da estatística dos sinais, como Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3], Análise de Componentes Independentes ICA (*Independent Component Analysis*) [4] e Redes Neurais Artificiais (RNA) [5], são frequentemente utilizadas na solução de problemas onde há a necessidade de processamento veloz, flexível e eficiente.

1.2 Motivação

Desde o final do século 19, quando foi descoberto o elétron, o estudo da física de partículas elementares de altas energias, ou simplesmente física de altas energias (HEP *High-Energy Physics*), teve um crescimento acentuado. Na década de 1950, com o uso dos aceleradores, foram descobertas centenas de novas partículas. A física

de partículas de altas energias pretende encontrar os componentes fundamentais da matéria e descrever suas formas de interação.

O LHC [2] é o maior e mais potente acelerador de partículas jamais construído e está em operação no CERN (Organização Européia para Pesquisa Nuclear) desde 2008. Ao operar na máxima capacidade, produzirá uma taxa de colisões que chegará a 40MHz. Entretanto, as assinaturas de interesse ocorrerão numa freqüência muito menor, o que faz do sistema de filtragem *online* um componente fundamental para os detectores.

O ATLAS [1] é um detector de propósito geral do LHC e está posicionado em um dos pontos de colisão. Entre os principais objetivos do ATLAS, pode-se destacar a busca do *bóson de Higgs*, que, segundo estudos teóricos, seria responsável por interagir com as partículas fornecendo-lhes massa [6]. A partícula de Higgs ainda não foi verificada experimentalmente.

Parcela importante das informações necessárias para a caracterização dos eventos é obtida do sistema de calorímetros, que, no ATLAS, é subdividido em 7 camadas. Os calorímetros são medidores de energia compostos por um grande número de sensores (células). Ao interagirem com o material do calorímetro, as partículas perdem energia (e consequentemente velocidade). As células dos calorímetros quantificam a energia perdida pelas partículas incidentes e a informação do perfil de deposição de energia é utilizada para a caracterização do tipo de partícula.

Os objetivos principais dos sistemas de filtragem *online* em experimentos de física de altas energias são maximizar a probabilidade de detecção (e consequente armazenamento) dos eventos de interesse, e minimizar a probabilidade de armazenar eventos não desejados (ruído de fundo ou falso-alarme). Em um ambiente como este, a alta dimensão dos dados, o intenso ruído de fundo e o curto tempo de resposta exigido são sérios entraves para o processamento e classificação *online* de eventos.

No ATLAS, o sistema *online* de filtragem (*trigger*) de eventos é composto por três níveis de seleção sequenciais. O ruído de fundo é gradualmente reduzido a cada nível de filtragem, esperando-se armazenar, em média permanente, uma taxa máxima de 200 Hz [7]. Considerando que a freqüência de colisões é 40 MHz, deve haver uma redução de 2×10^5 vezes.

No contexto dos diversos canais de interesse para a física no ATLAS, este trabalho

se encontra focado na discriminação elétron/jato (e^-/j). Os elétrons podem estar envolvidos em fenômenos como o decaimento do bóson de Higgs, supersimetria e a descoberta de novos bósons. Porém, em termos de calorimetria, alguns jatos podem apresentar um perfil de deposição de energia semelhante ao dos elétrons. Portanto, os jatos representam ruído de fundo no processo de identificação de elétrons. Apenas uma parcela dos candidatos a elétrons aceitos pelo primeiro nível são realmente elétrons; cabendo à filtragem de alto nível (segundo e terceiro níveis) reduzir ainda mais o ruído de fundo, mantendo a maior parte das assinaturas de interesse.

Considerando a alta taxa de eventos e a intensidade do ruído de fundo produzidos pelo LHC, a busca por algoritmos de filtragem *online* mais eficientes é uma tarefa importante. A redução no número de eventos não relevantes (ruído de fundo) armazenados em mídia permanente significa maior eficiência na análise *offline* dos eventos de interesse e redução do espaço (mídia) necessário para armazenamento.

Neste trabalho, estão sendo propostas alternativas para o algoritmo padrão de detecção de elétrons em uso atualmente no segundo nível de filtragem (L2) do detector ATLAS. Os algoritmos desenvolvidos apresentam maior eficiência de discriminação e tempo de processamento dentro das restrições do L2.

1.3 Trabalhos Anteriores

Considerando os desafios existentes no ambiente de filtragem *online* do ATLAS, no qual os sinais são adquiridos com fina segmentação, estão imersos em intenso ruído de fundo e as assinaturas de interesse são raras, técnicas avançadas de extração de características podem ser utilizadas para melhorar a eficiência de classificação.

No trabalho [8] foi inicialmente proposto o uso de um classificador neural supervisionado (arquitetura *Perceptron* de Múltiplas Camadas) para o canal elétron/jato do segundo nível de filtragem do detector ATLAS. Utilizando informação especialista a respeito do problema, os sinais medidos nos calorímetros são formatados em anéis concêntricos de deposição de energia. A formatação dos anéis preserva a informação discriminante do perfil de deposição de energia e compacta a informação (de 1000 células para 100 anéis).

Em [9] o sistema neural de identificação de elétrons (que ficou conhecido como *Neural Ringer*) foi implementado no sistema (*software*) de filtragem do ATLAS. Numa comparação de desempenho entre o Neural Ringer e o discriminador oficial do ATLAS (T2Calo) foi mostrado que o *Ringer* apresenta desempenho superior e é capaz de operar dentro da janela de tempo permitida para o segundo nível (L2).

No trabalho [10] alguns métodos de compactação como Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3] e Componentes Principais de Discriminação (PCD - *Principal Components of Discrimination*) [11] foram aplicados, em conjunto com o modelo linear da Análise de Componentes Independentes (ICA - *Independent Component Analysis*) [4], sobre os sinais em anéis como um pré-processamento para o classificador neural. A utilização destas técnicas de extração de características e compactação permitiu aumento do desempenho de classificação e redução do tempo necessário para tomada de decisão. Neste trabalho, foi realizada também uma implementação otimizada do *Neural Ringer* no sistema *online* de filtragem do ATLAS operando como uma sub-rotina do T2Calo. Neste nova versão o custo computacional foi reduzido e o Ringer utiliza uma parte do processamento realizado pelo T2Calo.

1.4 Objetivos

Os calorímetros são projetados para serem detectores lineares, porém, diversas fontes de não-linearidades podem surgir numa implementação prática [12]. Neste caso, um método não-linear de extração de características talvez seja mais indicado para o problema.

O principal objetivo do presente trabalho é avaliar o desempenho obtido pelo discriminador *Neural Ringer* quando os sinais em anéis são pré-processados por métodos de extração de características baseados no modelo não-linear da análise de componentes independentes (NLICA - *Nonlinear Independent Component Analysis*) [13].

Neste contexto, foram utilizados diversos modelos e algoritmos de estimação da NLICA e seus resultados foram comparados em termos da eficiência de discriminação e do tempo de processamento. Foi proposta também a segmentação dos processos

de extração de características e classificação, visando explorar adequadamente toda segmentação e granularidade disponíveis no detector.

O modelo da NLICA foi originalmente definido para realizar a extração de características de modo não-supervisionado. Ou seja, não há como garantir que a transformação seja útil para o problema de classificação (no sentido de revelar características discriminantes). Neste trabalho foram propostas modificações no modelo tradicional da NLICA visando a estimativa de componentes com maior poder de discriminação entre as classes.

1.5 Metodologia

O projeto dos discriminadores de partículas foi realizado a partir de um conjunto de eventos simulados. Estas simulações, por técnicas de Monte Carlo [14], consideram todas as características físicas do detector e do acelerador. Foram utilizados, também, conjuntos de dados experimentais obtidos na fase inicial de operação do LHC. Sabe-se que os raios cósmicos representam ruído de fundo para o canal elétron/jato, pois produzem mísulas no detector. Assim, os algoritmos propostos foram também testados para uma base de dados composta de eventos de raios cósmicos, visando verificar a capacidade de rejeição para esse sinal. Uma outra análise utilizou eventos de colisão obtidos recentemente na fase inicial de operação do LHC.

Foram aplicados diversos algoritmos para estimar o modelo não-linear das componentes independentes (NLICA). Visando explorar adequadamente as características do detector, o modo de executar as tarefas de extração de características e classificação foi variado entre as abordagens segmentada (onde o processamento é feito em cada camada do calorímetro) e não-segmentada (quando os sinais em anéis, gerados a partir de todas as camadas, são concatenados em um único vetor).

Entre os diversos modelos existentes para a estimativa da NLICA, neste trabalho foram utilizados os modelos sem restrição estrutural (através dos mapas auto-organizáveis) e o modelo pós não-linear (que restrige os mapeamentos não-lineares possíveis a uma mistura linear seguida de funções não-lineares aplicadas a cada componente desta mistura). A ICA Local, que é um modelo diretamente ligado ao da

NLICA também foi utilizada. Para cada modelo proposto foi realizado um estudo comparativo de desempenho com o Neural Ringer quando este opera diretamente sobre os sinais em anéis.

Considerando o sistema de classificação neural, foi proposta a utilização de classificadores especialistas na informação de cada camada do calorímetro. Diversos modos de combinar a informação deste conjunto de classificadores foram testados. Deste modo verificou-se que há redundância de informação entre as diversas camadas do calorímetro. Pode-se, então, obter o mesmo desempenho de discriminação utilizando-se apenas uma parcela da informação disponível e, assim, contribuir para a redução do tempo de processamento.

1.6 Conteúdo do Trabalho

No Capítulo 2 será apresentado o ambiente científico no qual o trabalho está sendo desenvolvido, contextualizando o detector de partículas ATLAS, o acelerador LHC e o CERN. Uma descrição dos sistemas de filtragem (*trigger*) *online* em experimentos de física de altas energias e será apresentada no Capítulo 3, com foco no detector ATLAS.

É descrito, no Capítulo 4, o processo de seleção de elétrons utilizando informações de calorimetria no contexto do sistema de filtragem do ATLAS. No Capítulo 5, serão mostrados os fundamentos teóricos das técnicas de extração de características que serão utilizadas para a otimização do sistema de filtragem do ATLAS.

A metodologia empregada no desenvolvimento deste trabalho é descrita no Capítulo 6. Os resultados obtidos serão apresentadas nos Capítulos 7 e 8, sendo o primeiro dedicado aos conjuntos de dados simulados e o segundo aos dados adquiridos experimentalmente. As conclusões e os futuros trabalhos são os tópicos abordados no Capítulo 9.

Nos Apêndices A e B são fornecidas as bases matemáticas para uma melhor compreensão, respectivamente, dos algoritmos de extração de características e classificação utilizados. No Apêndice C será descrito o algoritmo genético utilizado para algumas tarefas de otimização no decorrer da tese. Finalmente, no Apêndice D serão listadas as publicações produzidas com os resultados obtidos neste trabalho.

Capítulo 2

Física de Partículas e o Detector ATLAS

Os experimentos de Física de Partículas Elementares têm como principais objetivos a confirmação de modelos desenvolvidos teoricamente e a identificação de novos fenômenos. O experimento LHC [2] é o maior acelerador de partículas já desenvolvido e quando operando em máxima capacidade, serão gerados aproximadamente 10^9 interações por segundo. Os detectores são responsáveis por selecionar, dentro de um conjunto enorme, os raros eventos de real interesse. O projeto e a montagem do detector ATLAS (*A Thoroidal LHC Apparatus*) [1] foram conduzidos por uma colaboração de 36 países, conhecida como *ATLAS Collaboration*, contando com pesquisadores de mais de 150 universidades e centros de pesquisa [15]. Sendo um dos detectores de propósito geral do experimento LHC, o ATLAS tem formato cilíndrico e foi projetado para cobrir um ângulo sólido próximo a 4π ao redor da região de colisão.

2.1 Panorama Geral da Física de Partículas Elementares

A noção de que a matéria é composta por um conjunto de constituintes elementares existe há mais de 2000 anos, desde o tempo dos filósofos gregos [16]. No decorrer do século 20, a compreensão dos componentes elementares da matéria forneceu à comunidade científica mundial informações importantes a respeito das leis fundamentais

da natureza. O estudo da física de partículas elementares teve início no final do século 19, quando foi descoberto o elétron, experimentando um crescimento acen-tuado na década de 1950, quando foram descobertas centenas de novas partículas. A partir de sua criação em 1954, o CERN (Centro Europeu para Pesquisa Nuclear) contribuiu significativamente nesse processo.

Inicialmente, pensava-se que a matéria era constituída de partículas subatômicas, chamadas elétrons, prótons e nêutrons. Mais tarde, descobriu-se que os prótons e nêutrons são compostos de quarks. Hoje sabe-se que toda a matéria no universo é composta de léptons e quarks. Existem ainda outras partículas elementares que são responsáveis por promover a interação entre léptons e quarks. Juntamente com as descobertas experimentais, estudos teóricos possibilitaram o desenvolvimento do Modelo Padrão (*SM-Standard Model*) [6] que descreve e prevê, de forma unificada, o comportamento das partículas e das forças de interação.

As características e propriedades dos processos nucleares dependem da energia envolvida. A unidade de energia mais utilizada, neste contexto, é o elétron-volt (eV) e seus múltiplos ($10^6 \text{ eV} \rightarrow \text{MeV}$, $10^9 \text{ eV} \rightarrow \text{GeV}$ ou $10^{12} \text{ eV} \rightarrow \text{TeV}$). O elétron-volt é definido como a energia necessária para aumentar o potencial elétrico de um elétron em 1 volt (em unidades do Sistema Internacional temos: $1 \text{ eV} = 1,6 \times 10^{-19} \text{ J}$) [17]. Os fenômenos produzidos quando a energia das partículas é menor que 20 MeV são chamados de física a baixas energias. A faixa entre 20 MeV e 400 MeV corresponde à física a energias intermediárias, e finalmente, fenômenos com energia superior a 400 MeV são estudados na física a altas energias [18].

Os experimentos realizados a partir do início da década de 1970 foram fundamen-tais para o desenvolvimento e teste do Modelo Padrão, mas também despertaram dúvidas a respeito de questões que não são completamente respondidas. O SM (ver Figura 2.1), divide as partículas elementares em quarks e léptons. Existem seis ti-pos de léptons (elétron, mûon, tau e três neutrino) e seis tipos de quarks (*up, down, charm, strange, top e bottom*).

Atualmente são conhecidas quatro formas de interação (ou acoplamento) entre as partículas elementares, são elas eletromagnética, gravitacional, fraca e forte. A força gravitacional é predominante entre corpos massivos separados por longas distâncias. Entre partículas, onde a massa é da ordem de 10^{-27} Kg , a interação gravitacional

	Fermions			Bósons
Quarks	u up	c charm	t top	y fóton
	d down	s strange	b bottom	Z
Leptons	V_e elétron neutrino	V_μ múon neutrino	V_τ tau neutrino	W
	e Elétron	μ Múon	τ tau	g glúon
Ainda não confirmado		Bóson de Higgs		

Figura 2.1: Diagrama do Modelo Padrão, mostrando as partículas elementares incluindo o, ainda não confirmado, bóson de Higgs.

tem intensidade muito baixa. Para distâncias maiores que 10^{-13} cm, a força eletromagnética domina, enquanto que para distâncias menores, as interações forte e fraca se destacam. A intensidade relativa (em comparação com a interação forte) dos quatro tipos de interação é mostrada na Tabela 2.1 [19].

Existem algumas teorias que descrevem como as interações elementares ocorrem. A eletrodinâmica quântica (QED - *Quantum Eletrodynamics*) considera que os processos elétricos e magnéticos acontecem a partir da interação fundamental entre dois elétrons, com a troca de um fóton. O fóton é a partícula mediadora da interação eletromagnética.

De modo semelhantes são definidos os outros tipos de interação. Para a interação

Tabela 2.1: Intensidade relativa (em comparação com a interação forte) dos diversos tipos de interação.

Interação	Intensidade Relativa
Forte	1
Eletromagnética	10^{-2}
Fraca	10^{-5}
Gravitacional	10^{-39}

forte, as partículas mediadoras são os glúons. Os bósons W e Z mediam a interação fraca e o gráviton a interação gravitacional. A teoria da cromodinâmica quântica (QCD - *Quantum Chromodynamics*) [20], por exemplo, descreve como os quarks e glúons interagem para formar os hadrons (prótons e neutrons são exemplos de hadrons). Uma diferença entre a QED e a QCD é que na última, os quarks e glúons não são observados como partículas livres (mas apenas na forma de hadrons), em oposição aos elétrons e fótons da QED. Mesmo em colisões de alta energia, quando quarks são produzidos, eles se afastam rapidamente uns dos outros, e antes que possam ser observados, se convertem em “jatos” de hadrons (ou jatos hadrônicos) [17].

Considerando as partículas e formas de interação já verificadas experimentalmente e uma formulação atualmente aceita pela física teórica (conhecida como *Gauge Invariance*), todas as partículas elementares teriam massa nula. Esta previsão é contrária aos resultados experimentais e somente pode ser corrigida assumindo-se que existe um outro tipo de interação. Esta interação foi prevista pelo cientista inglês Peter Higgs em 1964, tendo como partícula mediadora o bóson de Higgs, que é responsável por fornecer massa às partículas [17]. A existência do bóson de Higgs é a mais importante previsão do modelo padrão ainda não verificada experimentalmente e sua busca é de máxima importância para a física de partículas.

Nas últimas décadas, os experimentos com aceleradores de partículas tornaram gigantescos empreendimentos envolvendo milhares de físicos e engenheiros, com contribuição financeira e intelectual de dezenas de países.

As partículas são injetadas na máquina por dispositivos que produzem uma fonte de alta intensidade e baixa energia. Os aceleradores usam força eletromagnética para aumentar a energia de partículas estáveis e carregadas eletricamente. Quanto às características construtivas, os aceleradores podem ser divididos em alvo-fixo ou colisionadores de feixes. Nos aceleradores de alvo-fixo, as partículas são aceleradas até a máxima energia, quando o feixe é retirado da máquina e direcionado a um alvo estacionário. Nos colisionadores, feixes de partículas são acelerados, em sentidos opostos, e quando a energia desejada é atingida, os feixes são colocados em rota de colisão em alguns pontos específicos do percurso. Quanto ao percurso dos feixes, os aceleradores podem ser lineares ou circulares [17].

Imediatamente após a colisão, uma grande quantidade de partículas elementares

é produzida. Algumas delas são estáveis, porém outras têm curtíssimo tempo de vida. Os elétrons e prótons, por exemplo, têm vida média superior a 10^{23} anos, enquanto os mísions podem ter vida média da ordem de 10^{-6} segundos [18]. Para que todos os eventos sejam corretamente identificados, um complexo sistema de detecção precisa ser construído.

A massa (m) das partículas é usualmente expressa em função da energia equivalente ($m = E/c^2$, onde c é a velocidade da luz e E a energia). Os bósons W e Z, por exemplo, têm massa, respectivamente, de 80 e 91 GeV/c² (onde $1\text{GeV}/c^2=1,78\times10^{-27}\text{ kg}$). Na Tabela 2.2, são mostradas as massas de algumas partículas importantes. Ainda não é possível determinar a massa esperada para o bóson de Higgs, porém, a partir do conhecimento adquirido com os experimentos operados antes do LHC, sua massa deve ser maior que 115 GeV/c² [6].

O acelerador LHC (*Large Hadron Collider* - Grande Colisionador Hadrônico) [2, 21] em operação no CERN (Centro Europeu para Pesquisa Nuclear) [22], será capaz de atingir energia de até 14 TeV e permitir a visualização do bóson de Higgs, caso o Modelo Padrão esteja correto.

Tabela 2.2: Exemplos de valores da massa (em energia equivalente) de algumas partículas.

Partícula	Massa (GeV/c ²)
Elétron	0,000511
Próton	0,938
Bóson W	80
Bóson Z	91
Bóson de Higgs ou outras novas partículas	>115

Além da busca pelo bóson de Higgs, existem outras questões que precisam ser respondidas em física de partículas. É um desejo antigo dos físicos, desde Einstein, a unificação das teorias sobre as forças de interação entre as partículas, incluindo no Modelo Padrão a força gravitacional. O estudo da física de partículas também é fundamental para o entendimento da natureza e origem do universo. Pretende-se, por exemplo, descobrir informações sobre a composição da matéria escura [23], super-simetria [24] e violação de CP (do inglês *Charge Parity*) [25].

Entre os laboratórios que conduzem experimentos em física de partículas pode-se destacar: CERN, DESY [26], KEK [27], Fermilab [28], SLAC [29] e Brookhaven [30], localizados respectivamente na Suíça, Alemanha, Japão e os três últimos nos Estados Unidos.

O CERN, fundado em 1954, é atualmente um dos maiores centros de pesquisas em física de partículas do mundo. Localizado em Genebra, Suíça, funciona com base num complexo sistema de colaboração internacional, que envolve centenas de instituições de pesquisa em centenas de países [22].

2.2 O Acelerador LHC

No CERN, foi projetado e construído o experimento LHC (*Large Hadron Collider* ou Grande Colisionador de Hádrons), que iniciou sua operação no final de 2008 [2, 31]. O LHC pode atingir níveis de energia de, aproximadamente, 14 TeV. O percurso do acelerador, localizado na fronteira franco-suíça, é aproximadamente circular, com 27 Km de comprimento, numa profundidade do solo que varia de 50 a 150 metros (ver Figura 2.2). Feixes de prótons são acelerados em sentidos opostos e direcionados para colisão nos centros dos detectores.

O LHC tem 6 detectores com propósitos diferentes: ATLAS [1], CMS [32], LHCb [33], LHCf [34], ALICE [35] e TOTEM [36]. O ATLAS e o LHCf estão localizados em Meyrin, Suíça, o CMS e o TOTEM em Cessy, França, o ALICE em St. Genis-Pouilly, França e o LHCb em Ferney-Voltaire, França (os experimentos TOTEM e LHCf não são mostrados na Figura 2.2 pois estão em locais próximos respectivamente ao CMS e ATLAS). O ATLAS (*A Toroidal LHC ApparatuS*) e o CMS (*Compact Muon Solenoid*) são detectores de propósito geral, enquanto os outros são dedicados a aplicações específicas, como o LHCb, que é dedicado a explorar informações sobre a física proveniente dos hádrons do tipo **b** produzida nas colisões do LHC.

Quando operando em máxima capacidade, o LHC irá produzir colisões de feixes de prótons a cada 25 ns e atingirá energia 7 vezes maior que o Tevatron, que é o acelerador de maior energia em operação atualmente, funcionando no Fermilab.

O número de colisões por centímetro quadrado produzidas por segundo é cha-

mado de luminosidade (L) [17]:

$$L = n \frac{N_1 N_2}{A} f \quad (2.1)$$

onde n é o número de feixes, N_1 e N_2 o número de partículas em cada feixe, A a área da seção transversal do feixe e f a frequência de colisão. Quanto maior a luminosidade do experimento, maior a quantidade de informação (partículas) gerada.

Quando operando numa luminosidade jamais alcançada por outros experimentos ($10^{34} cm^{-2}s^{-1}$), espera-se, em média, 25 choques próton-próton a cada cruzamento de feixe (colisão). Como a frequência de colisões é de 40 MHZ, a taxa de interações será da ordem de 1 GHz. No LHC, fatores como a alta taxa de interações, as altas doses de radiação, a multiplicidade de partículas e largas faixas de energia a cobrir, em conjunto com a necessidade de medições precisas, definiram novos padrões para o projeto dos detectores.

A seguir, serão descritas, de modo geral, as principais características do detector ATLAS.

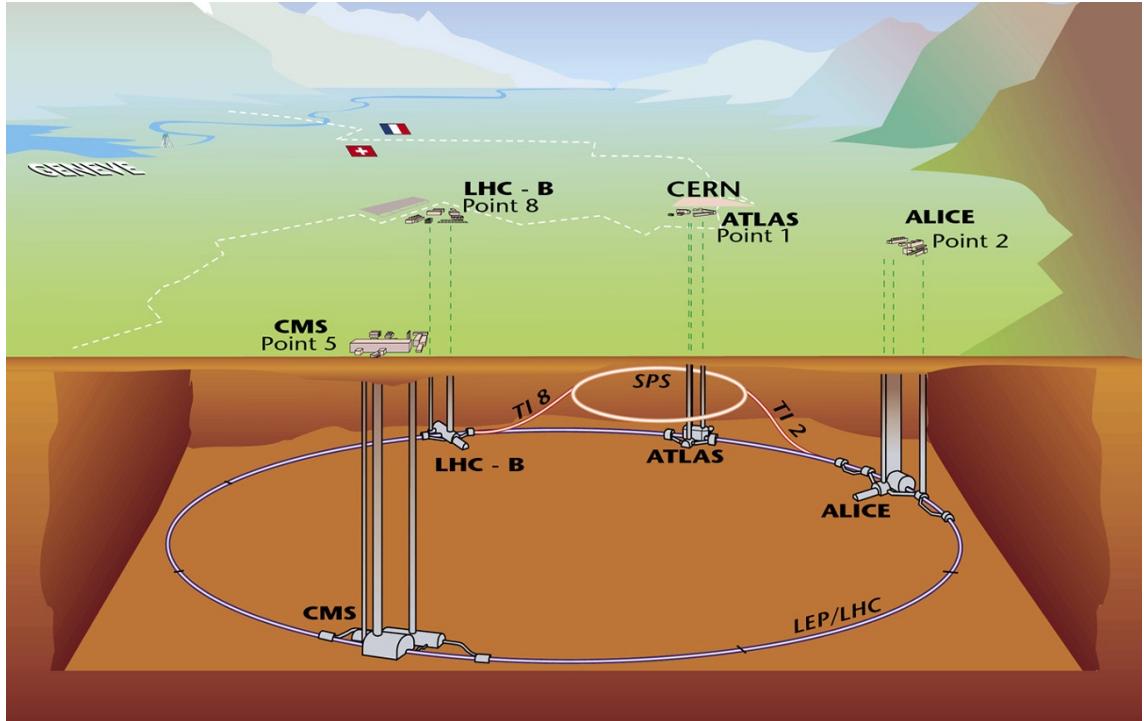


Figura 2.2: Mapa de localização dos detectores do LHC.

2.3 Características Gerais do Detector ATLAS

Os métodos de detecção em física têm como princípio básico promover interação entre as partículas em estudo e um material conhecido, produzindo informações sobre a natureza e as características da própria partícula. Os instrumentos que possibilitam a medição experimental destas quantidades físicas são chamados de detectores. Em particular, os detectores de energia são chamados de calorímetros. À medida que a energia envolvida aumenta, o sistema de detecção precisa ser mais sofisticado [18].

O ATLAS foi fruto do trabalho de uma grande colaboração, que envolveu milhares de físicos, engenheiros, técnicos e estudantes por um período de vinte anos de projeto, desenvolvimento, fabricação e instalação.

O detector tem 45 m de comprimento, 25 m de altura e pesa aproximadamente 7.000 toneladas, sendo dividido em subsistemas dispostos em camadas. Conforme ilustrado na Figura 2.3, os principais subsistemas são: detector de trajetórias (ou de traço), calorímetros eletromagnético e hadrônico e os detectores de múons.

A função dos detectores de trajetória é medir o momento das partículas eletricamente carregadas, a partir da curvatura de sua trajetória, quando imersos no campo magnético do solenóide central [37]. Caminhando do eixo central para as extremidades, em sequência, estão os calorímetros, que medem a energia depositada pelas partículas [38]. Na interação com as células do calorímetro, são produzidos chuveiros de partículas secundárias [12]. Num último estágio está o detector de múons, sistema dedicado à detecção destas partículas, que são as únicas que não ficam contidas nos calorímetros [39].

O sistema xyz de coordenadas do ATLAS é único para todos os subsistemas. Conforme mostrado na Figura 2.4, a direção do feixe do LHC define o eixo z , e os eixos x e y formam um plano transverso ao feixe. A direção positiva do eixo x é definida apontando do ponto de interação para o centro do anel do LHC, e o eixo y positivo aponta para cima. O ângulo azimutal é obtido a partir de:

$$\phi = \text{arctg}(x/y), \quad (2.2)$$

sendo que $\phi = 0$ corresponde ao eixo x positivo e ϕ aumenta no sentido horário.

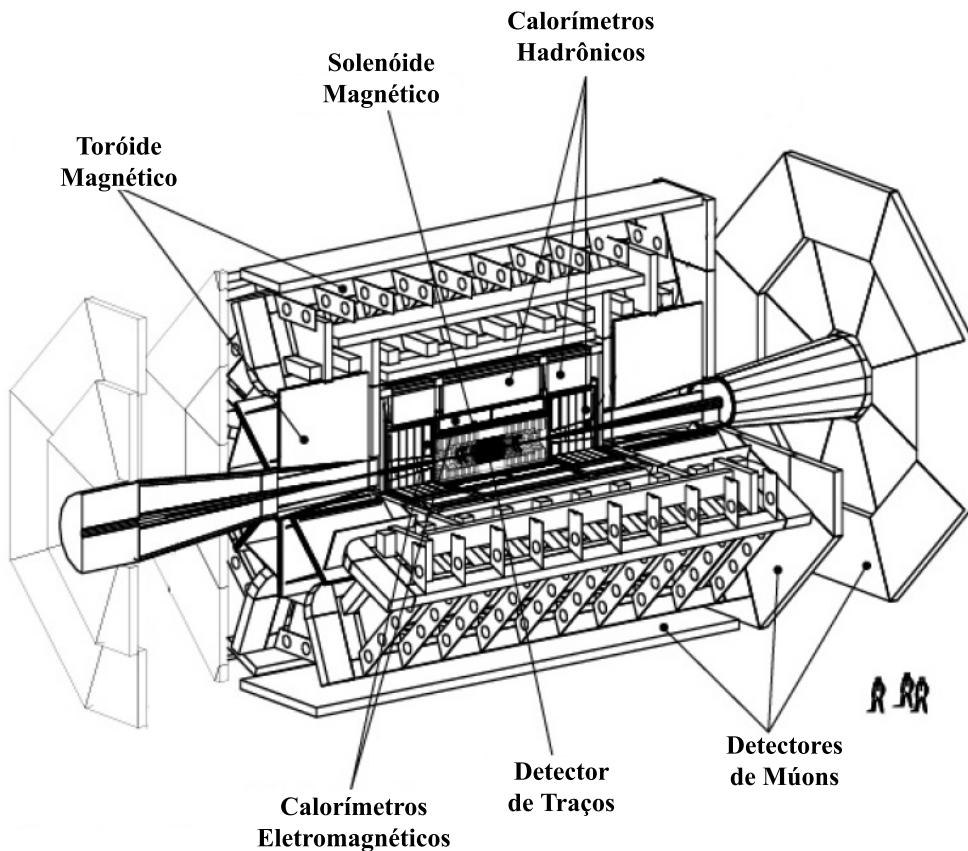


Figura 2.3: Diagrama esquemático do ATLAS.

O ângulo polar θ é medido a partir do eixo do feixe (eixo z positivo). O momento transverso p_T , a energia transversa E_T e a energia transversa perdida E_T^{miss} são definidas no plano xy . A pseudo-rapidez η é calculada a partir do ângulo θ de espllhamento em relação ao eixo z (ângulo de saída das partículas após a colisão) [7]:

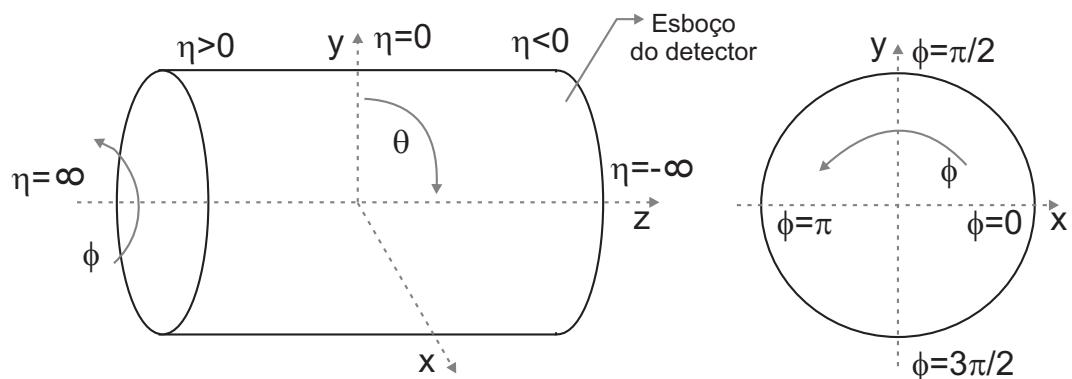


Figura 2.4: Eixo de coordenadas do ATLAS.

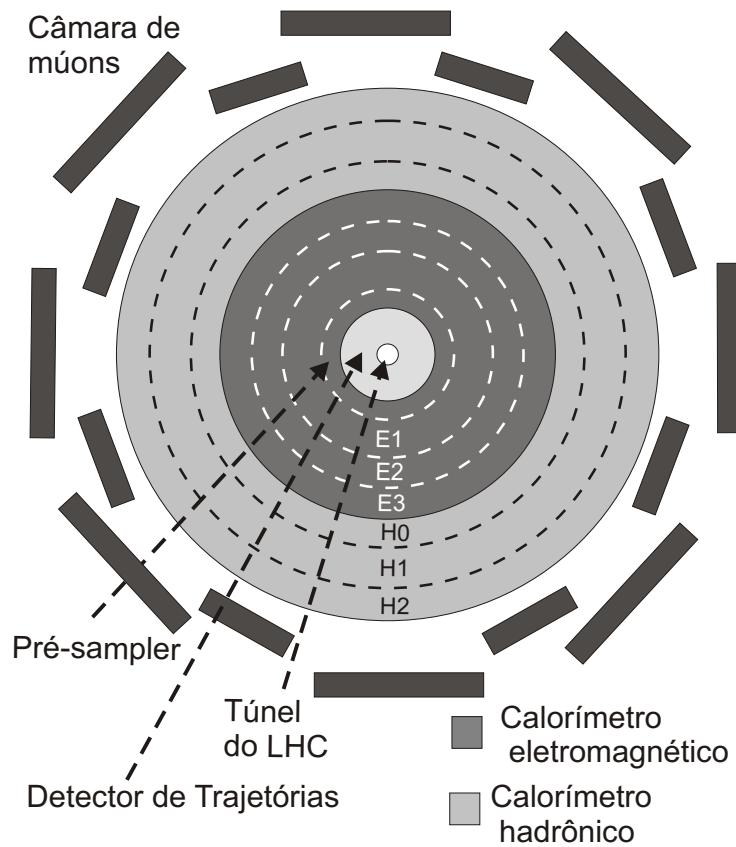
$$\eta = -\ln \operatorname{tg}(\theta/2). \quad (2.3)$$

A partir das definições das equações 2.2 e 2.3, define-se o eixo (η, ϕ) , onde o ângulo ϕ representa a rotação e η a direção de projeção das partículas após a colisão. Grandes valores da pseudo-rapidez indicam que a colisão das partículas não foi frontal, pois o ângulo de saída, após o choque, é pequeno, no limite quando $\theta \rightarrow 0$ então $\eta \rightarrow \infty$. Nesse tipo de colisão, como quase não houve choque, a produção de partículas elementares é pequena. O ATLAS foi projetado com baixa resolução para $\eta > 3$.

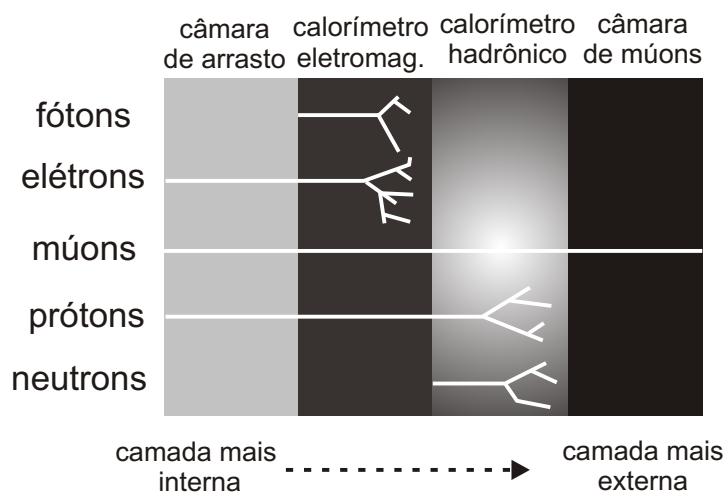
Após uma colisão, as partículas geradas interagem com o material do detector, perdendo energia e consequentemente velocidade. Na Figura 2.5(a) pode-se visualizar os subsistemas do ATLAS num corte transversal (paralelo ao plano xy). Percebe-se que a câmara de mísseis e o calorímetro hadrônico têm as maiores dimensões. Na figura 2.5(b) é mostrada a penetração e consequente visualização esperadas de algumas partículas nas camadas do detector. Não espera-se, por exemplo, deposição de energia de fótons ou elétrons além das camadas eletromagnéticas do calorímetro, pois estas partículas interagem intensamente com os materiais que compõem a seção eletromagnética, perdendo toda sua energia. As partículas hadrônicas (por exemplo: prótons e nêutrons) em geral interagem menos intensamente com o calorímetro eletromagnético e precisam das camadas hadrônicas para serem paradas. Os mísseis são partículas que perdem pouca energia nos calorímetros, necessitando de um sistema especial para serem detectados, o detector de mísseis.

O ATLAS foi projetado e construído considerando-se as condições experimentais produzidas pelas colisões do LHC e com o objetivo de obter informações importantes para responder às questões chave em física de partículas. Entre os principais critérios utilizados para o projeto do detector pode-se destacar [1]:

- Os elementos sensores e circuitos eletrônicos devem apresentar resposta rápida e resistência a altos níveis de radiação. A fina granularidade do detector é importante para reduzir a influência dos eventos sobrepostos.
- Excelente calorimetria eletromagnética para a identificação de elétrons e fótons, complementada por informações acuradas dos calorímetros hadrônicos, para medições de jatos hadrônicos e energia transversa E_T ;



(a)



(b)

Figura 2.5: Cortes (a) transversal e (b) axial do ATLAS.

- Eficiente sistema de identificação de trajetória para medição do momento;
- Alta precisão na identificação de múons;
- Alta aceitação em pseudo-rapidez (η) com cobertura quase total no ângulo azimutal (ϕ);
- Alta eficiência do sistema de filtragem (*trigger*), armazenando a maioria dos eventos físicos de interesse e reduzindo ao máximo o ruído de fundo (informação não relevante) produzido nas colisões do LHC.

Conforme mencionado no Capítulo 1, o estudo conduzido neste trabalho utiliza informações do sistema de calorímetros do ATLAS e propõe algoritmos para a otimização da detecção (*trigger*) de elétrons. Nas próximas seções, serão descritos o sistema de calorímetros do ATLAS e a importância da detecção de elétrons para o desempenho do experimento.

2.4 Principais Objetos de Interesse no ATLAS

Dentre os eventos gerados nas colisões do LHC, apenas uma pequena parte será útil para a caracterização dos processos da “nova física”. Com o LHC operando em alta luminosidade podem ocorrer até 10^9 interações por segundo, porém, os eventos de interesse são muito raros. Por exemplo a depender de sua massa, a taxa de ocorrência do bóson de Higgs pode variar de 0,01 a 0,1 Hz e eventos de supersimetria são esperados a 1 Hz [40]

Provar a existência do bóson de Higgs é um dos principais objetivos do LHC. Com o conhecimento adquirido até agora, não é possível determinar sua massa m_H , embora seu limite inferior ($m_H > 114 \text{ GeV}/c^2$) tenha sido determinado pelos resultados obtidos em outros aceleradores, como o LEP (*Large Electron Positron Collider*, acelerador que operou no CERN de 1989 a 2000) [41]. O limite superior esperado por estudos teóricos é $\sim 1 \text{ TeV}/c^2$ [42].

Considerando os diversos decaimentos possíveis para a partícula de Higgs, espera-se que o canal mais limpo para seu estudo aconteça se sua massa estiver aproximadamente na faixa $150 < m_H < 700 \text{ GeV}/c^2$ [1]. Neste caso, o Higgs pode apresentar

o decaimento em 2 bósons Z, com cada Z, por sua vez, decaindo¹ em dois léptons (elétrons ou múons):

$$H \rightarrow ZZ^{(*)} \rightarrow l^+l^-l^+l^- \quad (2.4)$$

Para este canal de busca do Higgs, é fundamental que os detectores tenham um sistema de filtragem capaz de identificar com alta eficiência elétrons e múons. Estas duas partículas, assim como os fótons, jatos e os taus, são importantes também para o melhor entendimento da supersimetria (SUSY - *supersymmetry*) [6]. Os taus também podem levar aos modelos de Higgs estendidos [43]. Outro fenômeno importante é a energia transversa perdida (E_T^{miss}), que geralmente é atribuído a partículas que passaram pelo detector e não foram detectadas. A E_T^{miss} pode ser identificada a partir da violação dos princípios de conservação de energia e momento.

Um resumo com os principais objetos de interesse no ATLAS e suas aplicações na física é mostrado na Tabela 2.3.

Tabela 2.3: Principais objetos de interesse no ATLAS, extraída de [44].

Objeto	Física de interesse
Elétron	Higgs, SUSY, dimensões-extra, novos bósons e quark top
Fóton	Higgs, SUSY e dimensões-extra
Múon	Higgs, SUSY, extra-dimensões, novos bósons e quark top
Jato	SUSY e ressonâncias
Jato + E_T^{miss}	SUSY e lépton-quarks
Tau + E_T^{miss}	Modelo de Higgs estendido e SUSY

A identificação *online* destes objetos dentro de um universo enorme de informações é realizada pelo sistema de filtragem (*trigger*). Conforme mencionado, um dos objetos de interesse na filtragem do ATLAS é o elétron. Para a identificação de elétrons, a informação obtida no sistema de calorímetros é muito importante. Em termos da calorimetria, as assinaturas de elétrons podem ser confundidas com perfis de energia gerados por alguns jatos hadrônicos (espacialmente concentrados nas camadas eletromagnéticas e com pouca energia nas hadrônicas). Considerando que a produção de jatos será muito frequente nas colisões do LHC, estes formarão

¹No decaimento de partículas elementares, parte da massa da partícula é convertida em energia e o restante em massa de outras partículas

um intenso ruído de fundo para a identificação de elétrons, tornando a discriminação elétron/jato (e^-/j) importante para o desempenho do detector.

2.5 O Sistema de Calorimetria do ATLAS

2.5.1 Breve Introdução à Calorimetria

Calorímetros [12] representam uma importante classe de detectores para medição de energia e posição da partícula. Durante o processo de absorção, as partículas interagem com o material dos calorímetros gerando partículas secundárias que, por sua vez, interagem também gerando outras partículas e assim por diante. Este processo é chamado de cascata ou chuveiro de partículas. Os calorímetros podem apresentar resposta muito rápida, da ordem de nano-segundos [12], e, por isso, são utilizados intensamente pelo sistema de filtragem online (*trigger*). As prováveis classes de partículas são identificadas a partir das características esperadas para o seu perfil de deposição de energia.

Calorímetros são detectores de absorção total. O processo de medição utilizado é destrutivo e as partículas não estão disponíveis após a passagem pelos calorímetros (com exceção aos mísions, que conseguem penetrar em grande quantidade de matéria e necessitam de um detector especial a câmara de mísions). Quando partículas atravessam matéria, elas em geral interagem e perdem assim uma parte de sua energia. Neste processo o meio é excitado ou aquecido (daí o termo calorimetria). Na prática absorção “total” significa 99 a 99,9 % da energia (ou até um pouco menos). A depender do tipo de partícula e da energia envolvida, a partícula pode exceder os limites do calorímetro (vazar) e assim interferir em outros detectores (como a câmara de mísions) [12].

Os calorímetros podem ser classificados em homogêneos ou amostradores. No calorímetro homogêneo todo o volume do detector é sensível às partículas e contribui para produção do sinal. No calorímetro amostrador, o material passivo interage com as partículas (absorvendo sua energia) e o material ativo produz o sinal [12].

A depender do tipo de partícula, a interação com o calorímetro ocorre de modo distinto. Partículas eletromagnéticas (EM), como elétrons e pósitrons, interagem com a matéria gerando um chuveiro de partículas menos energéticas e necessitam de

pequena quantidade de material para serem totalmente absorvidas. Os mísions, por outro lado, perdem sua energia muito lentamente necessitando de grande quantidade de matéria para a absorção total. As partículas hadrônicas interagem com a matéria através da força nuclear forte. O processo de interação é muito mais complexo que o EM e uma variedade muito grande de fenômenos pode ocorrer. Os hadrons podem, por exemplo, se comportar de modo semelhante a elétrons e mísions, ou se envolverem em uma interação nuclear e se transformarem em 15 novos hadrons. Uma parte da energia das interações hadrônicas não é visível (detectável) pelos calorímetros pois os hadrons neutros não ionizam o calorímetro e a energia é perdida em interações nucleares (não detectadas pelo calorímetro) [12].

Como as características dos chuveiros eletromagnéticos e hadrônicos são diferentes, na prática, são utilizados tipos de calorímetro específicos para estas classes de partículas [17]. O calorímetro eletromagnético é usualmente instalado internamente ao hadrônico. As partículas eletromagnéticas (ex: elétrons e fótons) apresentam perfil de deposição de energia que, em geral, é concentrado em torno do ponto de colisão. Tipicamente, as partículas eletromagnéticas são absorvidas completamente nos calorímetros eletromagnéticos. Os chuveiros hadrônicos apresentam formas variadas e iniciam sua interação com o calorímetro eletromagnético, mas, em geral, somente são completamente absorvidas nas camadas hadrônicas (mais externas).

Os calorímetros deveriam ser intrinsecamente lineares para a detecção de partículas EM. Por exemplo um par de elétrons de 10 GeV deveria gerar um sinal de mesma intensidade que um único elétron de 20 GeV. Na prática, desvios do comportamento linear (para partículas EM) podem ser observados na prática devido a fenômenos como [12]:

- Saturação das foto-multiplicadoras (PMT - *Photo-Multipliers*) - as PMT convertem luz dos calorímetros cintiladores em sinais elétricos, sua saturação implica em distorção não-linear dos sinais elétricos;
- Vazamento do chuveiro - com o aumento da energia algumas partículas podem extrapolar os limites do detector, havendo, neste caso, perda de parte da informação;
- Recombinação dos elétrons com íons do material ativo - se isso ocorrer a io-

nização não é detectada;

- Atenuação da luz - a luz emitida pelo material ativo (cintilante) pode ser atenuada antes de atingir as PMT.

Calorímetros homogêneos são intrinsecamente não-lineares para a detecção de hadrons e jatos. A fração EM de chuveiros hadrônicos depende da energia e varia bastante de evento para evento, tornando a resposta hadrônica não-constante em função da energia (tanto para calorímetros homogêneos como para os amostradores). A fração não EM (que produz interações nucleares) é, em geral, menor. Em qualquer calorímetro a resolução em energia para hadrons é pior que para elétrons de mesma energia. Isso se deve ao fato de ocorrerem flutuações na energia visível (detectável) aos calorímetros [12].

Diferente de outros tipos de detectores, a precisão dos calorímetros aumenta com a energia [12]:

$$\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{E}} \quad (2.5)$$

onde E é a energia incidente por partícula e σ_E a variação (desvio padrão) esperado na medição. Outras fontes de flutuações de menor importância contribuem com fatores de outra ordem como ruído eletrônico: $\propto 1/E$ (domina em baixa energia, principalmente em calorímetros de Argônio Líquido - LAr); e vazamento lateral do chuveiro: $\propto 1/(E)^{1/4}$. Existem ainda flutuações que são independentes da energia. Uma característica interessante é que as flutuações podem não ser simétricas em torno do valor médio. Um estudo detalhado a respeito das flutuações encontradas em calorímetros pode ser encontrado em [12].

Para o calorímetro do ATLAS, foi calculada experimentalmente em [38] a resolução esperada . Os valores encontrados foram $0,1/\sqrt{E}$ e $0,4/\sqrt{E}$, respectivamente para os calorímetros eletromagnético (de argônio líquido) e hadrônico (de telhas cintilantes). A seguir, o sistema de calorímetros do ATLAS será descrito mais detalhadamente.

2.5.2 Características dos Calorímetros do ATLAS

O sistema de calorímetros do detector ATLAS [1] é sub-dividido em 7 camadas [38], sendo 4 eletromagnéticas (PS, E1, E2, E3) e 3 hadrônicas (H0, H1 e H2), conforme

Figura 2.6. Cada camada apresenta diferente concentração de células detectoras por unidade de área (granularidade). O calorímetro eletromagnético (EM) é composto de finas folhas de chumbo separadas por dispositivos sensores de argônio líquido, cobrindo a região onde $|\eta| < 3,2$. As três camadas do calorímetro eletromagnético são divididas em barril (região central do detector, onde $|\eta| < 1,5$) e tampa (regiões mais externas onde $1,4 < |\eta| < 3,2$). Na região onde $|\eta| < 1,8$, imediatamente antes da primeira camada EM existe uma fina camada de argônio líquido chamada pré-amostrador (*presampler* ou PS). O pré-amostrador é importante para corrigir medições nas quais existe perda de energia no caminho até os calorímetros.

O calorímetro hadrônico envolve o eletromagnético. Na região onde $|\eta| < 1,7$, ele é composto de placas absorvedoras de aço separadas por telhas de material plástico cintilante. Quando as partículas atravessam as telhas elas emitem luz de intensidade proporcional à energia incidente [45]. O sinal luminoso é então convertido em elétrico através de placas foto-multiplicadoras. Esta parte do calorímetro hadrônico do ATLAS é conhecida como calorímetro de telhas (*Tile Calorimeter* ou simplesmente *TileCal*) e é dividida em barril ($|\eta| < 1,0$) e barril-estendido ($0,8 < |\eta| < 1,7$) [46]. Para a tampa do calorímetro hadrônico ($|\eta| > 1,5$) utiliza-se a tecnologia do argônio líquido.

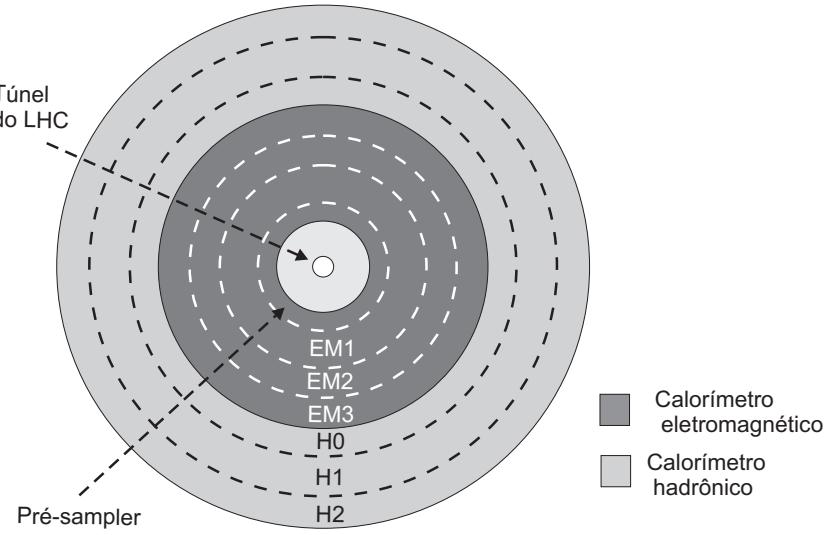


Figura 2.6: Disposição em camadas dos calorímetros do ATLAS.

A informação da energia depositada nas camadas do calorímetro, com fina segmentação, é muito importante para a caracterização das partículas. Após um evento ser aceito pelo sistema de filtragem, todas as informações relativas a este evento são

Tabela 2.4: Região de cobertura em η , granularidade e número de canais de leitura das camadas dos calorímetros.

Pre-amostrador	Barril	Tampa
Cobertura	$ \eta < 1,52$	$1,5 < \eta < 1,8$
Granularidade ($\Delta\eta \times \Delta\phi$)	$0,025 \times 0,1$	$0,025 \times 0,1$
Canais de Leitura	7808	1536 (ambos os lados)
Eletromagnético	Barril	Tampa
Cobertura	$ \eta < 1,475$	$1,375 < \eta < 3,2$
Granularidade ($\Delta\eta \times \Delta\phi$)		
Camada 1	$0,025/8 \times 0,1$	$0,025/8 \times 0,1$ a $0,1 \times 0,1$
Camada 2	$0,025 \times 0,025$	$0,025 \times 0,025$ a $0,1 \times 0,1$
Camada 3	$0,050 \times 0,025$	$0,05 \times 0,025$
Canais de Leitura	101760	62208 (ambos os lados)
Had. Telhas Cintilantes	Barril	Barril estendido
Cobertura	$ \eta < 1,0$	$0,8 < \eta < 1,7$
Granularidade ($\Delta\eta \times \Delta\phi$)		
Camadas 1, e 2	$0,1 \times 0,1$	$0,1 \times 0,1$
Camada 3	$0,2 \times 0,1$	$0,2 \times 0,1$
Canais de Leitura	5760	4092 (ambos os lados)
Had. Argônio Líquido	Tampa	
Cobertura	$1,5 < \eta < 3,2$	
Granularidade ($\Delta\eta \times \Delta\phi$)		
Camadas 1, 2 e 3	$0,1 \times 0,1$ a $0,2 \times 0,2$	
Canais de Leitura	5632 (ambos os lados)	

armazenadas em mídia permanente para posterior análise *off-line*. A granularidade, ou quantidade de células por unidade de área, varia entre as camadas do calorímetro. A Tabela 2.4 traz informações sobre a região de cobertura em η , granularidade e quantidade de canais de leitura (células sensoras) de cada camada do calorímetro.

Conforme ilustrado na Figura 2.7, percebe-se que a primeira camada eletromagnética apresenta mais fina segmentação em η possibilitando medição precisa do

ponto de colisão nessa coordenada. A segunda camada apresenta células detectoras quadradas e maior profundidade, absorvendo maior parcela da energia. A terceira camada, por sua vez, captura os detalhes do final do chuveiro eletromagnético.

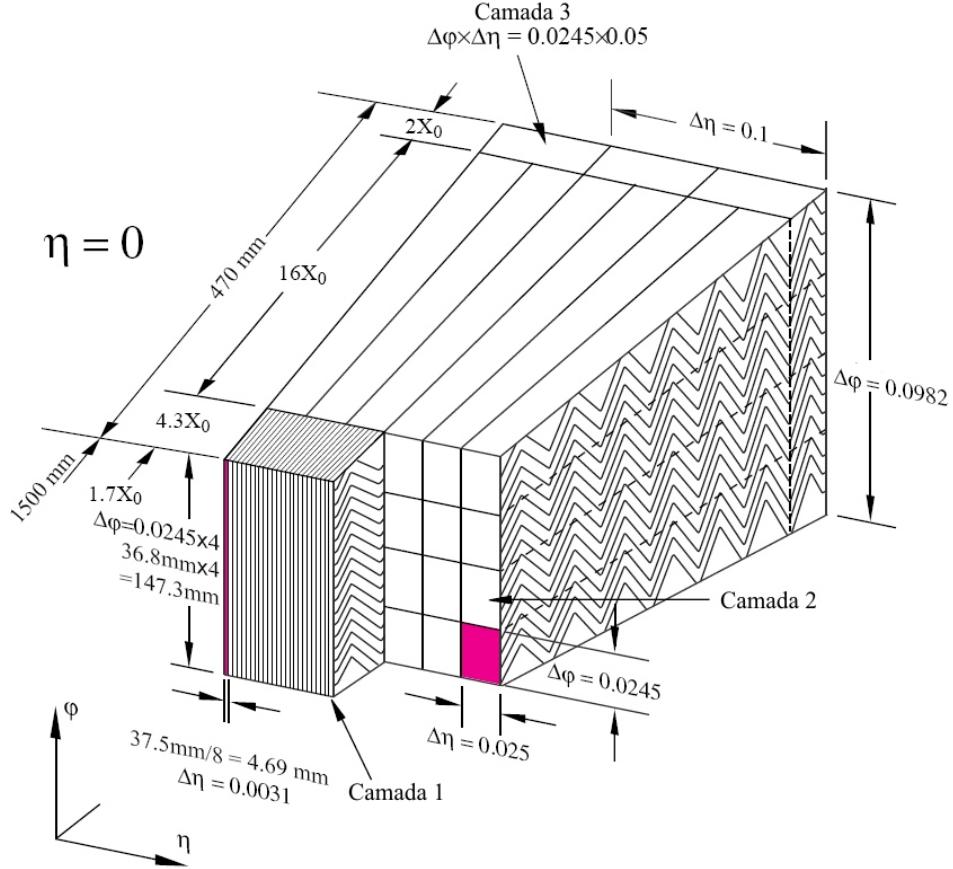


Figura 2.7: Granularidade e profundidade das camadas do calorímetro eletromagnético, extraído de [1].

Calorímetros são construídos utilizando estruturas modulares. Os cabos de transmissão de sinais e alimentação passam por espaços existentes entre os módulos (conhecidos como *cracks*). No ATLAS, na região onde $|\eta| \approx 1,5$ (interconexão entre barril e barril-estendido) existe uma descontinuidade nos calorímetros para passagem de cabos de alimentação e dados do detector de trajetória e do barril do calorímetro eletromagnético [1]. Nesta faixa, há menor quantidade de células detectoras e, consequentemente, baixa resolução nas medições obtidas (o que pode representar um problema para o sistema de filtragem).

2.5.3 Desempenho Esperado dos Calorímetros

Devido a limitações construtivas, a precisão nas medições dos calorímetros do ATLAS varia com a energia e a posição de interação. Uma característica particular do ATLAS é a grande quantidade de material instalado entre o ponto de colisão e o sistema de calorímetros, fato que provoca medição incorreta da energia de elétrons, por exemplo, em 119 GeV 5% da energia de elétrons é perdida neste material. A Figura 2.8 ilustra este problema. Uma fina camada detectora (chamada Pré-Amostrador ou *Pre-Sampler* - PS) foi instalada imediatamente antes da primeira camada do calorímetro eletromagnético numa tentativa de estimar a energia perdida pelas partículas antes de chegarem ao calorímetro. Os sinais medidos no PS são ponderados por um fator α e somados aos sinais medidos nas outras camadas para compor a energia total do evento [12]:

$$E_{tot} = E_{calo} + \alpha E_{ps}. \quad (2.6)$$

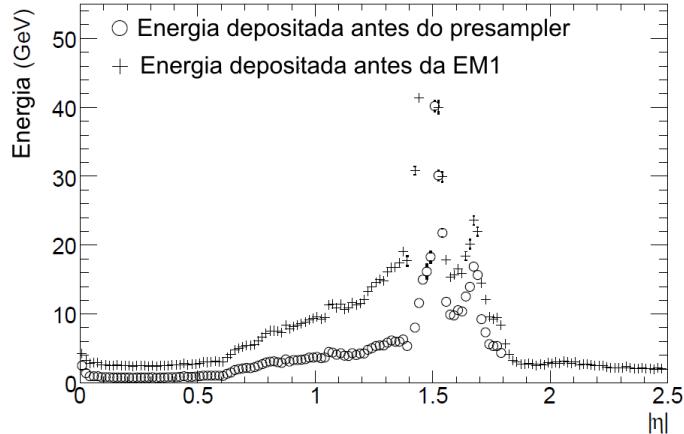


Figura 2.8: Energia perdida por elétrons antes do calorímetro, adaptado de [1].

Um outro fator que pode provocar medição incorreta nos calorímetros é o vazamento de energia de elétrons além da terceira camada eletromagnética. O calorímetro eletromagnético foi projetado para conter todo o chuveiro eletromagnético, porém, principalmente os eventos que incidem na região de transição entre o barril e o barril estendido, por encontrarem menor quantidade de material para interagirem, podem alcançar o calorímetro hadrônico. Essa região é conhecida como *crack*, fica localizada próximo a $\eta = 1.5$ e apresenta um número reduzido de células detectoras.

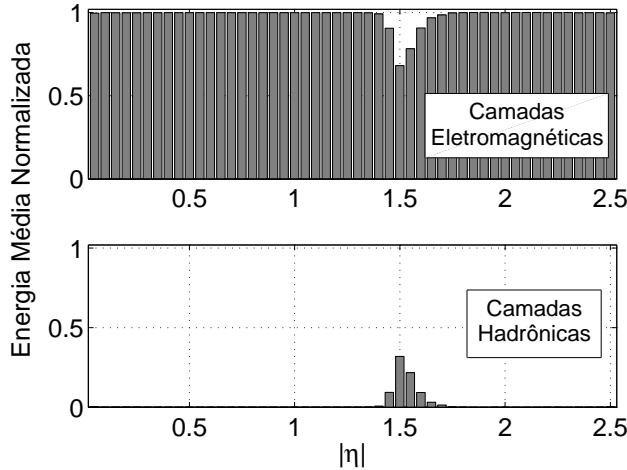


Figura 2.9: Energia normalizada média depositada nas seções eletromagnética e hadrônica em função de η para elétrons.

Conforme ilustrado na Figura 2.9, a energia de elétrons é quase totalmente concentrada no calorímetro eletromagnético, porém, quando a partícula interage próximo ao *crack*, uma parcela significativa de energia chega ao calorímetro hadrônico.

Numa tentativa de corrigir as medições de energia, fatores de ponderação dependentes de η são utilizados na estimativa da energia total do evento [1]:

$$E = s(\eta)[c(\eta) + w_0(\eta).E_{PS} + E_{EM1} + E_{EM2} + w_3(\eta).E_{EM3}] \quad (2.7)$$

sendo s um fator global de correção, c uma tendência (ou *bias*), w_0 corrige a energia perdida antes do calorímetro e w_3 corrige o vazamento longitudinal de energia, os termos E_{PS} , E_{EM1} , E_{EM2} e E_{EM3} representam a energia medida em cada camada do calorímetro eletromagnético (*presampler*, primeira, segunda e terceira camadas eletromagnéticas).

Os erros relativos ($\frac{\sigma_E}{E}$) esperados em função da energia e de η para assinaturas de elétrons são mostrados na Figura 2.10. Percebe-se que a precisão diminui para baixas energias e próximo ao *crack* ($\eta \approx 1.5$). Sempre que o erro de medição aumenta, a caracterização das partículas é prejudicada, provocando queda de desempenho dos algoritmos de filtragem de elétrons.

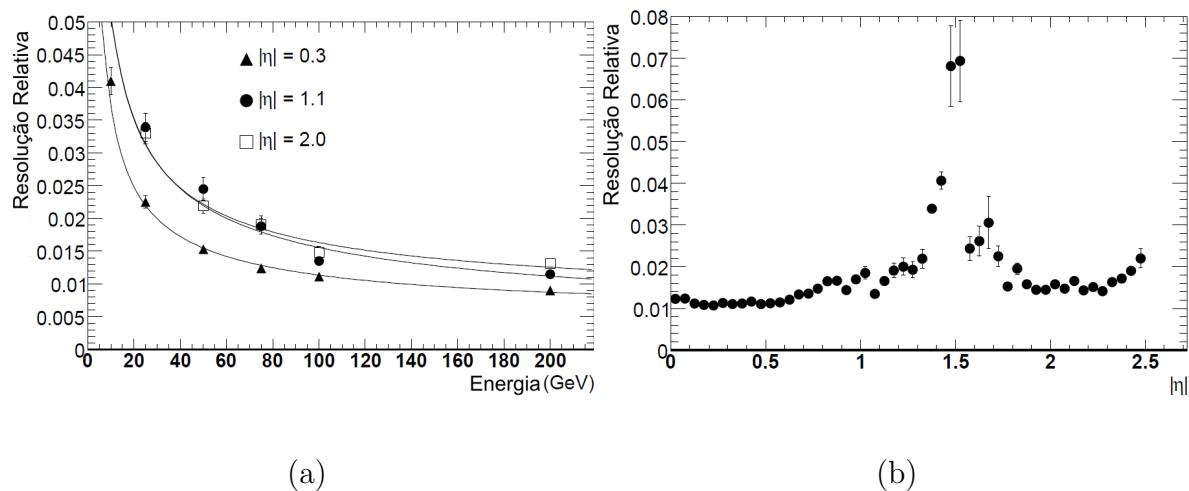


Figura 2.10: Erro relativo do calorímetro na medição da energia de elétrons para diferentes valores de (a) energia e (b) η , adaptado de [1].

Capítulo 3

Filtragem Online no ATLAS

Neste capítulo será apresentada uma breve introdução aos sistemas de filtragem online (*trigger*) em experimentos de física de altas energias (HEP - *High-Energy Physics*) e em seguida serão descritas as características gerais do sistema de filtragem e aquisição de dados (TDAQ - *Trigger and Data Acquisition*) do detector ATLAS.

3.1 Introdução aos Sistemas de Filtragem em HEP

A maioria dos fenômenos físicos que são estudadas atualmente em experimentos de física de altas energias são raros, pois a grande parte da informação produzida representa processos já conhecidos (identificados e estudados anteriormente em outros experimentos) [47]. Por exemplo, o LHC produzirá uma taxa de eventos da ordem de 10^9 Hz para alta luminosidade ($L = 10^{34}\text{cm}^{-2}\text{s}^{-1}$). Dependendo de sua massa, a taxa de produção esperada para o bóson de Higgs varia entre 10^{-1} e 10^{-2} Hz [40]. Neste caso a frequência de interesse é de 10^{10} a 10^{11} vezes menor que a taxa de eventos produzidos, o que significa dizer que todo o restante da informação produzida representa ruído de fundo para a identificação da física de interesse.

Conforme ilustrado na Figura 3.1, os sistemas de *trigger*, em geral, utilizam diferentes níveis hierárquicos de filtragem, onde os níveis mais baixos operam em janelas de tempo extremamente curtas e são responsáveis pela rejeição de eventos utilizando critérios mais simples e óbvios, enquanto que os níveis mais altos implementam análises mais complexas, pois dispõem de mais tempo para a tomada

de decisão. Como os níveis são hierárquicos (sequenciais), uma vez que o evento foi rejeitado em um dado nível ele não está mais disponível para análise nos níveis posteriores [47].

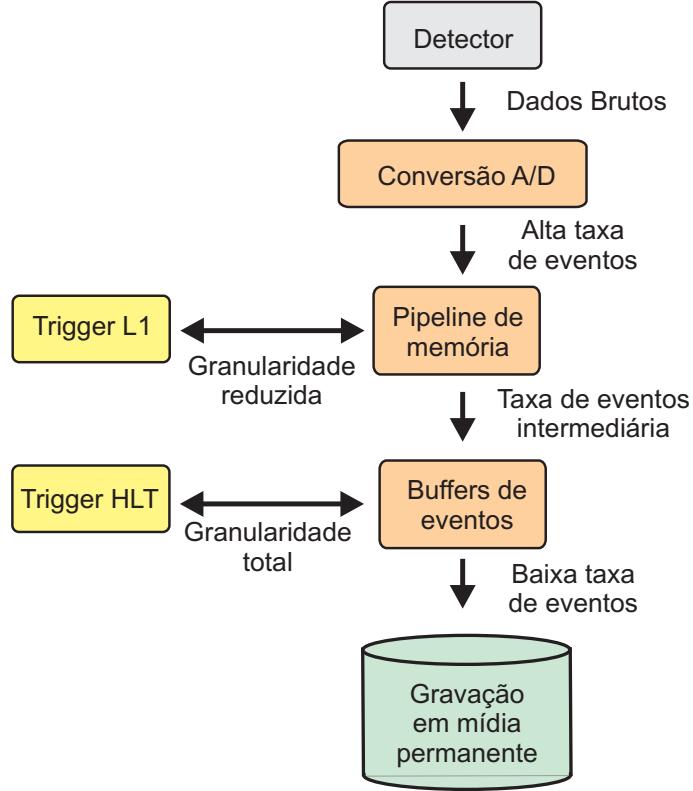


Figura 3.1: Diagrama em blocos do sistema de filtragem genérico para um experimento de HEP.

O primeiro nível de filtragem (*L1 - Level One*) tem disponível um tempo muito curto para tomada de decisão (da ordem de μs), sendo, tipicamente, implementado através de *hardware dedicado* [48], utilizando, por exemplo, *FPGA (Field Programmable Gate Arrays* [49]), *DSP (Digital Signal Processors* [50]) ou *ASIC (Application Specific Integrated Circuit* [51]). Os detectores ATLAS, CMS, CDF [52] e D0 [53], por exemplo, utilizam filtragem de primeiro nível em *hardware dedicado*.

A filtragem de alto nível (*HLT - high-level trigger*) pode ser implementada através de *software* e executado em paralelo por computadores pessoais (PCs). Este estágio dispõe de um maior tempo para produzir a decisão de aceitação ou rejeição e opera com uma menor taxa de eventos (uma vez que uma parcela do ruído de fundo já foi rejeitada pelo primeiro nível). As análises executadas no HLT envolvem operações mais complexas (comparadas ao L1) e podem até requerer a recomposição completa

do evento [48]. O detector H1 do acelerador HERA é uma exceção, pois no segundo nível de filtragem utiliza redes neurais artificiais implementadas em *hardware*, mais detalhes a respeito do sistema de detecção neural do H1 serão fornecidos na próxima seção.

As características de seleção de cada nível variam de acordo com o experimento em questão [47]. Em modernos experimentos de física de alta energia os detectores são divididos em sub-sistemas responsáveis pela detecção de classes específicas de assinaturas. Por exemplo os calorímetros podem ser utilizados para a identificação de candidatos a elétrons, fótons e jatos, já os mísseis necessitam de um detector específico (o sistema de mísseis). O sistema de filtragem utiliza informações destes sub-detectores para identificar as assinaturas da física de interesse.

Considerando os detectores ATLAS e CMS (os dois detectores de propósito geral do LHC), diferentes soluções foram utilizadas para resolver o problema de filtragem e aquisição de dados. No ATLAS, o nível 1 transmite para o nível 2 um sub-conjunto da informação total do detector (conhecido como região de interesse - RoI, do inglês *Region of Interest*) que contém as características necessárias para a identificação da assinatura em questão [54]. O CMS utiliza um sistema de filtragem de primeiro nível que envia toda a informação do evento ao HLT [55].

A utilização das RoI exige inteligência para selecionar adequadamente a informação necessária, mas a grande vantagem é que a taxa de transmissão de dados é reduzida, pois se um evento for rejeitado no nível 2 não há a necessidade de transmitir toda a informação referente ao mesmo. A solução utilizada no CMS exige maior largura de banda para transmissão de dados, porém não há a necessidade de inteligência para efetuar a seleção da RoI. No CMS o problema da largura de banda foi solucionado a partir da utilização de redes de transmissão de dados operando em paralelo e de modo relativamente independente [55].

3.1.1 Aplicações de Redes Neurais Artificiais

As redes neurais artificiais (RNA) [5] de treinamento supervisionado vêm sendo utilizadas em experimentos de física de altas energias, desde o início da década de 1990, com objetivo de auxiliar na detecção *online* dos eventos e também para caracterização dos fenômenos físicos de interesse em análises *offline*. Mais detalhes a

respeito das redes neurais artificiais (e sua utilização como classificador supervisionado) podem ser encontrados no Apêndice B.

Em geral, nas aplicações de filtragem *online* são utilizados classificadores neurais na arquitetura Perceptron de Múltiplas Camadas com treinamento supervisionado [5]. A escolha de redes neurais para aplicações em sistemas de *trigger* é motivada pelo fato do problema ser essencialmente de reconhecimento de padrões (identificação das assinaturas de interesse). Uma outra característica particular é que as assinaturas não tem estrutura temporal, representando as medições instantâneas dos diversos sub-detectores. Finalmente, a possibilidade de realizar implementação paralela (seja em *hardware* ou *software*) possibilita a redução do tempo de processamento [56].

O primeiro sistema de filtragem *online* (*trigger*) baseado em redes neurais foi projetado e testado no detector D0 do FermiLab para a identificação da trajetória de mísseis [57, 58]. O *trigger* neural não foi implementado no experimento, mas operou em paralelo com o algoritmo oficial, obtendo uma resolução 40 vezes melhor. Uma rede neural tipo *Perceptron* de Múltiplas Camadas (alimentada adiante) foi utilizada nesta aplicação. Ainda para o detector D0, conforme detalhado no trabalho [59], foram realizados estudos a respeito da aplicação de uma rede neural para a discriminação entre elétrons (produzidos em decaimentos do tipo $Z \rightarrow ee$) e jatos (jatos duplos com alto momento transverso) a partir de dados simulados. Neste trabalho, sinais simulados do perfil de deposição de energia (dados brutos provenientes da região adjacente ao pico de energia) foram utilizados como entradas para a rede neural. O *trigger* neural obteve eficiência de 90 %, desempenho muito superior ao algoritmo padrão do experimento, que identificou corretamente apenas 75 % dos elétrons.

O primeiro sistema de *trigger* neural operando em um experimento de HEP [60] foi implementado no segundo nível de filtragem do detector H1 (um dos detectores do acelerador HERA, que operou no laboratório DESY na Alemanha) [61]. Conforme descrito em [62], um classificador neural, utilizando funções de ativação do tipo sigmoidal, foi treinado para cada canal de filtragem existente no experimento. No H1, o sistema de *trigger* neural foi capaz de produzir a decisão em $8 \mu\text{s}$ e operou em paralelo com um sistema de detecção tradicional (baseado em cortes lineares

nos parâmetros de interesse), porém, em alguns momentos, o *trigger* neural operou sozinho.

Visando minimizar o tempo de processamento, as redes neurais dos detectores D0 e H1 foram implementadas em *hardware* dedicado [63]. Outras aplicações de rede neural implementadas em *hardware* podem ser encontradas em [64, 65, 66].

No experimento HEGRA (*High-Energy Gamma Ray Astronomy*) [67], composto por um conjunto de detectores de raios cósmicos instalados a 2000 metros de altitude, na ilha La Palma, Espanha, uma rede neural foi utilizada para separar eventos de raios cósmicos carregados eletricamente (γ -induzidos) do ruído de fundo composto por chuveiros induzidos por hadrons (que é 100 vezes mais frequente que a assinatura de interesse). O classificador neural foi alimentado por medições de um conjunto de 221 cintiladores, que possibilitam a reconstrução da direção da partícula primária. Diversas topologias de classificadores neurais foram testadas e foi atingida uma rejeição de 92% do ruído de fundo para aceitação de 60% da física de interesse.

Um *trigger* de segundo nível baseado em redes neurais e informação de calorimetria foi proposto para o detector ATLAS no trabalho [8]. Informações do perfil de deposição de energia nos calorímetros foram utilizados para produzir a identificação de elétrons a partir de uma rede neural. Em [68], numa sequência do trabalho anterior, foi desenvolvido para o segundo nível de filtragem do ATLAS o discriminador *Neural Ringer*, que utilizada informação especialista da física de interesse para pré-processamento dos dados brutos do calorímetro, produzindo-se anéis concêntricos de deposição de energia. A rede neural classificadora (arquitetura MLP) opera sobre os sinais em anéis. Foi obtida eficiência de discriminação superior ao discriminador padrão do ATLAS, que opera a partir de cortes lineares em parâmetros calculados do perfil de deposição de energia. Mais detalhes a respeito do *Neural Ringer* e suas extensões serão fornecidos na seção 4.3.

As redes neurais são aplicadas também na análise *offline* de eventos. Nas aplicações *offline* não existem restrições severas quanto ao tempo de processamento ou risco de “perder” eventos de interesse. Nesse caso trabalha-se com dados gravados em mídia permanente, com o objetivo de estimar, com precisão, as características de cada evento.

No detector ALEPH do experimento LEP, um classificador neural (arquitetura

MLP) foi utilizado para separar os quarks em três classes distintas (quarks b, c e leves). Após testes de desempenho, a topologia escolhida ($20 \times 20 \times 8 \times 3$) utilizou 20 variáveis de entrada, duas camadas escondidas e 3 neurônios na camada de saída (cada um correspondendo a uma das classes) [69]. Neste trabalho, foi obtida eficiência de classificação da ordem de 90 % para cada classe.

Redes neurais artificiais foram utilizadas também para a estimação da massa de top-quarks e eventos de raios cósmicos, respectivamente, no acelerador Tevatron do Fermilab [70] e no observatório Pierre Auger [71]. A massa de top-quarks é extremamente alta se comparado às demais partículas elementares (aproximadamente 40 vezes maior que o segundo quark mais pesado). O motivo para tal característica ainda não está completamente esclarecido na teoria, portanto a medição precisa da massa destes eventos (que são relativamente raros) é muito importante. Com o uso da rede neural em [70], foi obtida a medição mais precisa da massa de top-quarks (até a publicação do referido trabalho) para a assinatura utilizada. Em eventos de raios cósmicos, a estimativa da massa (que é realizada de modo indireto a partir de parâmetros da cascata desenvolvida na atmosfera) é importante na determinação da origem e da natureza dos raios cósmicos primários. No trabalho [71], eventos simulados com energia variando entre 10^{18} e 10^{19} eV foram utilizados para treinamento e teste da rede neural (arquitetura MLP). A rede foi alimentada a partir de parâmetros estimados dos eventos e obteve eficiência da ordem de 99%.

Um exemplo recente de aplicação de rede neural para análise *offline* pode ser encontrado em [72]. Neste trabalho, uma rede *Perceptron* de Múltiplas Camadas foi utilizada para realizar a reconstrução do ponto de interação de raios gama com um detector composto de cristais cintiladores. O treinamento da rede foi realizado de modo supervisionado a partir de eventos experimentais (para baixas energias) e simulados (por *Monte Carlo* para altas energias). Os resultados indicaram que a rede neural foi capaz de realizar a estimativa com alta precisão, superando métodos tradicionalmente utilizados (como os métodos do centroide [73] e da máxima verossimilhança [74]), atingindo precisão menor que 1 mm para baixas energias e aproximadamente igual a 2,1 mm para altas energias.

Mesmo com todos os exemplos de aplicações bem sucedidas de redes neurais, o seu uso está longe de ser uma unanimidade entre a comunidade de HEP. Uma carac-

terística particular do campo de aplicação é que há uma busca pelo entendimento de novos fenômenos, que são representados nos dados simulados (que em geral são utilizados no treinamento dos classificadores) por modelos teóricos aproximados. Um experimento pode concluir que um modelo teórico existente está incompleto ou até mesmo errado, nesse caso o treinamento do sistema ficaria comprometido. O uso de redes neurais é justificado pela facilidade de operação em alta dimensão (inúmeras variáveis são utilizadas no processo de identificação da física de interesse), porém, a dependência dos modelos teóricos é mais difícil de ser verificada ou corrigida nos classificadores não-lineares (em comparação com os cortes lineares mais usualmente utilizados em HEP) [60]. Essa característica particular da HEP talvez tenha produzido uma resistência maior ao uso de redes neurais em comparação a outros campos da ciência. No caso da aplicação no detector H1, toda a colaboração do experimento precisou se convencer que as redes não eram uma caixa preta misteriosa e um esforço conjunto foi feito no sentido de entender como as redes funcionam e como podem ser ajustadas de modo ótimo para cada problema [63]. De um modo geral, o uso de redes neurais está consolidado e bem aceito como ferramenta importante na análise *offline*, porém, no trigger online a situação ainda é ambígua, com grupos a favor e outros contra [60].

3.2 O Sistema de Filtragem *Online* do ATLAS

No sistema de filtragem *online* do ATLAS as estratégias de aceitação de eventos devem garantir que as informações de interesse não sejam perdidas, reduzindo ao máximo a quantidade de ruído de fundo (eventos não-relevantes) gravados em mídia permanente [75].

Considerando as condições de operação do LHC (40×10^6 colisões por segundo, frequencia de interações de 1 GHZ, eventos de interesse raros e imersos em intenso ruído de fundo) e sabendo que, a cada evento (colisão do LHC), no ATLAS são gerados aproximadamente 1,5 MByte de informação, é produzido no detector aproximadamente 60 TBytes de informação por segundo. Com a tecnologia disponível atualmente não é viável armazenar essa quantidade de informação. Mesmo que as leituras de todos os eventos fossem acumuladas, o processo de filtragem *offline* so-

bre toda essa massa de dados não seria possível. Assim, é necessário um sistema eficiente de filtragem *online*.

O sistema de filtragem do ATLAS (usualmente chamado de sistema de *trigger*) acessa informação dos três principais sub-detectores, o detector de trajetórias, os calorímetros e a câmara de mísseis. O *trigger* online opera em 3 níveis sequenciais de seleção de eventos: o nível 1 (L1), o nível 2 (L2) e o filtro de eventos (EF - *event filter*), sendo que os dois últimos juntos compõem a filtragem de alto nível (HLT - *High Level Trigger*) [44]. Cada nível é responsável pela rejeição de uma parcela das assinaturas não relevantes (ruído de fundo), refinando a decisão do nível anterior. A Figura 3.2 ilustra as principais características dos três níveis de filtragem de eventos no ATLAS.

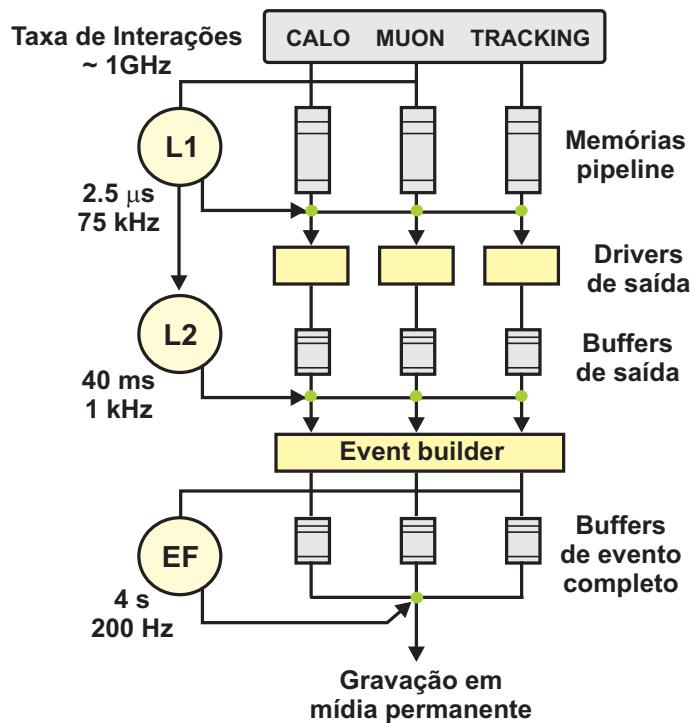


Figura 3.2: Diagrama em blocos do sistema de filtragem e aquisição de dados do ATLAS, extraído de [45].

O primeiro nível tem sérias restrições quanto ao tempo de processamento (latência máxima de $2,5\mu s$), recebendo a plena taxa de eventos do LHC como entrada. Esse nível é implementado em *hardware* dedicado, e para produzir a decisão rapidamente, usa apenas parte da resolução disponível ao detector. A combinação da baixa resolução com critérios de seleção simplificados resulta numa aceitação de

parcela considerável do ruído de fundo pelo primeiro nível, que deve ser gradualmente reduzida pelos níveis subsequentes. Até a decisão do primeiro nível quanto à aceitação ou rejeição do evento, toda a informação do detector relacionada a esse evento é mantida em um *pipeline* de memórias. O L1 entrega ao segundo nível a localização das áreas onde possivelmente aconteceram eventos de interesse, regiões estas conhecidas como RoI (*Regions of Interest*).

A seleção de eventos no L2 é feita através de *software* especializado, rodando em um conjunto de ≈ 700 PCs (Computadores Pessoais) dedicados operando em ambiente de processamento distribuído [54]. Neste nível, o tempo máximo para tomada de decisão é 40 ms e estão disponível informações (que podem ser descarregadas através pelos *drivers* de saída - ROD *Read-Out Drivers*) do detector de trajetórias, das câmaras de múons, assim como a total resolução dos calorímetros. Dispondo de um tempo maior e da total resolução do detector, o L2 utiliza critérios de seleção mais refinados (em comparação com o L1), reduzindo consideravelmente o ruído de fundo, sem perder muitas assinaturas de interesse.

O filtro de eventos é o último estágio do sistema de filtragem, recebendo uma taxa de eventos mais baixa, tendo latência de alguns segundos para tomada de decisão. O EF tem acesso a todo o evento (não somente à RoI como no L2) e utiliza técnicas semelhantes às da reconstrução *offline* operando num conjunto de ≈ 2000 PCs [54]. Os eventos aprovados nesse nível são armazenados em mídia permanente para posterior análise *offline*.

A estratégia de processamento sequencial permite que os eventos sejam rejeitados na primeira etapa possível, minimizando a necessidade de acesso a informações, e facilitando o ajuste e eventual modificação das estratégias de extração de características. Na Tabela 3.1 é apresentado um resumo das principais características dos 3 níveis de filtragem do ATLAS.

Um problema que afetará os algoritmos de extração de características é o efeito de empilhamento (do inglês *pile-up*), que ocorre quando há uma sobreposição de assinaturas em regiões do detector [12], ou seja, uma assinatura que ainda se desenvolve tem seu padrão de deposição de energia distorcido por uma nova que chega e se sobrepõe, gerando um ruído de fundo que pode atingir grande intensidade.

Para o projeto e teste dos métodos de extração de características, foram usa-

Tabela 3.1: Principais características do sistema de *trigger* do ATLAS, onde Te e Ts são, respectivamente, as taxas de eventos na entrada e na saída e $Cr = Te/Ts$ é o coeficiente de redução de eventos.

Nível	Te (Hz)	Ts (Hz)	Cr	Latência (s)	Implementação
L1	40×10^6	$75 \times (100)10^3$	533 (400)	$2,5 \times 10^{-6}$	<i>Hardware</i>
L2	$75(100) \times 10^3$	$3,5 \times 10^3$	21 (29)	40×10^{-3}	<i>Software</i>
EF	$3,5 \times 10^3$	≈ 200	17	≈ 4	<i>Software</i>

dos conhecimentos prévios adquiridos em outras experiências com aceleradores de partículas e eventos simulados através de técnicas de Monte Carlo [14]. As simulações utilizam modelos estocásticos que descrevem as interações, levando em conta as características físicas do acelerador e do detector, assim como os efeitos de cada nível de filtragem. Para o ATLAS, foram utilizados geradores de eventos para colisões próton-próton como HERWIG, ISAJET, GEANT e PYTHIA, descritos em [45] e [76]. Os algoritmos de classificação e extração de características foram projetados para os dados simulados e posteriormente adaptados para a realidade de operação quando do início da aquisição de dados.

3.2.1 Primeiro Nível de Filtragem

O primeiro nível de filtragem (L1) é responsável por reduzir a taxa de eventos de aproximadamente 40 MHz para 75 kHz (a taxa de saída do L1 pode ser aumentada até 100 kHz, a depender das condições de operação do detector), implicando numa redução da ordem de 500 vezes. A decisão do primeiro nível deve ser tomada até $2,5\mu s$ após o cruzamento de feixes (colisão) ao qual o evento está associado.

O L1 identifica as assinaturas básicas de interesse e, para tornar mais rápida a tomada de decisão, a granularidade dos subsistemas do detector é menos fina [75]. Por exemplo, considerando o sistema de calorímetros, que possui mais de 100.000 células detectoras, o L1 utiliza apenas a informação de 7000 “torres” de soma analógicas (as torres são obtidas somando-se a energia de células dentro de regiões de $0,1 \times 0,1$ em $\Delta\eta \times \Delta\phi$) [77].

A tomada de decisão quanto à aceitação ou rejeição de eventos no nível 1 é feita pelo processador central de *trigger* (*central trigger processor* - CTP), que combina

as informações disponíveis nos calorímetros, para a detecção de partículas eletromagnéticas e hadrônicas, e nos sub-detectores RPC (*Resistive Plate Chamber*) e TGC (*Thin Gap Chambers*) para a detecção de múons, conforme mostrado na Figura 3.3. O filtro dos calorímetros é dividido em três sub-sistemas (Pré-processador, Processador de Regiões e Processador de Soma de Energia/Jato). O pré-processador digitaliza os sinais medidos e envia as informações para os processadores de regiões (CP - *Cluster Processor*) e de soma de energia / jatos (JEP - *Jet/Energy Sum Processor*). O CP é responsável pela identificação dos candidatos a elétrons, fótons e τ -leptons e o JEP identifica candidatos a jatos e produz a soma global de energia do evento. Considerando o filtro de múons, cada um dos sub-detectores é responsável pela identificação dos candidatos a múons em uma região do detector, o RPC opera no barril e o TGC na tampa. Se a assinatura analisada satisfaz um critério de aceitação de algum dos sub-detectores (calorímetros, RPC ou TGC), então ela é aceita pelo L1 e enviada para uma análise mais criteriosa na filtragem de alto-nível.

Até a tomada de decisão de nível 1, toda a informação do evento é armazenada em memórias tipo *pipeline*. Quando um evento é aceito, as informações referentes a este são descarregadas para uso pelo nível 2 de filtragem. As informações dos eventos rejeitados são descartadas (não estando mais acessível para a filtragem de alto-nível).

O L1 é responsável por fornecer ao HLT informações sobre a posição (no plano η, ϕ) onde os eventos aceitos ocorreram, sinalizando assim as regiões de interesse (RoI) para análise no L2. O nível 1 também fornece outras informações importantes, como o critério utilizado para aceitação do evento e a identificação da colisão (cruzamento de feixe) ao qual o evento está associado. O sistema de filtragem do L1 é implementado utilizando hardware dedicado (através de FPGAs) conforme detalhado em [78].

Na Tabela 3.2 são mostradas as freqüências dos principais canais de filtragem esperadas na saída do L1 quando o LHC estiver operando em alta luminosidade ($L=10^{34}\text{cm}^{-2}s^{-1}$). Os principais objetos de *trigger* a serem identificados são candidatos a múons, elétrons/fótons (regiões eletromagnéticas), taus, jatos, e E_T^{miss} . Pode-se perceber que a taxa total desta simulação é da ordem de 40 kHz, aproximadamente 2 vezes menor que a freqüência de saída esperada para o nível 1 na

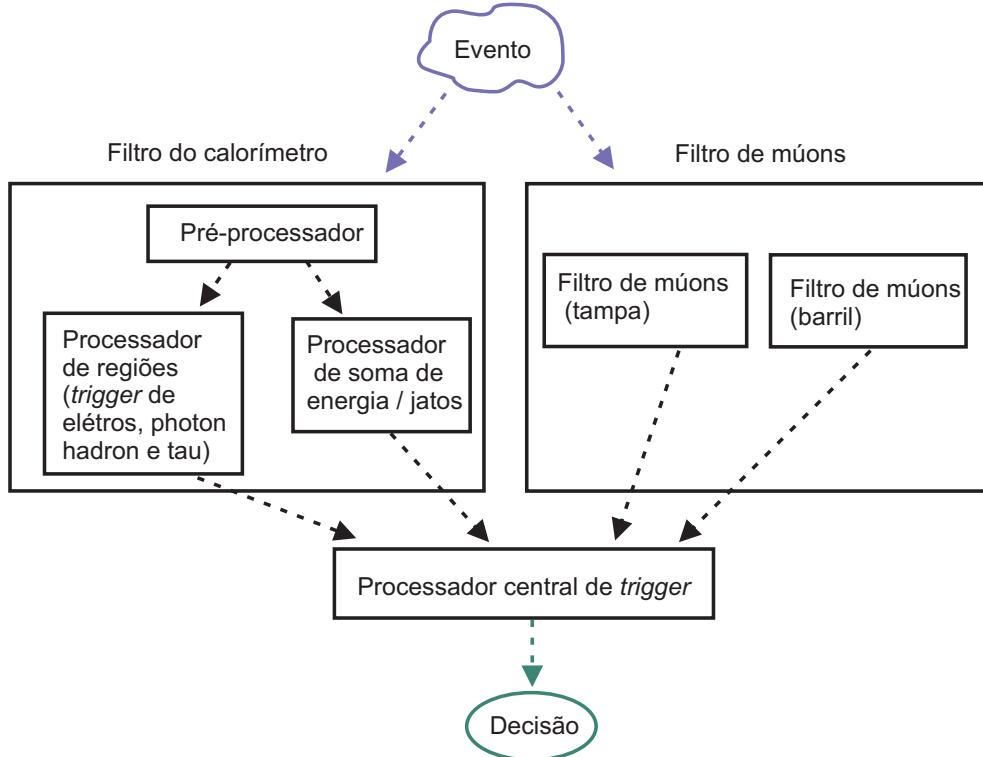


Figura 3.3: Diagrama em blocos do primeiro nível de filtragem.

operação do LHC [75].

3.2.2 Filtragem de Alto Nível

O sistema de filtragem de nível 2 (L2) e o filtro de eventos (EF) são responsáveis pela filtragem de alto nível (HLT - *high-level trigger*) do ATLAS. O nível 2 deve reduzir a taxa de eventos de 75 kHz (podendo chegar até 100 kHz) para $\approx 3,5$ kHz, tendo um tempo de latência de aproximadamente 10 ms para tomar a decisão [76]. O EF precisa diminuir a taxa de eventos de 3,5 kHz para 200 Hz. Os eventos que forem aceitos pelos três níveis de filtragem serão armazenados em mídia permanente para futura análise *offline*. O tempo para a tomada de decisão no filtro de eventos é de alguns segundos. O HLT é implementado em *software* e opera em um conjunto de PCs (≈ 2800) operando em paralelo [54]. Considerando os efeitos conjuntos do segundo nível e do filtro de eventos, a filtragem de alto nível deve reduzir em aproximadamente 500 vezes a taxa de eventos.

O segundo nível opera guiado pelas informações da RoI fornecidas pelo L1. Conforme mostrado na Figura 3.4, após um evento ser aceito pelo L1, as informações

Tabela 3.2: Freqüência esperada para os principais canais de *trigger* no primeiro nível de filtragem do ATLAS ($L=10^{34}\text{cm}^{-2}s^{-1}$) , extraída de [75].

Canal	Freqüência (kHz)
Um múon	4
Par de múons	1
Região eletromagnética	22
Par de regiões eletromagnéticas	5
Um jato	0,2
Três jatos	0,2
Quatro jatos	0,2
Jato e E_T^{miss}	0,5
Tau e E_T^{miss}	1
Múon e região eletromagnética	0,4
Outras condições	5
Total	≈ 40

(fragmentos) da RoI geradas por diferentes detectores são descarregadas no construtor de RoI (RoIB - *RoI Builder*). Com o uso das RoIs, apenas $\approx 2\%$ da informação do detector é necessária para produzir a decisão do L2, reduzindo consideravelmente a taxa de transmissão na rede de dados [54]. O RoIB agrupa os fragmentos e transmite o registro produzido para um supervisor do segundo nível (L2SV - *Level 2 supervisor*), que atribuirá a RoI recebida a uma unidade de processamento do L2 (L2PU - *Level 2 processing units*). A L2PU executa os algoritmos de seleção do segundo nível, que utilizam plena granularidade dos detectores e produzem a decisão (aceite ou rejeição) do L2. Enquanto a decisão do segundo nível é aguardada, todas as informações dos eventos aceitos pelo L1 são armazenadas nos ROBs (*Read-Out Buffers*). Quando a decisão do L2 é produzida, o L2PU informa ao L2SV, que em caso de aceite, envia as informações do evento completo (que estavam armazenadas temporariamente nos ROBs), através da rede do L2, para o Filtro de Eventos (EF). Quando o evento é rejeitado, as informações referentes a ele são descartadas.

As principais características desejadas para os algoritmos de filtragem no L2 são listadas a seguir [76]:

- alta eficiência ($> 95\%$) por RoI selecionada no L1;
- eficiência uniforme em η (o que é difícil pois o detector apresenta descontinuidades) e eficiência uniforme ou crescente com E_T ;
- redução do ruído de fundo minimizando a taxa de eventos classificados de forma incorreta (falso alarme);
- robustez em relação à luminosidade, ruído de medição, imperfeições de alinhamento e calibração.

Quando um evento é aceito pelo L2, o construtor de eventos (EB - *Event Builder*) coleta nos ROBs toda a informação do evento e a disponibiliza ao Filtro de Eventos (EF - *Event Filter*) para a última etapa da filtragem *online*. O evento completo é armazenado nas Fazendas de Entrada do Filtro de Eventos (EF *Sub-Farm Input -SFI*) [40]. O EF reduz a taxa de eventos a 200 Hz tendo disponível até 4 segundos para a tomada de decisão. Os algoritmos do EF analisam todo o evento (não se restringem apenas à RoI como no L2) e operam de modo semelhante à análise *offline* [54]. Quando o evento é aceito pelo EF, todas as informações referentes a ele são armazenadas (gravadas) em mídia permanente.

O HLT foi desenvolvido utilizando, sempre que possível, tecnologias padronizadas (disponíveis comercialmente) [79]. Todos os processadores utilizados são de uso geral (semelhantes aos utilizados em computadores pessoais) e praticamente toda a comunicação é feita através de redes *Gigabit Ethernet* [80]. As aplicações estão sendo desenvolvidas utilizando C++ [81]. Estas escolhas foram feitas considerando-se padronização, velocidade, confiabilidade e facilidade na reposição de equipamentos.

Mais informações sobre os componentes dos sistemas de filtragem podem ser encontradas em [75], [76], [45] e [43].

Desafios do HLT

Conforme mencionado anteriormente, a filtragem de alto nível pode utilizar toda a granularidade e precisão dos subdetectores do ATLAS. Então, percebe-se que, os algoritmos do HLT operam num espaço de decisão multidimensional, com restrições no tempo de processamento e rígidos padrões de eficiência. Deve-se notar também

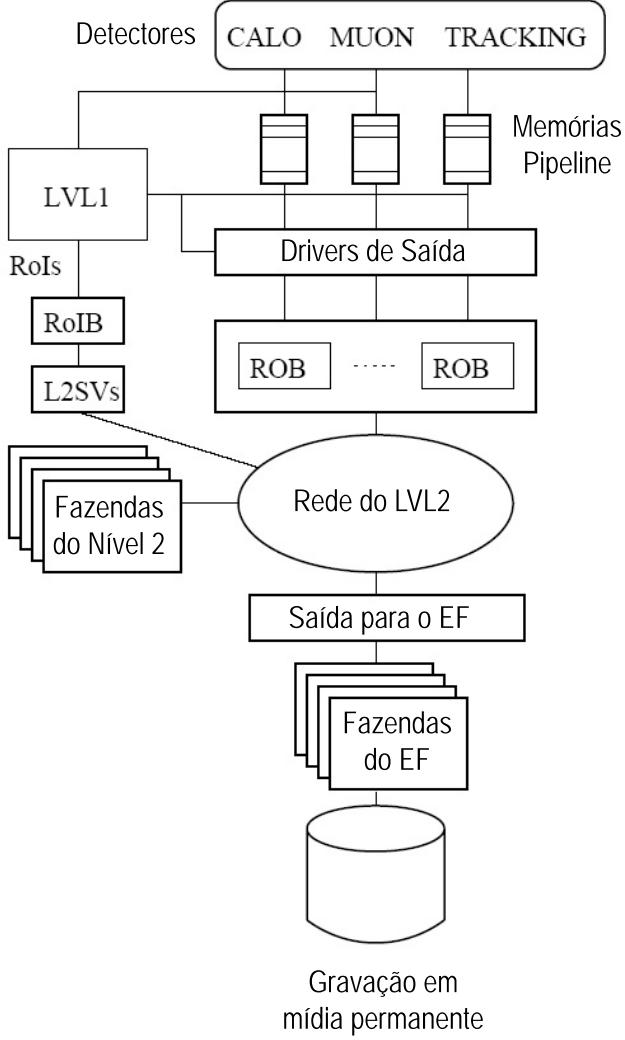


Figura 3.4: Diagrama em blocos do segundo nível de filtragem.

que, as características do sistema de filtragem podem mudar com o melhor conhecimento do detector, após os testes e o início de operação. Percebe-se, então, que o sistema de *trigger* de alto nível demanda esforços no sentido de propor, testar e comparar diferentes estratégias de seleção.

Embora existam algoritmos desenvolvidos pela colaboração do ATLAS para a filtragem de alto nível (nos diversos canais de interesse), pesquisas continuam sendo conduzidas com o objetivo de propor rotinas de filtragem alternativas que fornecem maior eficiência de discriminação da física de interesse e, ao mesmo tempo, maior rejeição do ruído de fundo.

Neste contexto, se encaixa este trabalho, que visa propor uma otimização ao Neural Ringer [9, 10] que é um discriminador alternativo para o canal elétron/jato

e apresenta desempenho superior ao T2Calo (discriminador oficial do ATLAS) e tempo de processamento dentro da janela permitida ao L2 (≈ 40 ms).

No próximo capítulo será descrito o processo de identificação de elétrons a partir de informações de calorimetria. Considerando o L2 do ATLAS, serão descritos os discriminadores T2Calo e *Neural Ringer*.

3.2.3 Plataforma de *Software* do Sistema de Filtragem

A colaboração do ATLAS desenvolveu um conjunto de softwares e ferramentas de controle (conhecidos como Athena) [82] que permitem aos membros da colaboração, independente de sua localização geográfica, acesso e análise dos dados gerados no detector.

O Athena, em sua estrutura modular, dispõe de rotinas para:

- Simulação (incorporando algoritmos como o Pythia, que simula a colisão dos feixes de prótons e o Geant, que simula a interação das partículas com a matéria e o comportamento dos detectores);
- Filtragem (*Trigger*);
- Reconstrução do evento;
- Análise da física.

O núcleo (*kernel*) do Athena é baseado no projeto Gaudi [83], desenvolvido originalmente para o detector LHCb e posteriormente adaptado para as necessidades do ATLAS. O Athena, na verdade, define uma estrutura (*framework*) de controle comum para todos os aplicativos e análises necessárias para a colaboração do ATLAS.

Entre os principais benefícios do uso de uma estrutura comum pode-se mencionar:

- Os desenvolvedores compartilham a mesma estrutura, onde podem inserir os seus próprios códigos, de acordo com sua necessidade específica;
- A comunicação e compatibilidade entre os diversos componentes é mais facilmente garantida;

- Facilidade para re-uso de código, poupando tempo no desenvolvimento de rotinas mais complexas.

O Athena está sendo utilizado no desenvolvimento e teste de algoritmos tanto para seleção de eventos (filtragem *online*), como para reconstrução e análise da física (modo *offline*).

Capítulo 4

Detecção de Elétrons a partir de Informações de Calorimetria no ATLAS

Neste capítulo serão descritos os principais aspectos do processo de filtragem *online* de elétrons no detector ATLAS. O conhecimento adequado da energia de elétrons em uma larga faixa (de alguns GeV a poucos TeV) é necessário para a busca pela partícula de Higgs no decaimento $H \rightarrow ZZ^* \rightarrow 4e$ (conforme previsto no Modelo Padrão) e para medições precisas de fenômenos não descritos pelo Modelo Padrão. A detecção de elétrons é também muito importante para os processos de calibração e alinhamento do detector durante a fase inicial de operação. Com a análise de eventos conhecidos (como por exemplo o decaimento $Z \rightarrow ee$) pode-se gerar informações valiosas para o melhor conhecimento das características do detector [84]. A discriminação destas partículas é baseada, principalmente, nas informações do sistema de calorímetros. Considerando informações do perfil de deposição de energia medido nos calorímetros, os elétrons podem ser confundidos com jatos hadrônicos (que neste caso representam ruído de fundo para o experimento). No LHC, a geração de jatos será bastante intensa (por exemplo, para energia de 40 GeV a relação de produção elétron/jato é da ordem de 10^{-5} [1]) o que tornará difícil a identificação da física de interesse. Tradicionalmente, a identificação de elétrons é realizada analisando-se o formato transversal do chuveiro e o vazamento para as camadas hadrônicas [85]. Estes parâmetros variam com a energia da partícula e nem sempre são capazes de

produzir a eficiência de discriminação desejada.

4.1 Filtragem de Elétrons no L1

Considerando as características desejadas para o primeiro nível de filtragem (L1), onde a taxa de entrada é da ordem de 10^6 eventos por segundo e uma taxa de redução da ordem de 500 vezes é exigida em no máximo $25\mu\text{s}$, para tornar mais rápido o processo de decisão, este nível opera com granularidade reduzida no sistema de calorímetros. No L1, as células dos calorímetros são somadas (através de somadores analógicos) para produzir os sinais conhecidos como torres de *trigger* (TT). O tamanho das torres de *trigger* é diferente para as camadas eletromagnéticas e hadrônicas. No calorímetro eletromagnético, cada TT cobre uma área de $0,1 \times 0,1$ no plano $\eta \times \phi$ (ver Figura 4.1), já no calorímetro hadrônico, cada TT representa uma região de $0,2 \times 0,2$ no plano $\eta \times \phi$.

A filtragem de elétrons no L1 é baseada em cortes lineares nos parâmetros do perfil de deposição de energia medido nos calorímetros. A definição dos critérios de seleção leva em conta o conhecimento das características típicas do perfil de objetos eletromagnéticos (elétrons e fótons), que geralmente apresentam:

- concentração ao redor do ponto de colisão (centro da região de interesse - RoI);
- alta concentração de energia na seção eletromagnética;
- baixa concentração de energia na seção hadrônica.

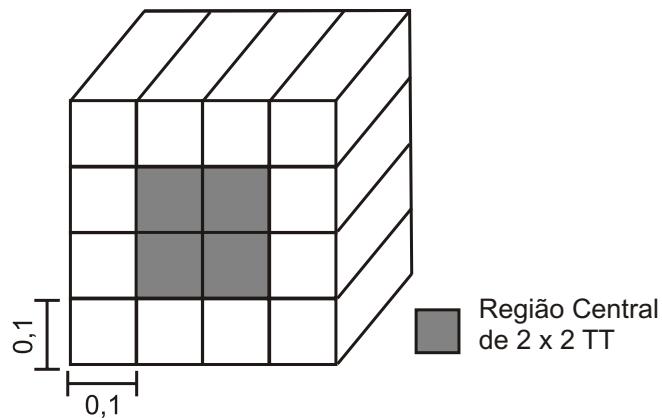


Figura 4.1: Janela deslizante analisada pelo L1 no calorímetro eletromagnético.

A partir da análise de uma janela deslizante, cobrindo uma região de 4×4 TT (16 no total), alguns critérios podem ser utilizados pelo L1 (a depender das características desejadas para a física de interesse pode-se variar os patamares de corte) para definir um possível candidato a elétron:

1. Os valores de energia das TTs da região central (de 2×2 TT) são somados dois a dois e utiliza-se o maior valor encontrado, que deve exceder um patamar de energia eletromagnética para o evento ser aceito neste critério;
2. O nível de energia na periferia (fora da região central de 2×2 TT) é calculado para verificar o isolamento em energia do objeto em questão. Para o evento ser aceito a energia periférica deve ser menor que o patamar estabelecido;
3. A energia depositada nas camadas hadrônicas é somada para verificar o vazamento de energia fora do calorímetro eletromagnético (isolamento hadrônico). Para o evento ser aceito a energia hadrônica deve ser menor que o patamar estabelecido.

Os patamares de seleção do nível 1 podem ser ajustados e os critérios combinados, a depender das características da física de interesse que se deseja estudar num dado momento de operação do detector.

Neste trabalho, foram utilizados sinais simulados do canal elétron/jato aprovados em dois cortes de primeiro nível diferentes. No primeiro conjunto, foi realizado apenas um corte em energia num patamar igual a 10 GeV, esse corte é conhecido como E10 (ou E10-*loose*, para indicar que não houve isolamento). No segundo conjunto, o patamar de corte em energia utilizado foi 15 GeV, e foram realizados também cortes por isolamento (periférico e hadrônico), produzindo os sinais do conjunto E15i (onde o sufixo i indica a utilização de cortes por isolamento). Mais detalhes a respeito das características dos conjuntos simulados E10 e E15i serão fornecidos no Capítulo 7.

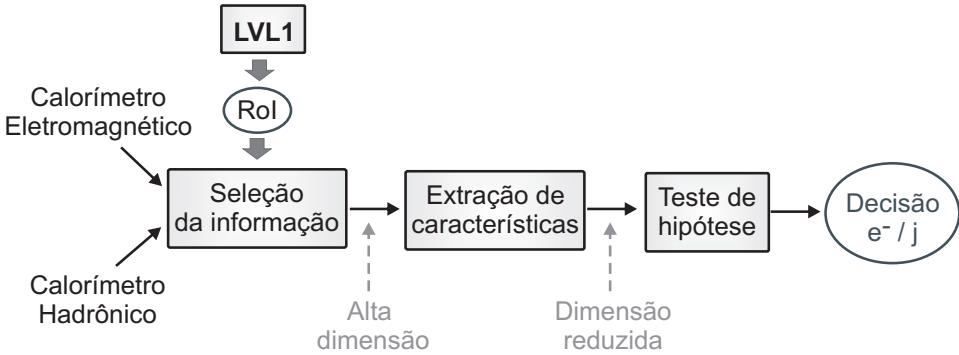


Figura 4.2: Processo de identificação de elétrons no L2.

4.2 Filtragem de Elétrons no L2 - Algoritmo T2Calo

Conforme mencionado anteriormente, o segundo nível de filtragem (L2) tem aproximadamente 40 ms para produzir a decisão (aceitação/rejeição) das assinaturas aprovadas pelo L1. O L2 opera apenas sobre a região de interesse (RoI) indicada pelo L1, que no caso da identificação de elétrons envolve os calorímetros e, em alguns casos, o detector de trajetória (*tracking system*).

O processo de discriminação de elétrons no L2 baseado em informações de calorimetria, conforme ilustrado na Figura 4.2, pode ser dividido em duas etapas distintas. Inicialmente é executada a extração de características (onde parâmetros discriminantes são estimados a partir dos dados brutos medidos nos calorímetros) e, em seguida, o teste de hipóteses (onde a discriminação propriamente dita é realizada a partir das características estimadas).

O **T2Calo** [76] é o algoritmo padrão no L2 do ATLAS para a identificação de objetos eletromagnéticos (elétrons ou fôtons) a partir da informação dos calorímetros. O T2Calo utiliza parâmetros que estimam a forma dos chuveiros de deposição de energia, operando de modo semelhante ao algoritmo de filtragem do L1, porém agora, toda a granularidade dos calorímetros está disponível.

O primeiro passo é refinar a posição do centro da RoI fornecida pelo L1, encontrando a célula de maior energia (célula mais quente) na segunda camada do calorímetro eletromagnético (η_1, ϕ_1). Essa posição será posteriormente refinada pelo cálculo da posição da média ponderada da energia em uma janela de 3×7 células

em (η, ϕ) .

Para a seleção dos objetos eletromagnéticos, o T2Calo estima os parâmetros descritos a seguir:

- Razão de Forma: $R_{shape} = E_{3 \times 7}/E_{7 \times 7}$, onde $E_{n \times m}$ é a energia depositada numa janela de $n \times m$ células em torno de (η_1, ϕ_1) na segunda camada eletromagnética;
- Razão de Energia: $E_{ratio} = (E_{1st} - E_{2nd})/(E_{1st} + E_{2nd})$, onde E_{1st} e E_{2nd} são o máximo e o segundo maior valor de energia encontrados na primeira camada eletromagnética numa janela de $\Delta\eta \times \Delta\phi = 0,125 \times 0,2$ em torno do centro da RoI (η_1, ϕ_1) ;
- A energia eletromagnética total: E_{EM} é calculada a partir da soma da energia concentrada em janelas de 3×7 células em torno de (η_1, ϕ_1) nas três camadas eletromagnéticas;
- A energia hadrônica total: E_{HAD} é calculada a partir da soma da energia concentrada em janelas de $\Delta\eta \times \Delta\phi = 0,2 \times 0,2$ em torno do centro da RoI nas três camadas hadrônicas.

Considerando que o perfil de deposição de energia dos elétrons é, em geral, mais concentrado ao redor do ponto de máximo e contido na seção eletromagnética do calorímetro, o T2Calo opera através de cortes lineares nos parâmetros estimados acima. Por exemplo, analisando os parâmetros R_{shape} , E_{ratio} e E_{EM} , os eventos são aprovados caso os valores calculados sejam maiores que um patamar de corte pré-estabelecido. Para E_{HAD} acontece o inverso, e um candidato a elétron é selecionado se o valor calculado para este parâmetro for menor que o patamar.

Embora a concentração em torno do centro da RoI e o vazamento hadrônico sejam características bastante úteis na identificação de elétrons, uma análise um pouco mais detalhada revela que, pelo fato dos jatos apresentarem uma grande flutuação em suas características, o uso apenas destes parâmetros pode não produzir o desempenho de discriminação desejado. Na Figura 4.3, considerando a razão $\frac{E_{HAD}}{E_{EM}}$ (num conjunto de eventos simulados para o L2 do ATLAS), percebe-se que há uma razoável superposição entre elétrons e jatos, essa limitação aumenta em baixa

energia. Comportamento semelhante é verificado na Figura 4.4, onde elétrons e jatos são comparados a partir da razão de forma ($R_{shape} = \frac{E_{3 \times 7}}{E_{7 \times 7}}$).

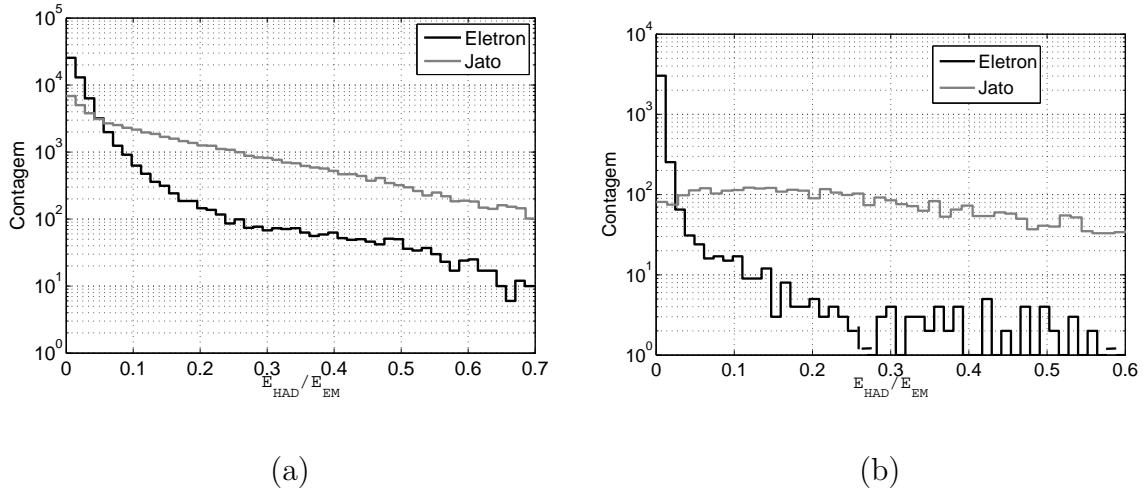


Figura 4.3: Distribuição de elétrons e jatos quanto à razão $\frac{E_{HAD}}{E_{EM}}$ para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.

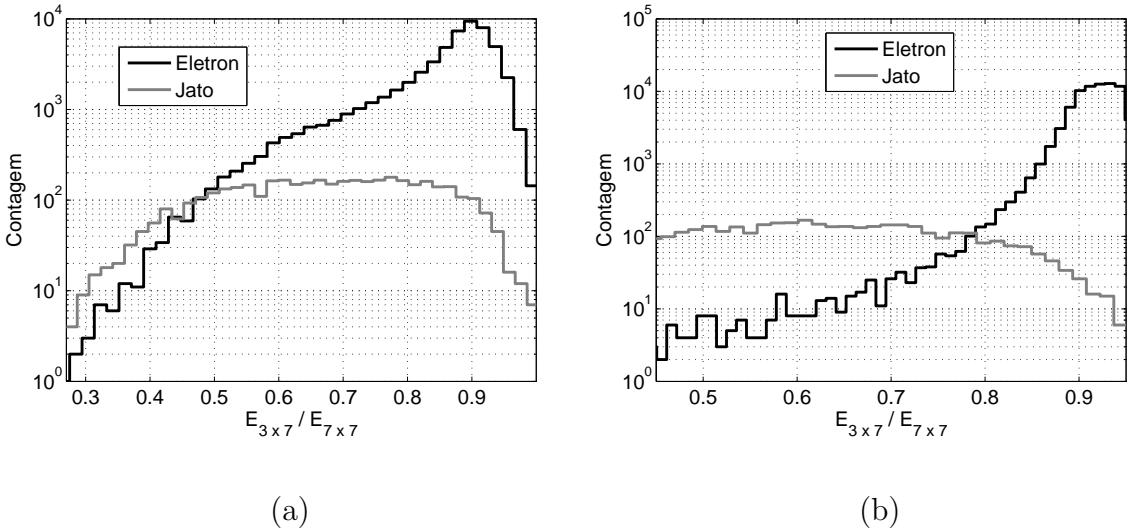


Figura 4.4: Distribuição de elétrons e jatos quanto à razão de forma ($R_{shape} = \frac{E_{3 \times 7}}{E_{7 \times 7}}$) para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.

Num ambiente complexo, como o que é gerado a cada colisão do LHC, onde o nível de ruído de fundo é extremamente alto e os eventos de interesse são raros, a busca por algoritmos de filtragem mais eficientes é muito importante. Aumentar o desempenho de discriminação significa menor quantidade de eventos não relevan-

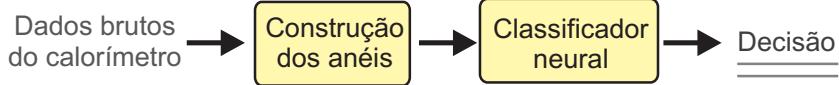


Figura 4.5: Fluxo de processamento do *Neural Ringer*.

tes armazenados mídia permanente, economizando recursos e facilitando a análise *offline*.

A seguir será descrito o *Neural Ringer*, que é um discriminador alternativo para o canal elétron/jato no L2 do ATLAS. Utilizando informação da física de interesse para o processo de extração de características e um classificador neural supervisionado para o teste de hipóteses, o *Neural Ringer* apresenta eficiência superior ao T2Calo e tempo de execução dentro da janela permitida para o L2.

4.3 Neural Ringer - Alternativa para Filtragem de Elétrons no L2

Um discriminador alternativo foi proposto inicialmente em [8] para o canal elétron/jato no segundo nível de filtragem do ATLAS. Conforme ilustrado na Figura 4.5, o processo de extração de características compreende um arranjo topológico dos sinais medidos no calorímetro (com a construção de anéis concêntricos a partir do perfil de deposição de energia), e o teste de hipóteses é realizado por um classificador neural supervisionado (arquitetura *Perceptron* de Múltiplas Camadas) [5]. Este discriminador ficou conhecido como *Neural Ringer*.

4.3.1 Extração de Características - Anelamento

Os sinais do perfil de deposição de energia utilizados para discriminação elétron/jato são medidos nas sete camadas dos calorímetros do ATLAS. Considerando a granularidade de cada camada, uma Região de Interesse (RoI) típica (de tamanho $0,4 \times 0,4$ em $\eta \times \phi$) é descrita por aproximadamente 1000 células [9].

Visando compactar os sinais dos calorímetros e, ao mesmo tempo, manter a interpretação física do perfil de deposição de energia, as células sensoras de cada camada são agrupadas em anéis concêntricos de deposição de energia. Conforme

ilustrado na Figura 4.6, o procedimento de geração dos sinais em anéis é executado em cada camada do calorímetro e envolve os passos a seguir:

1. Seleção das regiões de interesse ($0,4 \times 0,4$ em $\eta \times \phi$ ao redor da célula mais energética);
2. A célula mais energética da camada é considerada como o primeiro anel (e consequentemente o “centro” da ROI);
3. As células ao redor do primeiro anel (imediatamente adjacentes) definem o segundo anel e assim sucessivamente (até cobrir toda a janela da ROI);
4. A energia amostrada pelas células pertencentes a um dado anel são somadas produzindo o sinal de energia em anéis;
5. Este processo é repetido para todas as sete camadas do calorímetro;
6. Os sinais em anéis formados a partir das diferentes camadas são concatenados em um único vetor e normalizados.

É interessante notar que, devido à diferença de granularidade entre as camadas do calorímetro, um número diferente de anéis é gerado para cada uma delas. A tabela no final da Figura 4.6 mostra a distribuição dos anéis por camada. No processo de formatação é considerado um número fixo de anéis por camada, porém, dependendo da posição do centro da ROI, esta pode não estar completa (as células sensoras que compõem os anéis podem não existir caso a partícula interaja próximo a uma extremidade ou descontinuidade do calorímetro), neste caso, assume-se igual a zero as leituras das células que faltam. Outra característica a ser observada é que, em algumas camadas, as células não são quadradas, neste caso, falar em anéis é uma abstração (pois ao final do processo descrito acima podem não existir realmente anéis de células sensoras).

São mostrados na Figura 4.7, considerando a segunda camada eletromagnética (E2), os sinais medidos para um elétron e um jato típicos e os respectivos sinais em anéis gerados para esta camada. A Figura 4.8 mostra sinais em anéis (após concatenação de todas as camadas e normalização), respectivamente, para um elétron

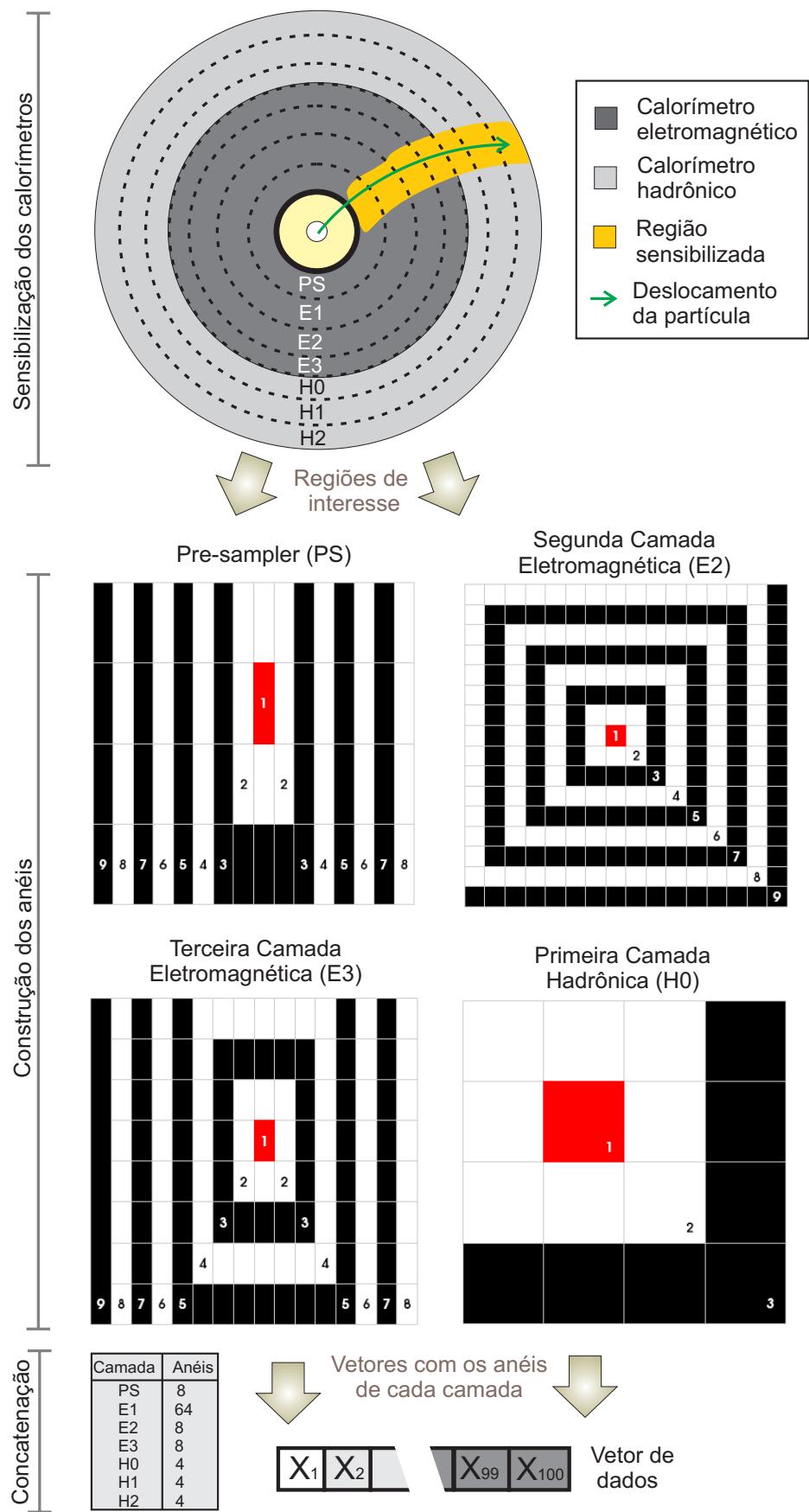


Figura 4.6: Diagrama do processo de construção dos anéis.

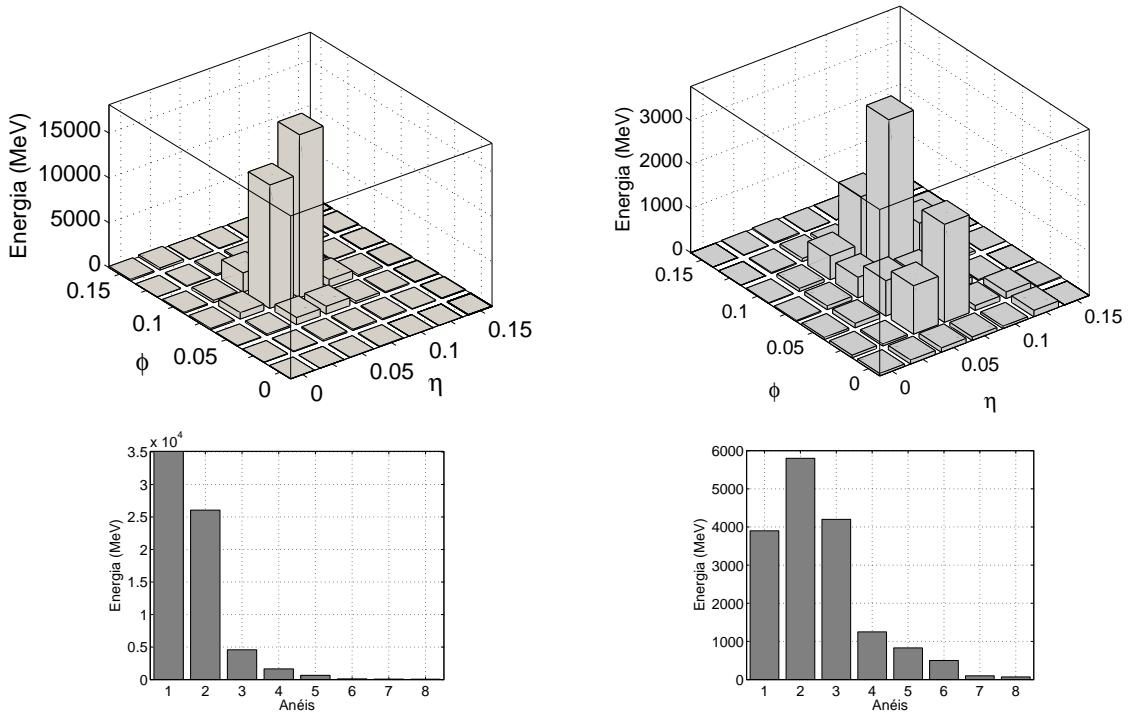


Figura 4.7: Sinais típicos medidos na camada EM2 (acima) e os respectivos sinais em anéis (abaixo), respectivamente para um elétron (esquerda) e um jato (direita).

típico, um jato típico e um jato com perfil semelhante ao de elétrons. Percebe-se que, o perfil de deposição medido para elétrons típicos apresenta pouco espalhamento (é contido em uma pequena região) e é concentrado nas camadas eletromagnéticas. Os jatos apresentam uma maior variação em suas características físicas. Essas partículas tipicamente depositam uma considerável quantidade de energia nas camadas hadrônicas e possuem perfil de deposição com pouca concentração espacial, ou seja, mais espalhados (ver Figura 4.8-b). Entretanto, alguns jatos (como o mostrado na Figura 4.8-c) podem apresentar perfil de deposição de energia semelhante ao de elétrons. Esse último tipo de assinatura hadrônica representa a maioria dos jatos aprovados como elétrons pelo L1, compondo um ruído de fundo de difícil identificação.

4.3.2 Normalização

Antes da utilização nos sistemas de classificação os sinais em anéis precisam ser normalizados. No trabalho [10], foram testadas diferentes formas de normalização

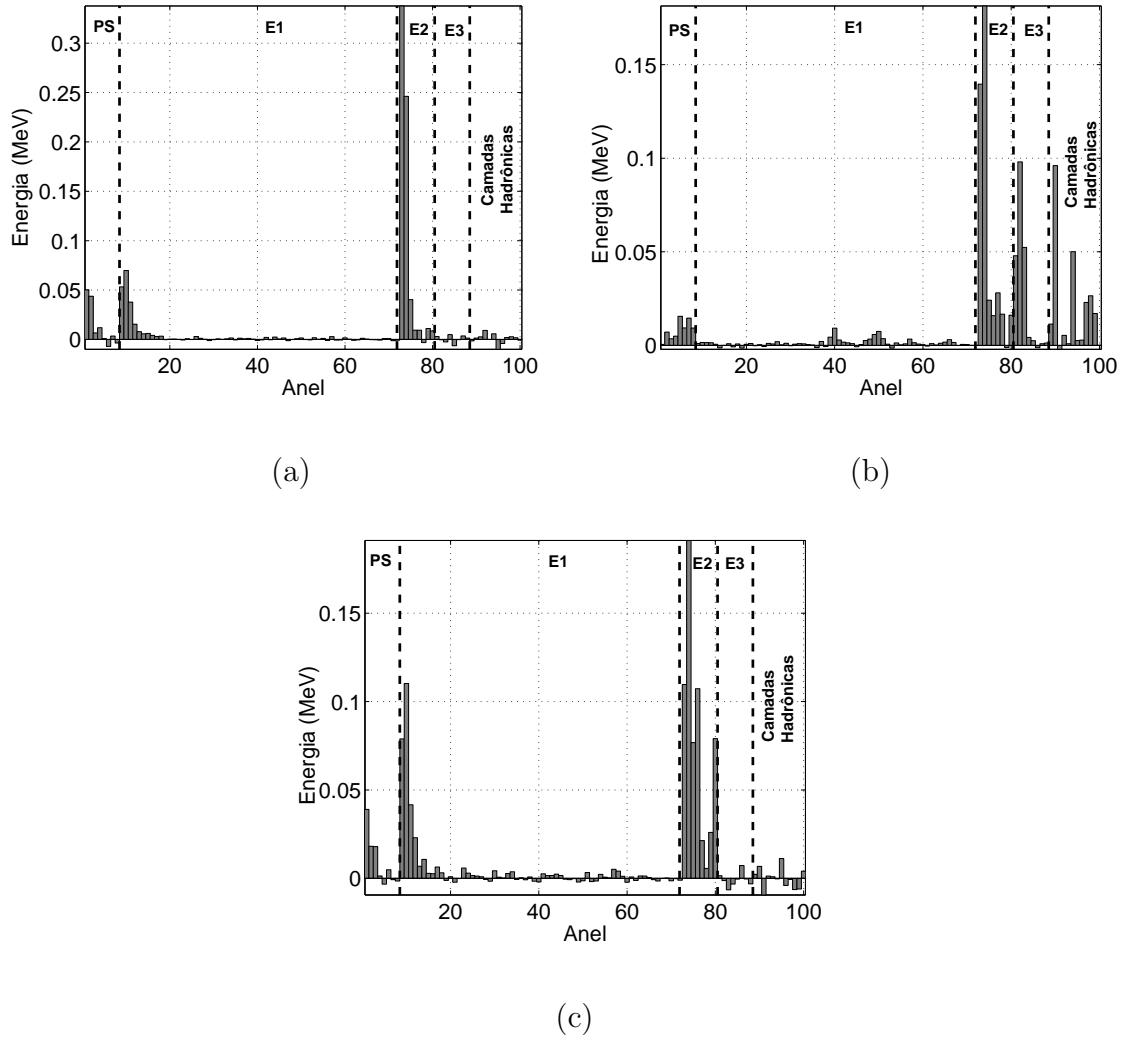


Figura 4.8: Sinais em anéis para (a) elétron típico, (b) jato típico e (c) jato com perfil semelhante ao de elétrons.

dos sinais em anéis como:

- **Energia total:** dividindo-se a energia de cada anel (Ea_i) pela energia total do evento:
$$E_{Ni} = \frac{Ea_i}{\sum_{i=1}^{100} Ea_i}. \quad (4.1)$$
- **Energia da camada:** dividindo-se a energia de cada anel pela soma da energia da camada a qual ele pertence;
- **Energia da seção do calorímetro:** dividindo-se a energia de cada anel pela soma da energia na seção do calorímetro (eletromagnética ou hadrônica) que contém o referido anel.

- **Sequencial:** esse procedimento de normalização visa amplificar as diferenças no perfil lateral dos chuveiros (realçando a energia distante do centro da ROI, que em alguns casos, pode ser útil para a discriminação). A energia do anel i da camada c (Ea_{ic}) é normalizada através de:

$$E_{Ni}^{\{c\}} = \frac{Ea_i^{\{c\}}}{E^{\{c\}} - \sum_{j=1}^{i-1} Ea_j^{\{c\}}} \quad (4.2)$$

onde $E^{\{c\}}$ é a energia total da camada c.

Após um estudo comparativo, concluiu-se que a normalização por energia total atendia um compromisso entre rapidez de execução e eficiência de discriminação, sendo assim, este procedimento foi adotado em todos os resultados simulados a serem mostrados nessa tese. A normalização sequencial produz eficiência semelhante à normalização por energia total e, embora tenha custo computacional um pouco mais elevado, é o procedimento em operação no detector atualmente. Portanto, nos conjuntos de dados experimentais (medidos no detector) este é o procedimento de normalização utilizado.

4.3.3 Teste de Hipóteses - Classificador Neural

Para o processo de discriminação propriamente dito (teste de hipóteses), o Neural Ringer utiliza um classificador neural supervisionado na arquitetura *Perceptron* de Múltiplas Camadas (MLP *Multi-Layer Perceptron*) [5]. Os classificadores neurais são capazes de produzir superfícies não-lineares de separação, contribuindo, em geral, para aumento na eficiência de discriminação quando comparados aos discriminadores lineares. Devido a sua estrutura paralelamente distribuída, uma rede neural é capaz de responder rapidamente quando um padrão de entrada é aplicado. Considerando as características expostas (alta eficiência e processamento veloz), os classificadores neurais se tornam uma opção interessante ao ambiente do segundo nível de filtragem *online* do ATLAS. Mais detalhes a respeito dos classificadores neurais na arquitetura MLP podem ser encontrados no Apêndice B.

4.3.4 Tempo de Execução

Um aspecto crucial nos algoritmos para o sistema de filtragem *online* é o tempo de execução. No caso do *Neural Ringer*, que é um algoritmo para o segundo nível de filtragem, a janela de tempo máxima permitida é de 40 ms. Uma análise detalhada do tempo de processamento requerido pelo *Neural Ringer* foi realizada no trabalho [10]. O algoritmo foi incorporado como uma sub-rotina do T2Calo, aproveitando parte do processamento realizado por este último algoritmo.

As diversas etapas necessárias para a execução do *Neural Ringer*, e seus respectivos tempos de execução no ambiente do L2, são listadas a seguir:

1. Inicialmente, são selecionadas as informações necessárias, o que é feito na etapa de seleção de região (*Region Selector*) do T2Calo, que solicita os fragmentos da RoI desejada. Essa etapa, por envolver um intenso fluxo de informação entre o pipeline de memória e o segundo nível de filtragem, corresponde a 40.7 % do tempo total (em média $0,4927 \pm 0,0787$ ms).
2. A seguir, o T2Calo realiza um pré-processamento nos sinais, onde são calculadas as suas 4 variáveis de decisão e os valores de E_T , η e ϕ com uma maior precisão que no L1. Este passo de processamento equivale a 11.6 % do tempo total de execução ($0,1408 \pm 0,0148$ ms). A partir daqui é iniciada a sequência de rotinas específicas do *Neural Ringer*.
3. Apenas parte da informação necessária para o *Neural Ringer* é utilizada pelo T2Calo, então, o restante precisa ser solicitado neste momento pelo *Region Selector* do *Neural Ringer*. Um novo fluxo de informação é estabelecido entre o L2 e as memórias de armazenamento temporário, consumindo cerca de 36 % do tempo total ($0,4375 \pm 0,0996$ ms).
4. Na etapa de pré-processamento do *Neural Ringer*, as informações da RoI são utilizadas na construção dos anéis para cada camada do calorímetro (conforme procedimento descrito na Seção 4.3.1). O tempo gasto nesta etapa corresponde a aproximadamente 8% do total ($0,0986 \pm 0,0165$ ms). Finalmente, a classificação dos sinais em anéis pela rede neural requer aproximadamente 3% do tempo total ($0,0387 \pm 0,0018$ ms).

Contabilizando todas as etapas de processamento chega-se a um tempo total de $1,2352 \pm 0,1288$ ms. Somando-se a isso a análise do perfil da trajetória da partícula medido no detector de traço (a identificação de elétrons no L2 pode envolver a análise de sinais de dois sub-detectores, os calorímetros e os detectores de traço), chega-se a $8,04 \pm 2,04$ ms, valor que ainda está muito abaixo do tempo máximo exigido para o nível 2, que é de 40 ms. Na Tabela 4.1, são mostradas as diversas etapas necessárias para a execução do *Neural Ringer*, juntamente com o tempo requerido por cada uma delas (em valores absolutos e também percentuais). A Figura 4.9 mostra a proporção entre os tempos gastos considerando todas as etapas de processamento, o que inclui também parte do T2Calo (no gráfico à esquerda) e considerando apenas as etapas específicas do *Ringer* (à direita).

Tabela 4.1: Tempo gasto nas diversas etapas necessárias para a execução *online* do *Neural Ringer*.

Etapas	Tempo (ms)	% do Total
<i>Region Selector - T2Calo</i>	$0,4927 \pm 0,0787$	40,7
Pré-processamento - T2Calo	$0,1408 \pm 0,0148$	11,6
<i>Region Selector - Ringer</i>	$0,4375 \pm 0,0996$	36,1
Pré-processamento (anelamento) - <i>Ringer</i>	$0,0986 \pm 0,0165$	8,2
Normalização	$0,0026 \pm 0,0015$	0,2
Classificação Neural	$0,0387 \pm 0,0018$	3,2
Total	$1,2109 \pm 0,1288$	100

Embora o *Neural Ringer* tenha, em comparação ao T2Calo, produzido um aumento no tempo de execução da ordem de 91% (de $0,6469 \pm 0,0802$ ms para $1,2352 \pm 0,1288$ ms, ou seja um aumento de 0,5883 ms), isso é compensado por uma eficiência de discriminação muito superior do *Ringer*, conforme será mostrado no Capítulo ??.

O excesso de tempo do *Ringer* em relação ao T2Calo é compensado no processamento do Filtro de Eventos, pois, como uma quantidade significativamente menor de ruído de fundo é enviado para o último nível, um menor tempo será gasto nesta etapa.

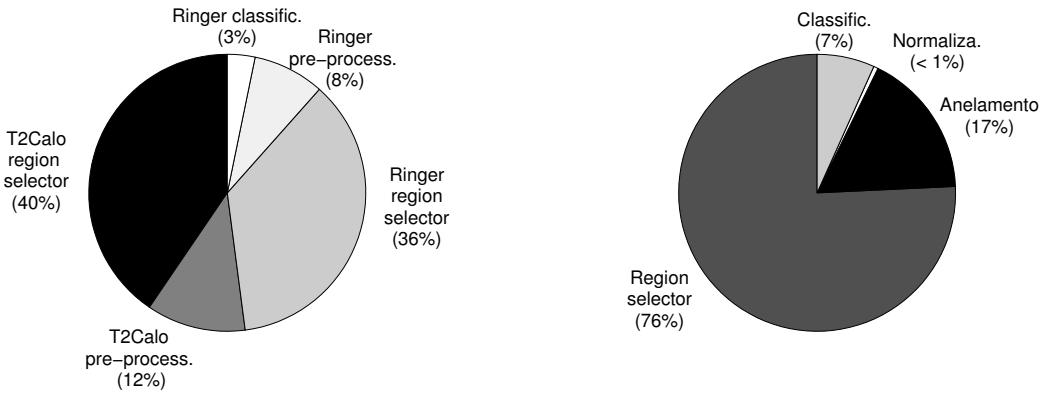


Figura 4.9: Tempos de processamento considerando todas as etapas (esquerda) e somente as etapas específicas do *Neural Ringer* (direita).

4.4 Extensões ao Neural Ringer

Com eficiência mais alta que o algoritmo padrão (T2Calo) e tempo de processamento dentro da janela aceita para o L2, o discriminador *Neural Ringer* mostrou-se uma opção bastante interessante para o problema da filtragem *online* de elétrons no ATLAS. Em sistemas neurais de classificação, sabe-se que o uso de uma etapa de pré-processamento adequada pode ser decisivo para o aumento da eficiência de discriminação [86]. Neste contexto, trabalhos vêm sendo desenvolvidos com o objetivo de melhorar a eficiência do Neural Ringer com a adição de uma etapa de pré-processamento ao classificador neural (ver Figura 4.10), na qual, os sinais em anéis são mapeados em um conjunto (em geral mais compacto) de características discriminantes. A seguir, serão descritos a importância da etapa de pré-processamento em classificadores neurais (em um caso genérico) e os trabalhos desenvolvidos neste sentido no contexto do projeto *Neural Ringer*.

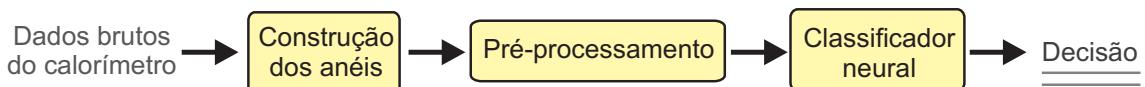


Figura 4.10: Fluxo de processamento das extensões ao *Neural Ringer*.

4.4.1 Importância do Pré-Processamento para Classificadores Neurais

Embora as redes neurais possam, essencialmente, realizar quaisquer mapeamentos não-lineares, pode-se observar que, o uso direto dos dados brutos como entradas geralmente produz um desempenho pior do que quando algum pré-processamento é aplicado [86].

Da teoria da informação, considerando duas variáveis \mathbf{x} e c (respectivamente o conjunto de características e os rótulos de classes), para qualquer transformação determinística $T(\cdot)$, a informação mútua (I) entre $T(\mathbf{x})$ e c tem como limite superior a informação mútua entre \mathbf{x} e c [87]:

$$I(T(\mathbf{x}); c) \leq I(\mathbf{x}; c). \quad (4.3)$$

Ou seja, nenhuma transformação é capaz de acrescentar informação, ao conjunto de características \mathbf{x} , a respeito das classes c .

Considerando o resultado da Equação 4.3, o que justifica o melhor desempenho dos classificadores após a etapa de pré-processamento, uma vez que ela não é capaz de adicionar informação a respeito do problema?

A resposta pode ser encontrada considerando-se que, os sistemas neurais de classificação são obtidos a partir de um processo iterativo de ajuste de pesos. Então, um pré-processamento eficiente, ou seja, capaz de revelar características discriminantes que estavam inicialmente ocultas nos dados brutos, pode se tornar decisivo para a obtenção de um classificador com melhor desempenho.

É importante notar que, todas as informações utilizadas para a discriminação estão presentes nos dados brutos. O pré-processamento, que neste contexto é conhecido como extração de características, é responsável apenas por uma transformação que torna as características discriminantes mais evidentes.

Um outro aspecto a ser analisado é a chamada “maldição da dimensionalidade”(do inglês *curse of dimensionality*) [86]. Sabe-se que, a utilização de um número elevado de entradas para o sistema classificador acaba dificultando o processo de treinamento. Isso ocorre pois, quanto maior a dimensão dos dados de entrada, maior a complexidade da rede neural (e consequentemente maior o número de parâmetros a serem ajustados), requerendo também maior informação estatística

(exemplos de treinamento) para o treino adequado do classificador.

A solução adotada, na maioria dos casos, é pré-processar os sinais com uma transformação que reduza a dimensionalidade do problema. Pois, entre os atributos disponíveis, alguns podem apresentar informações redundantes ou até mesmo irrelevantes para o processo de classificação, dificultando o processo de treinamento dos classificadores. Porém, é preciso escolher adequadamente a informação a ser descartada para minimizar a chance da perda de características relevantes para a discriminação das classes.

Resumindo, a transformação ótima deve, entre outras coisas, mapear os atributos disponíveis em um número reduzido de características, eliminar a redundância e manter toda a informação discriminante para o problema.

4.4.2 Pré-Processamento Linear

Voltando para o contexto do discriminador *Neural Ringer*, embora a construção dos anéis já seja responsável por uma considerável redução de dimensão (por um fator de 10 vezes), no trabalho [88] foi realizado um estudo detalhado a respeito da utilização da Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3] para compactação. A PCA foi aplicada aos sinais em anéis e o classificador neural treinado a partir dos componentes principais estimados. Com a PCA, foi alcançado um fator de redução de aproximadamente 4 vezes, mantendo desempenho semelhante ao Neural Ringer tradicional. Visando explorar toda a granularidade e segmentação disponível aos calorímetros do ATLAS, foi proposta a realização do processo de compactação de modo segmentado (em cada camada do calorímetro).

No trabalho [10], foi utilizada uma técnica de compactação mais adequada a problemas de classificação, as Componentes Principais de Discriminação (PCD - *Principal Components for Discrimination*) [11]. Foi realizado também um estudo referente à aplicação da Análise de Componentes Independentes (ICA *Independent Component Analysis*) [4] como uma etapa adicional de pré-processamento, aparecendo logo após a compactação. Os resultados mostraram que a combinação de ICA e PCD foi capaz de melhorar o desempenho em relação ao Neural Ringer tradicional, sem modificar significativamente o tempo de processamento necessário para a tomada de decisão.

4.4.3 Pré-Processamento Não-linear

Conforme descrito no Capítulo 2, os calorímetros, embora projetados para serem detectores lineares na identificação de elétrons, na prática, seu comportamento pode apresentar características não-lineares, devido a fenômenos como saturação de sensores, atenuação da luz (em calorímetros cintilantes), etc. Então, é possível que uma técnica não-linear de extração de características seja mais adequada ao problema.

Os estudos desenvolvidos no contexto desta tese, tiveram o objetivo de investigar a aplicação da versão não-linear da análise de componentes independentes (NLICA - *Nonlinear Independent Component Analysis*) [13, 89] para extração de características, como uma etapa de pré-processamento ao *Neural Ringer*.

Considerando que existem diversos algoritmos para a NLICA, neste trabalho foram realizados estudos com alguns destes métodos. Os diferentes modelos foram comparados, considerando eficiência de discriminação e tempo de processamento necessário para produzir a decisão. No próximo capítulo serão mostrados os fundamentos teóricos da análise de componentes independentes nos seus modelos linear e não-linear, bem como suas aplicações em problemas de extração de características.

Capítulo 5

Análise de Componentes Independentes

Neste capítulo, serão mostradas, de modo resumido, as principais características dos modelos linear e não-linear da análise de componentes independentes (respectivamente ICA e NLICA). Também serão brevemente apresentados os algoritmos utilizados para estimação da NLICA no contexto deste trabalho. A seguir, será realizada uma revisão bibliográfica das aplicações da análise de componentes independentes em problemas de extração de características, dando enfoque especial a Física de Altas Energias (HEP - *High-Energy Physics*). Em complementação a este Capítulo, no Apêndice A, são apresentados, de modo mais detalhado (com um maior formalismo matemático), alguns dos algoritmos mais utilizados na literatura para ICA/NLICA, juntamente com os diversos parâmetros empregados nesta rotinas para estimar o grau de independência entre os sinais.

5.1 Modelo Linear da ICA

Em muitos problemas de processamento de sinais multi-dimensionais¹ deseja-se encontrar uma transformação que, de algum modo, torne a estrutura essencial dos dados mais acessível [4]. Entre as técnicas lineares que buscam, através de pre-

¹Sinais multi-dimensionais são, em geral, produzidos por sistemas de medição com múltiplos sensores, mas também podem surgir a partir da aplicação de transformações (como a transformada de Fourier ou Wavelet) a sinais uni-dimensionais.

missas distintas, por uma nova representação para os sinais multi-dimensionais, pode-se mencionar a Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3], a Análise de Fatores (*Factor Analysis*) [90] e a Análise de Componentes Independentes (ICA - *Independent Component Analysis*) [4].

Entre as técnicas listadas acima, a análise de componentes independentes busca por uma transformação onde os componentes na saída são estatisticamente independentes. A ICA vêm sendo aplicada na solução de diversos problemas na área de processamento de sinais como cancelamento de ruído [91], sonar passivo [92], telecomunicações [93], reconhecimento facial [94] e biomédica [95].

- **Independência Estatística:** Considerando duas variáveis aleatórias (VAs) y_1 e y_2 , se elas são independentes, então o conhecimento de uma não traz nenhuma informação a respeito da outra. Matematicamente, y_1 e y_2 são independentes estatisticamente se e somente se [96]:

$$p_{y_1,y_2}(y_1, y_2) = p_{y_1}(y_1)p_{y_2}(y_2), \quad (5.1)$$

onde $p_{y_1,y_2}(y_1, y_2)$, $p_{y_1}(y_1)$ e $p_{y_2}(y_2)$ são respectivamente as funções de densidade de probabilidade (pdf - *probability density function*) conjunta e marginais de y_1 e y_2 [96].

Pode-se obter uma expressão equivalente à equação (5.1) se, para todas as funções $g(y_1)$ e $h(y_2)$ absolutamente integráveis em y_1 e y_2 , vale a igualdade:

$$\mathcal{E}\{g(y_1)h(y_2)\} = \mathcal{E}\{g(y_1)\}\mathcal{E}\{h(y_2)\} \quad (5.2)$$

Considerando que a estimativa das funções de densidade de probabilidade (necessárias na equação (5.1)) é um problema de difícil solução, pois, em geral os componentes independentes são desconhecidos, uma vantagem da equação (5.2) é que as pdfs não são necessárias. A definição de independência pode ser facilmente estendida para mais de duas variáveis aleatórias. Percebe-se que a independência é um princípio mais restritivo que a descorrelação (quando $g(x) = x$ e $h(y) = y$). O conceito de independência envolve o conhecimento de toda a estatística dos dados, sendo assim muito mais abrangente

que a descorrelação (utilizada pela PCA), que somente utiliza estatística de segunda ordem (variância).

Na ICA, considera-se que um sinal multi-dimensional $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ observado (ou medido) é gerado a partir da combinação linear das fontes independentes $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$. Na forma matricial e suprimindo-se o índice temporal t, pode-se escrever [97]:

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \dots & a_{NN} \end{bmatrix}}_{\mathbf{A}} \times \begin{bmatrix} s_1(t) \\ \vdots \\ s_N(t) \end{bmatrix}, \quad (5.3)$$

ou na forma matricial (omitindo o índice temporal):

$$\mathbf{x} = \mathbf{As}, \quad (5.4)$$

onde \mathbf{A} é a matriz de mistura.

O objetivo final da ICA é encontrar uma aproximação \mathbf{y} das fontes independentes, utilizando apenas os sinais observados \mathbf{x} . O vetor \mathbf{y} é definido por:

$$\mathbf{y} = \mathbf{Wx}, \quad (5.5)$$

sendo \mathbf{W} a matriz de separação. Se $\mathbf{W} = \mathbf{A}^{-1} \rightarrow \mathbf{y} = \mathbf{s}$, então o problema foi completamente solucionado.

Um problema clássico que pode ser solucionado usando-se ICA é conhecido como *cocktail-party problem* [4], e está formulado de forma simplificada, omitindo atrasos temporais e outros fenômenos físicos, como a existência de múltiplas reflexões, nas equações (5.6) e (5.7) (ver Figura 5.1). Considerando que numa sala existem duas pessoas falando simultaneamente e dois microfones em diferentes posições, os sinais gravados $x_1(t)$ e $x_2(t)$ são uma soma ponderada das fontes $s_1(t)$ e $s_2(t)$:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \quad (5.6)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t); \quad (5.7)$$

os coeficientes a_{ij} dependem das distâncias dos microfones às pessoas, e são os elementos da matriz de mistura \mathbf{A} , onde:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \quad (5.8)$$

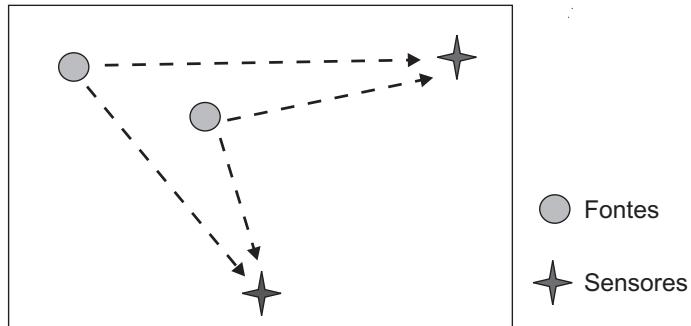


Figura 5.1: Diagrama do *cocktail party problem*.

Em um exemplo de aplicação da ICA, a Figura 5.2-a mostra as fontes $s_1(t)$ e $s_2(t)$, que foram misturadas linearmente, gerando os sinais $x_1(t)$ e $x_2(t)$ da Figura 5.2-b. Após a aplicação de um algoritmo para extração das componentes independentes (FastICA [98]), foram obtidas as curvas da Figura 5.2-c. Percebe-se que os sinais recuperados são cópias dos originais, a menos de fatores multiplicativos. Esta é uma das limitações inerentes ao modelo da ICA, não há como garantir o fator de escala (que pode ser positivo ou negativo) ou a ordem de extração dos componentes.

As técnicas de ICA foram desenvolvidas inicialmente para solucionar problemas de separação cega de sinais (BSS - *Blind Signal Separation*) semelhantes ao *cocktail-party problem*, porém, mais recentemente, surgiram outras aplicações interessantes, como extração de características, separação de fontes em telecomunicações e redução de ruído em imagens [4, 98]. Atualmente, ICA é aplicada com sucesso tanto para separação de sinais como para extração de características.

5.2 ICA Não-Linear

Em muitos problemas práticos, o modelo básico da ICA, onde os sinais observados são considerados combinações lineares e instantâneas das fontes, não representa

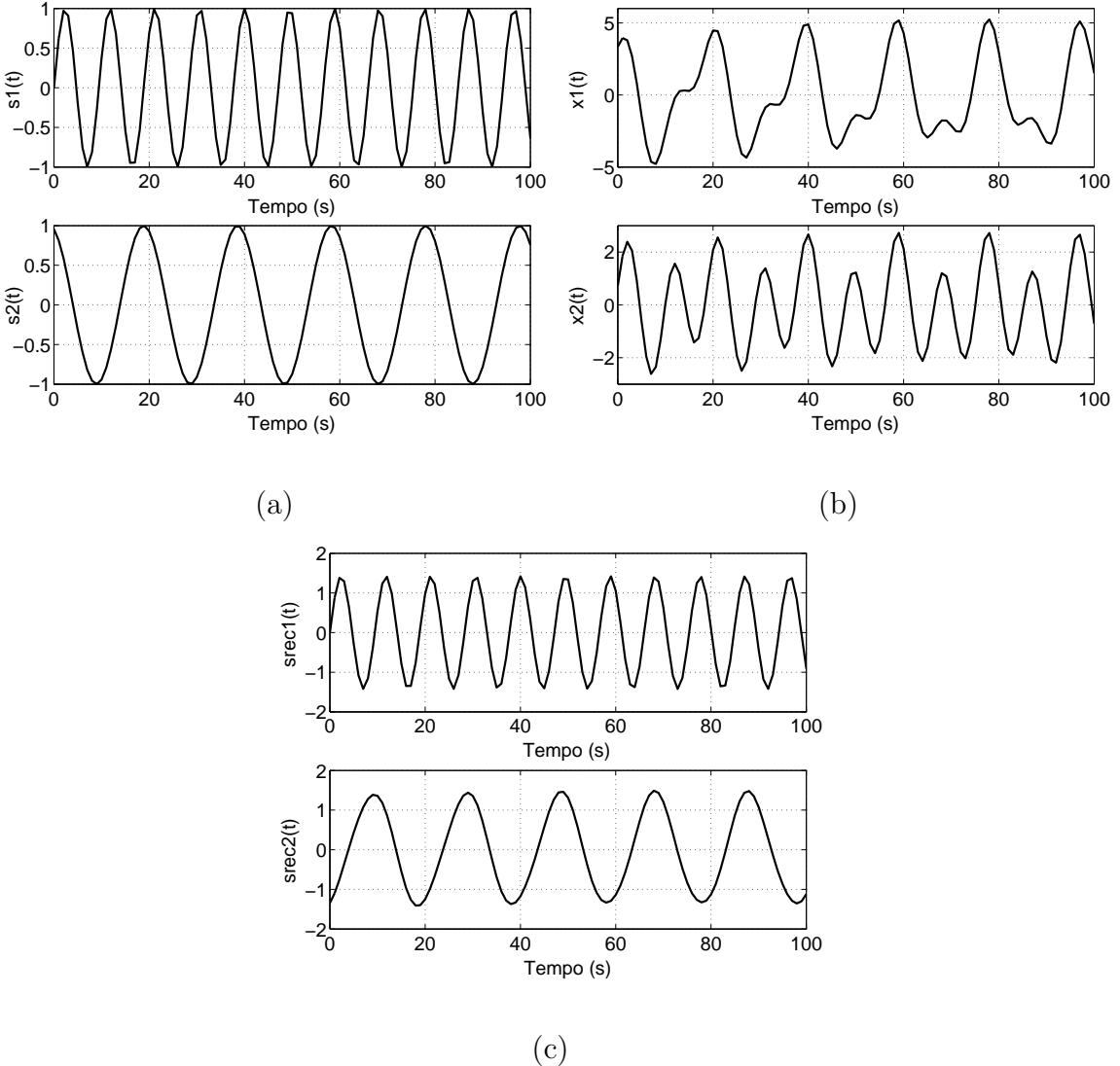


Figura 5.2: Sinais (a) fonte, (b) observados e (c) recuperados através da ICA.

corretamente o cenário real.

A equação (5.9) apresenta um modelo geral para as misturas não-lineares:

$$\mathbf{x} = \mathbf{F}(\mathbf{s}), \quad (5.9)$$

onde $\mathbf{F}(\cdot)$ é um mapeamento não-linear de $\mathbb{R}^N \rightarrow \mathbb{R}^N$, \mathbf{x} e \mathbf{s} são respectivamente os sinais observados e as fontes. A ICA não-linear consiste em encontrar o mapeamento $\mathbf{G}(\cdot): \mathbb{R}^N \rightarrow \mathbb{R}^N$ tal que os componentes de \mathbf{y} sejam estatisticamente independentes [89]:

$$\mathbf{y} = \mathbf{G}(\mathbf{x}). \quad (5.10)$$

Uma característica da NLICA é que o problema apresenta múltiplas soluções [99].

Se \mathbf{y}_1 e \mathbf{y}_2 são variáveis aleatórias independentes, é fácil provar que $f(\mathbf{y}_1)$ e $g(\mathbf{y}_2)$ (onde $f(\cdot)$ e $g(\cdot)$ são funções diferenciáveis) são também independentes [100]. Fica claro que, numa dada aplicação, sem o uso de restrições adequadas, infinitos mapeamentos inversos \mathbf{G} satisfazem a condição de independência entre os sinais estimados y_i , $i=1,\dots,N$. Se o objetivo do problema for realizar a separação das fontes de modo não-supervisionado ou cego (do inglês *Blind Signal Separation - BSS*), neste caso, deseja-se que os componentes y_i sejam aproximações das fontes independentes (\mathbf{s}) que produziram os sinais observados (\mathbf{x}). Então, informações a respeito do modelo de mistura ou das fontes devem ser conhecidas a priori. A NLICA vem sendo aplicada com sucesso em problemas como processamento de sinais de fala [101, 102], processamento de imagens [103, 104], predição de séries de ações em bolsas de valores [105] e processamento de sinais de um *array* de sensores [106, 107].

Em geral, o número de parâmetros a serem estimados num modelo de ICA não-linear é maior do que no caso linear. Os algoritmos de NLICA, se comparados aos de ICA, apresentam maior complexidade computacional e convergência mais lenta [99]. Considerando estas limitações, nas aplicações de NLICA deve-se verificar se existem restrições quanto ao aumento do custo computacional no processo de estimação dos componentes. Em problemas de separação cega de fontes, o algoritmo a ser utilizado deve ser escolhido utilizando informações a respeito do modelo de mistura.

Entre os algoritmos de NLICA propostos na literatura, uma classe de métodos impõe restrições estruturais ao modelo de mistura. Neste caso, pode-se garantir que os componentes estimados são aproximadamente iguais às fontes (a menos das indeterminações de fator multiplicativo e ordem de estimação dos sinais, assim como ocorre no modelo linear). Existem também alguns métodos (ou algoritmos) que não impõem restrições ao modelo de mistura, neste caso, o mapeamento não-linear dos dados observados em componentes independentes não garante a separação das fontes. Outro método diretamente relacionado com a NLICA, chamado de ICA Local, propõe uma etapa de agrupamento dos sinais em conjuntos de características semelhantes, que deve ser realizada antes da ICA. O agrupamento produz um mapeamento não-linear dos dados e a ICA (linear) estima os componentes independentes. Mais informações a respeito dos diversos algoritmos e modelos de NLICA serão fornecidas nas próximas seções.

5.2.1 Unicidade da Solução em NLICA

Conforme mencionado anteriormente, no caso não-linear, a independência estatística não é suficiente para garantir a separação das fontes. Se duas variáveis aleatórias y_1 e y_2 são independentes, então $p_{y_1,y_2}(y_1, y_2) = p_{y_1}(y_1)p_{y_2}(y_2)$. Para funções diferenciáveis f e g , pode-se provar que [108]:

$$p_{f(y_1),g(y_2)}(y_1, y_2) = p_{f(y_1)}(y_1)p_{g(y_2)}(y_2), \quad (5.11)$$

e então as variáveis $f(y_1)$ e $g(y_2)$ são também independentes. Esta indeterminação, diferente do fator de escala e da ordem de estimativa dos componentes (que são inerentes a ICA linear), não são aceitáveis em um problema de separação de fontes.

Estudos teóricos indicaram que a unicidade da solução da NLICA pode ser conseguida se o problema apresentar pelo menos uma das características a seguir [100]:

- O número de componentes é igual a dois. Deste modo, os sinais podem ser considerados como uma variável complexa.
- As pdf dos componentes independentes são limitadas a valores conhecidos.
- O mapeamento \mathbf{F} preserva o zero ($\mathbf{F}(0) = 0$) e é unívoco.
- O modelo de mistura é conhecido a priori e utilizado como informação para o algoritmo de estimativa das componentes independentes.

5.2.2 Modelos com Restrições Estruturais

Um caso especial da ICA não-linear são os métodos de estimativa dos componentes independentes que utilizam informações a respeito do modelo não-linear que gerou os dados observados. Estas informações se configuram em restrições estruturais ao mapeamento inverso (que é estimado pelo algoritmo). Entre os modelos com restrições estruturais pode-se destacar o de misturas **pós não-lineares** (PNL), que será descrito a seguir.

Misturas Pós Não-Lineares

O modelo de misturas pós não-lineares [108] é um dos mais utilizados na literatura, com aplicações em processamento de fala [109, 110], separação de sinais de áudio [111, 112] e processamento de imagens [113].

No modelo PNL, considera-se que inicialmente ocorre uma combinação linear das fontes (como no modelo básico de ICA), e as funções não-lineares f_i são aplicadas antes da observação dos sinais x_i :

$$x_i = f_i \left(\sum_{j=1}^n a_{ij} s_j \right). \quad (5.12)$$

É importante notar que as não-linearidades são aplicadas individualmente a cada componente da mistura linear (não são permitidas não-linearidades cruzadas). A Figura 5.3 ilustra o modelo de misturas PNL.

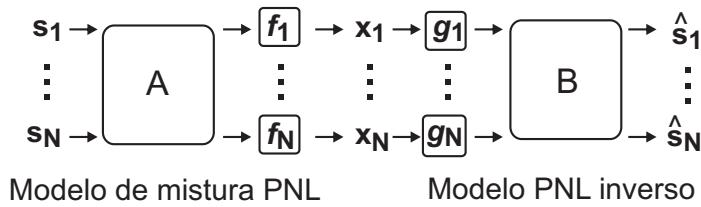


Figura 5.3: Diagrama do modelo de mistura PNL.

A consideração de que as misturas são pós não-lineares permite uma grande simplificação do problema, e as indeterminações existentes se tornam semelhantes às do caso linear. A modelagem através da equação (5.12) satisfaz grande parte dos fenômenos não-lineares, como, por exemplo, a modelagem da distorção de sensores num meio de propagação linear.

Diversos algoritmos foram propostos na literatura para estimação do modelo PNL. Um dos primeiros trabalhos [108], utiliza redes neurais para estimar cada função não-linear g_i e o ajuste dos pesos é feito a partir da minimização da informação mútua usando o método do gradiente (mais detalhes a respeito deste algoritmo podem ser encontrados no Apêndice A). Na estimação do modelo PNL usando um algoritmo do tipo gradiente descendente (que realiza uma busca local), o grande número de parâmetros e as características não-lineares do problema contribuem para a convergência em mínimos locais. Visando minimizar esta limitação, foram propostos algoritmos de estimação do modelo PNL usando métodos de busca global como algoritmos genéticos [114, 115], reconhecimento simulado (*Simulated Annealing*) e aprendizado competitivo (*Competitive Learning*) [109].

Outros Modelos de Misturas com Restrições Estruturais

Alguns modelos com restrições estruturais diferentes do PNL foram propostos na literatura. No trabalho [112], o modelo de mistura é definido por:

$$\mathbf{x} = \mathbf{A}_2 f(\mathbf{A}_1 \mathbf{s}), \quad (5.13)$$

sendo \mathbf{A}_1 e \mathbf{A}_2 matrizes quadradas e $f = [f_1, f_2, \dots, f_N]^T$ um mapeamento com funções não-lineares aplicadas a cada componente (assim como o modelo PNL, este também não permite não-linearidades aplicadas a mais de um componente). O modelo definido na Equação 5.13 e ilustrado também na Figura 5.4 é chamado Pós Não-linear Linear (PNL-L). O bloco linear \mathbf{A}_2 é executado após a aplicação das funções não-lineares, produzindo um modelo mais geral que o PNL. Nos trabalhos [116, 117] são propostos algoritmos baseados em redes neurais para a estimativa do modelo PNL-L.

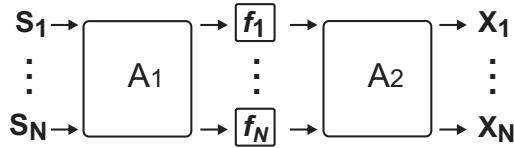


Figura 5.4: Diagrama do modelo PNL-L.

Em [117], um modelo estrutural chamado mono não-linearidade (ver Figura 5.5) foi proposto para o problema da NLICA. Neste modelo os sinais observados são gerados a partir de:

$$\mathbf{x} = f^{-1}(\mathbf{A}f(\mathbf{s})). \quad (5.14)$$

Este modelo (chamado de mistura de mono não-linearidade e ilustrado na Figura 5.5) é dito mais geral que o PNL, pois as funções não-lineares (f_i) podem ser aplicadas a mais de um componente. A análise deste modelo, a partir da teoria da análise funcional (*functional analysis*) [118], mostra que pode representar qualquer mistura com duas camadas de não-linearidades [117].

5.2.3 Algoritmos sem Restrições Estruturais

Se nenhuma restrição ao modelo de mistura é imposta, não há garantia que os componentes independentes estimados estejam relacionados com as fontes (ver Seção

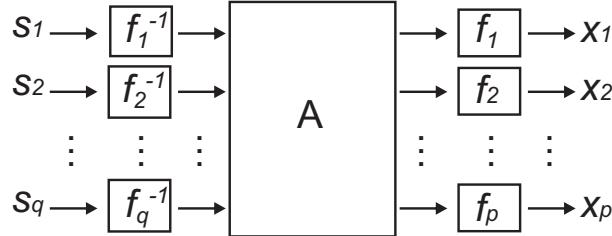


Figura 5.5: Diagrama do modelo da Mono não-linearidade.

5.2.1). Entre os métodos de NLICA sem restrições estruturais, pode-se destacar o uso de mapas auto-organizáveis e os métodos que utilizam inferência Bayesiana.

NLICA a Partir de Mapas Auto-Organizáveis

Uma das primeiras tentativas bem-sucedidas de estimar o modelo da NLICA foi realizada através de mapas auto-organizáveis (SOM - *Self-Organizing Maps*) [119]. Pode-se provar que as coordenadas y_1 e y_2 do neurônio vencedor no mapa (ver Figura 5.6) são independentes e aproximadamente uniformemente distribuídas [119]. Para estimar a NLICA, o SOM é treinado usando como entradas os sinais observados, e as coordenadas do neurônio vencedor correspondem a uma aproximação discreta dos componentes independentes (o número de níveis possíveis sinal estimado é igual ao número de neurônios utilizados para representá-lo).

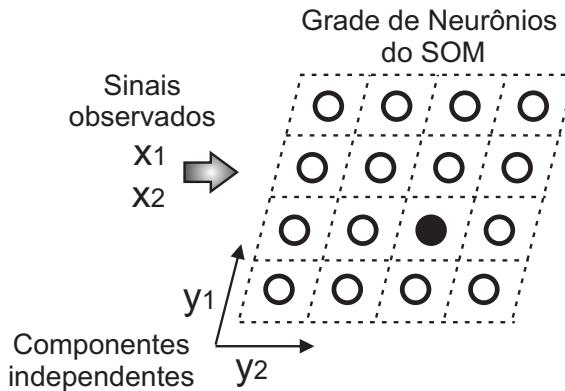


Figura 5.6: NLICA a partir de SOM.

Entre as desvantagens do método pode-se destacar [99]:

- O mapeamento é discreto (existe um número limitado de neurônios no mapa), então, algum tipo de regularização é necessária para produzir componentes

contínuos. Esse problema pode ser minimizado aumentando-se o número de neurônios do mapa (porém, isto aumenta o custo computacional para estimação do mapa).

- Os componentes a serem estimados devem ter pdf (função densidade de probabilidade) sub-gaussiana (quanto mais próxima da distribuição uniforme, melhor).
- O custo computacional para treinamento dos mapas aumenta rapidamente com o número de componentes independentes a serem estimados.

Diversas aplicações do SOM para estimação da NLICA podem ser encontradas na literatura, entre elas pode-se citar os trabalhos e [120, 103, 121], onde os objetivos eram, respectivamente, separação de imagens sobrepostas, remoção de ruído e visualização de dados multi-dimensionais.

Com uma formulação alternativa aos SOM, o Mapeamento Topográfico Geral (GTM - *Generative Topographic Mapping*) foi introduzido em [122], e apresenta princípios estatísticos mais fundamentados que o mapa SOM. O método GTM básico tem poucas vantagens práticas em relação aos Mapas Auto-Organizáveis, pois aqui as componentes independentes também são assumidas como processos uniformemente distribuídos e o espaço de características é formado a partir de uma grade retangular discreta m-dimensional. Porém, devido a sua formulação matemática, o GTM pode ser estendido para variáveis não uniformes.

O trabalho [123] propõe uma modificação à formulação básica onde são introduzidos coeficientes de ponderação que permitem a estimação de componentes independentes com qualquer tipo de distribuição. Os componentes são modelados como misturas de sinais Gaussianos e os parâmetros são estimados usando o algoritmo *Expectation Maximization* [98]. O treinamento do GTM envolve dois passos, a avaliação da probabilidade a *posteriori* e a adaptação dos parâmetros do modelo, neste sentido, o processo é semelhante ao utilizado pela abordagem da inferência Bayesiana, que será mostrada com mais detalhes a seguir. Não foram encontradas muitas aplicações do GTM na estimação da NLICA.

NLICA a partir de Inferência Bayesiana

Nos métodos baseados em inferência Bayesiana, considera-se que os sinais observados são gerados a partir de [124]:

$$\mathbf{x} = f(\mathbf{s}) + \mathbf{n} \quad (5.15)$$

onde \mathbf{n} é definido como ruído Gaussiano independente dos componentes a serem estimados.

Neste contexto, os componentes independentes são modelados como misturas de sinais de distribuição Gaussiana. Pode-se provar que, dado um número suficiente de Gaussianas, qualquer distribuição de probabilidade pode ser aproximada [124]. Uma variação deste método foi aplicada em [125] para o modelo linear da ICA. Em grande parte dos algoritmos Bayesianos para NLICA, redes neurais tipo MLP de duas camadas são treinadas para aproximar o mapeamento não-linear, neste caso, têm-se que [99]:

$$f(\mathbf{s}) = \mathbf{B}\Phi(\mathbf{As} + \mathbf{a}) + \mathbf{b} \quad (5.16)$$

Em um método de estimação Bayesiano, probabilidades *a posteriori* são associadas a cada modelo não-linear que, possivelmente, teria gerado os dados observados. Verificar uma quantidade tão grande de modelos não é possível na prática; então, os métodos Bayesianos para NLICA utilizam uma técnica chamada de “aprendizagem amostral” (EL - *ensemble learning*) [126]. Na EL, somente o conjunto mais provável de modelos é testado utilizando uma aproximação paramétrica que é ajustada à probabilidade *a posteriori* [127].

Métodos Bayesianos de NLICA foram propostos em [128] e [129]. No trabalho [130] foram realizados testes experimentais para comparar o desempenho dos modelos Bayesiano e pós não-linear (PNL) na estimação dos componentes independentes, as principais conclusões foram:

- os algoritmos PNL apresentam desempenho superior quando as misturas seguem o modelo PNL clássico (não-linearidades inversíveis e mesmo número de componentes independentes e sinais observados);
- o desempenho de ambos os métodos pode ser melhorada a partir da exploração da informação de mais misturas que componentes independentes;

- a principal vantagem do método Bayesiano é que mapeamentos mais genéricos podem ser produzidos (uma vez que não há restrições estruturais). Estes métodos geralmente apresentam maior custo computacional e necessitam de várias inicializações para obter uma solução ótima (podem apresentar problemas com mínimos locais da função custo).

No trabalho [107] um algoritmo Bayesiano de NLICA foi utilizado com sucesso para a separação de sinais medidos em um conjunto de sensores químicos.

O Algoritmo MISEP

O algoritmo MISEP [131] utiliza a minimização da Informação Mútua (ver Apêndice A) como estratégia para busca pelos componentes independentes. Esta rotina é considerada como uma extensão do método INFOMAX [4], podendo ser utilizado para estimar tanto o modelo linear quanto o não-linear da ICA. Na Figura 5.7 pode-se observar um diagrama do MISEP (para duas entradas e duas saídas), onde x_i e y_i são respectivamente os sinais observados e os componentes independentes estimados, o bloco $\mathbf{G}(.)$, no caso linear, aproxima a matriz de separação \mathbf{W} , e para a NLICA, deve fornecer uma aproximação do mapeamento não-linear inverso. As funções não-lineares ψ_i e as variáveis de saída z_i são utilizadas apenas no processo de treinamento. Após a convergência do algoritmo, as não-linearidades devem ser aproximações da função de probabilidade cumulativa (cdf - *cumulative probability function*) dos componentes independentes.

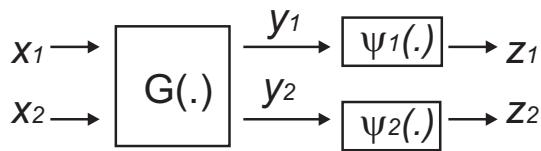


Figura 5.7: Diagrama do algoritmo MISEP.

Para a aplicação em NLICA, o bloco $\mathbf{G}(.)$ é estimado por uma rede neural (que pode utilizar tanto a arquitetura *perceptron* de múltiplas camadas - MLP como rede de funções de base radial - RBF). Como o objetivo é estimar a função de probabilidade cumulativa (cdf - *cumulative distribution function*), as saídas z_i são restritas ao intervalo $[0,1]$ e as ψ_i são limitadas a funções estritamente crescentes. Para estimação iterativa de cada função ψ_i , são utilizadas redes neurais MLP com

uma camada oculta (de neurônios sigmoidais) e uma camada de saída (linear). Estas redes tem uma entrada (y_i) e uma saída (z_i). O treinamento do modelo MISEP é feito a partir da maximização da entropia das saídas z_i , o que acaba produzindo a minimização da informação mútua dos componentes y_i , mais detalhes podem ser encontrados em [131]

O MISEP foi aplicado em processamento de sinais de áudio [131] e separação de imagens [132]. Foram propostas também, modificações ao algoritmo MISEP visando otimizar a estimativa dos componentes independentes quando as misturas seguem os modelos pós não-linear (PNL) [133] e pós não-linear-linear (PNL-L) [134].

5.2.4 ICA Local

Considerando um problema onde o conjunto de sinais multi-dimensionais apresenta grande variação em suas características estatísticas, o modelo linear da ICA pode não ser capaz de representar adequadamente os dados. Neste contexto, pode ser mais interessante tratar o conjunto de sinais de modo local, ou seja, em subconjuntos onde os elementos tem características semelhantes.

A ICA Local realiza a estimativa dos componentes independentes a partir de k subconjuntos dos dados (ver Figura 5.8). Conforme proposto em [135], um conjunto de dados de alta dimensão pode ser separado em sub-conjuntos (onde os elementos apresentam características semelhantes), através de algum algoritmo de agrupamento como o *k-means* [136] ou SOM [137], e modelos da ICA linear são então estimados para cada subconjunto. Caso não exista informação a priori a respeito do número de agrupamentos, metodologias foram propostas nos trabalhos [138, 139] para sua estimativa.

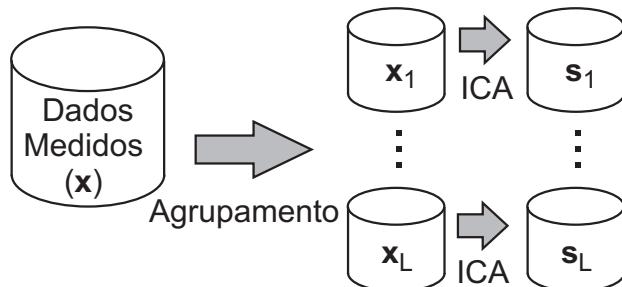


Figura 5.8: Diagrama do modelo da ICA local.

Na ICA Local, o agrupamento é responsável por uma representação não-linear dos dados, enquanto modelos de ICA linear aplicados a cada sub-conjunto (*cluster*) descrevem as características locais. A ICA local pode ser considerada como um compromisso entre os modelos linear e não-linear da ICA [99]. O objetivo é obter uma melhor representação dos dados (se comparado com o modelo linear da ICA), evitando os problemas computacionais do modelo não-linear [140]. Em diferentes abordagens, os agrupamentos podem ser montados com superposição, usando por exemplo fronteiras *fuzzy* [141, 142], ou sem superposição [140, 143].

Nos trabalhos [144, 145] a ICA Local foi aplicada para a estimação da informação mútua. A informação mútua [87] é uma importante ferramenta em diversas aplicações de processamento de sinais, especialmente na seleção de características. O modelo da ICA Local foi utilizado também em [146] para a remoção de ruído. O modelo proposto consiste em realizar k agrupamentos de versões atrasadas no tempo dos sinais medidos e estimar a ICA linear para cada um destes conjuntos. Através da abordagem proposta, foi alcançado um aumento da relação sinal/ruído da ordem de 10 dB.

5.3 Aplicações de ICA e NLICA para Extração de Características

Nessa Seção, serão resumidamente descritas algumas aplicações da análise de componentes independentes, nos seus modelos linear e não-linear (respectivamente ICA e NLICA), em problemas de extração de características.

No trabalho [147], ICA foi utilizada como pré-processamento para problemas de classificação em nove bases de dados diferentes (obtidas no repositório de bases de dados para aprendizado de máquina da Universidade da Califórnia - Irvine, CA, Estados Unidos [148]). Entre os problemas testados, estão a classificação de vinhos a partir de características físicas e químicas, a identificação da existência de câncer em amostras de tecido da mama, a identificação isolada de vogais independente do locutor e previsão de sobrevida de pacientes que sofreram ataque cardíaco a partir do resultado do eletrocardiograma. A transformação da ICA foi estimada através do algoritmo JADE [149]. Utilizando-se classificadores neurais (MLP), a eficiência

foi comparada para sinais sem pré-processamento, sinais branqueados e sinais após ICA. Em alguns casos (como na identificação de vogais) o uso da ICA produziu uma redução do erro de identificação (24,13% sem pré-processamento, 21,05% após o branqueamento e 20,77% após a ICA). Em outros casos, porém, a aplicação da ICA tornou mais difícil o problema de classificação (como no caso da identificação do cancer de mama, onde, sem pré-processamento, o erro foi de 2,55% e após a ICA aumentou para 2,63%). Analisando-se todos os resultados conclui-se que, nem sempre a aplicação da ICA contribui para um aumento na eficiência, esse fato é intensificado em problemas onde o modelo da mistura linear não se aplica (pois possivelmente existem não-linearidades envolvidas). O uso da ICA parece tornar mais suave a curva de erro de treinamento das redes neurais, contribuindo para a diminuição da quantidade de mínimos locais e, consequentemente, da probabilidade do treinamento ficar estacionado num desses mínimos.

Em [150], ICA foi novamente utilizada para detecção do cancer de mama a partir de imagens digitalizadas de mamografias. Nesse trabalho, os componentes independentes foram estimados através do algoritmo FastICA [151] e classificadores neurais (MLP) foram utilizados para produzir a decisão. As amostras disponíveis pertenciam a três classes distintas (normais, com alterações benignas e com alterações malignas). A ICA foi estimada a partir de pequenas janelas nas mamografias onde as classes de interesse eram mais facilmente identificadas. Foram obtidas eficiências de identificação da ordem de 99,9% para as amostras normais, 86,8% e 91,1% respectivamente para amostras com alterações benignas e malignas.

Microarranjos de DNA foram pré-processados por ICA em [152] para a classificação através de máquinas de vetor de suporte (SVM - *Support Vector Machines*) [5]. Os microarranjos de DNA são fragmentos genômicos que representam segmentos gênicos em particular. Nesse trabalho, o algoritmo FastICA foi utilizado para extrair características dos microarranjos (de quatro bases de dados distintas) com o objetivo de identificar a presença de diferentes tipos de tumores (de colo de útero, leucemia, de fígado e do sistema nervoso). As eficiências de identificação obtidas para os quatro tipos foram, respectivamente, 90%, 100%, 74% e 76% com pré-processamento por ICA, 86%, 94%, 74% e 72% com pré-processamento por PCA e 90%, 94%, 70% e 69% sem pré-processamento. Uma outra aplicação de ICA no

mesmo problema pode ser encontrada em [153].

Ainda na área biomédica, no trabalho [154], a ICA foi utilizada como pré-processamento para um mapeamento não-supervisionado de características oculares, com o objetivo de identificar a presença de glaucoma. A partir de padrões de um exame conhecido como *standard automated perimetry* (SAP), aplicou-se a ICA e o agrupamento (não-supervisionado) foi realizado sobre os componentes independentes estimados. Através dessa abordagem, 98,4% das assinaturas de olhos com padrão óptico normal foram corretamente classificadas e, considerando-se os olhos com glaucoma, o acerto foi de 68,6%.

A ICA foi utilizada em [155] para a análise de sinais de espectrometria eletrônica de perda de energia (EELS - *Electron Energy Loss Spectroscopy*). A EELS [156] pode ser empregada para medições precisas de espessura (com resolução da ordem de 0,1 nm), pressão e análise de composição química. A ICA, através do algoritmo SOBI [157], foi utilizada como ferramenta complementar de análise dos espectros eletrônicos produzidos. Com o uso da ICA, foi possível a análise simultânea de dois espectros misturados e eliminadas as escolhas subjetivas durante a análise (que sem o uso da ICA precisam ser feitas pelo usuário).

O modelo não-linear da ICA também é utilizado em problemas de extração de características, por exemplo, o trabalho [158] ilustra a aplicação da NLICA num problema de classificação de sinais de eletroencefalograma (EEG). O objetivo é a separação das diferentes atividades cerebrais independentes, porém como não há garantia que o processo de combinação é linear, utilizou-se o modelo da NLICA numa tentativa de modelar dinâmicas cerebrais não-lineares. O modelo pós não-linear (PNL) foi empregado para estimar os componentes independentes. A informação mútua foi utilizada como medida da independência e um algoritmo genético [159] buscou sua minimização. Múltiplos classificadores lineares (cada um treinado com um dos componentes estimados) foram utilizados para identificar os sinais provenientes do movimento da mão. Uma combinação das saídas dos múltiplos classificadores foi utilizada para produzir a decisão final. A eficiência de identificação a partir dos sinais medidos (sem pré-processamento) foi de 73,84%, aumentando para 74,61% e 77,95% quando utilizados, respectivamente, pré-processamento por ICA e NLICA.

A NLICA foi utilizada no trabalho [160] visando à extração de características de

séries temporais de ações para a previsão do índice diário de uma bolsa de valores. Para formar o vetor N-dimensional de entrada para a NLICA, foram utilizados a série com os valores de fechamento diário da bolsa e N-1 versões atrasadas desta série. O algoritmo MISEP [131] foi utilizado para estimar os componentes independentes. Um modelo de regressão por vetor de suporte (*Support Vector Regression*) foi utilizado para prever o comportamento da bolsa. Comparando a NLICA com pré-processamento por ICA e PCA, as eficiências obtidas foram, respectivamente, 80%, 75% e 79%. Também neste exemplo, o modelo da NLICA mostrou-se mais eficiente para evidenciar as características discriminantes do problema em questão.

Em Física de Altas Energias (HEP - *High-Energy Physics*) também são encontradas algumas aplicações de PCA e ICA para extração de características. Alguns desses trabalhos serão descritos brevemente na próxima seção. Para a NLICA, talvez por ser uma técnica ainda menos difundida (em comparação com PCA e ICA), não foram encontradas, até este momento, aplicações na área de HEP.

5.4 Aplicações em Física de Altas Energias e Áreas Correlatas

A partir do final da década de 1990, os métodos de aprendizado estatístico multi-variável vêm sendo aplicados com sucesso em problemas na área de física de alta energia.

Um dos primeiros trabalhos neste tópico [161], foi publicado em 1998 e utiliza mapas auto-organizáveis (SOM) para a classificação de eventos de raios gama em astronomia de alta energia. O SOM foi utilizado para otimizar a sensibilidade de um telescópio IATC (*Imaging Atmosferic Cherenkov Technique*²). Com o uso do SOM, foi possível aumentar a sensibilidade da técnica em aproximadamente 14%.

Em [162], mapas SOM foram aplicados para a separação de bósons W do ruído de fundo composto por jatos hadrônicos no experimento DELPHI do acelerador LEP

²Método através do qual raios gama de alta energia são detectados por telescópios na superfície terrestre. Ao entrarem na atmosfera, os raios gama interagem com a atmosfera e geram um chuveiro de partículas carregadas (conhecido como *Extensive Air Showers*). Devido a sua alta energia, as partículas carregadas produzem descargas luminosas (ou radiação de *Cherenkov*) de curta duração que são captadas pelos telescópios IACT.

no CERN. As entradas para o SOM foram, inicialmente, 90 variáveis físicas que descrevem cada evento (como momento, energia total, presença de agrupamentos de energia, etc). Com o sistema proposto, foi obtida uma probabilidade de detecção de 77,8% para um falso-alarme de apenas 1,3%. O trabalho contemplou também um estudo sobre a relevância das variáveis de entrada e ao final, utilizando-se apenas as 16 variáveis mais relevantes obteve-se desempenho semelhante.

No trabalho [163], o ruído de fundo gerado na aceleração do feixe de partículas do acelerador KEK-B foi rejeitado a partir de mapas auto-organizáveis modificados. O ruído de fundo resulta das colisões dos feixes de elétrons com moléculas de gás residual presentes no tubo de vácuo do acelerador. Se o vácuo fosse perfeito esse problema não aconteceria, mas em condições normais de operação o ruído é função linear da concentração do gás residual. Comparado com um classificador linear, o método baseado no SOM rejeitou 75% do ruído de fundo e classificou corretamente 97% do sinal de interesse, enquanto que o classificador linear obteve respectivamente 68% e 96%.

Redes SOM também foram utilizadas com sucesso para análise, classificação e monitoramento de sinais do telescópio OGLE (no Chile) [164]. Nesta aplicação foi realizado o mapeamento de medições fotométricas de anomalias como supernovas³ e microlentes gravitacionais (*gravitational microlens*)⁴. Os mapas foram treinados a partir de 8000 espectros e os valores aproximados de parâmetros como temperatura e gravidade na superfície são obtidos diretamente do mapa a partir do neurônio vencedor.

Num outro trabalho, mapas auto-organizáveis foram utilizados para a identificação de prováveis assinaturas de bósons de Higgs MSSM (*Minimal Supersymmetric Standard Model*) neutros e pesados [165]. Os mapas foram treinados a partir das mesmas variáveis dos algoritmos tradicionalmente utilizados para o problema, com uma base de dados composta de 80.000 eventos (igualmente divididos para treino e teste). Ao final, foi obtida eficiência de 73%, contra apenas 35% dos algoritmos tradicionais.

A análise de componentes principais (PCA) é um técnica de descorrelação e compactação bastante utilizada em diversas áreas do conhecimento. Em física de

³Corpos celestes originados após a explosão de estrelas

⁴Fenômeno relacionado com a curvatura da trajetória da luz ao passar perto de objetos massivos.

altas energias, PCA foi aplicada para a seleção de variáveis de entrada de um discriminador neural no trabalho [166]. Foram utilizados eventos de altas energias do acelerador LEP ($e^+e^- \rightarrow$ quark+antiquark \rightarrow hadrons). O objetivo é determinar o sabor dos quarks (b, c ou leve). Das 150 variáveis iniciais, após a transformação por PCA foram retidas as 20 mais energéticas e utilizadas para alimentar um classificador neural supervisionado. A mesma metodologia foi aplicada para outros métodos de seleção de varáveis como o teste F [167] e métodos de poda de rede neural [5]. A PCA apresentou uma das melhores eficiências entre os métodos lineares, classificando corretamente 73% das assinaturas.

Em [168], são apresentadas diversas aplicações em HEP onde a PCA é utilizada para extração de características e compactação. No trabalho [169], sinais ópticos de nebulosas planetárias são processados por PCA com o objetivo de extrair informações a respeito de suas características morfológicas. Numa outra aplicação em astrofísica, PCA é utilizada, em conjunto com ICA, para a remoção do ruído de fundo e de outras fontes de interferência, permitindo melhor visualização de dados de ventos e tempestades solares [170].

O trabalho [88] utiliza a PCA, de forma segmentada, para compactação de sinais de calorimetria do ATLAS, em seguida classificadores neurais realizam a decisão elétron/jato, conseguindo eficiência de classificação superior ao algoritmo tradicional utilizado para o problema.

A análise de componentes independentes (ICA) tem aplicação mais recente em HEP, sendo que um dos primeiros trabalhos foi publicado em 2005 e descreve a redução de ruído na análise de sinais do feixe de partículas do experimento BOOSTER do Fermilab [171]. Neste trabalho também é realizada uma comparação com um sistema semelhante baseado em PCA, e a ICA apresenta resultados melhores. No trabalho [172], ICA é utilizada para análise de dados multi-variados em experimentos de física atômica e nuclear. A aplicação de ICA proporcionou redução do ruído de fundo, permitindo melhor visualização do sinal de interesse.

ICA também foi aplicado com sucesso para separação de sinais em astrofísica de alta energia conforme detalhado a seguir. Em [173], ICA foi aplicada para a separação de imagens de fontes sobrepostas adquiridas pelo satélite Planck da Agência Espacial Européia. O objetivo é analisar as informações (mapas de radiação) gera-

dos pelo satélite, que são compostas da superposição de diversas fontes astrofísicas independentes. Neste trabalho foi desenvolvido um algoritmo eficiente para a estimação dos componentes independentes em ambientes não-estacionários e ruidosos. A partir deste método, foi obtida uma relação sinal/ruído da ordem de 20 dB, enquanto que aplicando diretamente um algoritmo de ICA (FastICA), relação foi aproximadamente 2 dB.

No trabalho [174], utiliza-se a análise de componentes independentes, em substituição aos filtros casados, para a decomposição de sinais astrofísicos simulados compostos pela combinação de moléculas elementares em estado congelado (*astrophysical ice mixtures*). Com a técnica proposta foi possível separar os espectros infra-vermelhos provenientes de moléculas de água, gás carbônico e monóxido de carbono.

Ainda na área de astrofísica, nos trabalhos [175, 176] ICA foi aplicado para a caracterização da radiação cósmica de fundo em microondas (CMB - *Cosmic Microwave Background*). A CMB é uma forma de energia eletromagnética que preenche todo o universo e foi inicialmente observada em 1965. A CMB é visualizada apenas por rádio-telescópios. A ICA mostrou-se bastante eficiente para redução da contaminação do sinal de interesse pelo ruído de fundo, com desempenho semelhante à técnica tradicionalmente utilizada para o problema.

No contexto da filtragem online do detector ATLAS (tema de estudo desta tese), um trabalho anterior [10], desenvolvido dentro do mesmo grupo de pesquisa, foi dedicado a estudar os efeitos do pré-processamento por PCA e ICA aplicados aos sinais dos calorímetros com o objetivo de otimizar o sistema neural de detecção de elétrons. Os resultados obtidos indicam que um pré-processamento adequado pode contribuir para aumentar a eficiência do discriminador.

A partir destes exemplos, percebe-se que, apesar da aplicação mais recente em física de altas energias e áreas correlatas, diversos problemas de extração de características, remoção de ruído, agrupamento não-supervisionado (*clustering*) e visualização vêm sendo resolvidos com a aplicação das técnicas estatísticas de processamento não-supervisionado de sinais.

5.5 Utilizando Informação das Classes na Estimação dos Componentes Independentes

Embora o modelo da ICA não tenha sido originalmente desenvolvido para extração (o propósito inicial da ICA era realizar a separação de fontes), conforme visto anteriormente em diversas aplicações, a ICA é uma alternativa interessante para esta tarefa, pois é capaz de transformar os sinais em um conjunto de componentes estatisticamente independentes, eliminando a redundância entre eles.

Quando a ICA é utilizada como pré-processamento para um problema de classificação, o objetivo é obter uma nova representação dos dados, de modo que as características discriminantes estejam mais evidentes. Conforme comentado anteriormente, não há garantia que a transformação por ICA/NLICA seja útil nesta tarefa. Considerando esta limitação, foram propostos na literatura métodos de estimação da ICA adaptados para considerar informação a respeito das classes (transformando a ICA num método supervisionado, ou semi-supervisionado).

No contexto da ICA, quando há necessidade de redução de dimensão (compactação), o método mais utilizado é a Análise de Componentes Principais (PCA - *Principal Component Analysis*, para mais detalhes a respeito ver o Apêndice A). A PCA é um método não-supervisionado que tem como objetivo projetar o conjunto de dados em componentes ordenados por energia (variância), não sendo adequado a problemas de classificação. Um modo alternativo à PCA para realizar a compactação, incluindo a informação das classes, é através do método conhecido como Componentes Principais de Discriminação (PCD - *Principal Components for Discrimination*) [11]. Essa abordagem será descrita mais detalhadamente na Seção 5.5.1.

A informação das classes também pode ser utilizada no contexto da ICA de diferentes modos. Um procedimento simples, proposto em [177], é a inclusão dos rótulos de classe como atributos de entrada para os algoritmos de estimação dos componentes independentes (ver Seção 5.5.2). Alternativamente, um método para estimação de componentes independentes e discriminantes foi desenvolvido no contexto desta tese e será mostrado na Seção 5.5.3.

5.5.1 Componentes Principais de Discriminação

Considerando um problema de classificação de padrões, o uso da PCA para compactação pode ser prejudicial, pois, os componentes menos energéticos (que são eliminados após a PCA) podem carregar informações discriminantes. Neste caso, pode-se utilizar técnicas de compactação mais adequadas. As componentes principais de discriminação (PCD - *Principal Discriminating Components*) [11, 178] são obtidas a partir da projeção dos sinais de entrada em um conjunto compacto que carrega informação importante para discriminação entre as classes.

Conforme proposto em [11], para um problema de classificação, o objetivo da PCD é obter uma projeção linear dos sinais de entrada $\mathbf{x} = [x_1, \dots, x_N]^T$ nos componentes $\mathbf{z} = [z_1, \dots, z_K]^T$ (com $K < N$) que maximizam a discriminação entre as classes (ou seja, z_i são os componentes principais de discriminação).

Considerando um problema de discriminação onde existem apenas duas classes possíveis, os PCD podem ser estimados a partir de uma rede neural (de arquitetura MLP - *Multi-Layer Perceptron*) [5] com uma camada oculta e um neurônio de saída, treinada para obter máxima discriminação entre as classes (saídas alvo: +1 para a classe 1 e -1 para a classe 2). Conforme indicado na Figura 5.9-a, uma rede neural com um neurônio na camada oculta é capaz de estimar o primeiro PCD, que é obtido a partir da projeção das entradas na direção dos pesos sinápticos do neurônio oculto:

$$z_1 = [x_1, \dots, x_N]^T \times [d_{1,1}, d_{1,2}, \dots, d_{1,N}] + b_1, \quad (5.17)$$

onde b_1 é o *bias* do neurônio. Adicionando-se mais neurônios ocultos consegue-se estimar os demais PCD (conforme ilustrado na Figura 5.9-b). O treinamento da rede neural pode ser feito com o congelamento dos pesos da camada de entrada correspondentes aos componentes já estimados, ou seja, na estimação do l -ésimo componente, os pesos $d_{i,j}$, com $i < l$ e $j = 1, \dots, N$ não são ajustados. Os demais pesos da rede são ajustados a cada novo componente estimado. A adição de neurônios continua até a estabilização (num valor máximo) da eficiência de discriminação.

No processo de estimação, estatística de ordem elevada é acessada a partir da utilização de funções de ativação não-lineares. Considerando que K componentes

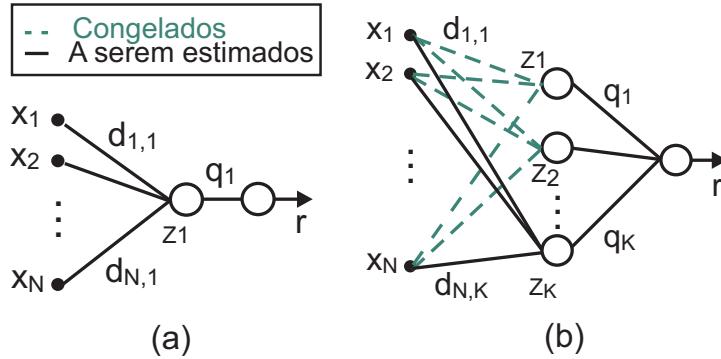


Figura 5.9: Modelos neurais para estimar (a) a primeira e (b) a k -ésima PCD.

principais foram estimados, têm-se:

$$\begin{bmatrix} z_1 \\ \vdots \\ z_K \end{bmatrix} = \underbrace{\begin{bmatrix} d_{11} & \dots & d_{1N} \\ \vdots & \ddots & \vdots \\ d_{K1} & \dots & d_{KN} \end{bmatrix}}_{\mathbf{D}} \times \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} + \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_K \end{bmatrix}}_{\mathbf{b}}, \quad (5.18)$$

ou, na forma matricial: $\mathbf{z} = \mathbf{Dx} + \mathbf{b}$.

Outros modelos que, de modo semelhante a PCD, utilizam redes neurais para extraír características discriminantes de um conjunto de sinais foram propostos nos trabalhos [179, 180, 181].

5.5.2 Utilizando os Rótulos de Classe como Sinais de Entrada para os Algoritmos de ICA

No trabalho [177] foi proposta a utilização dos rótulos de classes como entrada para os algoritmos de estimação dos componentes independentes. Conforme mostrado na Figura 5.10, num problema de classificação binária (com apenas duas classes), para cada exemplo de entrada é associado um novo atributo c com valor igual a 1 (para a classe 1) e -1 (para a classe 2). O bloco G pode ser utilizado para estimar os modelos linear ou não-linear da ICA, a depender do algoritmo de treinamento utilizado.

O parâmetro c é adicionado ao vetor de atributos original $\mathbf{x} = [x_1, x_2, \dots, x_N]$, gerando $\mathbf{x}_C = [x_1, x_2, \dots, x_N, c]$. Para treinamento dos algoritmos de ICA utiliza-se como entrada o vetor \mathbf{x}_C .

Como, num cenário prático de operação do sistema classificador, os rótulos de classe não estarão disponíveis, a informação das classes deve ser removida dos com-

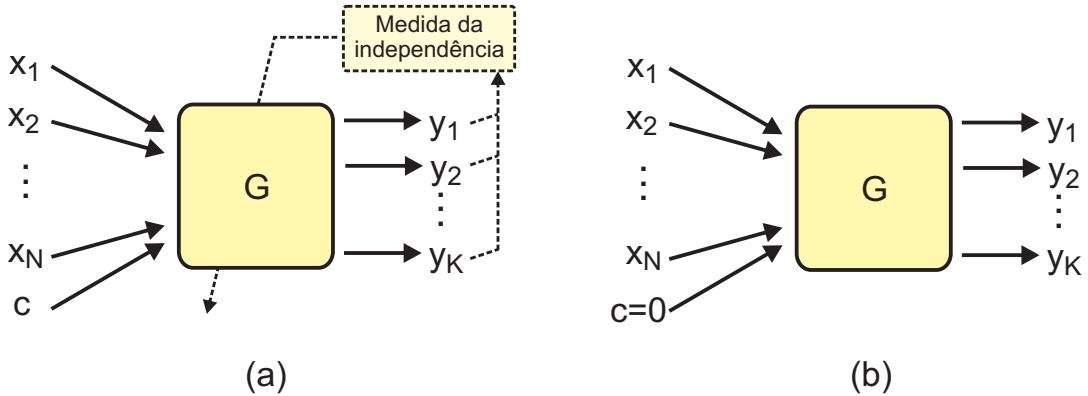


Figura 5.10: Diagramas de (a) treinamento e (b) operação dos algoritmos de ICA/NLICA utilizando informação das classes na entrada.

ponentes estimados. Isso pode ser feito removendo-se as conexões da entrada c ao modelo, ou substituindo c por um vetor de zeros.

Em um trabalho semelhante [182], o mesmo procedimento de treinamento foi adotado, porém o modelo da ICA utilizado é linear e quadrado (mesmo número de entradas e saídas, não havendo portanto redução de dimensão). Após a estimação dos componentes, aqueles que tem pesos de baixa amplitude são eliminados, produzindo um conjunto mais compacto de características discriminantes.

5.5.3 Proposta de Algoritmo para Estimação de Componentes Independentes e Discriminantes

No desenvolvimento desta tese, foi proposto um método alternativo para estimação de componentes independentes e discriminantes. No trabalho [183], foi apresentado um algoritmo de treinamento para um modelo pós não-linear (PNL) modificado. Este método será descrito a seguir.

Conforme ilustrado na Figura 5.11, um bloco de compactação foi adicionado ao modelo PNL, com o objetivo de transformar o conjunto de N atributos em K componentes (com $N < K$). Assim, a arquitetura proposta é adequada para o caso sobre-determinado (quando existem mais sinais observados do que fontes).

O modelo PNL modificado pode ser estimado a partir de dois procedimentos distintos. Uma abordagem possível é executar a estimação do bloco de compactação de modo independente do modelo PNL. Neste caso, a compactação se configura

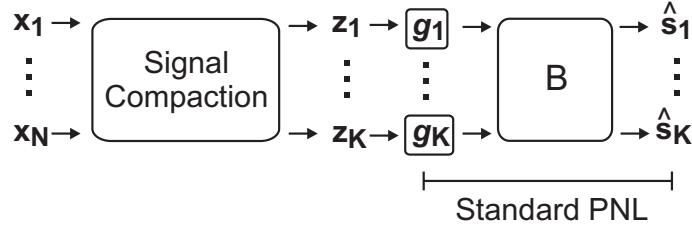


Figura 5.11: Modelo Pós Não-linear modificado.

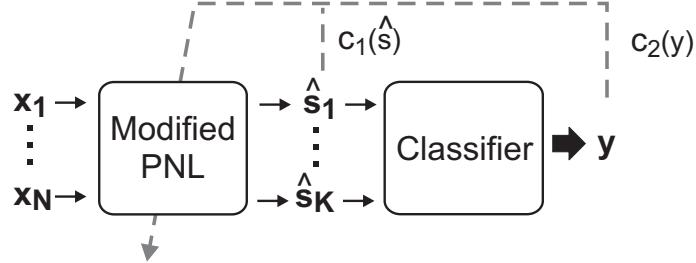


Figura 5.12: Procedimento de treinamento para o modelo pós não-linear modificado.

numa etapa de pré-processamento para a NLICA e pode ser executada, por exemplo através da transformação PCD.

De modo alternativo, o modelo PNL modificado pode ser estimado através do procedimento mostrado na Figura 5.12, que combina informação de duas funções custo diferentes $c_1(\hat{s})$, que mede a independência estatística entre os componentes estimados \hat{s} , e $c_2(y)$, que avalia a eficiência de discriminação produzida a partir de um discriminante linear (DL) [136], onde y é a saída do classificador.

Considerando um bloco de compactação linear, os componentes independentes estimados são descritos por:

$$\hat{s}_i = \sum_{j=1}^K b_{ij} g_j(z_j) \quad i = 1, \dots, K \quad (5.19)$$

onde $\mathbf{z} = \mathbf{Dx}$, \mathbf{D} é uma matriz retangular ($K \times N$) de compactação e os b_{ij} são elementos da matriz quadrada \mathbf{B} ($K \times K$).

A estimativa das não-linearidades $g_i(\cdot)$ é feita de modo semelhante ao proposto no trabalho [108], assim, cada função é aproximada por:

$$g_i(z_i) = \sum_{h=1}^{N_H} \beta_{hi} \tanh(\omega_{hi} z_i - \eta_{hi}) \quad i = 1, \dots, K \quad (5.20)$$

onde β_{hi} , ω_{hi} e η_{hi} são parâmetros a serem determinados.

O objetivo do método proposto é estimar o conjunto de parâmetros \mathbf{D} , \mathbf{B} , β_{hi} ,

ω_{hi} e η_{hi} que maximiza a função custo definida por:

$$c(\hat{\mathbf{s}}, \mathbf{y}) = \frac{\alpha_1}{c_1(\hat{\mathbf{s}}) + \alpha_3} + \alpha_2 c_2(\mathbf{y}) \quad (5.21)$$

sendo α_1 , α_2 e α_3 constantes a serem previamente escolhidas. É importante observar que o propósito de α_3 é limitar o primeiro termo da Equação 5.21, quando $c_1(\hat{\mathbf{s}}) \rightarrow 0$. Valores adequados para as três constantes serão indicados a seguir. O número de componentes independentes a serem estimados (K) precisa ser escolhido a priori. Na prática, se K for desconhecido, pode-se utilizar um procedimento semelhante ao descrito na Seção 5.5.1 para a escolha do número de componentes principais de discriminação (PCD).

A função custo que avalia a independência estatística ($c_1(\hat{\mathbf{s}})$) utiliza uma medida do cumulante de quarta ordem, semelhante à proposta no algoritmo JADE para a ICA [149]:

$$c_1(\hat{\mathbf{s}}) = \sum_{\substack{i,j=1 \\ i < j}}^K \sum_{l,m=1}^K \mathbf{cum}\{s_i, s_j, s_l, s_m\}^2 \quad (5.22)$$

sendo $\mathbf{cum}\{s_i, s_j, s_l, s_m\}$ o cumulante de quarta ordem [4]:

$$\begin{aligned} \mathbf{cum}\{s_i, s_j, s_l, s_m\} = & E\{s_i, s_j, s_l, s_m\} - E\{s_i, s_j\}E\{s_l, s_m\} \\ & - E\{s_i, s_l\}E\{s_j, s_m\} - E\{s_i, s_m\}E\{s_j, s_l\} \end{aligned} \quad (5.23)$$

Um modo para calcular a medida da independência baseada no cumulante de quarta ordem foi proposto em [184], e utilizado neste trabalho para estimar $c_1(\hat{\mathbf{s}})$. É interessante notar que $c_1(\hat{\mathbf{s}})$ é uma sempre não-negativa e zero para sinais independentes, então, maximizar a independência entre os componentes $\hat{\mathbf{s}}$ implica em minimizar $c_1(\hat{\mathbf{s}})$.

A função custo que avalia a eficiência de discriminação é o índice soma-produto (SP) normalizado, que para um problema de classificação em M classes é definido por:

$$c_2(\mathbf{y}) = \sqrt{\frac{\sum_{i=1}^M Ef_i}{M}} \times \sqrt[M]{\prod_{i=1}^M Ef_i} \quad (5.24)$$

onde Ef_i é a eficiência de discriminação obtida para a classe i . A função definida na Equação 6.1 varia no intervalo $[0,1]$ e alcança o máximo quando $Ef_i = 1$ para $i = 1, \dots, M$ (eficiência total). Uma característica de $c_2(\mathbf{y})$ é sua sensibilidade a degradação da eficiência para qualquer classe.

Considerando que $c_1(\hat{\mathbf{s}}) \geq 0$ e $0 \leq c_2(\mathbf{y}) \leq 1$, as constantes α_i , na Equação 5.21 são escolhidas para produzir $0 \leq c(\hat{\mathbf{s}}, \mathbf{y}) \leq 1$. Usando por exemplo, $\alpha_1 = \alpha_3/2$, $\alpha_2 = 0.5$ e $\alpha_3 = 0.001$, o mesmo fator de ponderação é dado para os termos de ambos, c_1 e c_2 .

Para a otimização de $c(\hat{\mathbf{s}}, \mathbf{y})$ foi utilizado um algoritmo genético (AG) simples, conforme mostrado em [159, 185], ao qual foram adicionados, elitismo, *crossover* uniforme e genocídio periódico. Os algoritmo genéticos são métodos de busca global capazes de apresentar desempenho satisfatório mesmo em problemas onde a superfície de erro é não-linear e apresenta ótimos locais. Uma outra característica do AG é que não há a necessidade do cálculo do gradiente da função de erro, pois opera diretamente com o resultado da função custo. Mais detalhes a respeito do AG utilizado são mostrados no Apêndice C.

Capítulo 6

Metodologia Proposta e Bases de Dados Utilizadas

Neste capítulo, será apresentada a metodologia proposta para otimizar o sistema neural de detecção de elétrons no segundo nível de filtragem do ATLAS.

O processo de discriminação é dividido em duas etapas distintas. Inicialmente, os sinais medidos são pré-processados (utilizando modelos relacionados com a NLICA) para que suas características discriminantes se tornem mais acessíveis. A seguir, as características extraídas são utilizadas como entrada para os classificadores propriamente ditos (que no escopo deste trabalho são classificadores neurais supervisionados na arquitetura MLP).

Considerando que cada camada do calorímetro do ATLAS tem propriedades físicas distintas, neste trabalho os sistemas de discriminação podem operar de dois modos. No modo segmentado o processamento é realizado em cada camada. No modo não-segmentado, os sinais de todas as camadas são concatenados num único vetor, que é utilizado como entrada para os discriminadores.

Este capítulo está organizado conforme descrito a seguir. Inicialmente serão apresentados os parâmetros utilizados para avaliar o desempenho dos diversos sistemas classificadores propostos. A seguir, serão descritas as topologias e algoritmos propostos para extração de características e, ao final, serão apresentados o procedimento de treinamento dos discriminadores e as bases de dados utilizadas.

6.1 Parâmetros de Avaliação do Desempenho

No contexto desse trabalho, para avaliar a eficiência dos discriminadores são utilizados a curva ROC (*Receiver Operating Characteristic*) [186], e o índice SP (soma-produto) [9].

A curva ROC mostra como as probabilidades de detecção¹ e falso alarme² (respectivamente PD e PF) variam com o patamar de decisão. A eficiência de um classificador pode ser estimada a partir da área sob a curva ROC. Quanto maior a área, mais eficiente é o discriminador.

O índice SP é definido por [68]:

$$SP = \sqrt{\frac{Ef_e + Ef_j}{2}} \times \sqrt{Ef_e \times Ef_j} \quad (6.1)$$

onde $Ef_e = PD$ e $Ef_j = 1 - PF$ são as eficiências obtidas, respectivamente, para elétrons e jatos.

O SP é utilizado como parâmetro para escolher o patamar de decisão “ótimo” para um dado discriminador. Variando-se o patamar de decisão em toda sua faixa de excursão (que no caso deste trabalho varia entre -1 e 1), calcula-se os valores do SP correspondentes. O SP máximo indica um patamar que apresenta alta eficiência para as duas classes.

Na Figura 6.1 são mostradas as distribuições das saídas de dois classificadores (fictícios) para duas classes distintas. Percebe-se, que no exemplo 1, há uma maior superposição entre as distribuições das duas classes (se comparado ao exemplo 2). Observando as curvas ROC (ver Figura 6.2) geradas para estes classificadores, percebe-se que o classificador do exemplo 2 produz uma maior área sob a curva ROC, e consequentemente maior eficiência de discriminação (o que já era esperado a partir da análise dos histogramas das saídas). Ainda na Figura 6.2, pode-se observar as curvas do patamar de decisão (Y) versus o SP para os dois exemplos. Percebe-se que os valores máximos do SP, em cada caso, aconteceram para valores distintos do patamar. Nas curvas ROC e SP são marcados os pontos de máximo SP e indicados os valores de PD, PF e do patamar ótimo para cada caso.

É importante notar que, o SP é sensível à degradação da eficiência de qualquer

¹Probabilidade de classificar corretamente os eventos de interesse.

²Probabilidade de classificar o ruído de fundo incorretamente como evento de interesse.

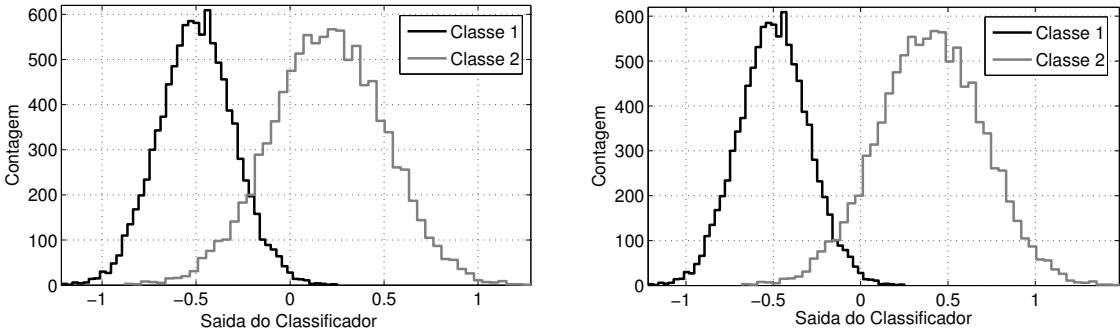


Figura 6.1: Histogramas das saídas de dois classificadores distintos, Exemplo 1 à esquerda e Exemplo 2 à direita.

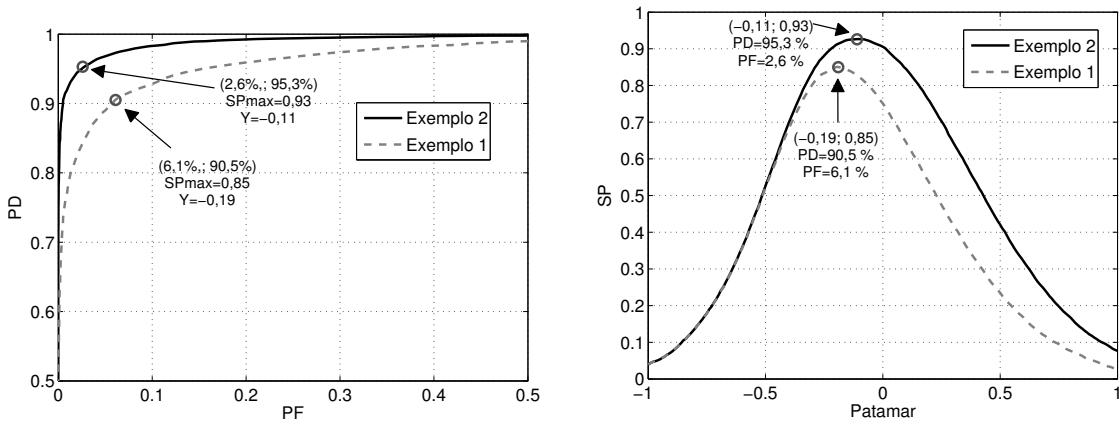


Figura 6.2: Curvas ROC (esquerda) e SP (direita) para dois classificadores distintos.

uma das classes, e ainda que, o máximo SP serve como uma métrica para comparação da eficiência de dois ou mais discriminadores (quanto maior o valor do máximo SP, mais eficiente é o classificador em termos da minimização do erro para as duas classes).

Adicionalmente a estes parâmetros, serão realizadas comparações de desempenho apresentadas como a variação da eficiência e do falso alarme em função de algumas variáveis físicas de interesse como a energia e a posição de interação no detector (considerando os eixos η e ϕ do sistema de coordenadas do ATLAS). Estas análises permitem validar os resultados obtidos considerando conhecimento especialista do problema.

Para expressar os resultados da eficiência de classificação (considerando os parâmetros máximo SP, PD e PF) será apresentado, para cada parâmetro, o valor médio e o desvio padrão associados, considerando-se as diversas inicializações de

treinamento através do procedimento de validação cruzada [5] descrito a seguir:

1. Os sinais disponíveis (para cada classe) são divididos de modo aleatório em 12 grupos (com aproximadamente a mesma quantidade de eventos em cada um).
2. Sorteia-se 4 grupos (em cada classe) para compor o conjunto de treino, 4 para o conjunto de validação e outros 4 para o de teste.
3. O procedimento de treinamento é realizado.
4. O resultado é armazenado e o treinamento re-iniciado até completar 10 iniciizações.

Os algoritmos de extração de características (NLICA) também são re-iniciados a cada sorteio assim como os classificadores neurais.

Para as análises em termos dos parâmetros físicos de interesse (energia, η e ϕ) são calculados histogramas onde, em geral, são mostradas barras de erro calculadas como $100/\sqrt{N}$, onde N é o número de eventos contidos em cada região analisada ao longo das faixas dinâmicas de cada parâmetro.

6.2 Extração de Características

Relembrando o que foi descrito no Capítulo 4, neste trabalho os sinais medidos nos calorímetros são pré-processados, em cada camada, para formação de anéis concêntricos de deposição de energia. Deste modo, considerando uma região de interesse (RoI) de tamanho fixo, são produzidos para as sete camadas dos calorímetros um total de 100 anéis.

Conforme mencionado no início deste capítulo, foram utilizados dois modos distintos para extração de características:

- O modo não-segmentado opera sobre os sinais em anéis gerados para todas as camadas do calorímetro (não há distinção entre as camadas). Para cada evento, as informações dos 100 anéis são concatenadas num único vetor de características. Este vetor é então utilizado como entrada para os algoritmos de aprendizado estatístico, este procedimento é ilustrado na Figura 6.3.

- Alternativamente, as características discriminantes podem ser estimadas separadamente (de modo segmentado) para os anéis produzidos em cada camada do calorímetro. Este procedimento, em geral, produz resultados de mais fácil interpretação física, pois sabe-se que cada camada do calorímetro possui características distintas como o tipo dos sensores e a granularidade das células detectoras. A Figura 6.4 ilustra este procedimento.

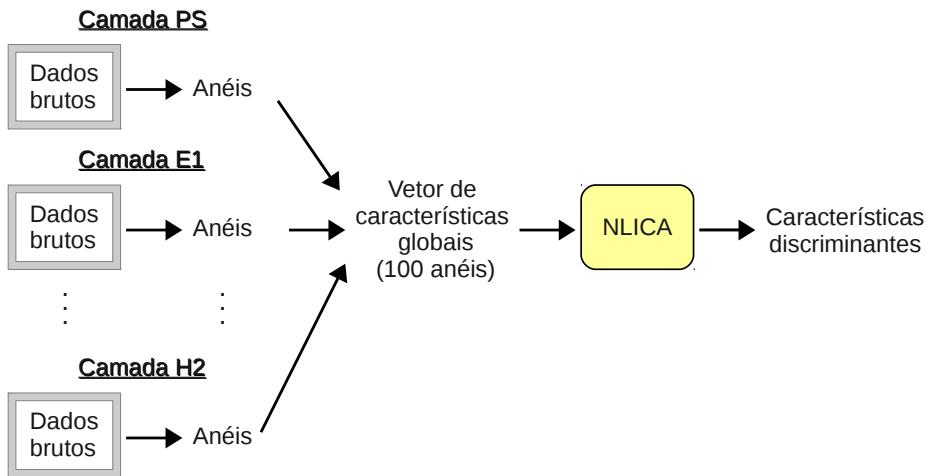


Figura 6.3: Processo de extração de características no modo não-segmentado.

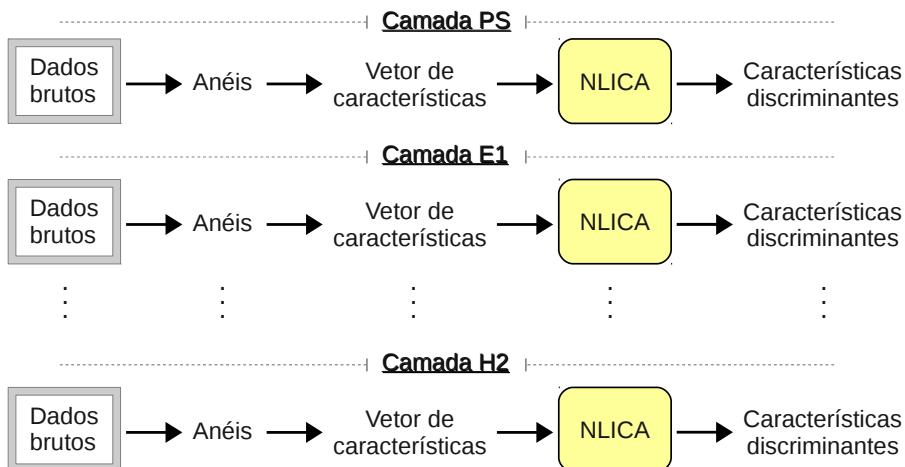


Figura 6.4: Processo de extração de características no modo segmentado.

Uma vantagem da abordagem segmentada, que será explorada neste trabalho, é a possibilidade de redução no tempo de processamento dos discriminadores. Conforme discutido anteriormente no Capítulo 4, grande parte do tempo de processamento do

Neural Ringer é dedicado ao processo de seleção de dados e produção dos sinais em anéis. Convém notar que, devido à configuração do sistema de filtragem, as informações de cada camada são solicitadas em bloco (ou seja, a transferência dos dados envolve sempre toda a informação da RoI numa certa camada).

Como as camadas do calorímetro são sobrepostas, e o evento que se desenvolve interage com todas elas, pode ocorrer algum tipo de redundância nas informações produzidas nos diversos segmentos. Neste trabalho, será realizado um estudo visando identificar as camadas que apresentam informação redundante, e verificar qual a economia obtida no tempo de processamento se as camadas redundantes forem descartadas.

6.2.1 Algoritmos de Extração de Características

Este trabalho está focado na utilização do modelo não-linear da análise de componentes independentes (NLICA) para extrair características relevantes do sinais dos calorímetros para a discriminação elétron/jato no segundo nível de filtragem do ATLAS.

Considerando os diversos algoritmos e modelos existentes para a estimação da NLICA (mostrados no Capítulo 5), no contexto deste trabalho, foram utilizados três configurações:

- um algoritmo para o modelo sem restrições estruturais, neste caso foram utilizados mapas auto-organizáveis - SOM;
- algoritmos que estimam o modelo pós não-linear (PNL);
- adicionalmente, foi utilizado o modelo da ICA Local.

Foram utilizados também algumas das modificações que incluem informação supervisionada no processo de estimação dos componentes independentes (conforme detalhado na seção 5.5). Neste contexto, foram testadas duas possibilidades:

- utilização dos rótulos de classe como entrada para os algoritmos;
- compactação através da análise de componentes de discriminação (PCD).

As configurações listadas acima foram escolhidas com o objetivo de explorar os principais modos de estimativa da NLICA (que compreendem os modelos com restrições estruturais e os modelos livres), através de algoritmos com aplicação difundida na literatura (PNL e SOM).

Metodologia para Estimação da ICA Linear

No trabalho [10], foi realizado um estudo, com dados simulados, a respeito da utilização do modelo linear da análise de componentes independentes para extração de características dos sinais em anéis. De modo bastante completo, foram abordadas diversas configurações para a estimativa dos componentes independentes. A configuração que apresentou melhor resultado, em termos da eficiência de discriminação, comprehende as etapas listadas a seguir:

- Compactação dos sinais em anéis de modo segmentado a partir da análise de componentes principais (PCD).
- Estimação da ICA através do algoritmo JADE [149].
- Classificação a partir de um discriminador neural alimentado pelos componentes independentes.

Em comparação com o algoritmo FastICA [151] (outro algoritmo bastante utilizado para a estimativa dos componentes independentes no modelo de mistura linear), o JADE mostrou, em trabalhos anteriores [187, 188], melhores resultados em problemas de extração de características e separação de fontes.

Na comparação que será apresentada nesta tese, será considerado apenas o modo de estimativa da ICA que produziu melhor resultado.

Metodologia para Estimação da ICA Local

Na ICA Local, conforme mencionado no Capítulo 5, inicialmente é executada uma etapa de agrupamento não-supervisionado (*clustering*), onde o conjunto de sinais é mapeado em dois grupos. O número de agrupamentos a ser utilizado foi escolhido considerando que o problema é de classificação binária (em duas classes, elétrons e jatos).

O algoritmo utilizado para o agrupamento foi o *k-means* [136]. O *k-means* é um dos métodos agrupamento mais utilizados na literatura [189, 190, 191]. Uma outra alternativa para a realização do agrupamento seria o uso de mapas auto-organizáveis. Para os sinais utilizados, o desempenho dos dois métodos foi semelhante, porém, o esforço computacional exigido pelo *k-means* foi menor.

Após o agrupamento, os componentes independentes são estimados em cada grupo (novamente foi utilizado o algoritmo JADE). Na aplicação à filtragem *online* de elétrons no ATLAS, o objetivo é a separação em duas classes (elétrons e jatos), então, foi realizado o agrupamento em dois grupos (*clusters*) através do algoritmo *k-means*.

Após a divisão nos *clusters*, foi aplicado um algoritmo de ICA linear (JADE) para estimar os componentes independentes locais. Em seguida, classificadores neurais MLP de duas camadas (um para cada agrupamento) foram treinados a partir destes componentes. Neste trabalho, todas as etapas de processamento do discriminador baseado em ICA Local (agrupamento, ICA e classificação) foram realizadas de modo não-segmentado.

É importante notar que, devido às diferentes concentrações dos eventos de cada classe nos agrupamentos, os patamares que otimizam o desempenho local (considerando os agrupamentos individualmente) podem não ser ótimos quando o objetivo é maximizar a eficiência no conjunto de todos os eventos. Deste modo, precisa-se otimizar o SP (global) variando-se os patamares locais (Y_1 e Y_2). Para realizar esse procedimento, foi utilizado um algoritmo genético (AG) [159], onde o espaço de busca é composto por Y_1 e Y_2 e o objetivo é maximizar o SP. Uma das vantagens deste método de otimização é que não há a necessidade de encontrar as derivadas da função custo (que no caso é o produto SP) em função de Y_1 e Y_2 , além disso no AG, a busca é realizada globalmente (mais detalhes no Apêndice C). Um diagrama do treinamento do discriminador baseado em ICA Local é mostrado Figura 6.5.

Metodologia para Treinamento dos Mapas Auto-Organizáveis

Neste trabalho, os mapas auto-organizáveis foram utilizados por estarem diretamente relacionados com o paradigma da NLICA³. Porém, para estimar N compo-

³O SOM foi um dos primeiros algoritmos propostos para a NLICA, mais detalhes em [89]

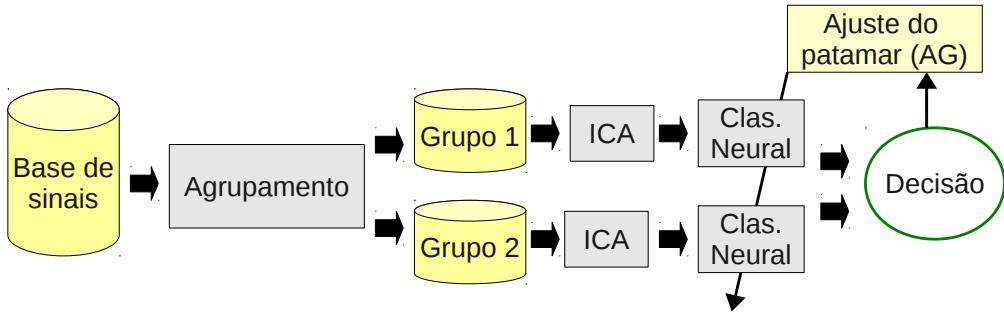


Figura 6.5: Diagrama do discriminador baseado em ICA Local.

nentes independentes, é preciso utilizar um mapa auto-organizável N-dimensional. Assim, para o problema em questão, onde seriam estimados algumas dezenas de componentes, este método seria proibitivamente lento. Um outro problema observado é que a precisão na estimativa também diminui com o aumento da dimensionalidade.

Uma solução encontrada foi a utilização de mapas bi-dimensionais, que, no contexto da NLICA, produzem uma estimativa de dois componentes independentes não-lineares (considerando um mapeamento sem restrições estruturais). Os mapas auto-organizáveis bi-dimensionais também são bastante úteis para a visualização e análise de dados de alta dimensão (monitoração *online* do *trigger* ou análise *offline* dos eventos salvos em mídia permanente).

Para treinamento dos mapas foram utilizadas diferentes abordagens, que podem ser classificadas de dois modos:

quanto à segmentação dos dados - em não-segmentado (considerando todos os 100 anéis como entrada) e segmentado (utilizando, separadamente, os sinais em anéis de cada camada do calorímetro, neste caso, são treinados 7 mapas, um para cada);

quanto ao uso de informação supervisionada no treinamento - em não-supervisionado (sendo este o procedimento tradicional de treinamento e referido no texto simplesmente como SOM), ajustado por LVQ [192] (neste caso, após o treinamento não-supervisionado, os pesos do mapa são ajustados, usando informação dos rótulos de classe, por um algoritmo de quantização vetorial por aprendizado - LVQ, para mais detalhes ver o Apêndice A) e supervisionado (este modo de treinamento foi descrito na Seção 5.5 e consiste

em utilizar os rótulos de classe como uma variável de entrada adicional). A Figura 6.6 ilustra os processos de treinamento dos SOM utilizando informação supervisionada



Figura 6.6: Diagramas dos processos de treinamento dos mapas auto-organizáveis utilizando informação a respeito das classes através de LVQ (esquerda) e rótulos de classe como entradas (direita).

Considerando que serão utilizados mapas bi-dimensionais, ainda é necessário escolher o número de neurônios a serem utilizados e sua distribuição na grade (quadrada, retangular, etc). Para guiar este processo, que usualmente é realizado de modo empírico (por tentativas) [193, 194], foram utilizados dois parâmetros: o erro de reconstrução médio e o índice soma-produto (SP), conforme descritos a seguir:

Erro de reconstrução médio (ERM) - Os mapas auto-organizáveis podem ser considerados como um método de quantização vetorial, onde os pesos dos neurônios formam o *codebook* (para mais detalhes ver o Apêndice A). Na operação do SOM, cada sinal \mathbf{x}_A apresentado na entrada é associado ao neurônio \mathbf{w}_c mais próximo (considerando, por exemplo, a distância euclidiana, este neurônio é chamado de “neurônio vencedor”). Os sinais associados a cada neurônio formam um agrupamento centrado em \mathbf{w}_c e podem ser aproximados pelo próprio neurônio. A partir desta correspondência, é gerado um erro de aproximação (ou reconstrução) definido por:

$$ER = |\mathbf{x}_A - \mathbf{w}_c|. \quad (6.2)$$

O erro de reconstrução médio é uma métrica que pode ser utilizada para indicar a qualidade da aproximação obtida pelo SOM e tende a diminuir com o aumento do número de neurônios do mapa (pois, neste modo, o tamanho médio dos agrupamentos associados a cada neurônio também tende a diminuir).

Índice soma-produto (SP) - Conforme definido na Equação 6.1, o produto SP está sendo utilizado como parâmetro para avaliar o desempenho dos classificadores. Associando-se cada neurônio do SOM à classe que mais o ativa, pode-se utilizar o mapa auto-organizável diretamente como classificador. O SP é calculado a partir das probabilidades de detecção e falso alarme produzidas pelo mapa.

No estudo do tamanho ótimo dos mapas, foram utilizados mapas quadrados ($L \times L$) e retangulares ($L \times P$), variando-se o número de neurônios de 2 a 100. Para o treinamento, foi utilizada a função de vizinhança gaussiana e aprendizagem em batelada. A largura da vizinhança é linearmente decrescida durante o processo de treinamento.

Para os mapas retangulares, a proporção entre L e P foi escolhida a partir da razão entre as energias retidas pelo primeiro e segundo componentes principais⁴:

$$\frac{L}{P} = \frac{E_{PC1}}{E_{PC2}}, \quad (6.3)$$

onde E_{PCi} é a energia do componente principal i . Para o conjunto de treino, obteve-se: $\frac{E_{PC1}}{E_{PC2}} \sim 2$, portanto os mapas retangulares utilizados tem a relação: $L=2P$.

Na escolha da dimensão ótima, o objetivo é manter um compromisso entre:

- Alto SP (consequentemente alta eficiência na discriminação);
- Baixo EMR (indicando boa representação dos dados);
- Reduzido número de neurônios (para garantir a rápida operação do sistema de classificação).

A partir dos mapas auto-organizáveis obtidos por diferentes abordagens de treinamento, foram calculadas as projeções dos sinais de entrada \mathbf{x} no espaço de características, fazendo:

$$\mathbf{u} = \mathbf{W}\mathbf{x} \quad (6.4)$$

onde \mathbf{W} é a matriz de pesos sinápticos do mapa, conforme definido com mais detalhes no Apêndice . Nos discriminadores baseados na extração de características por mapas auto-organizáveis, os classificadores são alimentados pelas projeções \mathbf{u} .

⁴Esta escolha tem o objetivo de melhor “acomodar” os dados em duas dimensões

Metodologia para Estimação do Modelo PNL

Considerando o modelo pós não-linear (PNL), foi comentado no Capítulo 5 que, com o aumento da dimensionalidade do problema (número de sinais envolvidos), há também um rápido aumento do custo computacional dos algoritmos e uma maior degradação na precisão da estimação dos componentes independentes.

O algoritmo utilizado neste trabalho para estimar a NLICA através do modelo PNL foi proposto no trabalho [108] e utiliza redes neurais para estimar as funções não-lineares (conforme descrito mais detalhadamente no Apêndice A). O treinamento é realizado através de um algoritmo gradiente que busca maximizar a independência entre os componentes estimados na saída do modelo. Este algoritmo foi escolhido por apresentar um custo computacional reduzido (se comparado a outros algoritmos que utilizam métodos de busca globais, como propostos em [114, 109]), e uma boa precisão na estimação dos componentes independentes, conforme observado no trabalho [89].

Com o objetivo de evitar problemas na estimação dos componentes independentes, devidos à alta dimensionalidade do problema, o algoritmo para o modelo PNL foi aplicado apenas no modo segmentado, pois assim, ao invés de utilizar os 100 anéis, a dimensão do problema é reduzida para o número de anéis de cada camada ($PS \rightarrow 8$, $E1 \rightarrow 64$, $E2 \rightarrow 8$, $E3 \rightarrow 8$, $H0 \rightarrow 4$, $H1 \rightarrow 4$, $H2 \rightarrow 4$).

A estimação dos componentes independentes pelo modelo PNL foi realizada de dois modos distintos, conforme descrito a seguir:

Utilizando uma etapa de pré-processamento para compactação - adicionalmente, foi incluído um pré-processamento para a redução da dimensão utilizando dois algoritmos de compactação: PCA (Análise de Componentes Principais - *Principal Component Analysis*) e PCD (Componentes Principais de Discriminação - *Principal Discriminating Components*) [11]. A compactação (assim como a estimação dos componentes independentes) também foi realizada de modo segmentado. Um diagrama do método proposto para a estimação dos componentes independentes através do modelo PNL é mostrado na Figura 6.7.

Através de um algoritmo PNL modificado - conforme descrito na seção 5.5.3

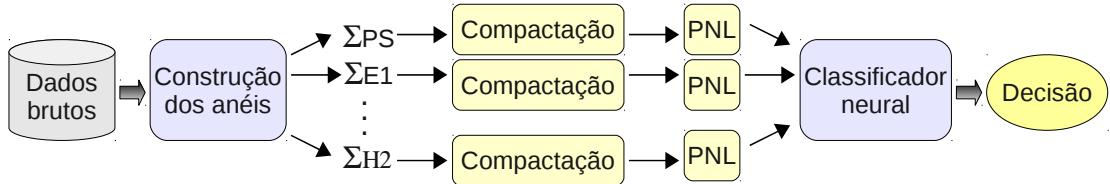


Figura 6.7: Diagrama do discriminador baseado no modelo PNL da NLICA.

e em [183], no desenvolvimento deste trabalho foi proposto um algoritmo para estimação do modelo pós não-linear da NLICA que utiliza informação supervisionada no processo de treinamento e estima, de modo simultâneo os blocos de compactação e do modelo PNL mostrados na Figura 6.7. É utilizado um algoritmo genético que busca otimizar uma função custo que avalia a independência entre os componentes e a discriminação obtida ao alimentar um classificador linear com os componentes estimados. O número de componentes ótimo para um dado problema é estimado de modo semelhante ao utilizado na compactação por PCD (ou seja com o gradual acréscimo de componentes até a estabilização da eficiência de discriminação próximo ao seu valor máximo). Uma vantagem do modelo proposto é que as tarefas de compactação e estimação dos componentes independentes são realizadas de modo simultâneo, pelo mesmo algoritmo. Por outro lado, o uso do algoritmo genético contribui para o aumento do custo computacional para treinamento do modelo (na operação *online*, após o treinamento, o custo computacional dos discriminadores baseados no modelo PNL é idêntico).

Conforme mencionado anteriormente (e descrito com mais detalhes no Apêndice A), na estimação das não-linearidades do modelo PNL são utilizadas redes neurais (cada função não-linear é aproximada por uma rede MLP de duas camadas com uma variável de entrada e um neurônio de saída com função de ativação linear). Um parâmetro que precisa ser escolhido antes do treinamento é o número de neurônios ocultos das redes aproximadoras. Quanto maior for este número, maior a capacidade da rede neural em aprender mapeamentos com alto grau de não-linearidade.

No contexto deste trabalho, o número de neurônios ocultos das redes aproxima-

doras foi escolhido através de tentativas. Inicialmente, foi determinado que todas as redes teriam o mesmo número de neurônios ocultos. A seguir, foram treinados diferentes modelos PNL variando-se o número de neurônios (iniciando-se por redes de um único neurônio oculto e adicionando-se neurônios gradualmente). Para cada novo modelo treinado, os componentes independentes foram calculados e utilizados para alimentar os discriminadores neurais. Deste modo, foi escolhida a configuração que apresentou maior eficiência de discriminação.

6.3 Classificação

Neste trabalho, foram utilizados classificadores neurais tipo perceptrons de múltiplas camadas (MLP) [5]. Mais detalhes sobre classificação de sinais e a implementação de classificadores neurais supervisionados podem ser encontrados no Apêndice B. Os classificadores, assim como os algoritmos de extração de características, podem operar nos modos não-segmentado e segmentado.

De modo semelhante ao descrito na seção 6.2.1, na classificação não-segmentada, as características estimadas são concatenadas num único vetor, que é usado como entrada para o classificador neural supervisionado (ver Figura 6.8).

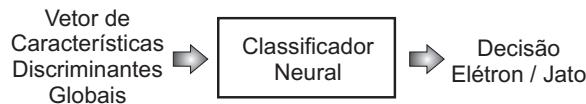


Figura 6.8: Decisão utilizando classificador global.

Quando as características são extraídas de modo segmentado (para cada camada), há a possibilidade de realizar o processo de classificação também de modo segmentado, ou seja, treinando classificadores especialistas nas características estimadas em cada uma das camadas. Neste caso, são utilizadas sete redes MLP, (uma para cada camada do calorímetro), conforme mostrado na Figura 6.9.

6.3.1 Motivação para o Uso de Classificadores Segmentados

No problema em estudo, o tempo de processamento é um fator crucial, e a maior parte do esforço computacional é exigido pelas rotinas de seleção de dados e anela-

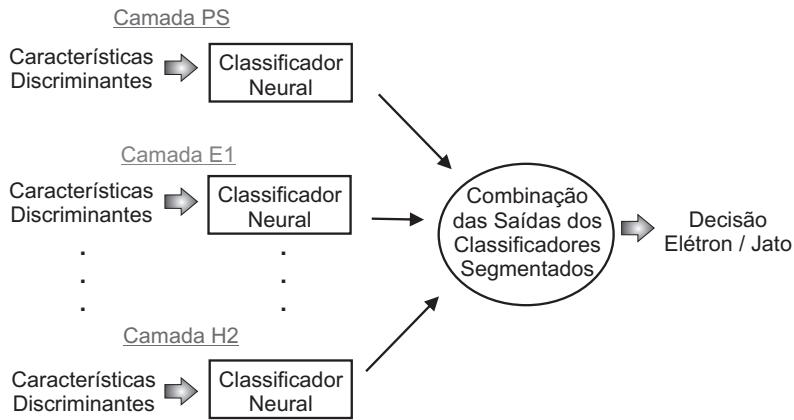


Figura 6.9: Decisão utilizando classificadores segmentados.

mento (geração dos sinais em anéis).

Uma característica particular do problema é que a seleção da informação das regiões de interesse no sistema de filtragem do ATLAS é realizada de uma única vez em cada camada, ou seja, mesmo que uma camada tenha poucos anéis relevantes, para acessá-los é necessário solicitar os dados de toda camada (mais detalhes serão apresentados na Seção 8.3.1).

Considerando o exposto, a realização dos processos de extração de características e classificação de modo segmentado (por camada do calorímetro) permite verificar a influência (relevância) de cada uma das sete camadas no processo de discriminação.

Uma estimativa da relevância da i -ésima camada no processo de identificação de elétrons pode ser obtida a partir do procedimento descrito a seguir:

- substitui-se (no conjunto de teste) os anéis da camada i pelos seus valores médios (calculados no conjunto de treino);
- propaga-se o conjunto de teste (com a camada i substituída pela média) por todas as etapas do discriminador (pré-processamento e classificação)
- calcula-se a eficiência obtida.

Quanto menor for o máximo SP calculado para a camada i substituída pela média, mais relevante esta camada é para o problema.

Alternativamente, o treinamento de classificadores especialistas na informação de cada camada do calorímetro (classificadores segmentados) também pode indicar a relevância das informações de cada camada. Neste caso, quanto maior o máximo SP

calculado para o classificador especialista na camada i , mais relevante é a informação desta camada para a identificação de elétrons.

Após o treinamento dos classificadores especialistas, pode-se também combinar as informações obtidas dos diversos segmentos, visando maximizar o desempenho de discriminação através da utilização de uma maior parte da informação disponível.

Um problema que surge na utilização de múltiplos classificadores é como combinar suas saídas para produzir a decisão final, o modo como este tópico foi tratado neste trabalho será descrito na próxima seção.

6.3.2 Combinação de Múltiplos Classificadores

A depender do tipo de saída escolhida para os classificadores, sendo variáveis contínuas (com excursão de -1 a 1) ou variáveis discretas (rótulos de classe), a combinação pode ser realizada através de estratégias distintas [195].

Considerando K classificadores com saídas contínuas u_k , uma forma usualmente utilizada para combinação é a média das saídas:

$$\mu(\mathbf{x}) = \sum_{k=1}^K u_k(\mathbf{x}) \quad (6.5)$$

Considerando que os múltiplos classificadores podem apresentar valores de eficiência diferentes, pode-se utilizar fatores de ponderação α_k , proporcionais à eficiência do discriminador k :

$$\mu(\mathbf{x}) = \sum_{k=1}^K \alpha_k u_k(\mathbf{x}), \quad (6.6)$$

assim, maior importância na combinação é dada aos discriminadores mais eficientes

Outra forma para a combinação de classificadores de saídas contínuas u_k é o cálculo da média geométrica:

$$\mu(\mathbf{x}) = \sqrt[K]{\prod_{k=1}^K u_k(\mathbf{x})} \quad (6.7)$$

Alternativamente, considerando que a saída dos múltiplos classificadores é o rótulo de classe associado ao vetor de entrada \mathbf{x} , um método muito utilizado para combinação das informações é a votação da maioria [195]. Neste caso, também podem ser utilizados fatores de ponderação, caso as eficiências dos classificadores

sejam diferentes. Deste modo, o voto de um classificador mais eficiente tem mais influência na decisão final.

Considerando que, neste trabalho os classificadores são redes MLP, é possível utilizar as saídas dos neurônios ocultos dos diversos classificadores como entradas para um outro classificador responsável por combinar a informação obtida de modo segmentado. Deste modo, conforme ilustrado na Figura 6.10, é formada uma rede MLP de 3 camadas de neurônios, porém, a camada de entrada não é totalmente conectada. O procedimento de treinamento desta rede neural compreende as etapas a seguir:

1. Treinamento dos classificadores segmentados;
2. Com as redes treinadas, calcula-se as saídas dos neurônios ocultos;
3. Concatena-se as saídas dos neurônios ocultos dos classificadores segmentados em um único vetor;
4. Uma rede neural combinadora (com duas camadas) é treinada a partir deste vetor de características.

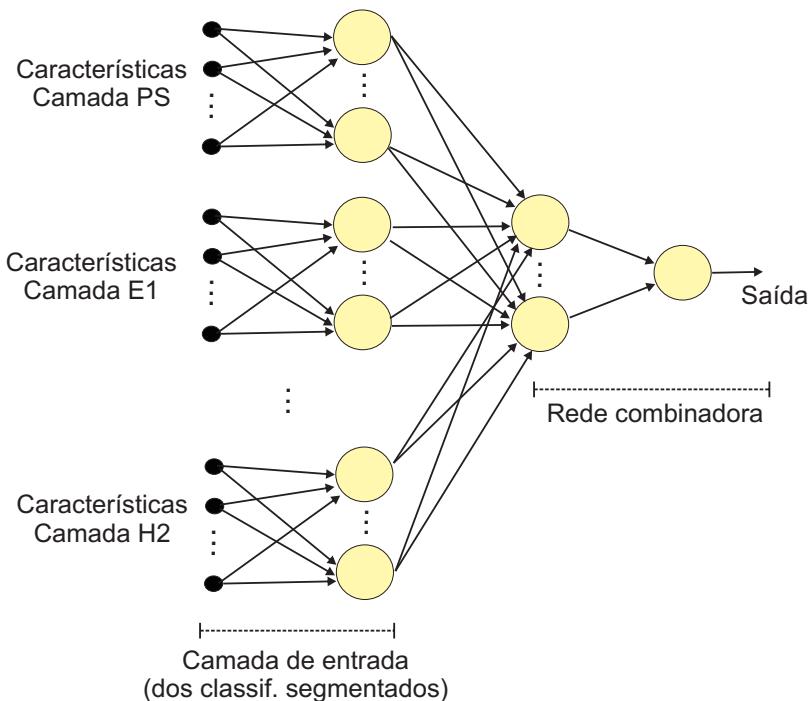


Figura 6.10: Modelo de rede combinadora para os classificadores segmentados.

6.4 Especificações de Treinamento

Em geral, foram utilizadas redes com uma camada oculta e um neurônio na camada de saída. Todos os neurônios utilizam funções de ativação tipo tangente hiperbólica.

Para o treinamento dos classificadores, foi utilizado o algoritmo RPROP (*Resilient Back-propagation*) [196]. E a inicialização dos pesos foi realizada através do algoritmo *Nguyen-Widrow* [197] Ao final do treinamento, são apresentados os pesos sinápticos da configuração de melhor desempenho para a figura de mérito utilizada (de acordo com o critério *save the best*).

Num trabalho anterior[10], foi realizado um estudo detalhado a respeito dos critérios de treinamento ótimos para o problema da seleção de elétrons no L2 do ATLAS considerando parâmetros como a figura de mérito para treinamento do classificador, o número ótimo de eventos na batelada, o número de épocas que garante a convergência do treinamento e o critério de parada. A Tabela 6.1 resume os resultados obtidos, que também foram adotados como valores de referência neste trabalho.

Tabela 6.1: Parâmetros de treinamento utilizados.

Parâmetro	Valor
Figura de mérito	SP
Tamanho da batelada	6.000 eventos por classe
Número de épocas	10.000
Critério de parada	Máximo número de épocas
Núm. Neurônios Ocultos	10

6.5 Bases de Dados

Neste trabalho, para avaliação da metodologia proposta, serão utilizadas bases de dados de características distintas. Toda fase de projeto e teste dos discriminadores foi iniciada antes do LHC entrar em operação, então, para estas tarefas foram empregados sinais simulados. A plataforma de *software* do sistema de filtragem do ATLAS (ATHENA) permite a produção de dados simulados onde, através de algoritmos de Monte Carlo, são consideradas as características físicas das colisões do LHC e a interação com o detector ATLAS (e seus diferentes subdetektore). É

possível também configurar diferentes estratégias de seleção para o primeiro nível de filtragem, simulando diversos cenários de operação do detector.

Com o início das colisões do LHC (a partir do final de 2009), foi possível adquirir dados experimentais. Um cronograma para a operação do acelerador foi fixado, compreendendo um gradual aumento da energia e da luminosidade dos feixes de prótons. Os sinais experimentais analisados foram obtidos em colisões onde a energia total é aproximadamente 7 GeV (3,5 GeV em cada feixe) e a luminosidade máxima chegou a aproximadamente $10^{29} \text{ cm}^{-2} \text{s}^{-1}$. Comparando com os valores nominais de projeto para o LHC (energia igual a 14 GeV e luminosidade de $10^{34} \text{ cm}^{-2} \text{s}^{-1}$), percebe-se que ainda há uma considerável diferença entre as duas configurações. Sabendo que os sinais simulados foram produzidos considerando as condições nominais do acelerador, é possível que as características físicas dos sinais experimentais (neste estágio de operação) ainda não sejam perfeitamente refletidas nos conjuntos simulados.

Outra classe de sinais utilizada para teste dos discriminadores propostos são as assinaturas de raios cósmicos. Essas partículas, originadas fora da atmosfera terrestre, são altamente energéticas e capazes de penetrar na crosta terrestre atingindo o ATLAS a 100 metros de profundidade e interagindo com os calorímetros. É necessário verificar a robustez dos algoritmos propostos quanto a rejeição das assinaturas de raios cósmicos, pois eles também podem constituir o ruído de fundo na identificação de elétrons.

A seguir, será apresentada uma breve descrição dos conjuntos de sinais utilizados. Inicialmente serão abordados os dados simulados, e em sequencia, os sinais experimentais.

6.5.1 Dados Simulados

Os eventos simulados utilizados foram produzidos por simulações de Monte Carlo [14] pela colaboração do ATLAS, utilizando os simuladores *Pythia* (que produz as colisões) e *Geant* (que realiza a interação dos eventos com o detector) [45]. Os sinais de interesse são elétrons isolados distribuídos uniformemente na faixa $7 < E_T < 80$ GeV (onde E_T é a energia transversa). O ruído de fundo é constituído de jatos com pelo menos um componente eletromagnético com $E_T > 17$ GeV.

Neste trabalho foram utilizados dois conjuntos distintos de dados simulados:

- Ao primeiro conjunto, foi aplicado um corte de nível 1 que simula a etapa inicial de operação do detector, onde a seleção do L1 é mais branda. A filtragem de primeiro nível realiza apenas um corte linear na energia total do evento, deixando passar somente eventos com energia maior que 7 GeV (os eventos de baixa energia são eliminados). Considerando a imprecisão do primeiro nível no cálculo da energia total do evento, as assinaturas que chegam ao L2 após este corte estão concentradas em energias maiores que 10 GeV e, por isso, são denominadas E10 (ou E10-*loose*). Este corte é mais utilizado para teste do sistema de filtragem no período inicial de operação do detector, pois é pouco provável a identificação de assinaturas relevantes, uma vez que, uma grande quantidade de eventos de falso alarme é aprovada pelo nível 1.
- O segundo corte utilizado é mais seletivo, se assemelhando aos que serão utilizados num estágio mais avançado da operação do detector, onde o objetivo é eliminar a maior parte do falso alarme. Neste caso, o nível 1 utiliza uma combinação de critérios como:
 - energia total maior que 13 GeV;
 - corte de vazamento para as camadas hadrônicas: a energia nas camadas hadrônicas é calculada e, se ela for maior que um valor limite (neste caso 1 GeV), o evento é rejeitado pelo L1;
 - corte de isolamento em energia: efetuado a partir da relação entre a energia no centro ($\sum E_C$) e na periferia ($\sum E_P$) do perfil de deposição de energia da segunda camada eletromagnética, se $\frac{\sum E_P}{\sum E_C}$ for maior que um limiar pré-estabelecido o evento é rejeitado pelo L1.

As assinaturas aprovadas neste corte do primeiro nível estão concentradas na faixa de energia maior que 15 GeV, sendo chamadas de E15i (o sufixo i indica que foram utilizados critérios de isolamento).

Na Tabela 6.2, são apresentados o número de assinaturas de cada conjunto, juntamente com os eventos que passaram pela seleção do L1. Conforme esperado, o corte E15i produziu uma redução muito maior do ruído de fundo (da ordem de 93 %), se comparado ao E10 (que eliminou apenas 55 % dos jatos). Em contrapartida, o corte E15i eliminou uma quantidade significativa de elétrons (52 %), enquanto que

o E10 preservou 98 % deles. Embora as características físicas dos eventos dos dois conjuntos (E10 e E15i) sejam as mesmas, a diferença entre a quantidade inicial de assinaturas é justificada pelo fato de terem sido gerados por versões distintas do *software* do sistema de filtragem (Athena).

Na Figura 6.11, são mostradas as distribuições em energia, η (pseudorapidez) e ϕ (ângulo azimutal) para elétrons e jatos dos conjuntos E10 e E15i após os cortes do L1. No gráfico da energia dos elétrons, percebe-se claramente a diferença no limiar de corte em energia para os dois conjuntos. Considerando a energia dos jatos, o corte E15i elimina a maioria das assinaturas com energia mais alta (pois nestas assinaturas os cortes por vazamento hadrônico e isolamento em energia são mais eficientes). Sendo assim, os jatos E15i têm um ligeiro deslocamento em relação a energias mais baixas, se comparados aos do conjunto E10. As distribuições em η e ϕ são semelhantes para os dois conjuntos.

Nas Figuras 6.12 e 6.13 são mostrados os eventos médios (média e desvio padrão de cada anel) para elétrons e jatos após os cortes E10 *loose* e E15i. Pode-se observar que os perfis dos elétrons são semelhantes nos dois casos. Por outro lado, os jatos E15i apresentam uma menor energia hadrônica e menor espalhamento (o que já era esperado, uma vez que foram efetuados cortes considerando estes parâmetros) do que os jatos E10, sendo mais parecidos com os elétrons e portanto representando um ruído de fundo de mais difícil identificação.

A correlação entre os 100 anéis para sinais de elétrons e jatos dos dois conjuntos simulados é mostrada na Figura 6.14. Pode-se observar que, para elétrons, há uma forte correlação entre os primeiros anéis das camadas eletromagnéticas (tanto dentro das camadas, como entre diferentes camadas). Nos jatos, a correlação é estendida para as camadas hadrônicas.

Tabela 6.2: Composição das bases de dados utilizadas antes e depois do corte de primeiro nível.

	E10			E15i		
Assinatura	Inicial	Após L1	Corte (%)	Inicial	Após L1	Corte (%)
Elétron	479.902	470.282	2.0	294.040	140.824	52.11
Jato	711.046	314.843	55.7	198.841	13.160	93.4

Embora a correlação mostre, de certa forma, a dependência entre os anéis, apenas os momentos estatísticos de primeira e segunda ordem são considerados no cálculo desta variável. Um parâmetro capaz de explorar todos os momentos é a informação mútua. A informação mútua $I(\cdot)$ entre duas variáveis aleatórias escalares u e v é definida como [98]:

$$I(u, v) = H(u) + H(v) - H(u, v), \quad (6.8)$$

onde $H(\cdot)$ é a entropia [87].

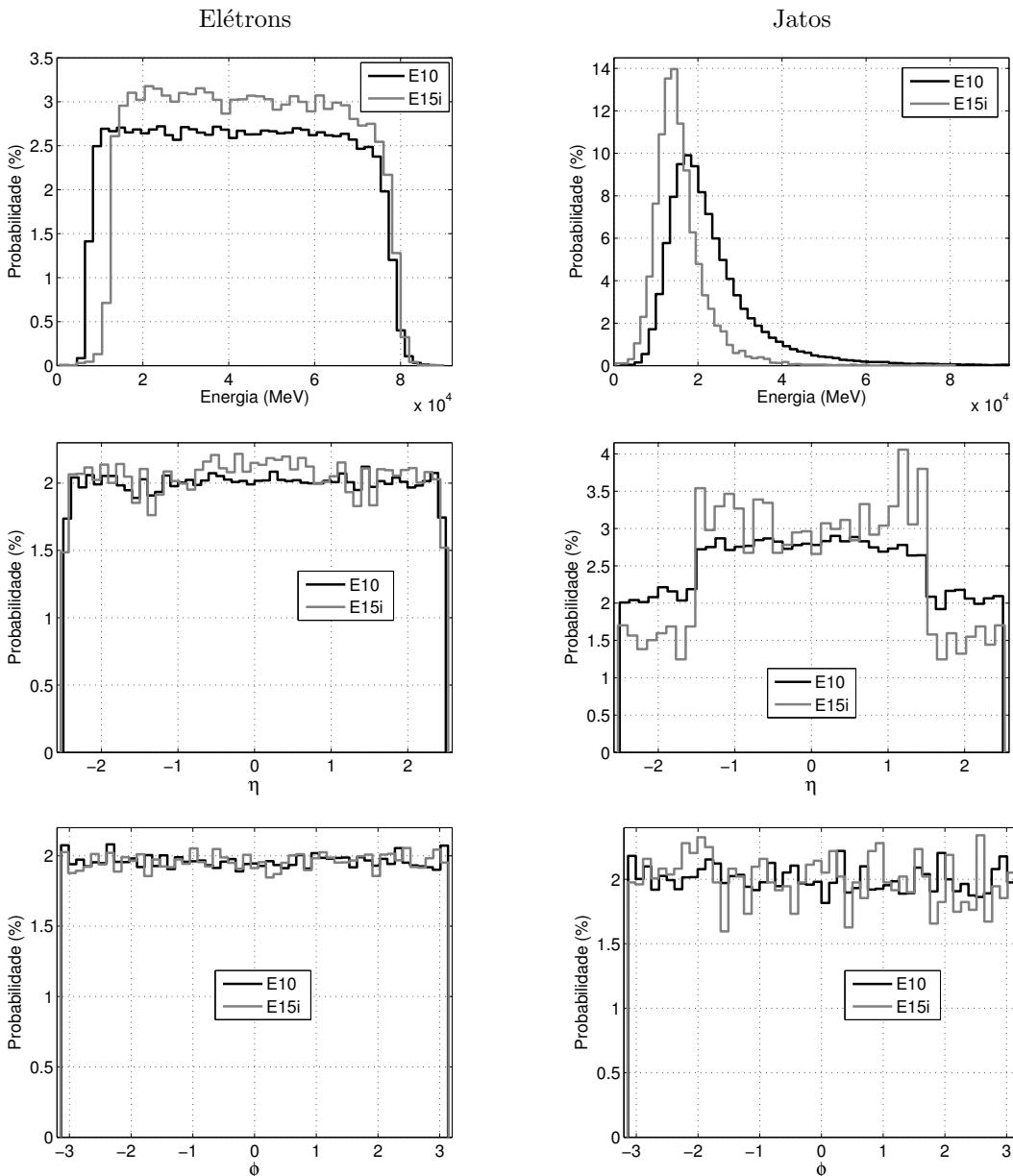


Figura 6.11: Distribuição em energia, η e ϕ das assinaturas de elétrons (esquerda) e jatos (direita) aprovadas pelos cortes de primeiro nível tipo E10 e E15i.

A informação mútua entre duas variáveis é nula quando elas são independentes, porém, não há um limite superior fixo (como no caso da correlação). Nesta análise, a entropia foi estimada através do método de *kernel* [198] e foi utilizado o procedimento de normalização proposto no trabalho [199] para limitar a informação mútua ao intervalo $[0,1]$. Observando a Figura 6.14, pode-se perceber que a dependência entre os primeiros anéis das camadas eletromagnéticas se confirma, além disso, é possível observar também alguma relação entre as camadas eletromagnéticas e hadrônicas (que não estava evidente na correlação). Ficou evidente também, que a dependência entre os anéis de jatos aumenta no conjunto E15i (em relação aos

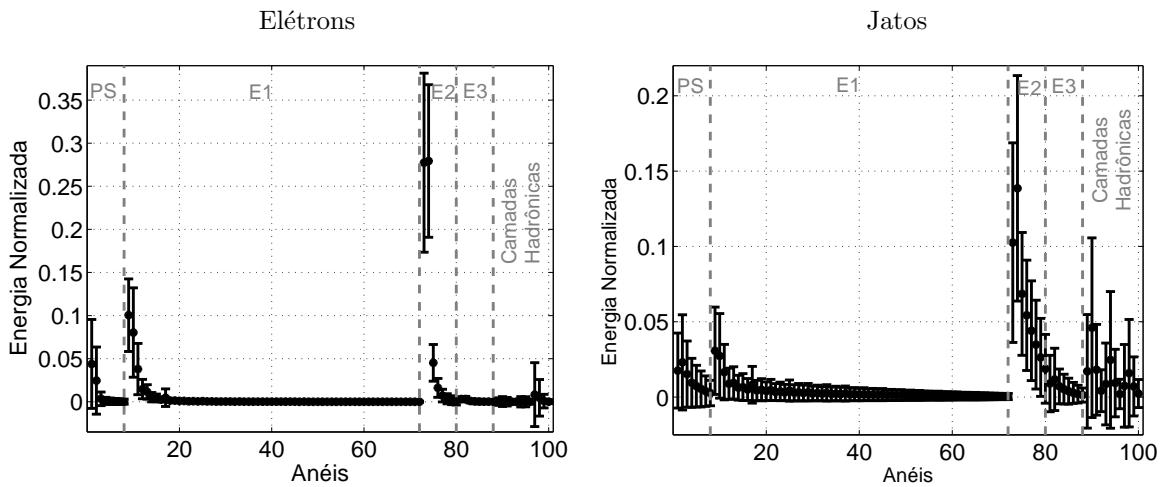


Figura 6.12: Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E10.

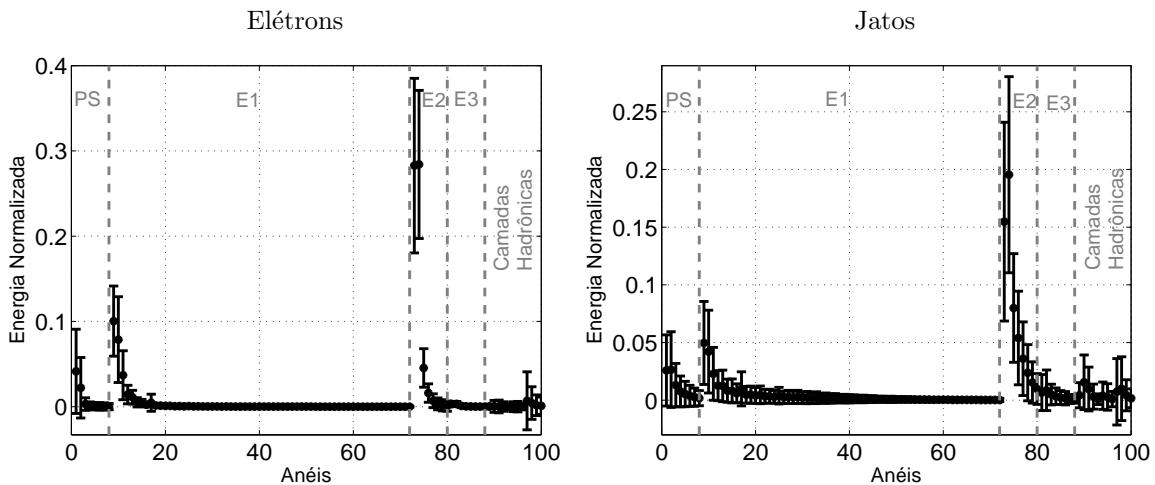


Figura 6.13: Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E15.

jatos do E10). Isso pode ser justificado pois, devido à seleção mais exigente no primeiro nível, os jatos do conjunto E15i são mais semelhantes a elétrons do que os do conjunto E10.

Conforme comentado no Capítulo 4, o treinamento de classificadores neurais pode ser otimizado com a redução da dependência entre as variáveis de entrada. Neste contexto, a aplicação da ICA/NLICA como um pré-processamento para o classificador *Neural Ringer* pode contribuir para um aumento na eficiência de discriminação.

6.5.2 Dados Experimentais

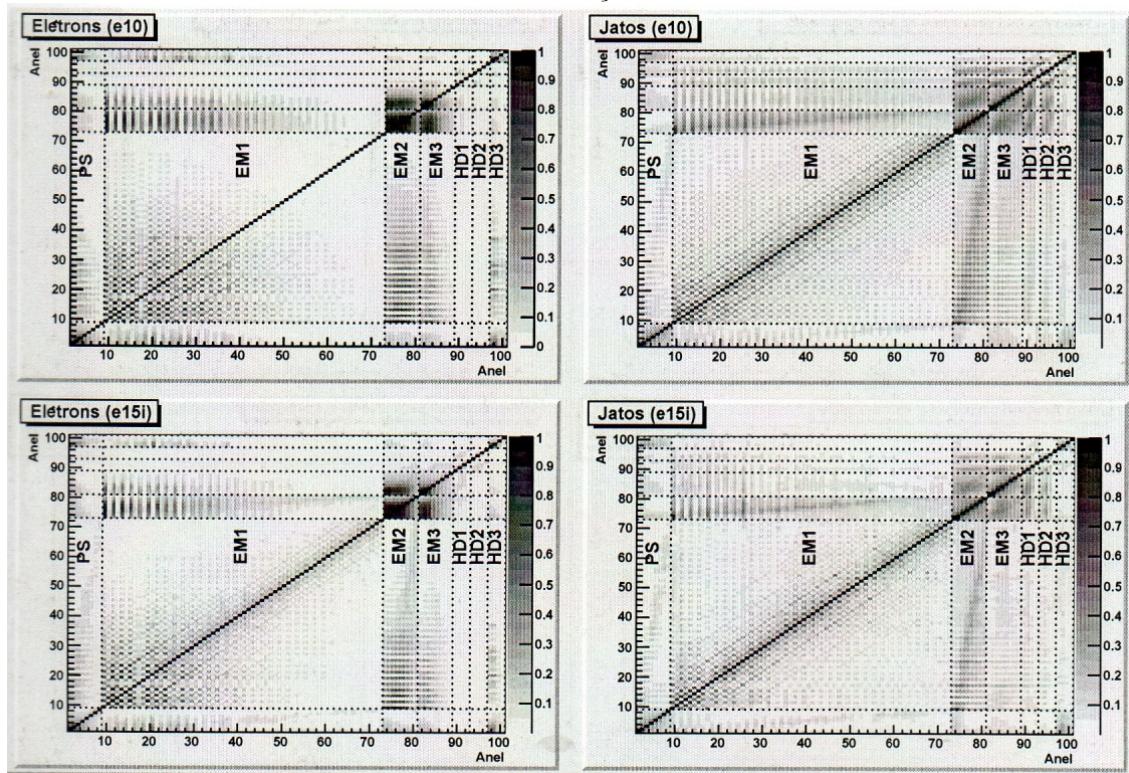
A partir do final de 2009, o LHC entrou novamente em operação, após uma interrupção para reparos (em setembro de 2008 foi identificado um vazamento de gás Hélio do sistema de refrigeração para o túnel do LHC, comprometendo a segurança do acelerador). Foi estabelecido um cronograma para o comissionamento do acelerador, que contemplou a gradual elevação na energia e na quantidade de prótons em cada feixe (o que é diretamente proporcional à luminosidade produzida).

Ainda em novembro de 2009 o LHC atingiu 1,18 TeV e definiu um novo recorde mundial (que até aquele momento era de 0.98 TeV, alcançado pelo acelerador Tevatron no Fermilab). No início de 2010 a energia foi elevada para 3 TeV e atualmente está na faixa de 7 TeV (que embora seja 7 vezes maior que a energia do Tevatron, é apenas metade da energia de projeto para o LHC que é de aproximadamente 14 TeV).

Na Figura 6.15, pode-se visualizar um evento (colisão) do LHC onde foram gerados um elétron e um anti-elétron num provável decaimento de um bóson Z. Nos dois ângulos de corte é possível visualizar a sensibilização dos calorímetros a partir da interação com o elétron e o anti-elétron (marcados em amarelo), que se deslocaram em direções opostas. Na parte superior esquerda é possível visualizar a energia depositada na segunda camada eletromagnética pelas duas partículas.

Para avaliar a eficiência dos métodos propostos, foram utilizados dois tipos de sinais experimentais: assinaturas de raios cósmicos e sinais medidos em colisões do LHC. As principais características destes conjuntos serão mostradas a seguir.

Correlação



Informação Mútua

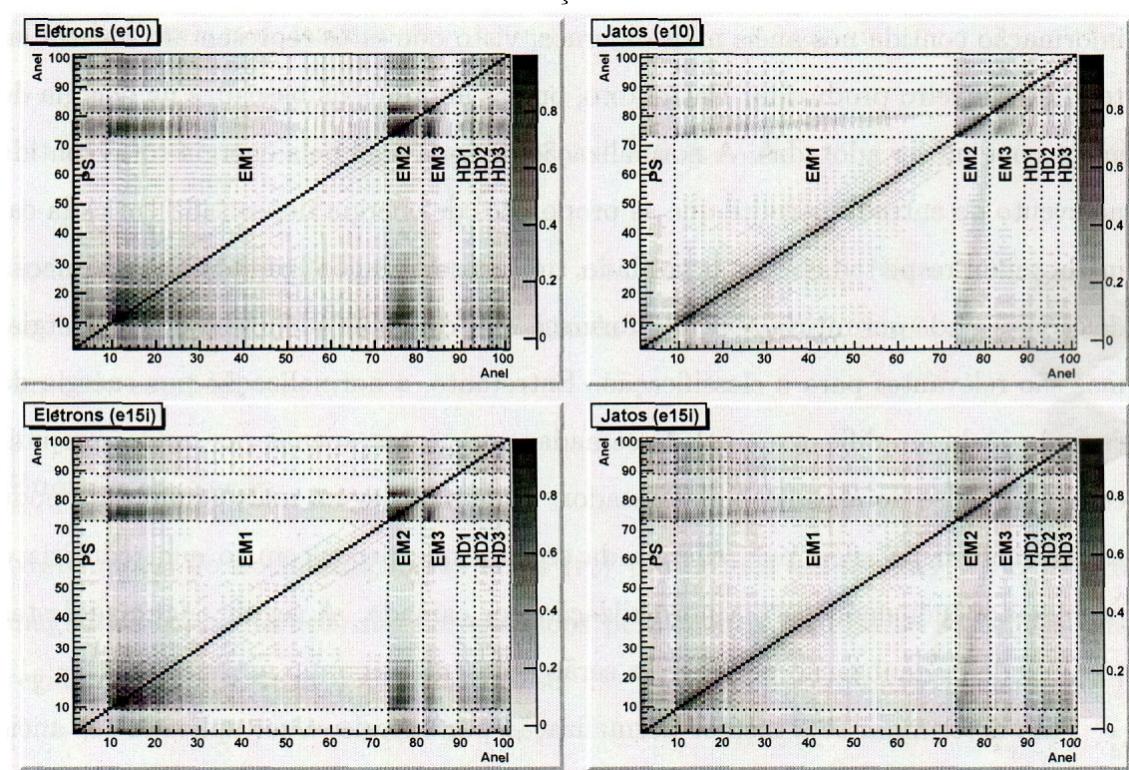


Figura 6.14: Correlação (acima) e informação mútua (abaixo) entre os 100 anéis, retirada de [10].

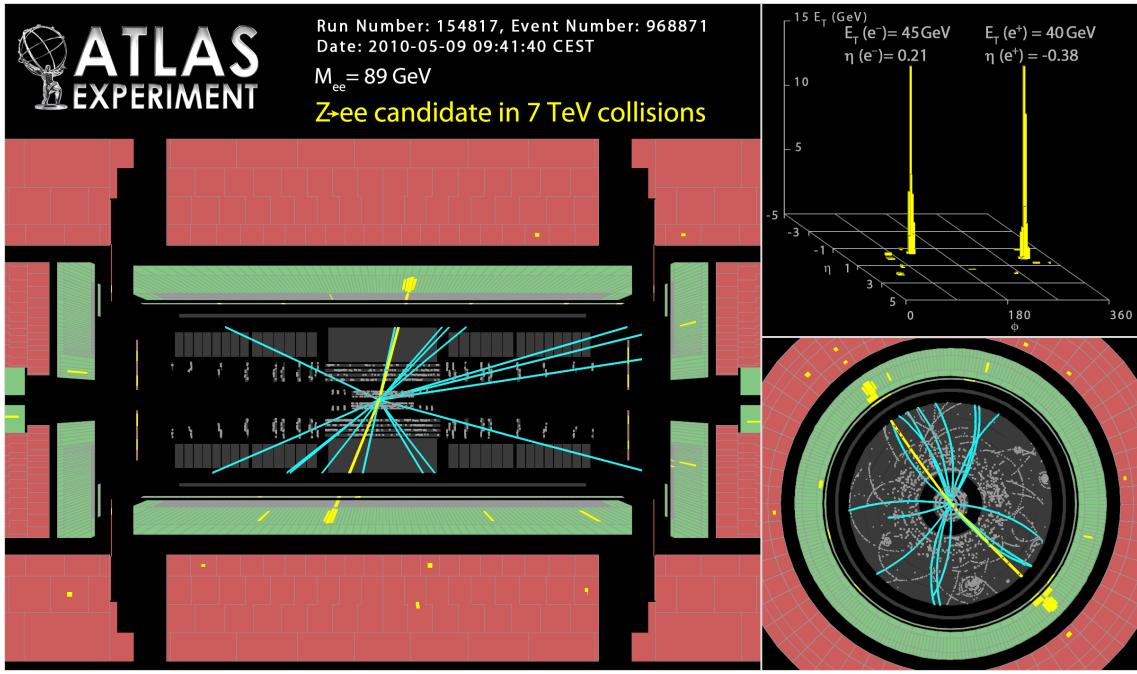


Figura 6.15: Visualização de um evento do LHC candidato ao decaimento de um bóson Z em dois elétrons (elétron e anti-elétron), extraído de [15].

Sinais de Raios Cósmicos

Os raios cósmicos [200] são partículas originadas no espaço, que se deslocam com velocidade próxima a da luz. Ao penetrarem na atmosfera terrestre, interagem com os átomos que a constituem produzindo uma “cascata” de novas partículas menos energéticas (conhecidas como raios cósmicos secundários que são compostos de diversas partículas, inclusive mísões [201]). O poder de penetração dos raios cósmicos é alto, podendo atingir o detector ATLAS (instalado a uma profundidade de aproximadamente 100 m) e interagir com o material dos calorímetros. Os raios cósmicos que chegam ao detector são compostos essencialmente por mísões. A energia destas partículas pode ser alta, atingindo até 10^{20} eV.

Alguns estudos conduzidos no ATLAS estão interessados na análise dos raios cósmicos, porém, considerando a detecção de elétrons, os raios cósmicos constituem uma fonte de ruído de fundo, que deve ser eliminada (ou pelo menos atenuada) pelo sistema de filtragem. Em momentos quando o LHC está desligado (não havendo portanto outra fonte de sinal para os calorímetros), foram coletadas diversas assinaturas originadas por raios cósmicos, que serão utilizadas para verificar a robustez do sistema de filtragem a essa fonte de ruído de fundo. O conjunto utilizado é composto

de 26.347 eventos adquiridos durante o ano de 2009.

Na Figura 6.16, pode-se observar a distribuição destes eventos para diferentes valores de energia, η e ϕ . Percebe-se que, embora concentrados abaixo de 100 GeV, estes eventos podem atingir energia bastante elevada (até a ordem de 400 GeV, ocorrendo eventos isolados em 500 GeV e ~ 1 TeV). Devido à proveniência (direção de chegada) da radiação cósmica que atinge o detector, a distribuição em η não é uniforme, havendo uma maior probabilidade destas partículas incidirem perpendicularmente ao detector (produzindo valores baixos de η). Considerando a distribuição em ϕ , percebe-se uma maior uniformidade.

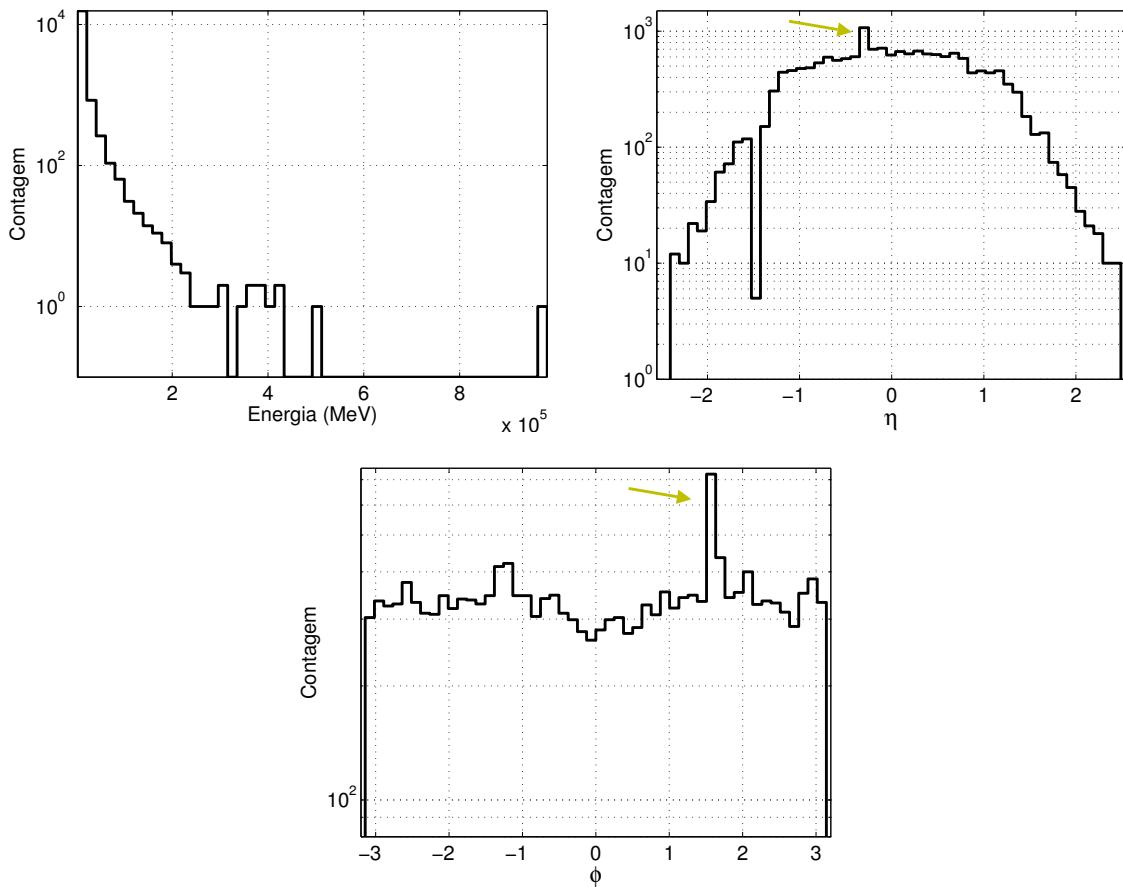


Figura 6.16: Histogramas em energia η e ϕ dos eventos de raios cósmicos.

Nos histogramas em η e ϕ , pode-se visualizar um grande pico em cada um deles, correspondendo à região em torno de $(\eta; \phi) = (-0.3; 1.6)$. Analisando-se o perfil de deposição de energia nesta região, observou-se uma grande quantidade de eventos que apresentam energia na camada E2, porém, nas outras apenas ruído (um evento deste tipo é mostrado na Figura 6.17). É muito improvável que um evento físico

real atravesse sete camadas do calorímetro e deposite energia em apenas uma delas. Neste contexto, concluiu-se que, estes eventos (chamados de eventos “fantasmas”) foram gerados por problemas nas células da segunda camada eletromagnética do calorímetro, que, por algum motivo desconhecido, acusam a deposição de energia mesmo quando nenhum evento foi recebido.

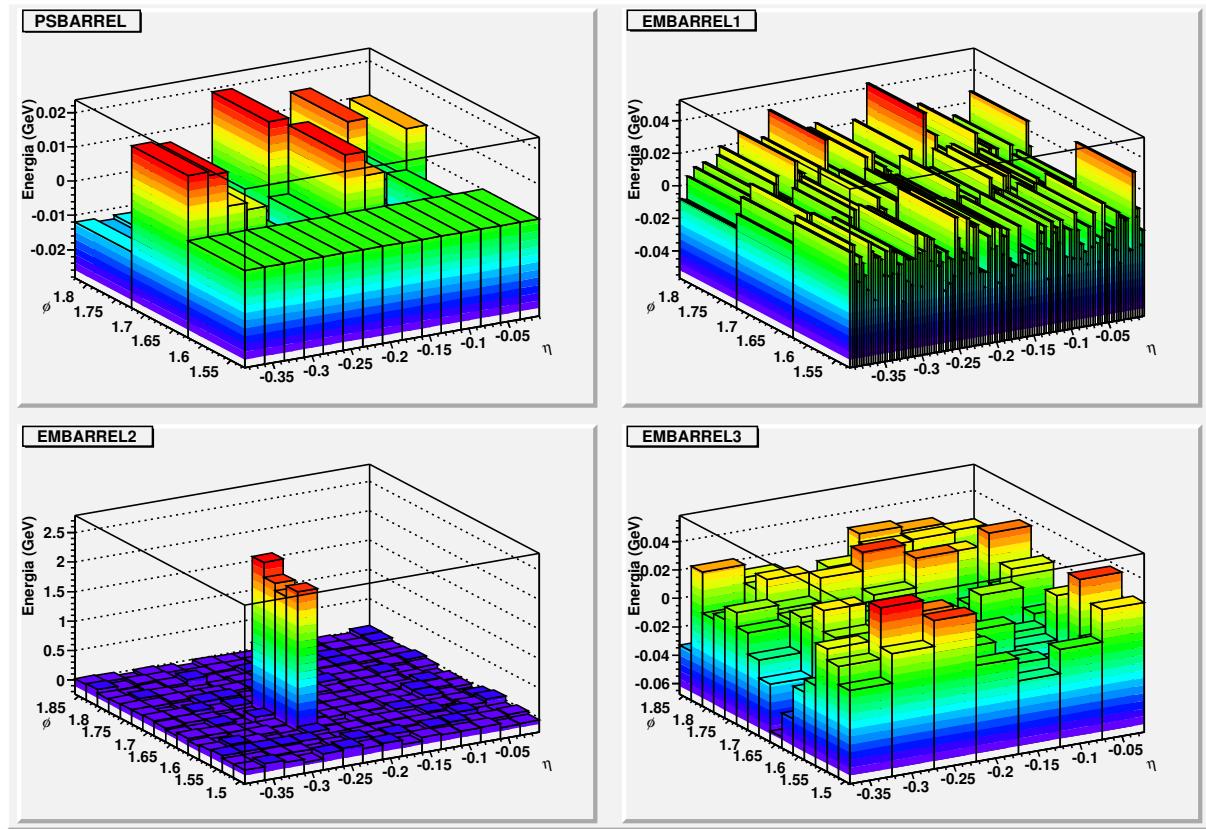


Figura 6.17: Exemplo de evento fantasma em $(\eta; \phi) = (-0.3; 1.6)$, retirado de [10].

Considerando a física envolvida no processo de interação dos raios cósmicos (que são compostos essencialmente por mísseis), espera-se que a maior parte dos eventos seja facilmente rejeitada pelos discriminadores. Porém, em alguns casos raros, mísseis podem interagir com o calorímetro produzindo fótons e pares $e^- e^+$. Estes eventos seriam capazes de produzir um perfil de deposição de energia semelhante ao de elétrons, confundindo o sistema de filtragem.

Sinais de Colisões do LHC

Em sua fase inicial de operação, o LHC irá operar com energias mais baixas do que a nominal (que é de aproximadamente 14 TeV). Nos dados mais recentes, a energia

máxima das colisões está limitada em 7 TeV, 3,5 TeV por cada feixe.

Outra característica importante é a luminosidade (que indica a quantidade de partículas por centímetro quadrado do detector a cada segundo), pois quanto maior a luminosidade, maior a probabilidade de ocorrência de eventos de interesse (e também de ruído de fundo). Na Figura 6.18, pode-se observar a variação das taxas de aceitação de assinaturas eletromagnéticas pelo primeiro nível em função da luminosidade. Fica claro que, as taxas aumentam, de modo aproximadamente linear, com a luminosidade.

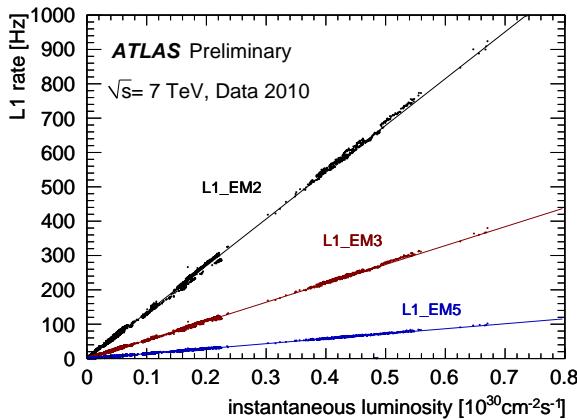


Figura 6.18: Taxas de aceitação de assinaturas eletromagnéticas no L1 em função da luminosidade [202].

As colisões de prótons do LHC realizadas em 2010⁵, contemplaram um gradual aumento da luminosidade. Nos primeiros meses de operação, a luminosidade foi mantida baixa e ocorreram muitos períodos de operação (*runs*) não-estáveis (nos quais, por algum problema de ordem técnica as colisões não produziram informações de interesse). À medida que a equipe de operação do acelerador ganhou maior experiência, a luminosidade foi aumentada, permitindo a produção de uma maior

⁵ As colisões de prótons no ano de 2010 começaram em 30 de março e encerraram em 4 de novembro, quando foi dado início ao programa de colisões de íons de chumbo (átomos de chumbo dos quais foram retirados todos os elétrons Pb⁸²⁺). As colisões com íons pesados produzem características semelhantes às que supostamente ocorreram no Big Bang, permitindo o estudo de acontecimentos que se suscederam à origem do universo. O LHC tem um detector especialmente projetado para lidar com estas colisões o ALICE, porém, os detectores ATLAS e CMS também irão adquirir informações destas colisões. No início de dezembro está previsto o desligamento do LHC por aproximadamente dois meses para manutenção de rotina.

quantidade de informações de interesse para os detectores.

Os eventos utilizados nas análises apresentadas nesta tese foram produzidos em *runs* da ordem de $10^{29} cm^{-2}s^{-1}$ (a luminosidade de projeto para o LHC é $10^{34} cm^{-2}s^{-1}$). Como a taxa de produção de eventos foi baixa, foram gravados todas as assinaturas aprovadas pelo primeiro nível (ou seja, a filtragem de alto nível não estava operando). As assinaturas gravadas foram, então, submetidas à análise *offline*. Visando a seleção de partículas eletromagnéticas (elétrons e fótons), a filtragem de primeiro nível realizou um corte na energia total em 2 GeV (L1EM2), gerando assinaturas de segundo nível concentradas em energias maiores que 3 GeV (assinaturas do tipo E3).

De modo distinto aos sinais simulados (que obtêm do próprio software simulador a certificação do tipo de partícula produzida), nos sinais obtidos experimentalmente nas colisões do LHC não há a caracterização prévia do tipo de partícula. Neste caso, para avaliar o desempenho dos discriminadores propostos, foram utilizadas informações obtidas da reconstrução *offline*, que indicam a probabilidade de um certo evento ser ou não um elétron, conforme descrito a seguir.

Critérios de Identificação de Elétrons na Análise *Offline*

O procedimento padrão para seleção de elétrons através da análise *offline* no ATLAS consiste em uma série de cortes lineares em variáveis capazes de discriminar o sinal de elétrons dos diversos tipos de ruído de fundo [203].

Ao todo, em torno de 20 variáveis são calculadas a partir das informações dos calorímetros e dos detectores de traço. Estes parâmetros podem ser consideradas individualmente ou combinadas em três conjuntos de referência para produzir três tipos de cortes pré-estabelecidos com diferentes graus de rejeição do ruído de fundo [204]:

- *Loose* - este critério se propõe basicamente a rejeitar a maior parte do ruído de fundo, nele são consideradas variáveis como a forma do chuveiro de deposição de energia na segunda camada eletromagnética e o vazamento hadrônico. Não é possível distinguir elétrons de fótons, pois este critério não considera informações do detector de traços.

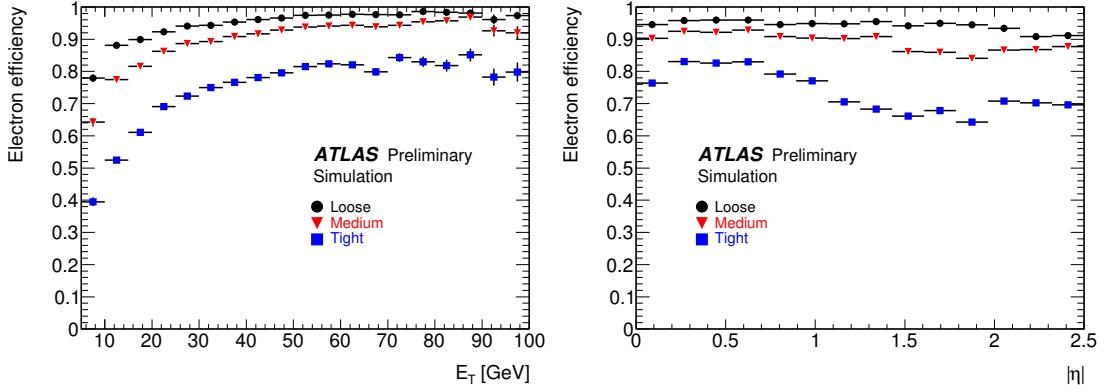


Figura 6.19: Eficiência dos cortes do *offline* em função de η e da energia, retirado de [202].

- *Medium* - apresenta uma maior rejeição de píons e de fótons que se converteram em elétrons⁶, pois, além das variáveis utilizadas no corte *loose*, considera também, a forma do chuveiro na primeira camada eletromagnética, a isolação lateral e uma análise básica das informações do detector de traços.
- *Tight* - além dos parâmetros utilizados no critério *electron medium*, este corte utiliza toda a precisão disponível no detector de traços, buscando, com precisão, a coincidência entre as posições do traço e do *cluster* identificado no calorímetro. Deste modo, é possível obter uma maior eficiência na identificação de elétrons e na eliminação de fótons.

A eficiência dos critérios de seleção pode variar em função da energia e da posição de interação no detector. Conforme mostrado na Figura 6.19, pode-se observar que, em troca de uma maior redução na aceitação de falso alarme, o critério *tight* produz menor eficiência na identificação de elétrons, principalmente em energia menor que 30 GeV e em $1 < |\eta| < 2$.

Nos trabalhos [204, 205], foi apresentado um estudo onde um conjunto simulado (por Monte Carlo) foi submetido aos critérios da seleção *offline* do ATLAS. Na Tabela 6.3 são mostradas as composições dos conjuntos original (sem nenhum corte) e após cada um dos cortes (para a associação dos tipos de assinaturas foi utilizada a “verdade” obtida do simulador). Pode-se observar que, a maioria ($\sim 82\%$) das

⁶Ao interagirem com o material do detector de traços, alguns fótons podem se transformar em elétrons, constituindo outra fonte de ruído para a identificação de elétrons isolados.

assinaturas aprovadas no corte *loose* são hadrons, $\sim 16\%$ são elétrons produzidos a partir da conversão de fótons durante a interação com o detector e apenas $1,6\%$ são elétrons. Nos critérios *medium* e *tight* a concentração de elétrons é gradualmente aumentada, enquanto que a de partículas hadrônicas vai sendo reduzida, de modo que, 48% das assinaturas aprovadas no critério *tight* são elétrons e a parcela de jatos hadrônicos foi reduzida a aproximadamente 30% .

Nos resultados apresentados nesta tese, foram considerados como elétrons os eventos aprovados no critério *electron tight* do *offline* (estes eventos tem maior probabilidade de serem realmente elétrons). Para o conjunto dos jatos, foram utilizados os eventos que não passaram no critério *electron loose* do *offline* (são eventos com probabilidade muito baixa de serem elétrons).

Run 1

O primeiro conjunto de sinais experimentais de colisões analisado foi obtido de uma *run* do dia 25 de maio de 2010⁷. A energia de cada feixe foi de 3,5 TeV (totalizando 7 TeV) e a luminosidade máxima igual a $2,1 \times 10^{29} cm^{-2}s^{-1}$.

Os 5.550.044 eventos aprovados no corte EM2 do primeiro nível (eventos que depositaram mais de 2 GeV de energia nos calorímetros), e consequentemente salvos em mídia permanente, pois a filtragem de alto nível não estava operando, foram submetidos aos critérios de seleção da análise *offline*. Aproximadamente 9.000 assinaturas foram aceitas pelo *electron tight* (e, então, consideradas nas análises como elétrons). Um total de aproximadamente 2.000.000 de assinaturas foram reprovadas

⁷A *run* 155697 foi iniciada às 19h 38min do dia 24/05/2010 e finalizada às 12h 12min do dia 25/05/2010.

Tabela 6.3: Composição das assinaturas de sinais simulados aprovados em diferentes critérios da análise *offline*, retirado de [204].

Critério	Composição (%)		
	Hádrons	Elétrons/conversões	Elétrons
Nenhum	82,1	16,3	1,6
Loose	67,5	28,4	4,1
Medium	66,5	22,8	10,7
Tight	29,6	22,4	48,1

no critério *electron loose* (e, assim, consideradas como jatos).

Na Figura 6.20, pode-se observar que, os eventos (para ambas as classes) estão concentrados numa faixa de energia abaixo de 15 GeV. Analisando as distribuições em η , percebe-se que, a concentração em torno de $|\eta| \sim 1,5$ (região do *crack*) é bastante reduzida (com uma menor quantidade de sensores nesta área do detector os eventos não são caracterizados adequadamente e a probabilidade de aceitação no corte do L1 é menor). Os eventos aprovados no critério *electron tight* também apresentam baixa concentração para $|\eta| < 0,5$ (o que pode ser explicado por uma menor eficiência deste critério de seleção nesta região do detector, conforme mostrado anteriormente na Figura 6.19). As distribuições em ϕ tem um comportamento mais uniforme.

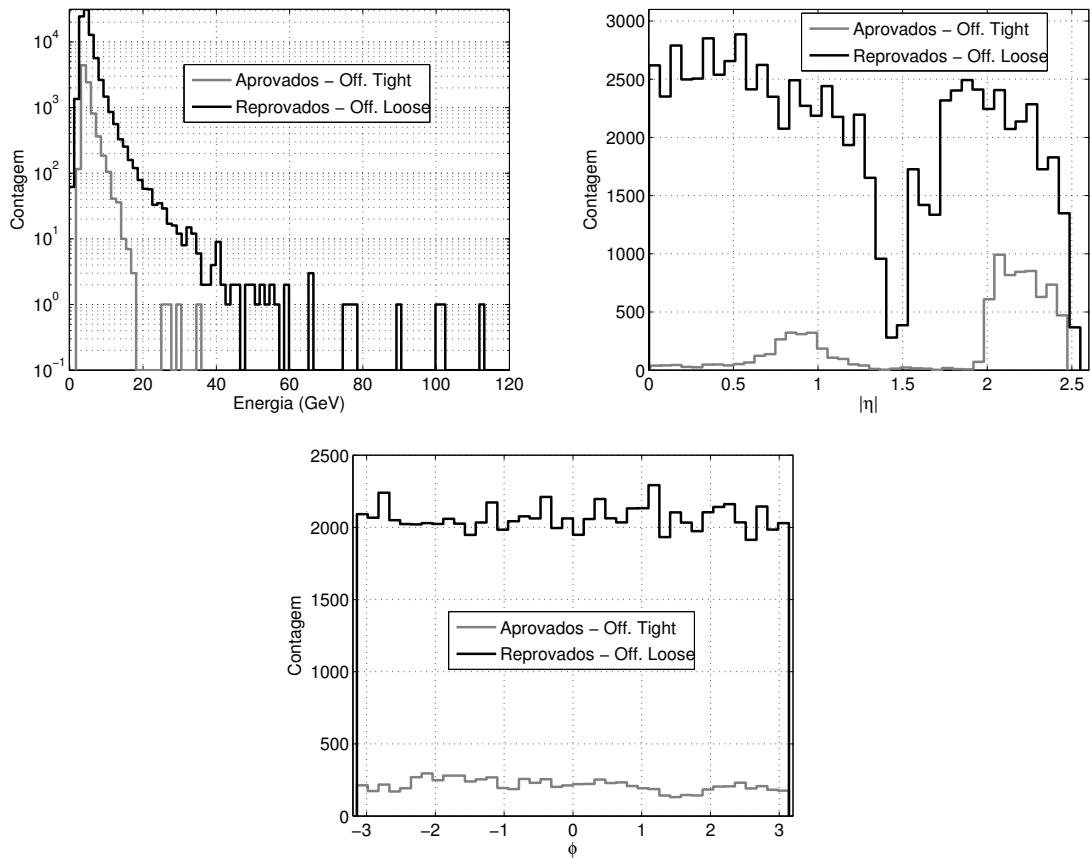


Figura 6.20: Histogramas em energia η e ϕ dos eventos de colisões do LHC.

Na Figura 6.21 são mostrados os eventos médios para as assinaturas aprovadas no critério *electron tight* (“elétrons”) e as reprovadas no critério *electron loose* (“jatos”). Como a implementação do *Ringer* na plataforma de software do ATLAS

utilizava, no momento da aquisição dos dados, a normalização sequencial, os eventos são mostrados neste formato. Pode-se observar que, os eventos se aproximam das características esperadas para as duas classes (semelhante ao observado nos sinais simulados). Uma particularidade observada nos eventos de “elétrons” é uma maior concentração de energia hadrônica se comparados aos sinais experimentais (que pode ser originada de fatores como a maior influência do ruído de medição em sinais de baixa energia e a possível presença de algumas assinaturas hadrônicas no conjunto aprovado no critério *electron tight*).

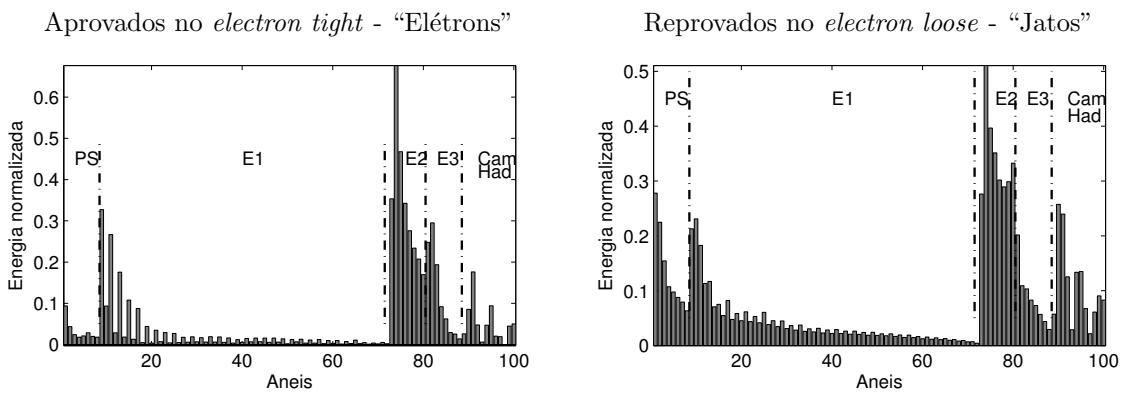


Figura 6.21: Sinais experimentais médios de “elétrons” e “jatos”.

Run 2

O segundo conjunto de sinais experimentais de colisões do LHC foram extraídos da run do dia 05 de julho de 2010⁸. As configurações do sistema de filtragem foram semelhantes ao da primeira *run* (todas as assinaturas aprovadas no primeiro nível foram salvas em mídia permanente). A energia das colisões foi mantida em 7 TeV, porém a luminosidade máxima foi aumentada para $6,1 \times 10^{29} cm^{-2}s^{-1}$ (três vezes maior que na *run 1*).

Devido a restrições de ordem operacional, apenas uma parte desta *run* está certificada pelos operadores do ATLAS para utilização (marcadas como *ready for physics*). Considerando este trecho, as assinaturas aprovadas no critério EM2 do primeiro nível (assinaturas que depositaram mais de 2 GeV nos calorímetros) foram submetidas à análise *offline* e aproximadamente 12.000 delas foram aceitas no

⁸A run 158582 começou às 01h 22min do dia 05 de julho de 2010 e terminou às 13h 59min do mesmo dia.

critério *electron tight* (e portanto consideradas como “elétrons”). Aproximadamente 1.200.000 assinaturas foram reprovadas no critério *electron loose* e portanto consideradas como “jatos”.

Nas Figuras ... pode-se observar as características destes eventos, e ...

- semelhança com o outro conjunto? - diferenças?

Capítulo 7

Resultados - Dados Simulados

A seguir, serão apresentados os resultados obtidos com a aplicação da metodologia proposta aos sinais dos calorímetros do ATLAS obtidos a partir de simulações por técnicas de Monte Carlo [14].

Conforme mostrado no Capítulo 6, serão utilizados dois conjuntos simulados distintos de assinaturas de segundo nível, aos quais foram aplicados diferentes cortes no primeiro nível. Nos dados simulados existe a informação do tipo de partícula correspondente a cada assinatura (fornecida pelo simulador), permitindo o treinamento supervisionado dos algoritmos de filtragem.

Para cada um dos conjuntos, inicialmente serão abordados os discriminadores existentes para o problema (T2Calo e *Neural Ringer* e *Ringer* com pré-processamento por ICA linear). Em seguida, serão considerados os classificadores com pré-processamento por NLICA, nos diversos modelos utilizados neste trabalho (SOM, PNL e ICA Local). Adicionalmente, será realizado um estudo a respeito da relevância de cada camada do calorímetro para o processo de discriminação de elétrons.

7.1 Conjunto E10

Nesta Seção, será utilizado o conjunto de sinais simulados que produziu as assinaturas de segundo nível do tipo E10 *loose* (energia maior que 10 GeV e sem considerar critérios de isolamento).

7.1.1 Resultados com os Discriminadores Existentes

No contexto da seleção de elétrons no segundo nível de filtragem online do ATLAS existem dois algoritmos implementados no sistema de *software* do detector, o T2Calo (discriminador padrão adotado pela colaboração do ATLAS) [1] e o *Neural Ringer* [9]. Adicionalmente, no trabalho [10] foi proposto o uso de ICA como pré-processamento para o discriminador *Neural Ringer*. A seguir serão apresentados os resultados obtidos com estes classificadores.

Na Tabela 7.1¹, os três discriminadores são comparados a partir do máximo SP e dos valores das probabilidades de detecção (PD) e falso alarme (PF), que indicam respectivamente as parcelas das assinaturas de elétrons identificados corretamente e jatos classificados como elétrons. Percebe-se que os discriminadores que utilizam o pré-processamento dos sinais dos calorímetros por anéis topológicos (*Neural Ringer* e ICA+*Ringer*) apresentam desempenho superior. É interessante notar que o pré-processamento por ICA possibilitou diminuição do falso alarme em 0,35 ponto percentual em comparação com o classificador que opera diretamente sobre os anéis (*Neural Ringer*). Nesta comparação foram utilizados os melhores resultados obtidos na aplicação da ICA como pré-processamento, que aconteceram em conjunto com a compactação por PCD, conforme detalhado no artigo [206].

Tabela 7.1: Comparaçāo de desempenho entre o T2Calo, *Neural Ringer*, e o discriminador neural com pré-processamento por ICA para o conjunto E10. São mostrados o máximo SP, os valores de PD e PF que produzem o máximo SP e PF para PD=92,5%.

Discriminador	Máx. SP×100	PD (%)	PF (%)	PF _{PD=92,5%}
T2Calo	91,99	92,50	8,55	8,55
Ringer	97,99 ± 0,18	98,77 ± 0,35	2,85 ± 0,11	1,10 ± 0,08
ICA	98,25 ± 0,27	99,01 ± 0,42	2,50 ± 0,09	1,00 ± 0,12

¹Os desvios mostrados correspondem à flutuação estatística obtida no processo de treinamento por validação cruzada.

***Neural Ringer* × T2Calo**

Na Figura 7.1, os discriminadores T2Calo e *Neural Ringer* são comparados quanto à eficiência na identificação de elétrons e ao falso alarme (jatos classificados como elétrons) em função da energia, η e ϕ . Nestes gráficos (assim como nas demais figuras semelhantes desta tese) a margem de erro (em %) é calculada como $100/\sqrt{N}$, onde N é o número de eventos contido na região analisada [207]. Considerando a energia, o *Neural Ringer* supera o T2Calo principalmente para energia mais baixa (onde os parâmetros calculados pelo T2Calo não são eficientes para separar as duas classes). Considerando a pseudo-rapidez (η), percebe-se que o T2Calo é fortemente influenciado pelos *cracks* (descontinuidades nos calorímetros em torno de $\eta = 1,5$). Os eventos que interagem perto dos *cracks* não são caracterizados corretamente, pois há uma menor quantidade de célula sensoras. Esta região também afeta o desempenho do *Neural Ringer*, porém com menor intensidade. Dos histogramas em ϕ , percebe-se que, para os dois discriminadores, tanto a eficiência como o falso alarme são distribuídos de modo aproximadamente uniforme (o que já era esperado, uma vez que, o detector é simétrico nessa coordenada).

Considerando o comparativo de eficiência (acerto×erro) mostrado na Figura 7.2, pode-se observar que, em conjunto, os dois classificadores identificam corretamente $\sim 92\%$ dos elétrons e $\sim 90\%$ dos jatos. Os erros conjuntos representam respectivamente $\sim 0,5\%$ e $\sim 1,2\%$. Aproximadamente 7% das assinaturas para as duas classes são identificadas corretamente pelo *Ringer*, porém indevidamente classificadas pelo T2Calo, enquanto que, no caso contrário, os índices foram de $\sim 0,5\%$ e $\sim 1\%$ respectivamente para elétrons e jatos.

Os histogramas da contagem percentual em energia e η dos eventos agrupados de acordo com o desempenho dos discriminadores são mostrados na Figura 7.3. Nestes gráficos, os valores percentuais foram calculados considerando, em cada faixa (*bin*) do histograma, a parcela dos eventos que pertence a uma das quatro classes de acerto/erro dos discriminadores (definidas anteriormente na Figura 7.2). Por exemplo, dos eventos de elétrons que tem energia menor que 4 GeV (e consequentemente pertencem ao primeiro *bin* do histograma), nenhum foi classificado corretamente pelo T2Calo, $\sim 30\%$ foi identificado pelo *Ringer* e $\sim 70\%$ foi classificado incorretamente por ambos os discriminadores. Deste modo, pode-se concluir que, elétrons

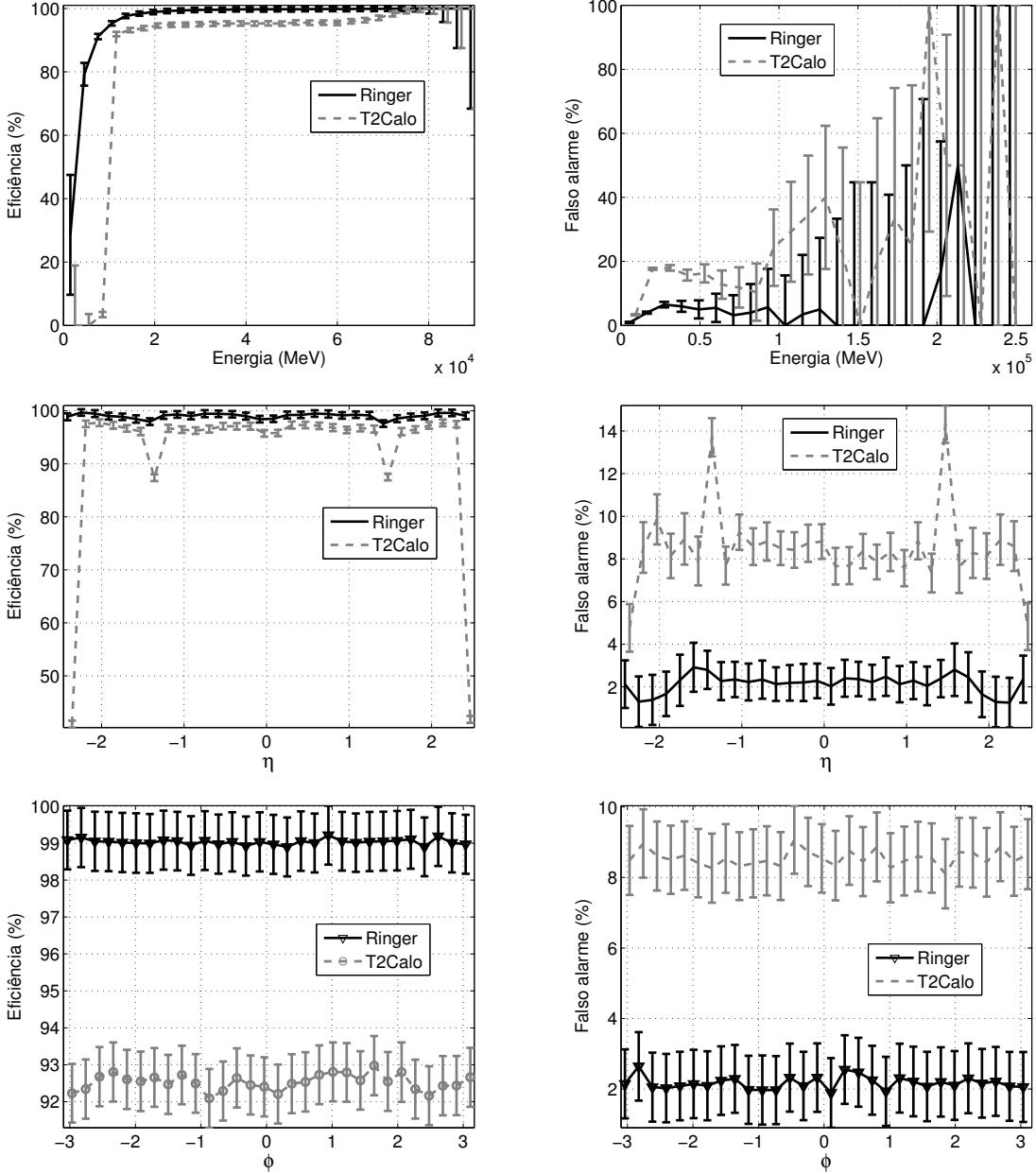


Figura 7.1: Eficiência (esquerda) e falso alarme (direita) em energia, η e ϕ para o *Neural Ringer* e o T2Calo.

de energia muito baixa são de difícil identificação para ambos os discriminadores (embora o *Ringer* tenha vantagem sobre o T2Calo).

Pode-se observar que, a probabilidade de detecção de elétrons aumenta com a energia. Para energias um pouco maiores ($7 < \text{Et} < 10$), o *Neural Ringer* é capaz de classificar corretamente elétrons que são perdidos pelo T2Calo. Com os jatos ocorre o contrário, em baixa energia são quase todos identificados por ambos os discriminadores. Para $\text{Et} \sim 20$ GeV e $\text{Et} > 80$ GeV há uma queda na eficiência do

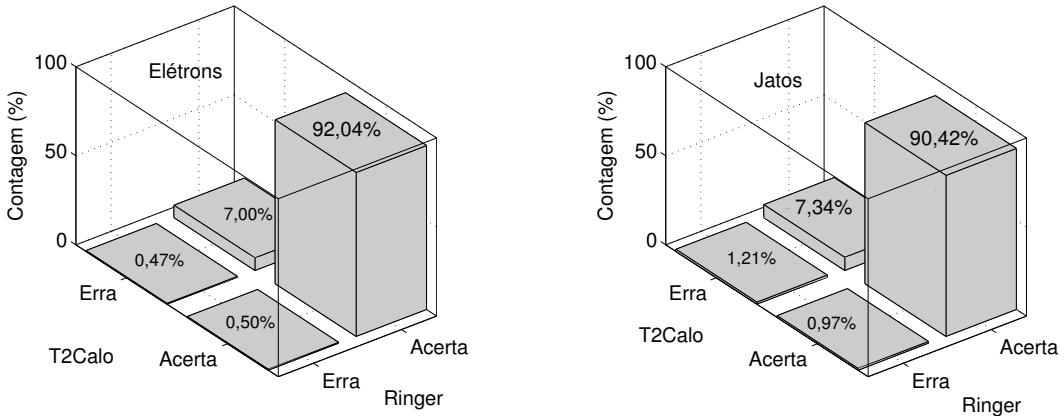


Figura 7.2: Contagem (%) das intersecções acerto×erro dos discriminadores *Neural Ringer* e *T2Calo* para elétrons (esquerda) e jatos (direita).

T2Calo que não afeta o *Ringer*.

Considerando os histogramas em η , observa-se que o *T2Calo* apresenta queda no desempenho próximo ao *crack* ($\eta \sim 1,5$), para ambas as classes. Em $|\eta| > 2,5$ (região próxima ao túnel do LHC, onde a granularidade do calorímetro é gradualmente reduzida, conforme mostrado na Tabela 2.4), o *T2Calo* apresenta uma severa queda de desempenho na identificação de elétrons. O *Neural Ringer* é mais robusto a estas limitações do calorímetro.

Continuando a análise de desempenho dos discriminadores, foram calculados os eventos médios de elétrons e jatos para os diferentes conjuntos (ver Figura 7.4). Pode-se observar que, o aparecimento de maior energia hadrônica, em conjunto com menor energia eletromagnética, dificulta a identificação dos elétrons pelo *Neural Ringer*. Os elétrons classificados incorretamente pelo *T2Calo* apresentam (na média) um pico de energia afastado do centro da RoI (célula mais energética) na camada E1.

Considerando os eventos de jatos, percebe-se que ambos os discriminadores possuem maior probabilidade de erro quando as assinaturas apresentam características semelhantes às de elétrons típicos (aumento da energia eletromagnética e diminuição da energia hadrônica).

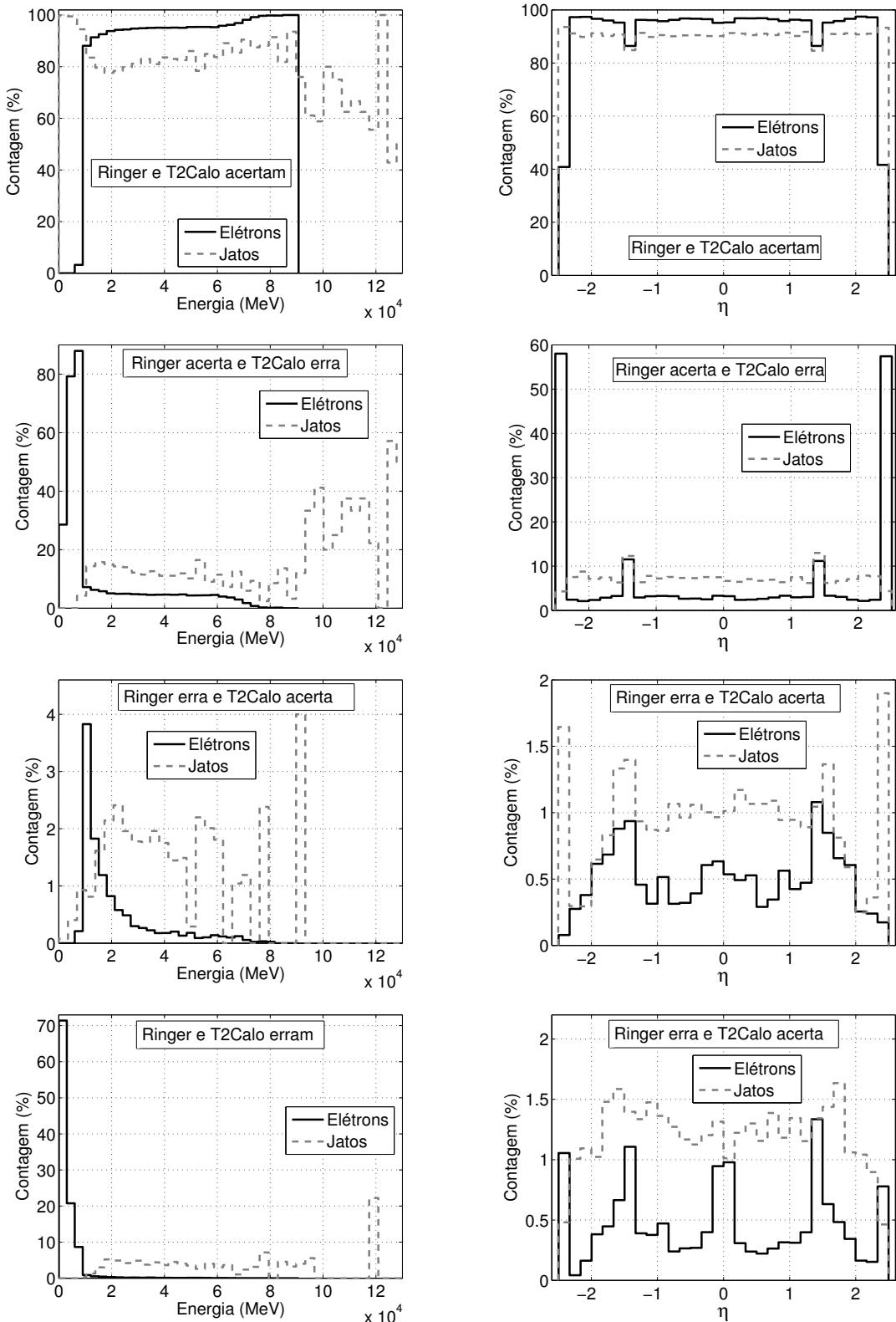


Figura 7.3: Histogramas da contagem (em %) em energia (esquerda) e η (direita) para assinaturas de elétrons e jatos de acordo com as decisões dos discriminadores T2Calo e *Neural Ringer*.

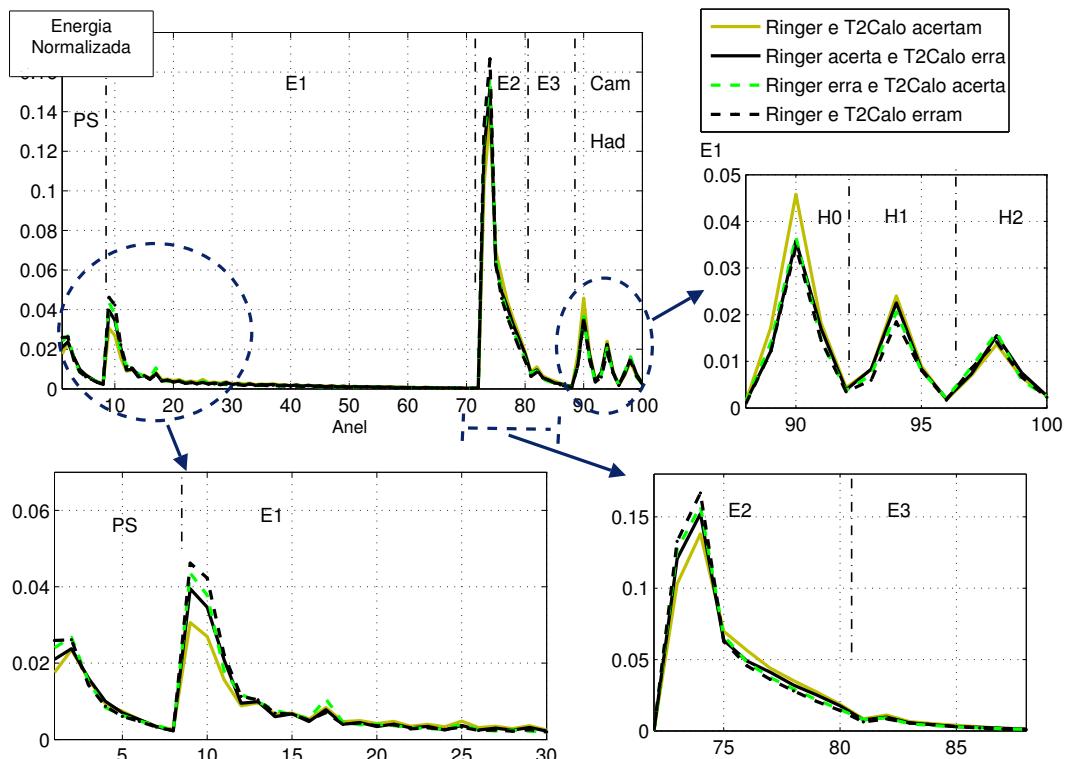
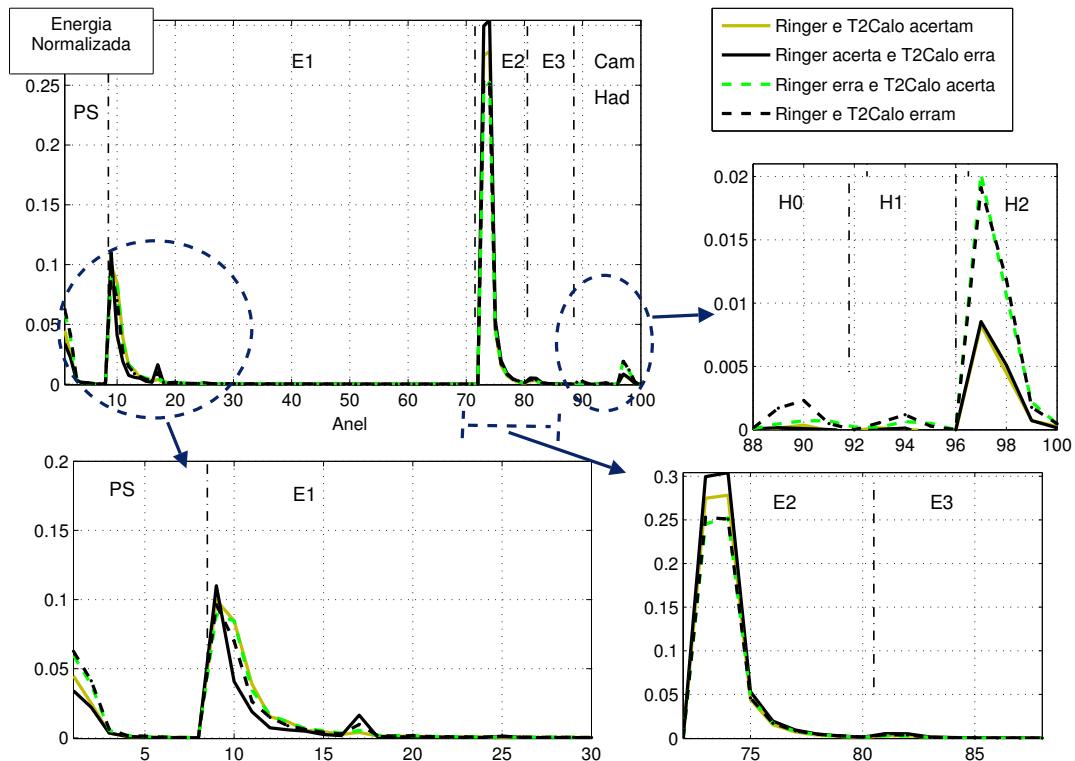


Figura 7.4: Comparação (*Neural Ringer* \times T2Calo) dos eventos médios para elétrons (acima) e jatos (abaixo).

Anéis × ICA

As curvas ROC obtidas para os discriminadores *Neural Ringer* e para o discriminador neural operando sobre os sinais em anéis projetados nos componentes independentes são mostradas na Figura 7.5. Pode-se observar que, o uso da ICA (segmentada e com compactação por PCD) produz um aumento na eficiência (considerando a área sob a curva ROC).

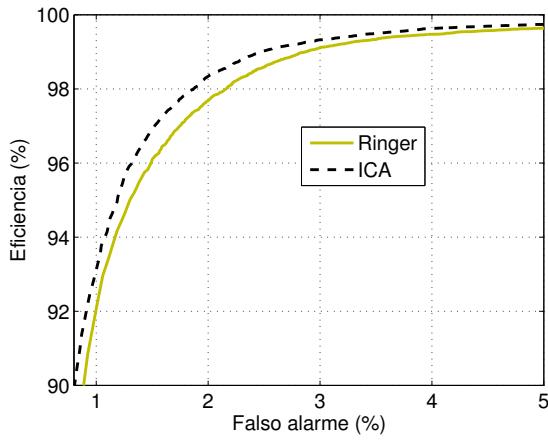


Figura 7.5: Curvas ROC para os discriminadores *Neural Ringer* e ICA+MLP.

Considerando o gráfico de dispersão das saídas das redes neurais alimentadas diretamente pelos sinais em anéis e pelos componentes independentes (ver Figura 7.6), pode-se observar que há uma grande concordância entre os dois discriminadores, tanto em termo dos acertos como dos erros.

Em termos quantitativos, conforme ilustrado na Figura 7.7, observa-se que, a intersecção entre os elétrons identificados corretamente pelos dois discriminadores corresponde a 98,32 % das assinaturas disponíveis. Com o pré-processamento por ICA, é possível identificar corretamente 0,69 % dos elétrons perdidos pelo *Neural Ringer*, porém, para 0,45 % dos elétrons o uso da ICA não é vantajoso em termos da eficiência de discriminação. As assinaturas de elétrons classificadas incorretamente por ambos discriminadores correspondem a 0,54 % dos sinais disponíveis.

Na Figura 7.8, é realizada uma comparação entre os discriminadores neurais operando diretamente sobre os anéis (*Ringer*) e sobre os anéis pré-processados por ICA em termos da eficiência e do falso alarme em energia e de η (as distribuições em ϕ são aproximadamente uniformes, então, foram omitidas desta análise).

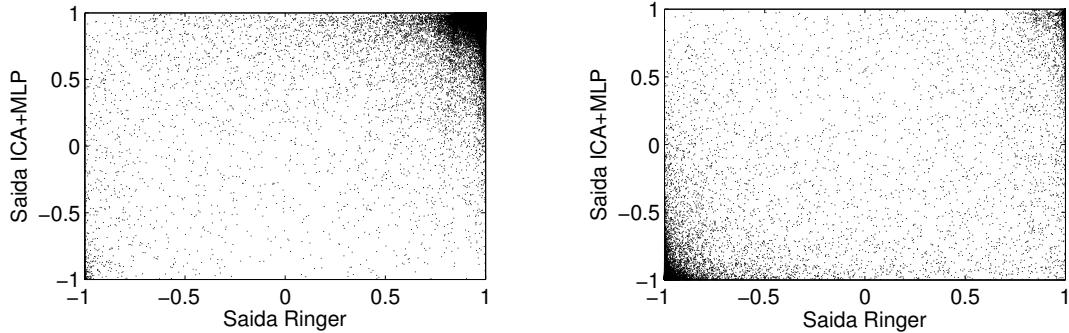


Figura 7.6: Dispersão das saídas dos discriminadores para elétrons (esquerda) e jatos (direita).

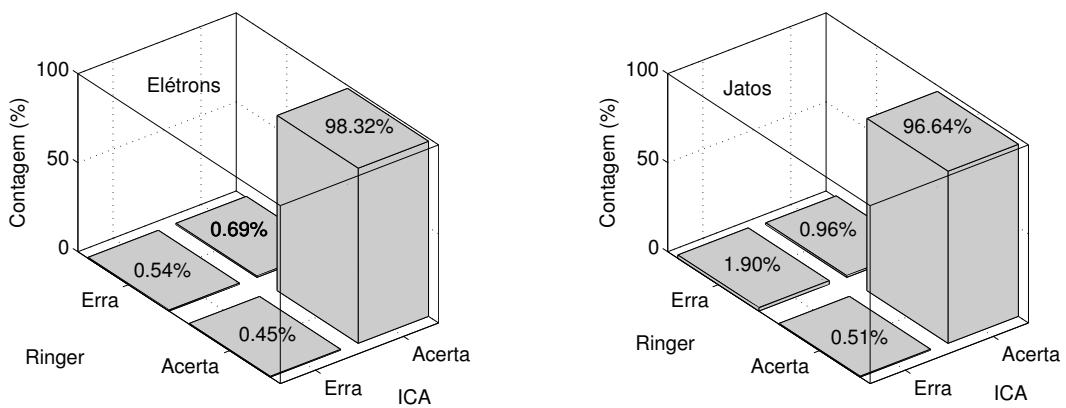


Figura 7.7: Contagem (%) das intersecções acerto×erro dos discriminadores neurais operando diretamente sobre os anéis (*Ringer*) e com pré-processamento por ICA para elétrons (esquerda) e jatos (direita).

Considerando a eficiência em energia, observa-se que, as duas abordagens apresentam desempenho semelhante para $E_t > 50$ GeV. Para energias muito baixas ($E_t < 10$ GeV) o *Ringer* apresenta desempenho melhor, porém, na faixa intermediária o ICA é superior. Analisando em η , o pré-processamento por ICA contribui para o aumento da eficiência principalmente em torno do *crack* ($|\eta| \sim 1,5$) e reduz o falso alarme região central do gráfico ($|\eta| < 1,5$).

Uma análise comparativa mais detalhada pode ser realizada a partir da Figura 7.9, onde percebe-se mais claramente que o pré-processamento por ICA produz uma aumento na eficiência de identificação de elétrons principalmente para energia entre 10 e 40 GeV e para $|\eta| \sim 1,5$ (região do *crack*). Na identificação dos jatos, o uso da ICA facilita a identificação para $|\eta| < 1,5$.

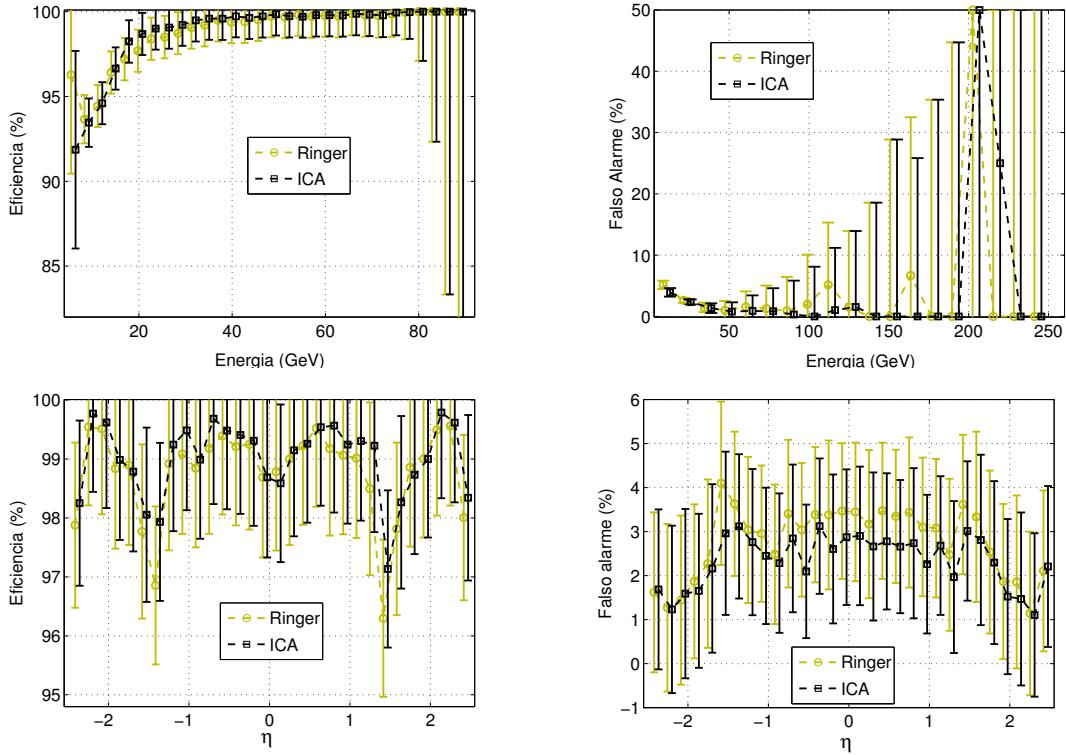


Figura 7.8: Eficiência (esquerda) e falso alarme (direita) em energia e η para o discriminador neural operando diretamente sobre os anéis (*Ringer*) e com pré-processamento por ICA.

7.1.2 Resultados com Pré-Processamento por NLICA

Nesta seção, serão apresentados os resultados obtidos com os discriminadores que utilizam pré-processamento por NLICA. O texto será dividido em três partes, inicialmente serão apresentados os resultados obtidos com a ICA Local, em seguida serão abordados os métodos baseados em mapas auto-organizáveis (SOM) e no modelo pós não-linear (PNL).

ICA Local

A extração de características por ICA Local, conforme mencionado no Capítulo 6, é dividida em duas etapas: agrupamento não-supervisionado (*clustering*) e projeção nos componentes independentes estimados localmente (para cada agrupamento).

A Figura 7.10 ilustra a divisão das assinaturas das diferentes classes em cada *cluster*. Pode-se observar que, o primeiro grupo é composto por aproximadamente 94 % dos elétrons e 11 % dos jatos e o restante dos eventos (6 % dos elétrons e 89 %

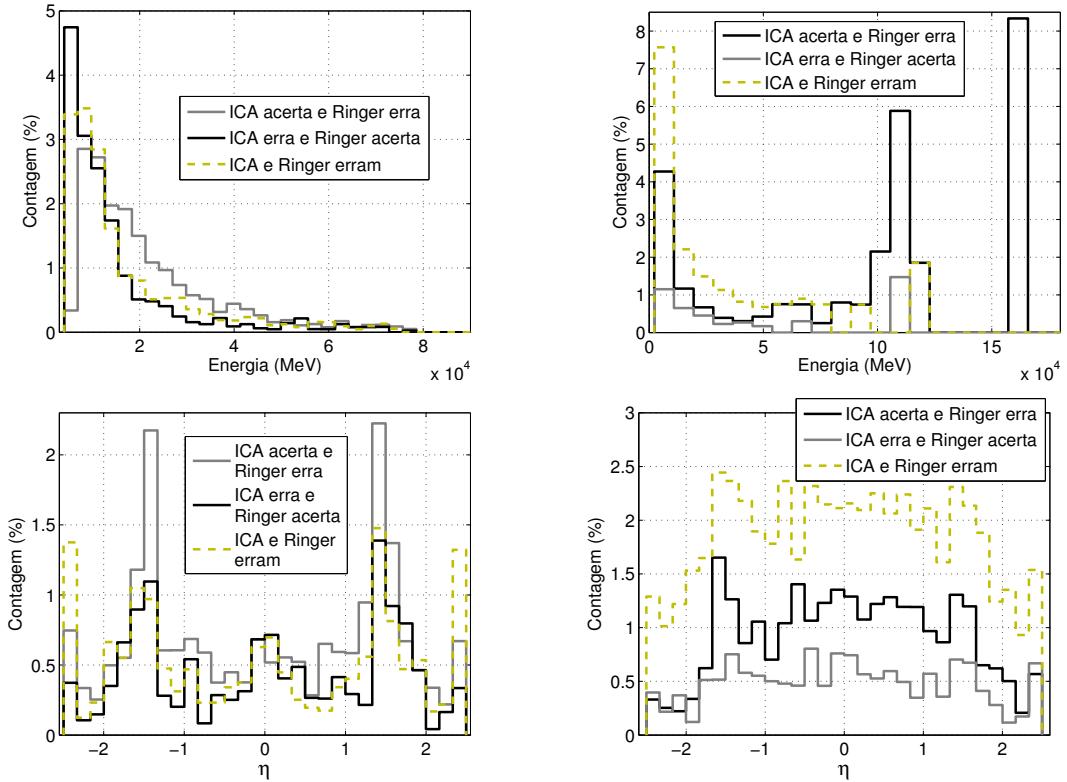


Figura 7.9: Comparativo de desempenho entre o discriminador neural operando diretamente sobre os anéis (*Ringer*) e com pré-processamento por ICA para elétrons (esquerda) e jatos (direita).

dos jatos) é designado ao segundo grupo.

Na Figura 7.11, são mostradas as distribuições em energia e η , para os eventos dos *clusters* 1 e 2. Pode-se observar que, os elétrons do grupo 1 apresentam distribuição aproximadamente uniforme em energia e η , exceto na região do *crack* ($|\eta| \sim 1,5$). Os elétrons do grupo 2 estão mais concentrados em energias mais altas e em $|\eta| > 1,5$ (com picos em $|\eta| \sim 1,5$). Este fato pode ser justificado, pois nesta região há uma menor concentração de células dos calorímetros, produzindo sinais com menor granularidade. Os jatos do grupo 1 apresentam maior concentração em energias menores que 30 GeV e $|\eta| < 2$, enquanto que os do segundo agrupamento são mais energéticos.

Considerando os perfis de deposição médios para elétrons e jatos nos agrupamentos e também em toda a base de sinais (ver Figura 7.12), percebe-se que o perfil médio dos elétrons do *cluster* 1 é muito semelhante ao de perfil médio dos elétrons calculado para todo o conjunto E10. Por outro lado, os elétrons associados

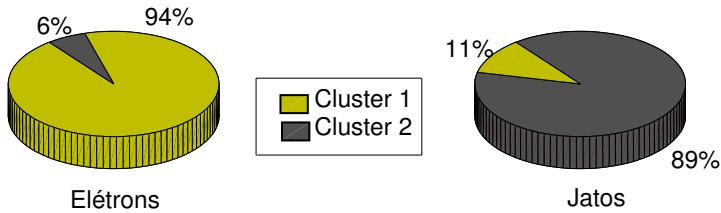


Figura 7.10: Probabilidade de agrupamento nos *clusters*.

ao segundo agrupamento tem perfil médio um pouco diferente, apresentado maior energia hadrônica (por isso foram agrupados num conjunto composto na sua maioria por jatos). Para os jatos ocorre o oposto, com a concentração dos sinais com perfil semelhante a jatos típicos no *cluster* 2.

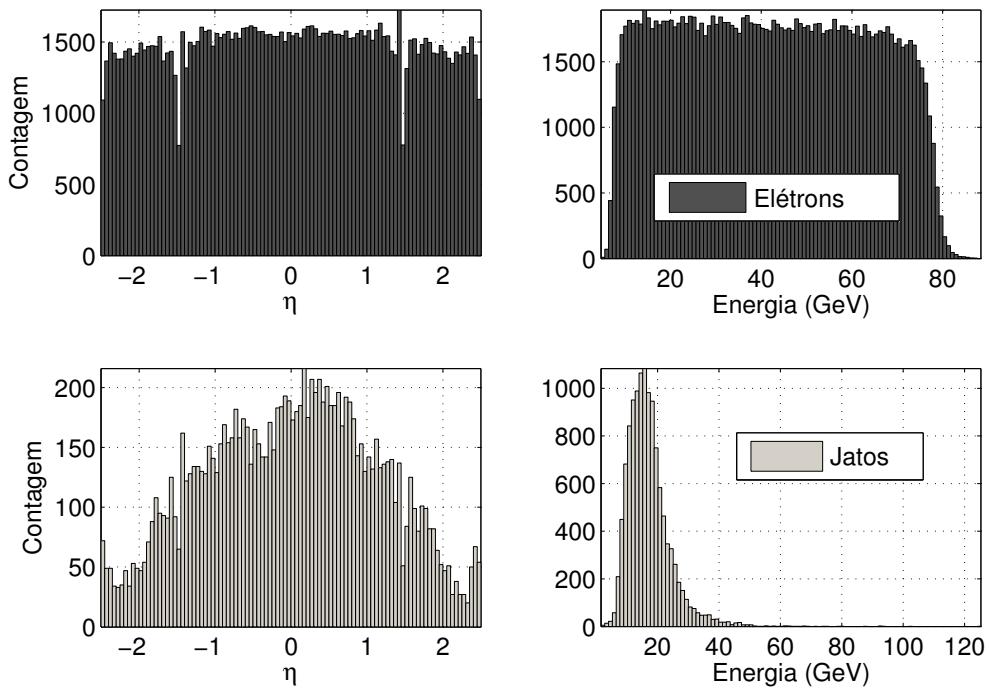
As curvas ROC obtidas para os classificadores treinados a partir dos componentes independentes locais são mostradas na Figura 7.13. Pode-se observar que o classificador do *cluster* 1 apresenta, para a mesma probabilidade de detecção, em geral, um maior falso alarme. Este fato indica que, a separação elétron/jato neste agrupamento é uma tarefa mais difícil que para o *cluster* 2.

Para a escolha da topologia (número de neurônios na camada oculta) dos classificadores locais, foram realizadas diversas inicializações do treinamento, variando-se o número de neurônios (de 3 a 23). Conforme mostrado na Figura 7.14, no primeiro *cluster* foram necessários 15 neurônios ocultos para obter o máximo desempenho, enquanto que no segundo agrupamento, o valor ótimo encontrado foram 10 neurônios.

Para finalizar o projeto do discriminador baseado na combinação de ICA Local e classificadores MLP, é preciso escolher os patamares de decisão dos dois classificadores. Inicialmente, foram utilizados os patamares de decisão escolhidos a partir do máximo SP de cada classificador local, ou seja, considerando apenas os eventos de cada agrupamento (que são respectivamente $Y_1 = 0,660$ para $SP_1 = 0,944$ e $Y_2 = -0,588$ para $SP_2 = 0,986$). Assim, foi obtida uma probabilidade de detecção igual a 98,0 % para falso alarme de 1,7 % ($SP=0,981$).

Realizando o ajuste dos patamares de corte através do algoritmo genético visando maximizar o SP global (calculado para todos os sinais do conjunto E10), foram escolhidos novos patamares ($Y_1 = 0,106$ e $Y_2 = -0,103$), assim, obteve-se uma

Cluster 1



Cluster 2

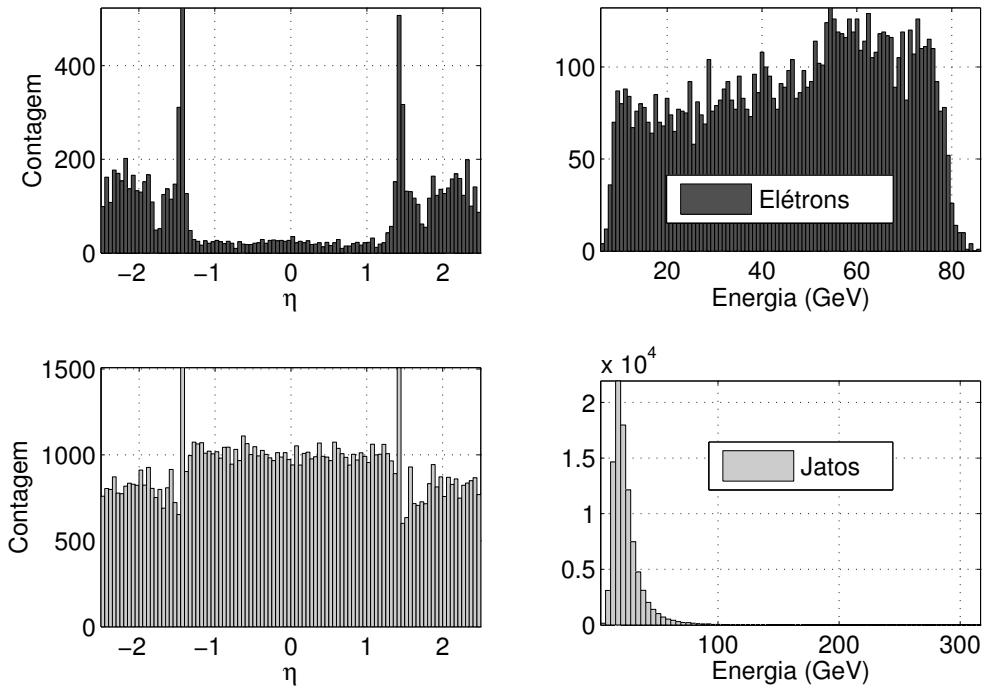


Figura 7.11: Distribuições em energia e η dos eventos nos *clusters* 1 (acima) e 2 (abaixo).

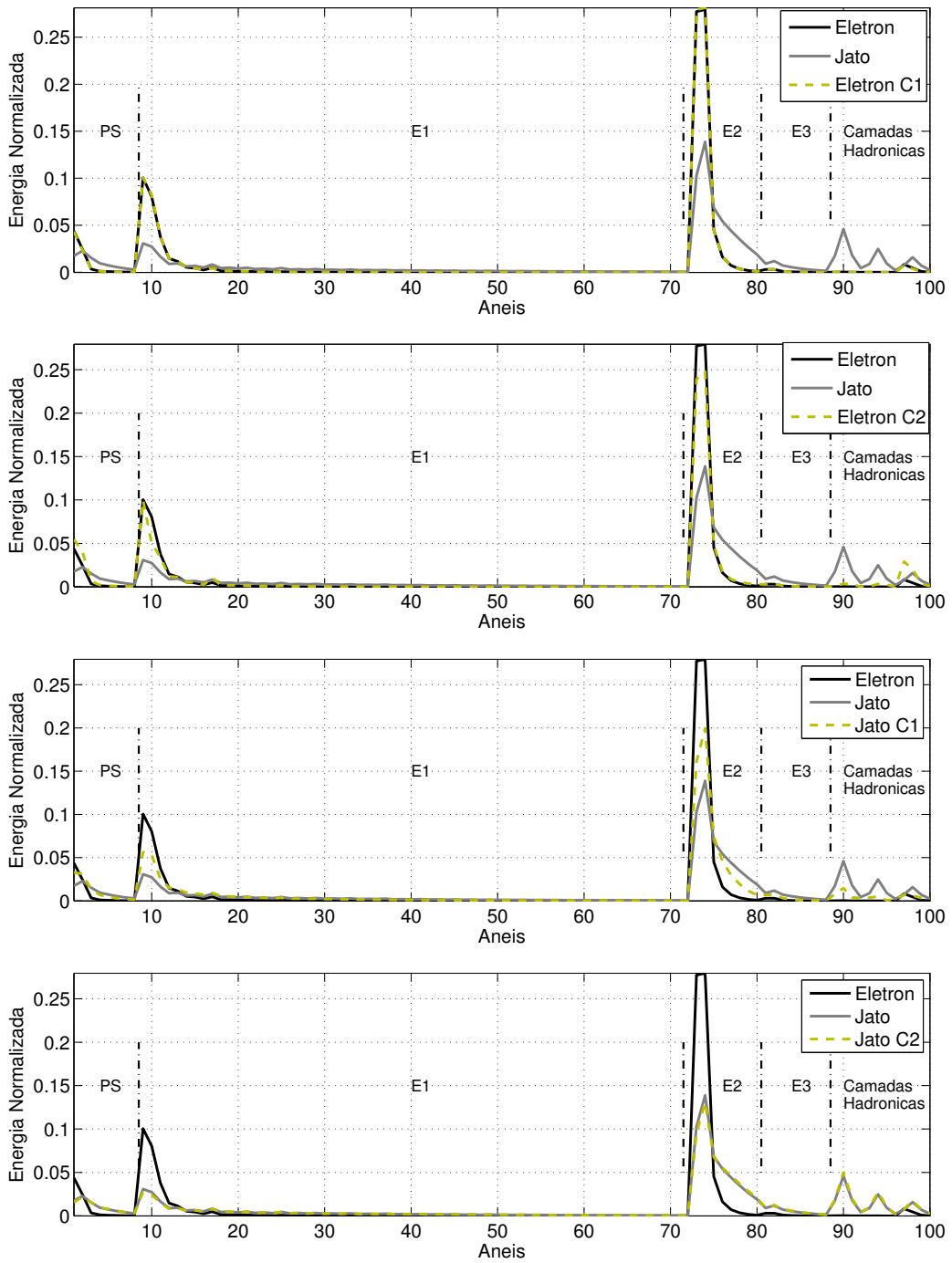


Figura 7.12: Eventos médios de elétrons e jatos calculados em cada *cluster*, comparados aos perfis médios para todo o conjunto E10.

probabilidade de detecção igual a 99,6 %, para falso alarme de 1,5 %, o que produz $SP=0,990$, otimizando o desempenho do discriminador.

A Figura 7.15 mostra um comparativo entre os discriminadores operando diretamente sobre os anéis e com pré-processamento por ICA e ICA Local. Pode-se perceber que o desempenho da ICA Local é superior (no caso da ICA Local não é

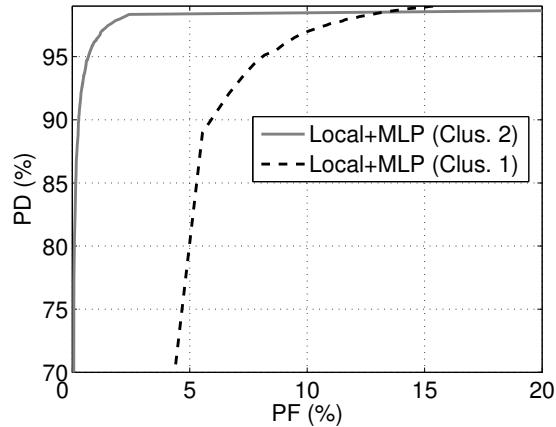


Figura 7.13: Curvas ROC obtidas para os classificadores neurais nos diferentes *clusters*.

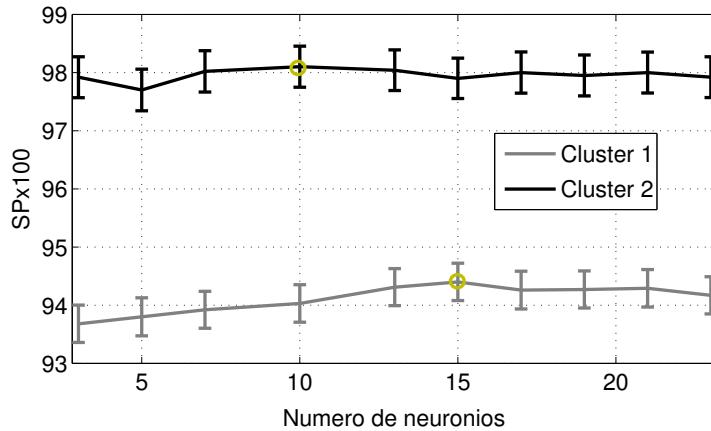


Figura 7.14: Variação do máximo SP com o número de neurônios ocultos.

possível obter uma curva ROC, os pontos mostrados correspondem aos melhores pa-
tamares locais obtidos pelo AG). A Tabela 7.2 apresenta um resumo dos resultados.

A Figura 7.16 mostra um comparativo da eficiência e falso alarme (em η , ϕ e energia) entre os discriminadores neurais operando sobre os anéis (*Neural Ringer*), os componentes independentes (ICA) e os componentes independentes locais. Pode-se observar que, o pré-processamento por ICA Local produz maior eficiência e menor falso alarme para todas as grandezas analisadas, considerando especificamente a pseudo-rapidez (η), há uma considerável melhora do desempenho próximo à região do crack ($|\eta| \sim 1,5$).

A partir da análise das taxas conjuntas de acerto e erro mostradas na Tabela 7.3, pode-se observar que, há uma grande concordância entre os discriminadores. Por

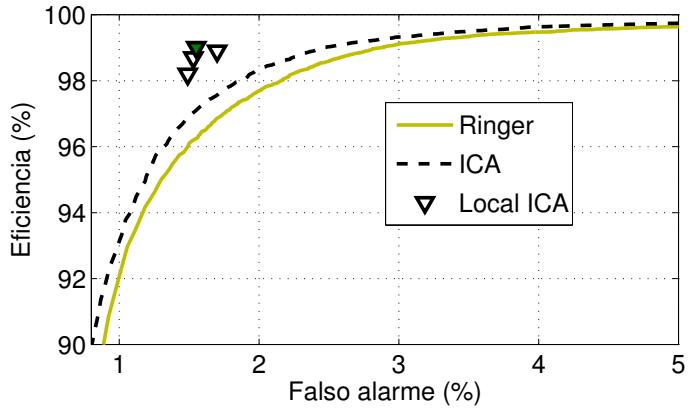


Figura 7.15: Diagrama do discriminador baseado em ICA Local.

exemplo, os discriminadores baseados no pré-processamento por ICA (linear e local) classificam corretamente 98,89 % dos elétrons e 97,41 % dos jatos. A superioridade do ICA local fica evidente quando analizamos as taxas de acerto×erro. Pode-se ver que, 1,03 % dos jatos são classificados corretamente através do processamento por ICA local e perdidos no caso da ICA linear, enquanto que o oposto (Local ICA erra e ICA acerta) acontece para apenas 0,19 %.

As características dos eventos classificados corretamente por um dos discriminadores e incorretamente por outro (considerando *Ringer*, ICA linear e ICA local) podem ser observadas em detalhes a partir dos histogramas da Figura 7.17. Nestes gráficos, de modo semelhante ao realizado na Figura 7.3, os valores percentuais foram calculados considerando, em cada faixa (*bin*) do histograma, a parcela dos eventos que pertence à classes de acerto/erro em questão.

Pode-se observar que, existem regiões onde há uma maior probabilidade de erro de identificação. Por exemplo, a maior parte dos erros para elétrons são concentrados

Tabela 7.2: Comparaçāo de desempenho entre o discriminador baseado na ICA Local e outros classificadores para o conjunto E10. São mostrados o máx. SP, os valores de PD e PF que produzem o máx. SP e PF para PD=92,5%.

Discriminador	Máx. SP×100	PD (%)	PF (%)	PF _{PD=92,5%}
T2Calo	91,99	92,50	8,55	8,55
Ringer	97,99 ± 0,18	98,77 ± 0,35	2,85 ± 0,11	1,10 ± 0,08
ICA	98,25 ± 0,27	99,01 ± 0,42	2,50 ± 0,09	1,00 ± 0,12
Local	99,05 ± 0,28	99,60 ± 0,43	1,50 ± 0,08	0,85 ± 0,09

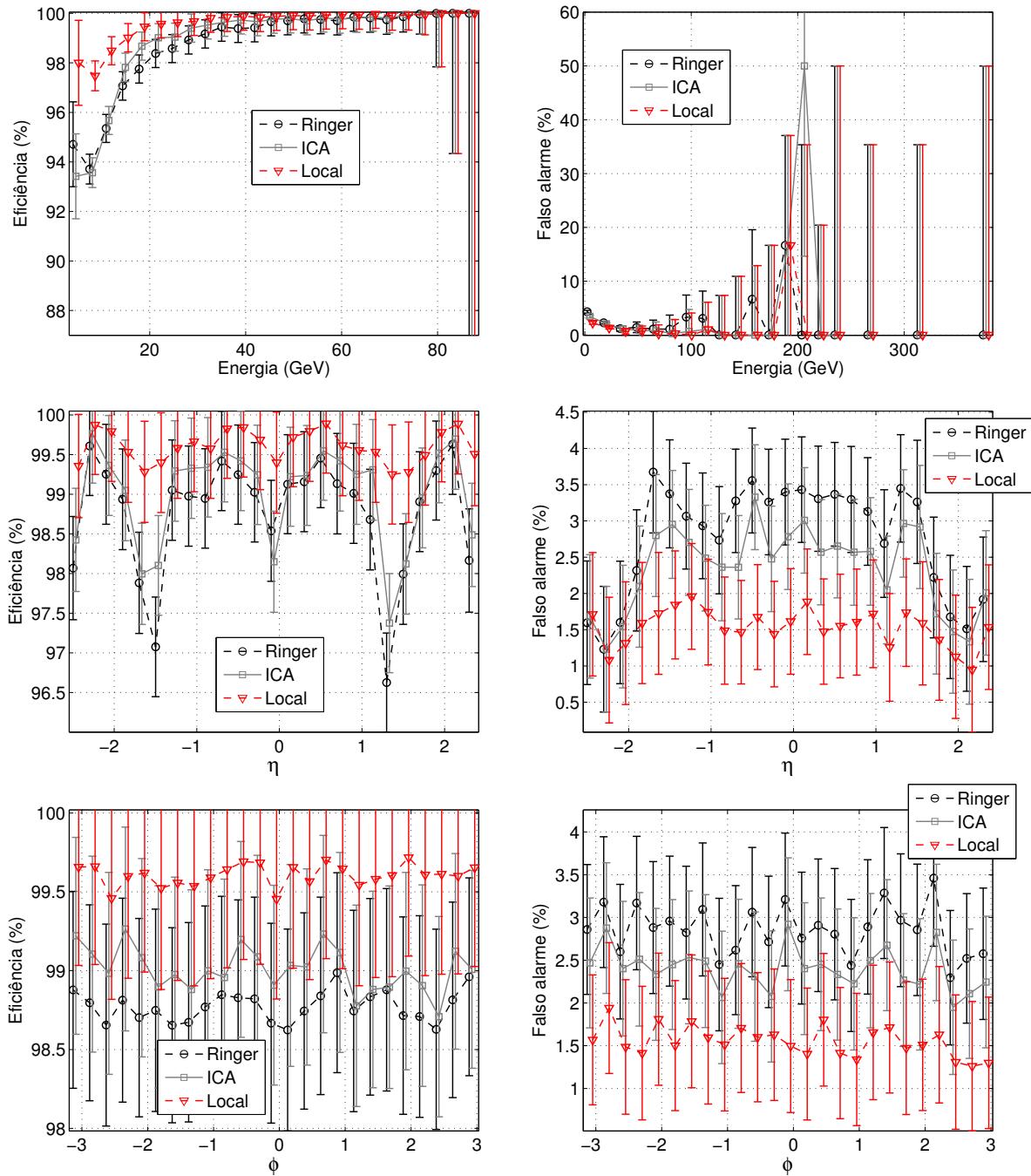


Figura 7.16: Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores *Neural Ringer* ICA+MLP e Local ICA+MLP.

Tabela 7.3: Taxas conjuntas (%) de acerto e erro para diferentes discriminadores.

		ICA acerta		ICA erra		Ringer acerta		Ringer erra	
		Elétron	Jato	Elétron	Jato	Elétron	Jato	Elétron	Jato
Local	Acerta	98,89	97,41	0,72	1,03	98,63	97,00	0,98	1,45
	Erra	0,13	0,19	0,27	1,37	0,15	0,15	0,24	1,40

em energias menores que 30 GeV (conforme comentado anteriormente no Capítulo 4 em baixas energias as características dos perfis de deposição de elétrons e jatos são mais semelhantes), e em três faixas em η ($\eta \sim 0$, $|\eta| \sim 1,5$ e $|\eta| \sim 2,5$). Nestas regiões em η existem interconexões entre os módulos do calorímetro eletromagnético. Em $\eta = 0$ são conectadas as duas metades do barril². Em $|\eta| \sim 1,5$ há uma descontinuidade para passagem de cabos de alimentação e dados (conhecida como *crack*) e em $|\eta| = 2,5$ há a junção dos dois blocos cilíndricos que compõem a tampa³.

Considerando as regiões descritas acima, pode-se observar que o discriminador baseado em ICA local consegue identificar corretamente uma maior quantidade de eventos que os demais classificadores. Por exemplo, observa-se que, em baixa energia (~ 5 GeV) com a ICA local é possível identificar corretamente $\sim 4\%$ dos elétrons perdidos pelo *Ringer* e o contrário só é verdade para $\sim 1\%$ destes eventos. Na comparação ICA local com ICA (linear) estas taxas são respectivamente $\sim 4\%$ (havendo um pico de $\sim 10\%$ para a primeira faixa de energia) e $\sim 0,7\%$. O uso da ICA local também produz melhores resultados nas três faixas em η onde existem as interconexões entre os módulos do calorímetro

Para os jatos observa-se que com a ICA local há um considerável aumento da eficiência em $|\eta| < 1,5$ (com acentuado destaque em torno do *crack*). Considerando a energia, o pré-processamento por ICA local produz uma maior taxa de identificação em quase toda a faixa de energia, com destaque para baixas energias (< 20 GeV) e também próximo a 100 GeV.

²O barril do calorímetro eletromagnético é dividido em duas partes (*half-barrels*), um deles cobre a faixa $0 < \eta < 1,475$ e o outro $-1,475 < \eta < 0$.

³As tampas do calorímetro eletromagnético são divididas em dois módulos coaxiais o *inner wheel* e o *outer wheel*, que cobrem respectivamente as regiões de $2,5 < |\eta| < 3,2$ e $1,375 < |\eta| < 2,5$.

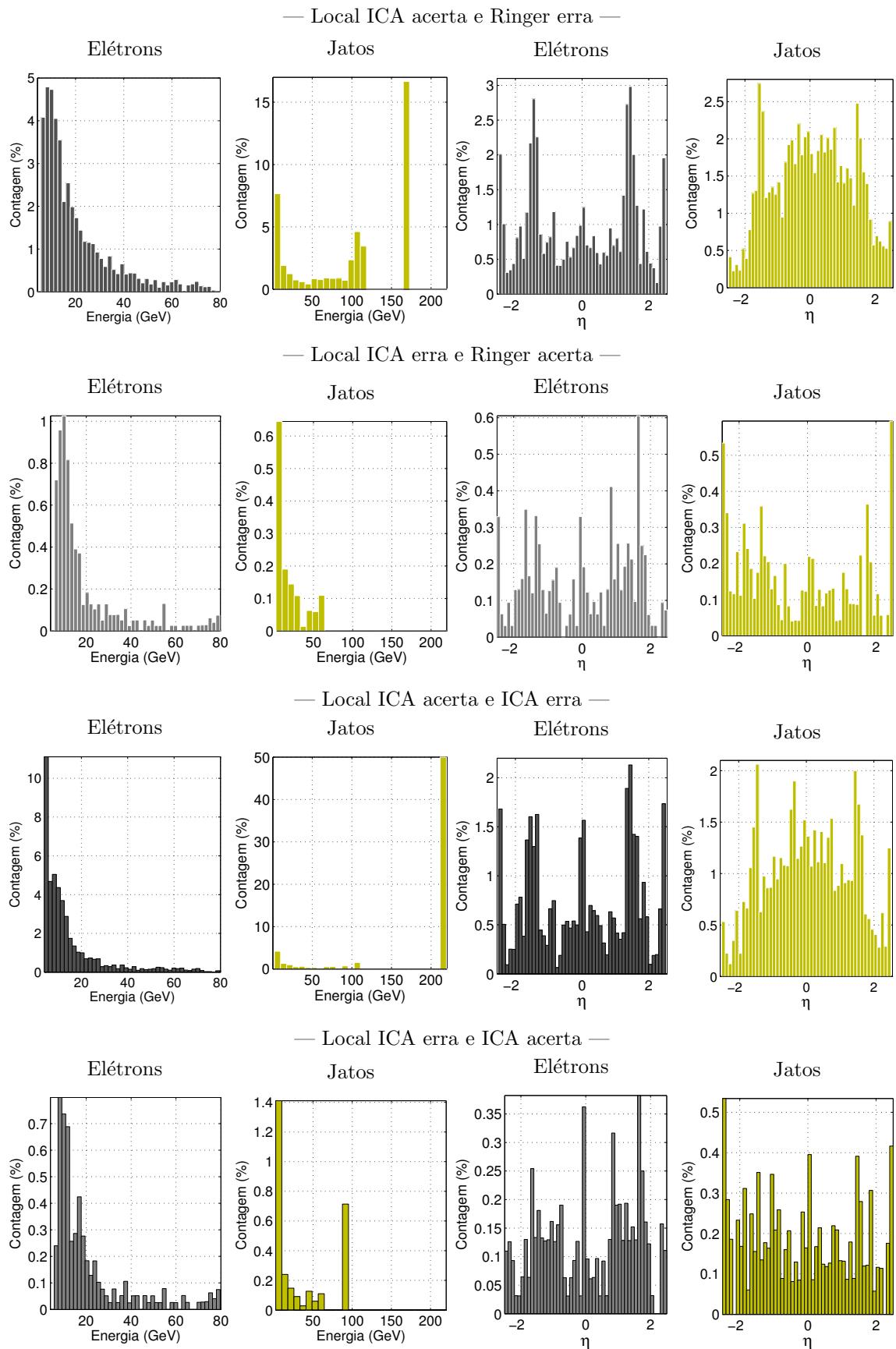


Figura 7.17: Comparativo de eficiência entre os discriminadores *Neural Ringer*, ICA+MLP e Local ICA+MLP.

Mapas Auto-Organizáveis - SOM

No pré-processamento por mapas auto-organizáveis foram utilizadas abordagens de treinamento que podem ser divididas conforme a segmentação da informação em segmentada e não-segmentada e conforme a utilização de informação supervisionada em não-supervisionada, ajustada por LVQ (quantização vetorial por aprendizado) e supervisionada.

Em todos os casos, é necessário escolher o número de neurônios a serem utilizados em cada mapa. Neste trabalho, conforme descrito anteriormente no Capítulo 6, para esta escolha foram utilizados o erro médio de reconstrução (ERM) e o índice SP calculados para cada mapa. A Figura 7.18 mostra os valores calculados para ERM e SP, variando-se as dimensões de mapas não-segmentados com treinamento não-supervisionado. De modo semelhante, os resultados obtidos para os mapas segmentados são apresentados na Figura 7.19 (desta vez para os mapas supervisionados). Resultados semelhantes foram obtidos para as outras abordagens de treinamento. Deste modo, foram escolhidos os mesmos tamanhos para todas as configurações.

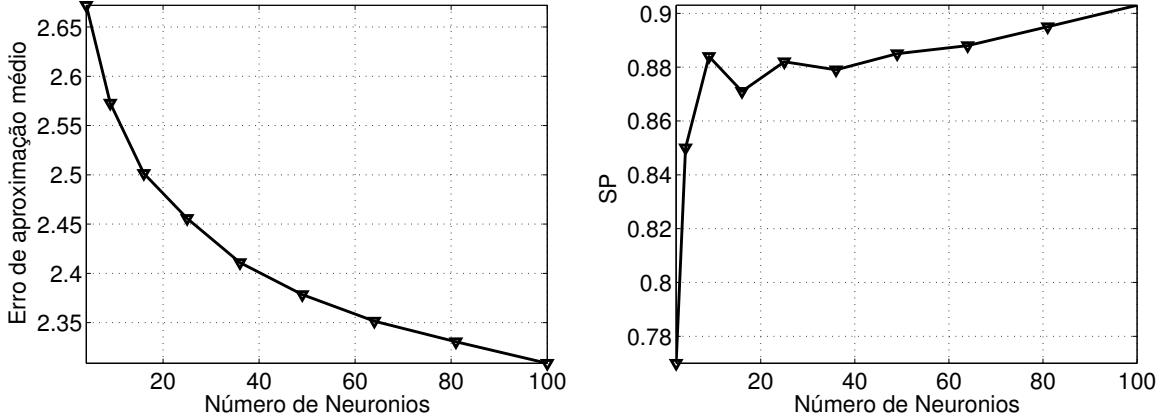


Figura 7.18: Erro de reconstrução médio e SP calculados variando-se o número de neurônios do mapa (para mapas não-segmentados treinados de modo não-supervisionado).

Na escolha da dimensão ótima dos mapas, o objetivo é manter um compromisso entre: alto SP, baixo ERM e pequeno número de neurônios. Então, buscou-se o mapa i que atende às restrições:

- $ERM_i - ERM_{\text{Min}} < 25\% \Delta_{ERM}$ (onde ERM_{Min} é o mínimo valor do ERM

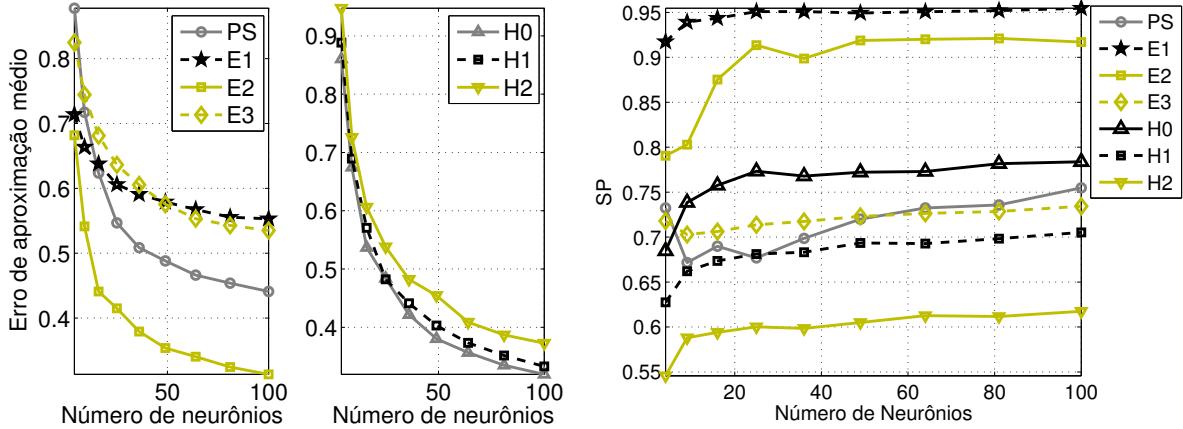


Figura 7.19: Erro de reconstrução médio e SP calculados variando-se o número de neurônios do mapa (para mapas segmentados treinados de modo supervisionado).

considerando todos os mapas treinados e Δ_{ERM} a variação do ERM);

- $SP_{\text{Max}} - SP_i < 20\% \Delta_{SP}$ (onde SP_{Max} é o máximo valor do SP e Δ_{SP} a variação do SP);
- Ter o menor número de neurônios entre os mapas que satisfazem às restrições acima.

Pois, desta forma, para o mapa i escolhido garante-se que, o ERM está entre os 25 % menores e o SP entre os 20 % maiores.

A Tabela 7.4 mostra os mapas escolhidos nas abordagens segmentada e não-segmentada. É importante notar que, os métodos que utilizam informação das classes no treinamento (SOM+LVQ e SOM Sup.) produzem maiores valores do SP, sendo que o SOM+LVQ supera o SOM supervisionado apenas nas camadas PS, H1 e H2. Na Figura 7.20 pode-se observar a probabilidade de ativação dos neurônios de um mapa treinado para a camada E2 antes e depois da LVQ, percebe-se que, o ajuste supervisionado é capaz de reduzir a região de confusão do mapa, onde ambas as classes apresentam considerável probabilidade de ativação.

Utilizando-se as projeções $\mathbf{u} = \mathbf{W}x$ (onde \mathbf{W} é a matriz de pesos do mapa) como entradas para discriminadores neurais, foram obtidas as curvas ROC mostradas na Figura 7.21. Pode-se observar que, o pré-processamento pelo mapa supervisionado não-segmentado produziu uma pequena melhora no desempenho se comparado ao *Neural Ringer*, porém, para a região de baixo falso alarme, o pré-

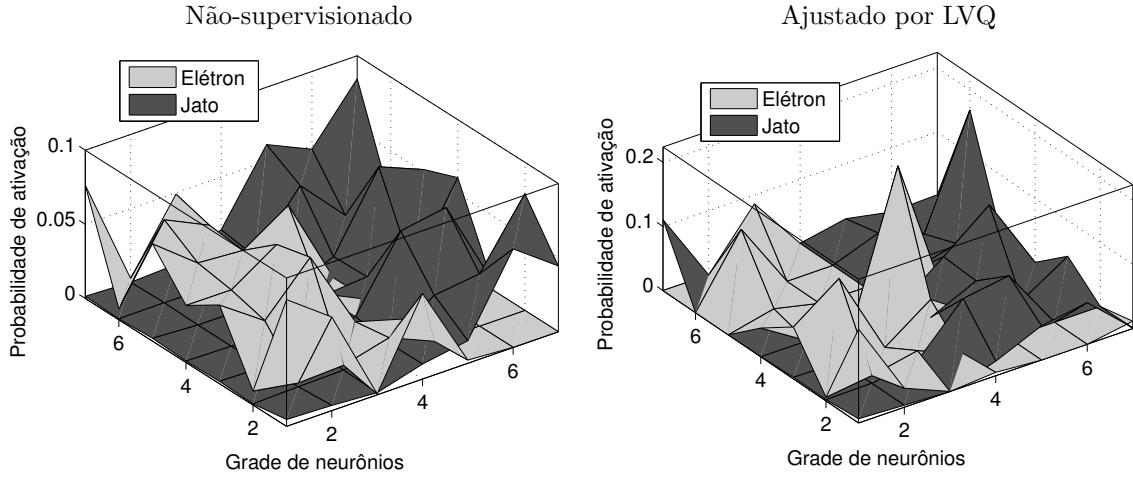


Figura 7.20: Probabilidade de ativação dos neurônios antes (esquerda) e depois da LVQ (direita) num mapa segmentado treinado para a camada E2.

processamento por ICA é um pouco mais eficiente. Nas demais abordagens do SOM, o pré-processamento não foi capaz de produzir maior eficiência do que o *Neural Ringer*. Um resumo dos resultados é apresentado na Tabela 7.5.

Considerando a abordagem de treinamento do SOM que apresentou maior eficiência de discriminação (SOM Supervisionado), um comparativo com os discriminadores baseados nos componentes independentes e nos anéis, em termos da eficiência e do falso alarme para diferentes valores de η , ϕ e energia, é apresentado na Figura 7.22. Pode-se observar que, em termos da discriminação de elétrons, o uso tanto da ICA como do SOM produzem maior eficiência se comparados ao discriminador neural operando sem pré-processamento (*Neural Ringer*). Analisando-se a aceitação do falso alarme, O discriminador baseado em ICA apresenta o melhor

Tabela 7.4: Máximo SP($\times 100$) para diferentes estratégias de treinamento dos mapas (aqui o SP é calculado considerando a operação do SOM diretamente como um classificador).

Camada	SOM Segmentado							SOM não-seg.
	PS	E1	E2	E3	H0	H1	H2	
Dimensões do mapa	5×5	7×7	6×6	6×6	6×6	6×3	4×4	12×6
Não-supervisionado	70,31	84,12	82,77	69,78	74,15	57,24	56,13	91,10
Ajustado por LVQ	72,76	91,02	88,30	72,98	76,32	68,45	60,23	92,67
Supervisionado	72,54	95,30	92,21	73,18	77,09	67,33	59,82	96,14

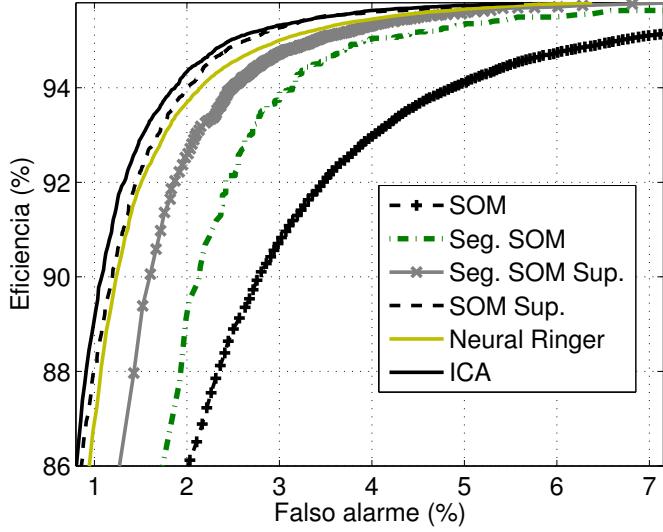


Figura 7.21: Curvas ROC obtidas para os diferentes modos de treinamento do SOM após a classificação por uma rede MLP.

resultado, enquanto que o uso do SOM não produz significativa redução do falso alarme em comparação ao *Neural Ringer*. Pode-se perceber também que o ganho obtido pelo SOM ocorre, de modo semelhante a ICA Local, em torno do *crack* para $|\eta| \sim 0$ e $|\eta| \sim 2,4$ e na faixa de energia menor que 40 GeV.

SOM como ferramenta para visualização

Os mapas auto-organizáveis são muito úteis para visualização de sinais multi-dimensionais, sendo que, aplicações em diferentes domínios podem ser encontradas em [208, 209]. No caso dos sinais do calorímetro do ATLAS, neste trabalho está sendo proposta a utilização de uma representação em duas dimensões onde o neurônio vencedor é destacado através de um contorno mais intenso e, para cada neurônio da grade, são associados os parâmetros a seguir:

- probabilidade de ativação por cada classe (para o conjunto de treino);
- classe dominante (que tem maior probabilidade de ativação no conjunto de treino);
- semelhança com o sinal de entrada (indicada através do preenchimento dos neurônios, de modo que o neurônio vencedor é completamente preenchido e os demais são preenchidos de modo inversamente proporcional ao erro de reconstrução do neurônio para o sinal em questão).

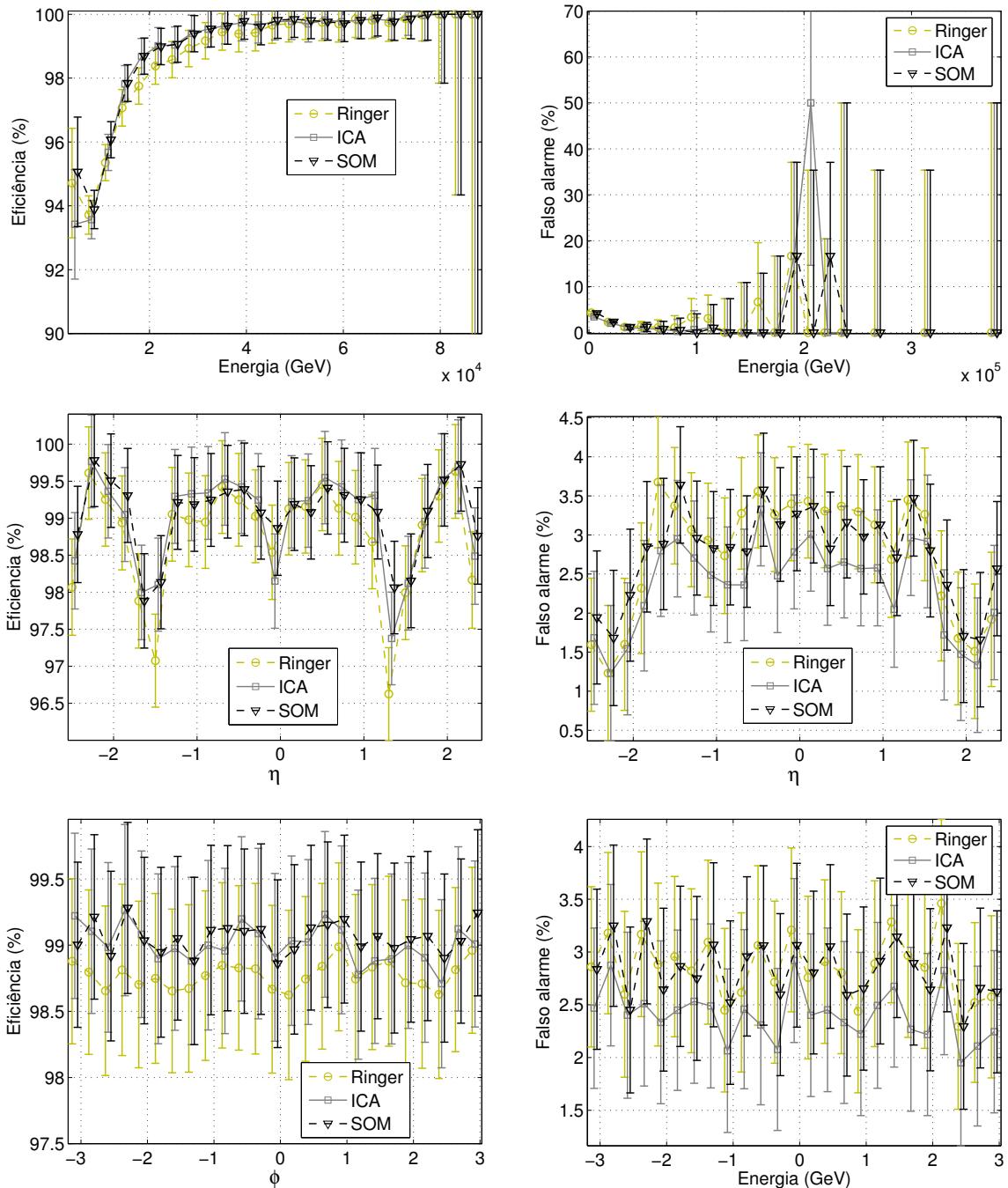


Figura 7.22: Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores *Neural Ringer* ICA+MLP e SOM+MLP.

Tabela 7.5: Comparaçāo de desempenho para diferentes abordagens de treinamento dos mapas auto-organizáveis para o conjunto E10. São mostrados o máximo SP, os valores de PD e PF que produzem o máximo SP e PF para PD=92,5%.

Discriminador	Máx. SP × 100	PD (%)	PF (%)	PF _{PD=92,5%}
T2Calo	91,99	92,50	8,55	8,55
Ringer	97,99 ± 0,18	98,77 ± 0,35	2,85 ± 0,11	1,10 ± 0,08
ICA	98,25 ± 0,27	99,01 ± 0,42	2,50 ± 0,09	1,00 ± 0,12
SOM	96,01 ± 0,23	94,8 ± 0,33	2,75 ± 0,10	2,40 ± 0,11
SOM Seg.	97,47 ± 0,31	98,00 ± 0,47	3,05 ± 0,19	1,96 ± 0,09
SOM+LVQ	98,01 ± 0,26	98,90 ± 0,36	2,87 ± 0,17	1,09 ± 0,15
SOM Seg.+LVQ	97,81 ± 0,19	98,52 ± 0,40	2,90 ± 0,11	1,94 ± 0,10
SOM Sup.	98,15 ± 0,22	99,10 ± 0,36	2,83 ± 0,17	1,02 ± 0,13
SOM Sup. Seg.	97,75 ± 0,18	98,01 ± 0,45	2,50 ± 0,13	1,46 ± 0,10

Na Figura 7.23 pode-se visualizar os eventos de um elétron e um jato típicos através da representação 2-D do mapa de características (para um mapa supervisionado não-segmentado de dimensões 4×2). Observa-se que os padrões para as duas classes são bem distintos, com concentração em lados opostos do mapa.

A intensidade das saídas de cada neurônio também pode ser visualizada a partir de um gráfico 3D, conforme ilustrado na Figura 7.24 (para um mapa segmentado de tamanho 7×7 treinado para a camada E1). A concentração em lados opostos do mapa também facilita a identificação das classes de cada evento.

Modelo com Restrições Estruturais - PNL

Conforme descrito anteriormente, para a estimação dos componentes independentes através do modelo PNL foi utilizada a abordagem segmentada associada a um bloco de pré-processamento para compactação da informação.

Considerando a compactação por PCA, num problema de classificação é importante notar que não se garante que, as componentes mais energéticas (que normalmente são retidas após a PCA) carreguem toda a informação discriminante para o problema. Neste trabalho, para escolha do número de componentes principais, foi calculado o valor do máximo SP obtido treinando-se classificadores neurais com os componentes estimados, conforme mostrado na Tabela 7.6.

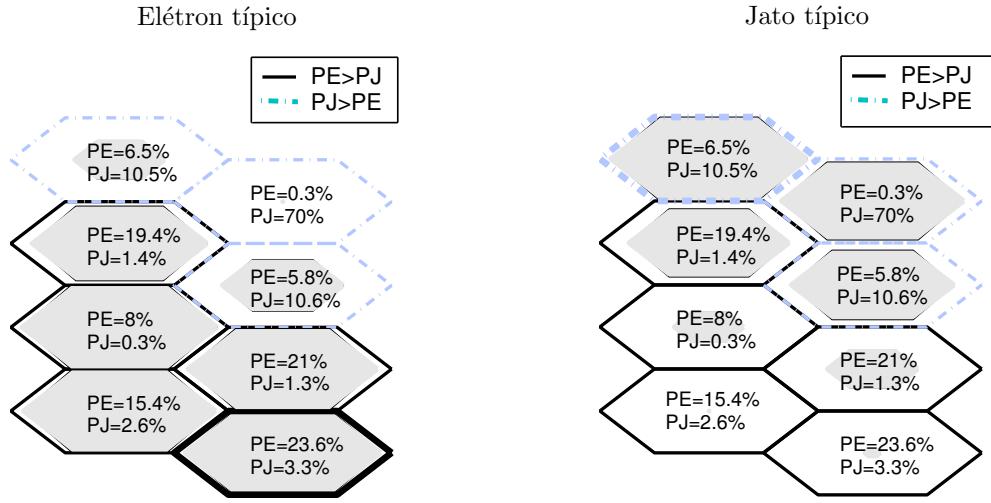


Figura 7.23: Visualização 2-D de eventos de elétron (esquerda) e jato (direita) típicos através do SOM.

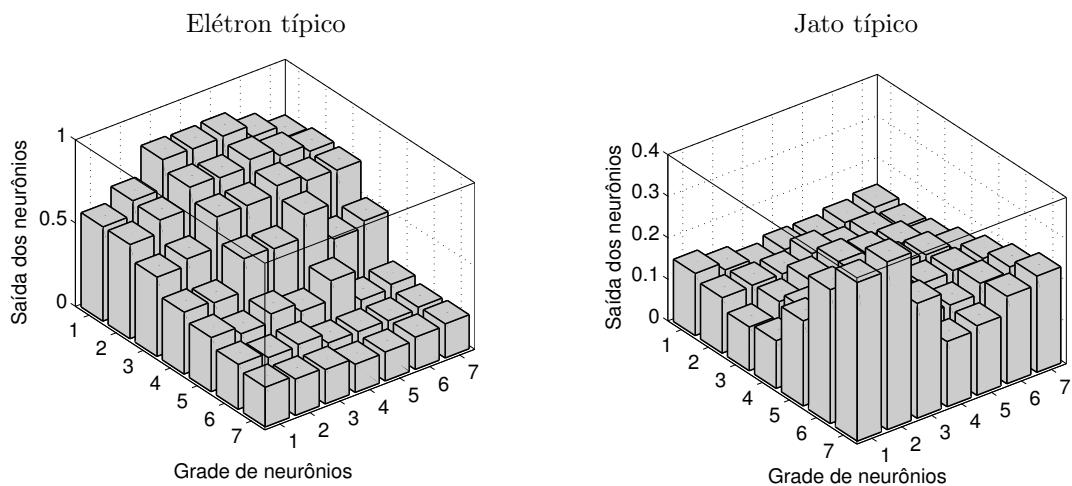


Figura 7.24: Visualização 3-D de eventos de elétron (esquerda) e jato (direita) típicos através do SOM.

Tabela 7.6: Número de componentes preservados (N) e máximo SP($\times 100$) para diferentes níveis de retenção de energia por PCA.

	Retenção de energia (%)										
	100	95	90	85	80	70	60	50	40	30	20
N	100	83	72	61	53	41	31	23	12	9	7
SP	97.7	97.6	97.4	97.5	97.6	97.4	97.2	97.4	96.7	96.2	94.6

Pode-se observar que o nível de retenção de energia pode ser reduzido a $\sim 50\%$ e, ainda assim, o máximo SP continua aproximadamente igual ao obtido com a

utilização de todos os componentes. Se a energia continuar sendo reduzida, o desempenho de discriminação decresce de modo mais acentuado.

Conforme descrito na Seção 5.5.1, a análise de componentes principais (PCD) fornece uma alternativa à PCA para compactação em problemas de classificação. Na PCD, os sinais de entrada são projetados em direções que produzem máxima discriminação entre as classes. As direções discriminantes são encontradas a partir de uma rede neural MLP treinada de modo supervisionado. O número de PCDs adequado a cada problema é encontrado a partir de um procedimento (descrito mais detalhadamente na Seção 5.5.1) no qual há um gradual acréscimo no número de componentes até o ponto em que a adição de um novo componente não contribui para aumentar a eficiência de discriminação.

No algoritmo PNL modificado, o bloco de compactação é estimado de modo simultâneo ao modelo PNL, através do procedimento descrito no Capítulo 5. Para estimar o ótimo número de componentes após a compactação é adotado um procedimento semelhante ao realizado para a PCD (gradual acréscimo no número de componentes até não haver ganho significativo na eficiência de discriminação com a inclusão de mais componentes).

Na Tabela 7.7 são mostrados os números de componentes estimados para cada camada do calorímetro após a compactação por PCA e PCD e para o algoritmo PNL modificado. Pode-se observar que, a PCA é capaz de produzir uma maior redução no número de componentes se comparado às demais.

Tabela 7.7: Número de componentes estimados a partir da compactação por PCA (50% de retenção de energia), PCD e pelo modelo PNL modificado.

Layer	PS	E1	E2	E3	H0	H1	H0	Total
Rings	8	64	8	8	4	4	4	100
PCA (50%)	3	10	1	3	2	2	2	23
PCD	6	4	5	4	3	3	3	28
PNL Mod.	5	8	6	4	3	3	2	31

Os algoritmos utilizados para estimar o modelo de mistura PNL (tanto no modelo tradicional como no modificado) utilizam redes neurais MLP de duas camadas para estimar as funções não-lineares. Um parâmetro que precisa ser escolhido adequadamente no modelo é o número de neurônios na camada oculta destas redes

neurais, pois assim é possível controlar o grau de não-linearidade no modelo PNL estimado. Considerando o exposto, foram realizadas diversas inicializações do discriminador baseado no modelo PNL⁴, variando-se o número de neurônios ocultos das redes aproximadoras. Na Tabela 7.8, pode-se observar que, com o aumento do número de neurônios (e consequentemente do grau de não-linearidade permitido ao modelo), o desempenho de discriminação piora.

Tabela 7.8: Comparaçāo entre discriminadores baseados no pré-processamento através do modelo PNL.

Discriminador	Ringer	PNL (2 neur.)	PNL (3 neur.)	PNL (4 neur.)	PNL (5 neur.)
Máx. SP×100	97,99	98,30	97,51	95,83	94,27
PF p/ PD = 97%	1,10±0,09	0,94±0,105	1,21±0,13	3,39±0,14	4,13±0,18

Esse fato pode ser justificado considerando-se que, pelas características construtivas dos calorímetros, não espera-se que o grau de não-linearidade nos sensores seja elevado. Assim, utilizando um número elevado de neurônios ocultos, o modelo fica super-dimensionado e não é capaz de estimar adequadamente os componentes independentes (produzindo baixa eficiência).

A eficiência de discriminação obtida com o treinamento de classificadores neurais alimentados pelos componentes independentes estimados pelo modelo PNL é mostrada na Figura 7.25. Pode-se observar que, as abordagens que consideram a informação supervisionada durante a extração de características (PNL+PCD e PNL modificado) apresentam maior eficiência. Um resumo dos resultados obtidos é mostrado na Tabela 7.9.

Tabela 7.9: Comparaçāo de desempenho para diferentes discriminadores, corte E10.

Discriminador	Máx. SP×100	PD (%)	PF (%)	PF _{PD=92,5%}
T2Calo	91,99	92,50	8,55	8,55
Ringer	97,99 ± 0,18	98,77 ± 0,35	2,85 ± 0,11	1,10 ± 0,08
ICA	98,25 ± 0,27	99,01 ± 0,42	2,50 ± 0,09	1,00 ± 0,12
PNL+PCA	97,67 ± 0,23	98,61 ± 0,30	3,30 ± 0,11	1,15 ± 0,16
PNL+PCD	98,30 ± 0,29	99,00 ± 0,37	2,40 ± 0,10	0,94 ± 0,12
PNL Mod.	98,55 ± 0,21	99,34 ± 0,41	2,29 ± 0,15	0,70 ± 0,08

⁴Os resultados mostrados se referem à configuração com compactação por PCD, porém, para os outros modelos o comportamento se mostrou bastante semelhante.

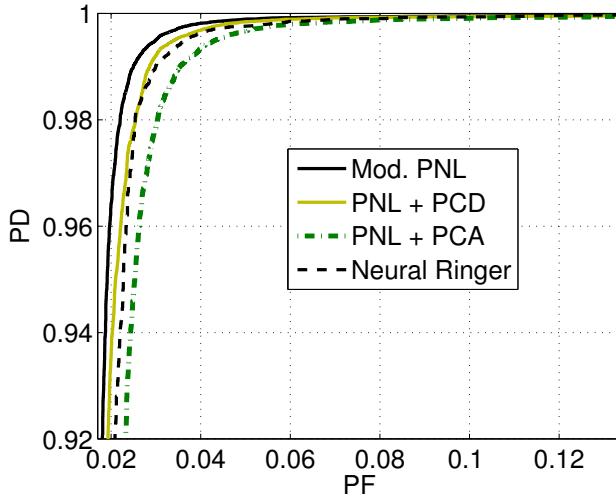


Figura 7.25: Curvas ROC para os discriminadores baseados no modelo PNL.

Considerando as variáveis físicas como a posição de interação (em termos das coordenadas η e ϕ) e a energia dos eventos, observou-se que, o ganho obtido com o discriminador baseado no modelo PNL (considerando a abordagem de maior eficiência de discriminação, ou seja o modelo PNL Modificado) aconteceu, de modo semelhante aos outros modelos da NLICA utilizados, principalmente em regiões onde a caracterização dos eventos é mais difícil (nas interconexões do calorímetro e em baixa energia).

7.1.3 Estudo da Relevância por Camada

Inicialmente, para os discriminadores propostos, foi avaliado o desempenho de discriminação quando os anéis de uma dada camada são substituídos pelo seu valor médio no conjunto de treino. Conforme mostrado na Figura 7.26, as informações das diferentes seções do calorímetro influenciam de modo distinto na eficiência de discriminação. Deste modo, a camada E1 (primeira camada eletromagnética) é a que apresenta maior relevância (pois sua substituição pelo valor médio ocasiona a maior queda de eficiência para todos os discriminadores), seguida pela camada E2. Considerando as camadas hadrônicas H1 e H2 e a terceira camada eletromagnética (E3), observa-se que, a queda de eficiência é bem pequena com sua substituição pela média (indicando pouca relevância de discriminação).

É interessante notar que, o discriminador baseado no pré-processamento por ICA

Local é o que mais depende da camada E1, enquanto que, utilizando-se a ICA, é observada a menor queda de desempenho para a mesma camada.

Continuando a análise da relevância por camada, a rede neural classificadora foi re-treinada retirando-se a informação de algumas camadas do calorímetro, neste caso, foram utilizados diretamente os sinais em anéis (sem pré-processamento), porém, resultado semelhante foi encontrado para os casos com pré-processamento por ICA/NLICA. Conforme mostrado na Figura 7.27, pode-se observar que, a retirada das camadas E3, H1 e H2 individualmente quase não provoca queda no SP, e mais ainda, retirando-se ao mesmo tempo as camadas H1 e H2 há uma ligeiro aumento na eficiência de discriminação (indicando que, na verdade, estas camadas introduzem informação capaz de confundir o classificador, o que pode ser justificado considerando que estas são as últimas camadas do calorímetro e, como, pouca energia é depositada nelas, seu sinal é, em muitas vezes, dominado por ruído de medição). É possível até retirar ao mesmo tempo três camadas (PS, H1 e H2) e, ainda assim, o desempenho é pouco afetado.

Em uma outra abordagem, com o objetivo de melhor explorar a segmentação da informação presente no problema, foram treinados classificadores neurais utilizando como entradas apenas os sinais de cada uma das camadas (classificadores segmentados). A Figura 7.28 mostra o desempenho obtido pelos classificadores segmentados em termos das curvas ROC⁵ e do máximo SP. Novamente pode-se observar que a primeira camada eletromagnética é a mais relevante, seguida da camada E2. As camadas menos relevantes foram novamente E3, H1 e H2.

Na Figura 7.29 é feita uma comparação entre as saídas do classificador segmentado da camada E1 (que produziu maior eficiência de discriminação) e as saídas dos discriminadores treinados para as demais camadas. Pode-se observar que, para a identificação de elétrons, apenas a camada E2 é capaz de complementar a decisão da E1 de modo significativo, há uma pequena concentração de eventos próximos a (-1;1) no gráfico E1×E2, que indicam eventos identificados corretamente pelo discriminador da camada E2 e incorretamente pela E1.

Considerando a identificação de jatos, pode-se observar uma maior influência

⁵Estes resultados se referem aos discriminadores operando diretamente sobre os anéis, o comportamento com a adição de pré-processamento por ICA/NLICA é bastante semelhante.

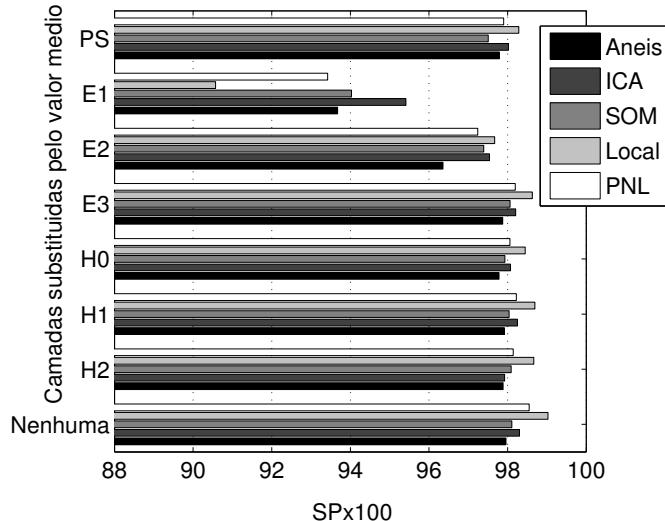


Figura 7.26: Desempenho de classificação obtido quando as camadas são substituídas pelo seu valor médio.

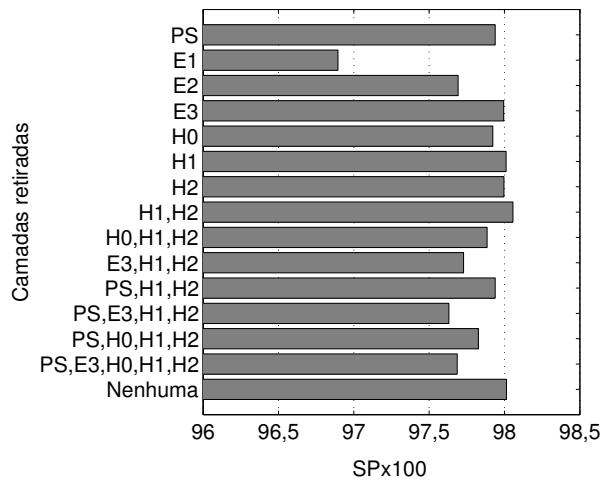


Figura 7.27: Desempenho de classificação obtido re-treinando o classificador neural retirando-se parte da informação.

de outras camadas como PS, E2, E3 e H0 a partir de uma maior concentração de eventos próximos a (1;-1) (erro do discriminador da camada E1 e acerto dos demais).

A partir das saídas dos classificadores especialistas (segmentados), foram testados diversos métodos de combinação de classificadores como média (aritmética e geométrica) das saídas, votação e também o uso de uma rede neural combinadora (conforme descrito na Seção 6.3.2) para a composição da informação segmentada na composição da decisão elétron/jato. Os modos de combinação ainda podem uti-

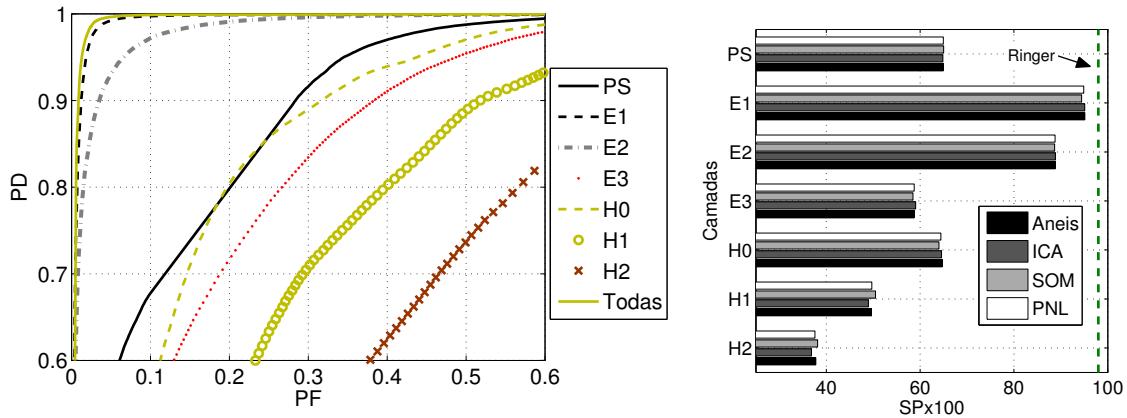


Figura 7.28: Desempenho de classificação obtido por classificadores treinados com a informação de apenas uma das camadas em termos das curvas ROC (esquerda) e do máximo SP (direita), onde a linha tracejada vertical indica o desempenho obtido pelo discriminador neural operando sobre todas as camadas.

lizar, como fator de ponderação, os valores máximos do SP calculados para cada classificador segmentado. Um comparativo da eficiência de cada um dos métodos de combinação é mostrado na Figura 7.30 em termos das curvas ROC. No caso da combinação por votação, não é possível traçar a curva ROC, pois as saídas dos classificadores são consideradas como variáveis discretas (indicando a classe associada a cada assinatura). Pode-se observar que, realizando-se a combinação dos classificadores segmentados através da média ponderada e da rede combinadora é possível obter melhor eficiência que o *Neural Ringer*.

Para determinação do número de neurônios ocultos na rede combinadora, foram realizadas 10 inicializações, variando-se o número de neurônios ocultos. A cada inicialização foram utilizados conjuntos de treino/teste/validation diferentes. Conforme mostrado na Figura 7.31, pode-se observar que o máximo valor do SP foi obtido para cinco neurônios ocultos⁶.

Sabendo que a combinação dos classificadores segmentados utilizando a rede combinadora é capaz de produzir um aumento na eficiência de discriminação, e considerando a existência de informações confundidoras em algumas camadas (conforme

⁶O número relativamente pequeno de neurônios ocultos da rede combinadora (que para os outros discriminadores estava na faixa dos 10 neurônios) indica que o problema de classificação para a rede combinadora é mais simples. Isso pode ser justificado pois os classificadores segmentados já resolveram parte do problema.

mostrado anteriormente nesta Seção), a rede combinadora foi re-treinada, retirando-se a informação de alguns dos classificadores segmentados. Conforme mostrado na Figura 7.31, utilizando a rede combinadora é possível obter desempenho ligeiramente melhor que o *Neural Ringer* utilizando-se apenas a primeira e segunda camadas eletrômagéticas. Essa redução no uso da informação, conforme será mostrado na

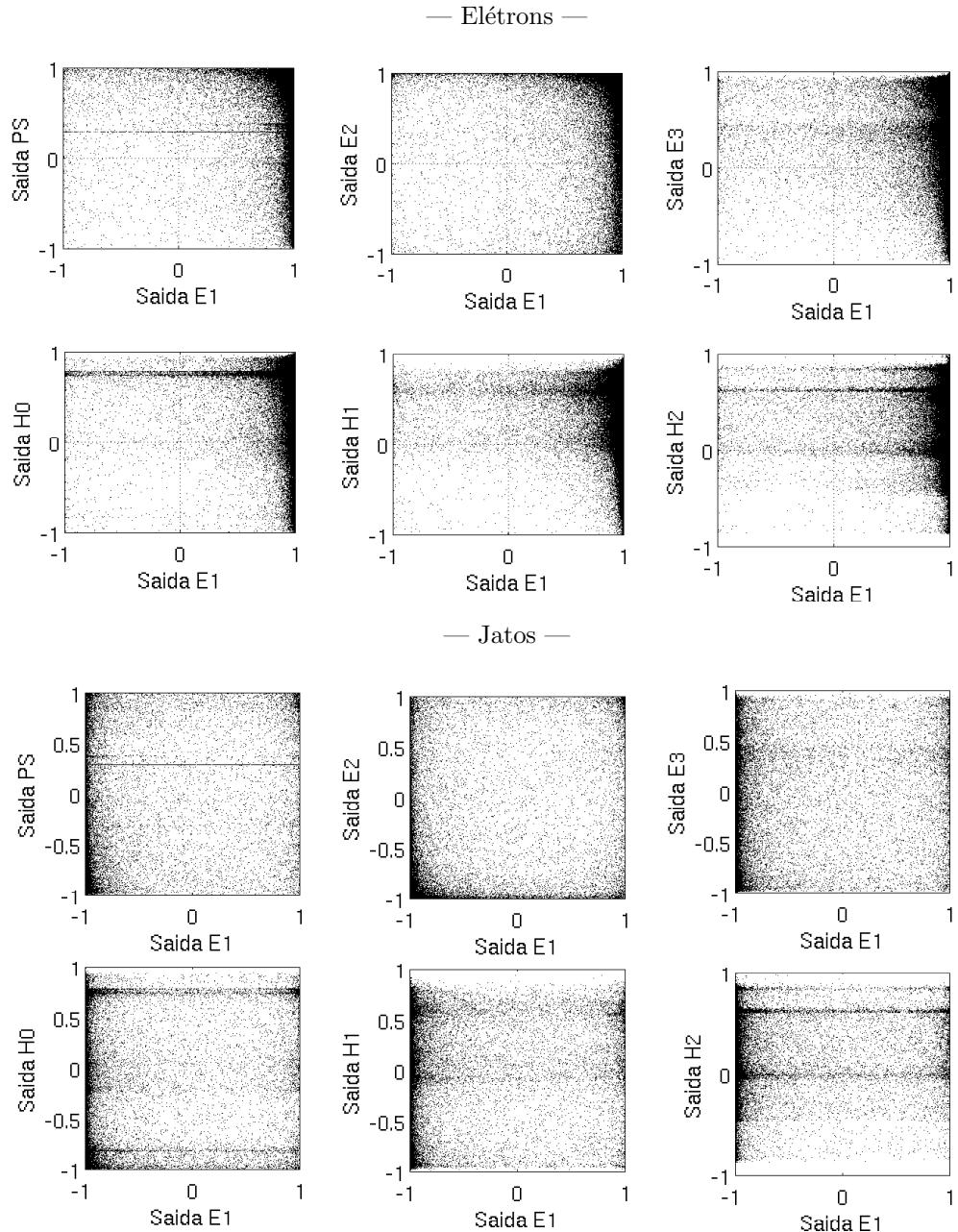


Figura 7.29: Gráficos de correlação entre a saída do classificador treinado para a camada E1 e para as demais camadas, para assinaturas de elétrons (acima) e jatos (abaixo).

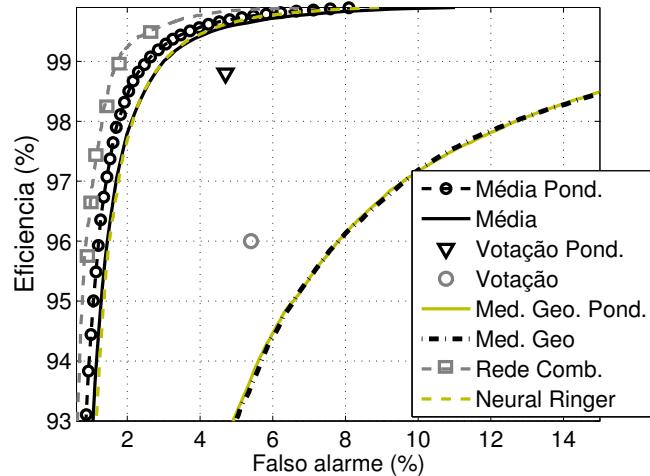


Figura 7.30: Curvas ROC obtidas para os diversos métodos de combinação de classificadores.

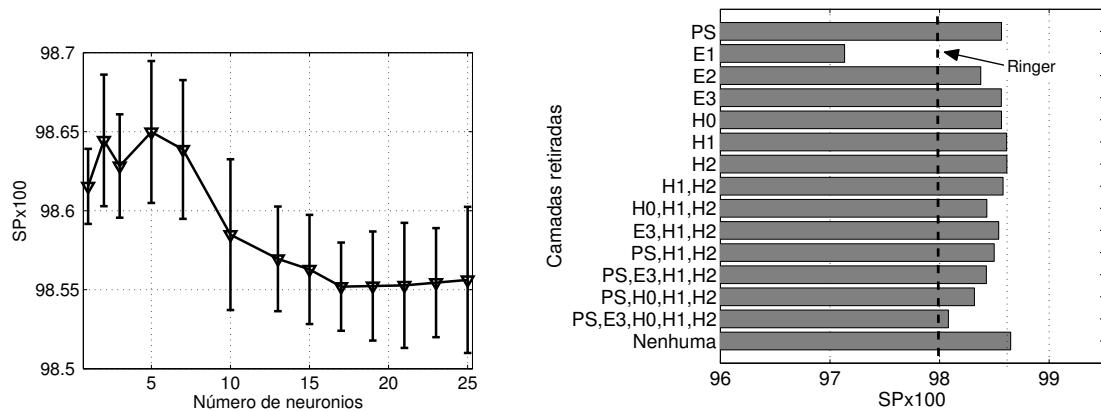


Figura 7.31: Variação do máximo SP com o número de neurônios ocultos para a rede combinadora (esquerda) e máximo SP obtido através da rede combinadora reirando-se, de modo segmentado, parte da informação.

Seção 8.3.1, corresponde a uma economia de aproximadamente 20 % do tempo total de processamento.

Um comparativo da eficiência para diferentes valores de η e energia é apresentado na Figura 7.32. Pode-se observar que, o uso da rede combinadora produz maior eficiência e menor falso alarme para quase todas as regiões em η e principalmente em baixas energias.

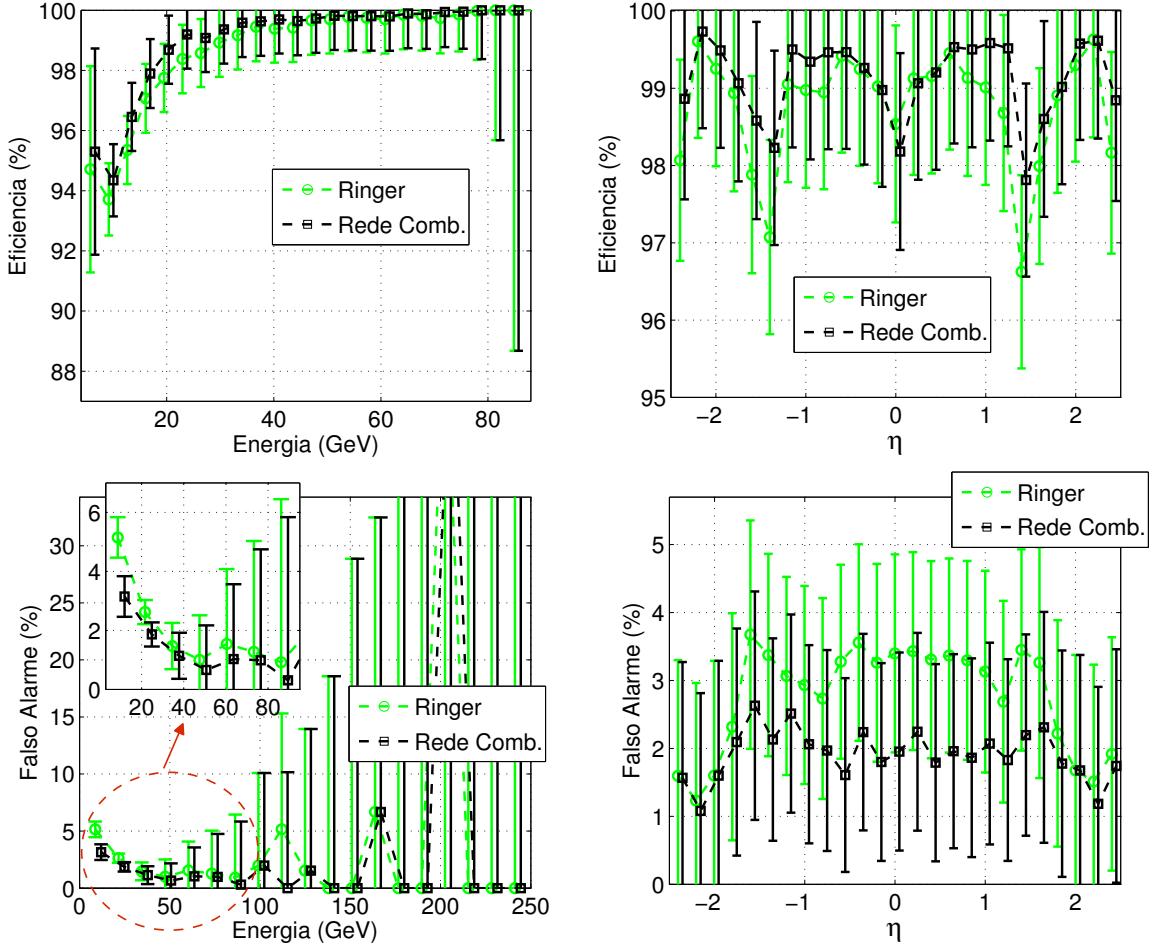


Figura 7.32: Eficiência (acima) e falso alarme (abaixo) em energia e η .

7.1.4 Comparação com Discriminadores Lineares

Conforme mostrado nos resultados obtidos até aqui, as técnicas de pré-processamento utilizadas tornaram a informação relevante mais acessível, facilitando a discriminação elétron/jato. Sob esta perspectiva, é importante avaliar se com a utilização dos sinais pré-processados (mapeados nos componentes independentes) o problema se torna linearmente separável. Deste modo, discriminadores de Fisher [136] foram alimentados com os componentes independentes estimados.

A Figura 7.33 mostra as curvas ROC obtidas para as diversas abordagens de pré-processamento através de ICA/NLICA (para o ICA Local, é mostrado apenas o ponto ótimo encontrado pelo algoritmo genético, para o SOM foi considerado o SOM supervisionado não-segmentado e para o PNL o pré-processamento por PCD). Pode-se observar que, para um mesmo pré-processamento, os discriminadores não-lineares (neurais) apresentam sempre melhor desempenho do que os lineares (Fisher).

É importante notar que, em alguns casos, como no pré-processamento por ICA, o uso do discriminador linear produz, para uma eficiência de 97 % um aumento de aproximadamente um ponto percentual no falso alarme, de modo que o uso do discriminante de Fisher poderia ser justificado para redução do custo computacional na operação *online*. O bom desempenho do classificador linear baseado na ICA pode ser justificado por haver maior separação entre as classes nos componentes independentes em comparação com os sinais em anéis, conforme indica a Figura 7.34.

Considerando o pré-processamento por NLICA, o uso dos discriminadores lineares provoca uma queda mais severa no desempenho. Uma explicação para isto pode ser encontrada analisando-se, por exemplo, um par de componentes independentes obtidos pelos modelos ICA (linear) e PNL para a camada E2, conforme mostrado na Figura 7.35. Pode-se observar que, que o mapeamento não-linear (obtido pela NLICA) produz classes separadas por uma fronteira não-linear, que pode ser identificada pelo discriminador neural, mas não por um classificador linear.

7.1.5 Comentários e Discussão

Um resumo dos resultados obtidos com o conjunto de dados simulados de assinaturas do tipo E10 é apresentado na Tabela 7.10. Pode-se observar que, entre os algoritmos

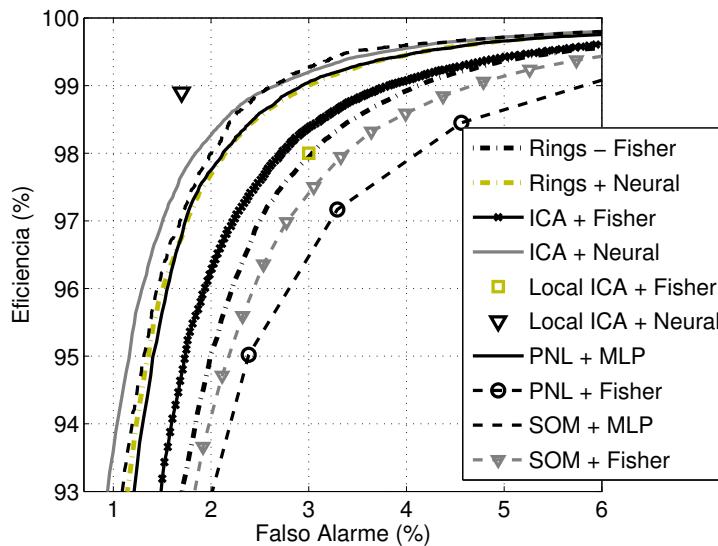


Figura 7.33: Curvas ROC para discriminadores lineares (*Fisher*) e não-lineares (MLP).

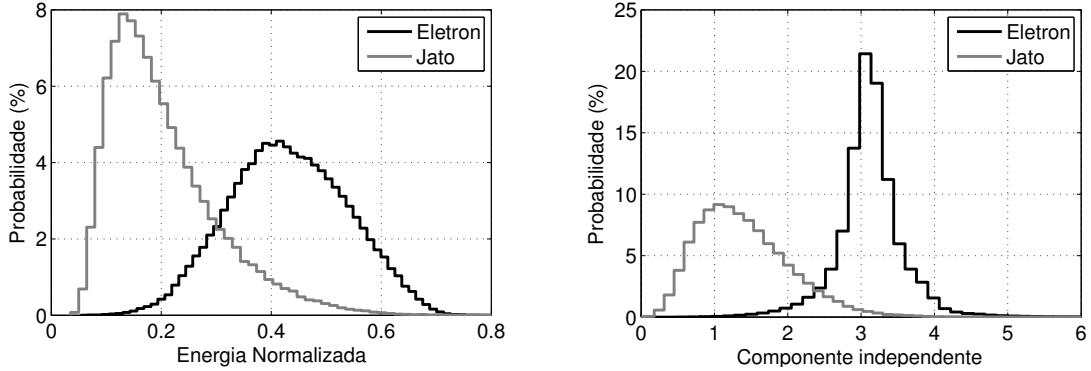


Figura 7.34: Histogramas do terceiro anel da camada E2 (esquerda) e de um dos componentes independentes (ICA linear) estimados para a mesma camada (direita).

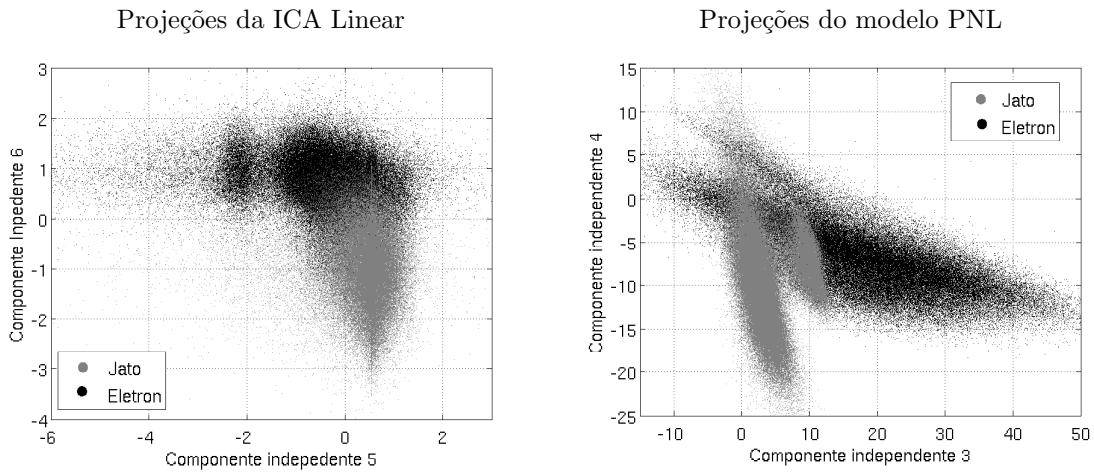


Figura 7.35: Gráficos de dispersão entre componentes independentes estimados por ICA linear (esquerda) e pelo modelo PNL (direita).

utilizados para estimação da NLICA, o de ICA Local apresentou maior eficiência de discriminação, seguido do modelo PNL modificado. Ficaram evidentes, também, os benefícios da segmentação dos processos de extração de características e classificação, pois, com a aplicação da rede combinadora, foi possível obter desempenho superior ao da rede neural convencional (utilizada no *Neural Ringer*). Foi possível observar ainda que, mesmo retirando-se a informação de 5 das 7 camadas (o que pode produzir uma economia de aproximadamente 20 % do tempo total de processamento) a rede combinadora apresenta desempenho equivalente ao do *Neural Ringer*.

Considerando o desempenho dos diferentes modos de estimação da NLICA utilizados neste trabalho, observaram-se algumas características em comum:

- Aumento da eficiência de discriminação em eventos baixa energia, que usual-

Tabela 7.10: Comparaçāo dos resultados obtidos para diferentes discriminadores no conjunto E10 considerando o māximo SP($\times 100$), as probabilidades de decisāo (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=92,5%.

Discriminador	Máx. SP$\times 100$	PD (%)	PF (%)	PF_{PD=92,5%}
T2Calo	91,99	92,50	8,55	8,55
Ringer	97,99 \pm 0,18	98,77 \pm 0,35	2,85 \pm 0,11	1,10 \pm 0,08
ICA	98,25 \pm 0,27	99,01 \pm 0,42	2,50 \pm 0,09	1,00 \pm 0,12
SOM	96,01 \pm 0,23	94,8 \pm 0,33	2,75 \pm 0,10	2,40 \pm 0,11
SOM Seg.	97,47 \pm 0,31	98,00 \pm 0,47	3,05 \pm 0,19	1,96 \pm 0,09
SOM+LVQ	98,01 \pm 0,26	98,90 \pm 0,36	2,87 \pm 0,17	1,09 \pm 0,15
SOM Seg.+LVQ	97,81 \pm 0,19	98,52 \pm 0,40	2,90 \pm 0,11	1,94 \pm 0,10
SOM Sup.	98,15 \pm 0,22	99,10 \pm 0,36	2,83 \pm 0,17	1,02 \pm 0,13
SOM Sup. Seg.	97,75 \pm 0,18	98,01 \pm 0,45	2,50 \pm 0,13	1,46 \pm 0,10
Local	99,05 \pm 0,28	99,60 \pm 0,43	1,50 \pm 0,08	0,85 \pm 0,09
PNL+PCA	97,67 \pm 0,23	98,61 \pm 0,30	3,30 \pm 0,11	1,15 \pm 0,16
PNL+PCD	98,30 \pm 0,29	99,00 \pm 0,37	2,40 \pm 0,10	0,94 \pm 0,12
PNL Mod.	98,55 \pm 0,21	99,34 \pm 0,41	2,29 \pm 0,15	0,70 \pm 0,08
Rede Comb.	98,65 \pm 0,34	99,12 \pm 0,34	1,82 \pm 0,09	0,61 \pm 0,07
Rede Comb. E1+E2	98,08 \pm 0,26	98,46 \pm 0,33	2,30 \pm 0,12	0,94 \pm 0,10

mente sāo mais influenciados pelo ruído de medição inerente ao calorímetro;

- Aumento da eficiência em regiões onde o sistema de calorímetros apresenta descontinuidades como o *crack* (para passagem de cabos) e as interconexões dos diferentes módulos do barril e da tampa do calorímetro eletromagnético.
- Reduçāo do falso alarme aproximadamente uniforme em η , ϕ e energia.

Considerando especificamente o pré-processamento por mapas auto-organizáveis, pode-se utilizar a resposta aos mapas como instrumento de visualização *online* do sistema de filtragem.

7.2 Conjunto E15i

Em comparaçāo com o conjunto E10, o problema de discriminação das assinaturas E15i é uma tarefa bem mais difícil, pois, com o corte realizado pelo nível 1 bem

mais severo, as assinaturas de jatos que chegam ao segundo nível têm características mais semelhantes às de um elétron médio (conforme mostrado anteriormente na Figura 6.13). Assim, é importante verificar o desempenho dos métodos propostos neste novo contexto, repetindo as análises realizadas anteriormente.

7.2.1 Resultados com os Discriminadores existentes

Em um outro trabalho [10], desenvolvido dentro do mesmo grupo de pesquisa, foi realizado um estudo sobre os métodos de pré-processamento linear aplicados aos sinais em anéis do conjunto E15i. Foram consideradas técnicas de compactação como análise de componentes principais (PCA) e componentes principais de discriminação (PCA), em conjunto com o modelo linear da análise de componentes independentes (ICA). Deste modo, serão apresentados aqui apenas um resumo com os principais resultados e conclusões obtidos no referido trabalho.

A Tabela 7.11 apresenta uma comparação dos resultados obtidos com o T2Calo, *Neural Ringer* e com o discriminador neural operando sobre os componentes independentes (modelo linear). Para este conjunto de dados simulados tipo E15i, as eficiências de discriminação apresentaram uma considerável redução em comparação aos sinais E10 (devido à maior semelhança entre os perfis de elétrons e jatos). Considerando o pré-processamento por ICA linear, de modo análogo ao observado no conjunto E10, a melhor configuração para estimativa dos componentes independentes foi através da abordagem segmentada e com compactação por PCD.

Tabela 7.11: Comparaçāo dos resultados obtidos para oT2Calo, *Neural Ringer* e o discriminador neural operando em conjunto com a ICA (linear) para o conjunto E15. São mostrados o máxmo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.

Discriminador	Máx. SP $\times 100$	PD (%)	PF (%)	PF _{PD=97,2%}
T2Calo	87,93	97,34	20,30	20,30
Ringer	94,94 \pm 0,12	96,69 \pm 0,27	6,79 \pm 0,30	7,31 \pm 0,25
ICA	95,17 \pm 0,10	96,86 \pm 0,52	6,50 \pm 0,39	6,81 \pm 0,31

A Figura 7.36 mostra um comparativo entre as eficiências em η e energia para o discriminador baseado nos componentes independentes (com compactação por PCD)

e o T2Calo. Pode-se observar que, o pré-processamento por ICA em conjunto com o classificador neural, produzem uma maior eficiência na identificação de elétrons principalmente em baixas energias e próximo às regiões de interconexões entre os módulos do calorímetro ($\eta \sim 0$, $|\eta| \sim 1,5$ e $|\eta| \sim 2,5$). O falso alarme é reduzido em toda a faixa de η e, principalmente, para energias entre 15 e 50 GeV.

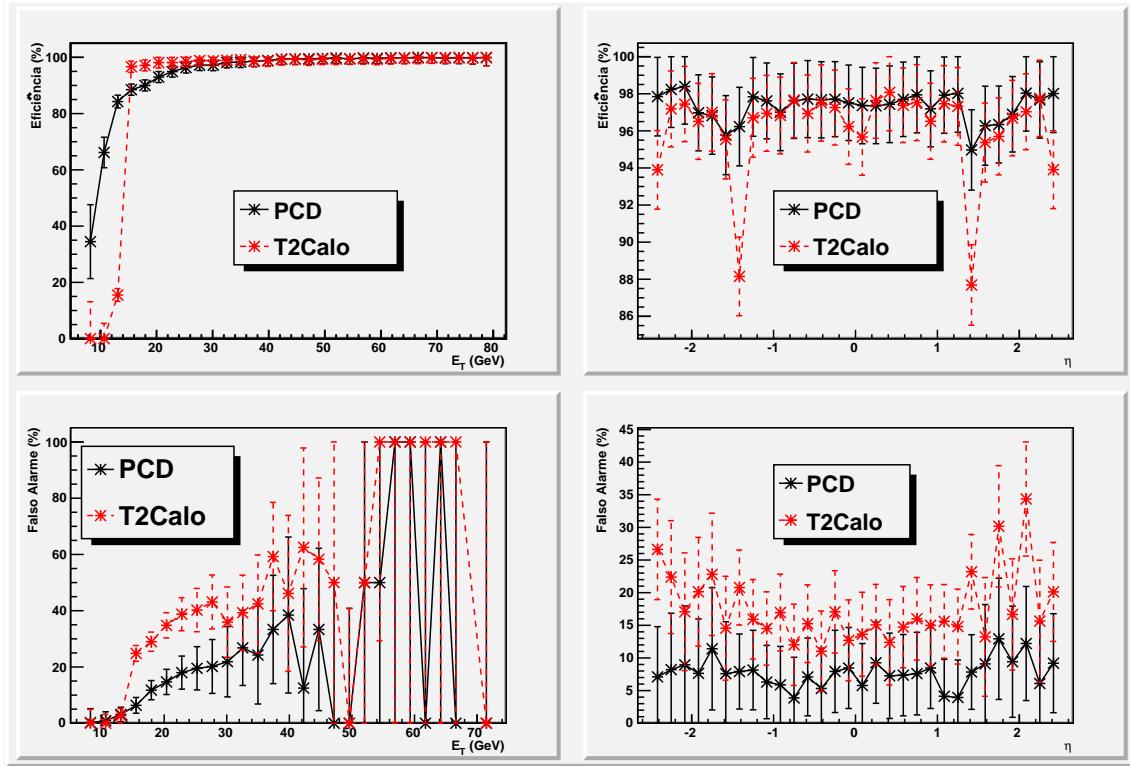


Figura 7.36: Comparativo da eficiência em energia (esquerda) e η (direita) para o T2Calo e o discriminador baseado no pré-processamento em ICA com compactação por PCD, retirada de [10].

7.2.2 Resultados com Pré-Processamento por NLICA

A seguir, serão apresentados os resultados obtidos repetindo-se para o conjunto E15i a metodologia aplicada anteriormente ao conjunto E10 para estimativa da NLICA.

ICA Local

De modo semelhante ao realizado para o conjunto E10, os sinais foram agrupados em dois *clusters* através do algoritmo *k-means*. A distribuição das assinaturas de elétrons e jatos nos dois *clusters* é mostrada na Figura 7.37. Pode-se observar que

quase todos os elétrons (99 %) e 48 % dos jatos são associados ao primeiro grupo. O restante das assinaturas é designada ao segundo agrupamento.

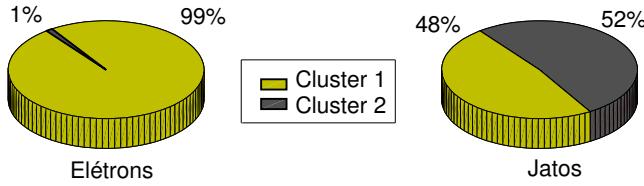


Figura 7.37: Divisão dos eventos de elétrons e jatos nos dois agrupamentos.

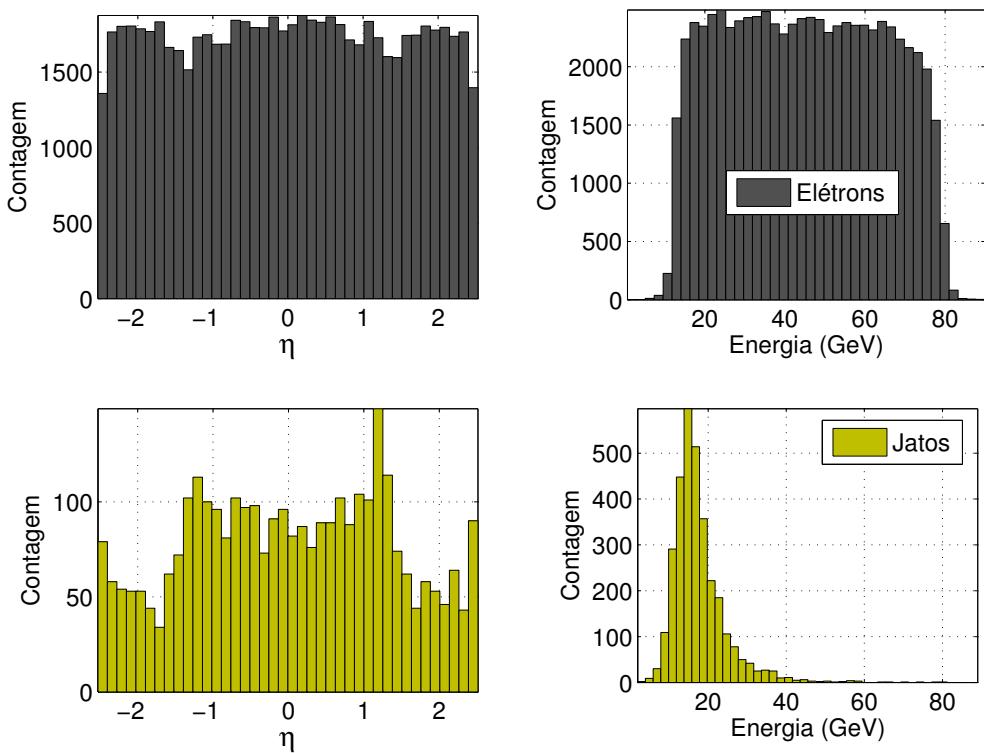
Considerando as distribuições em η e energia dos eventos agrupados em cada um dos *clusters* (ver Figura 7.38), pode-se observar que, o comportamento em η é bastante semelhante ao obtido para o conjunto E10, onde tanto os jatos, como a pequena parte dos elétrons associada ao agrupamento 2 estão concentrados em torno do *crack*. Por outro lado, diferente do observado para o conjunto E10, o perfil dos jatos do grupo 1 (que estão agrupados juntamente com a maioria dos elétrons) é bastante semelhante aos do grupo 2. Uma possível explicação para esse fato é que, nos sinais E15i, a seleção de nível 1 foi bastante severa e deixou passar apenas uma estatística reduzida para jatos.

Os perfis médios de elétrons e jatos nos dois agrupamentos são mostrados na Figura 7.39. Pode-se observar que o agrupamento 1 concentra os elétrons típicos e os jatos parecidos com eles (apresentando maior energia eletromagnética), enquanto que, o segundo agrupamento reúne os elétrons com características mais semelhantes a jatos (elétrons com menor concentração de energia na segunda camada eletromagnética e um pequeno pico de energia na camada hadrônica⁷).

Considerando a metodologia para projeto do discriminador baseado na ICA Local, após o agrupamento, um algoritmo de ICA linear (JADE) foi aplicado a cada um dos grupos e, em seguida, discriminadores neurais foram treinados a partir dos com-

⁷Uma justificativa para a existência de energia hadrônica em elétrons é que estas assinaturas, conforme mostrado na Figura 7.38, interagiram com o calorímetro eletromagnético na região do *crack* e por isso encontraram menor quantidade de material para absorver sua energia, ficando um resíduo que atinge a seção hadrônica

Cluster 1



Cluster 2

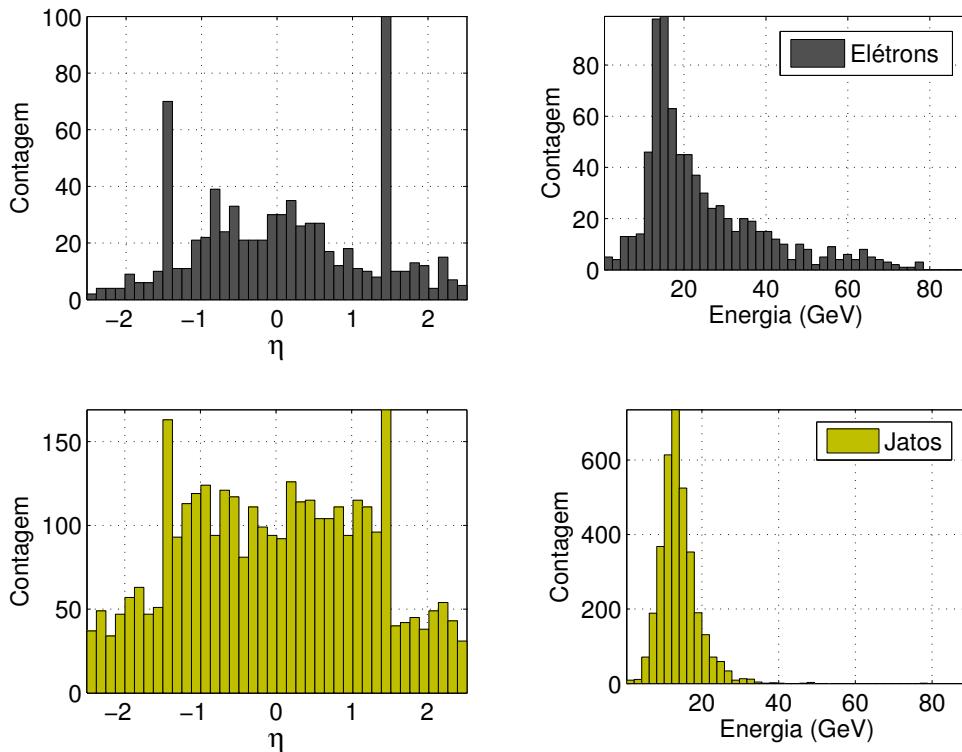


Figura 7.38: Distribuições em energia e η dos eventos nos *clusters* 1 (acima) e 2 (abaixo).

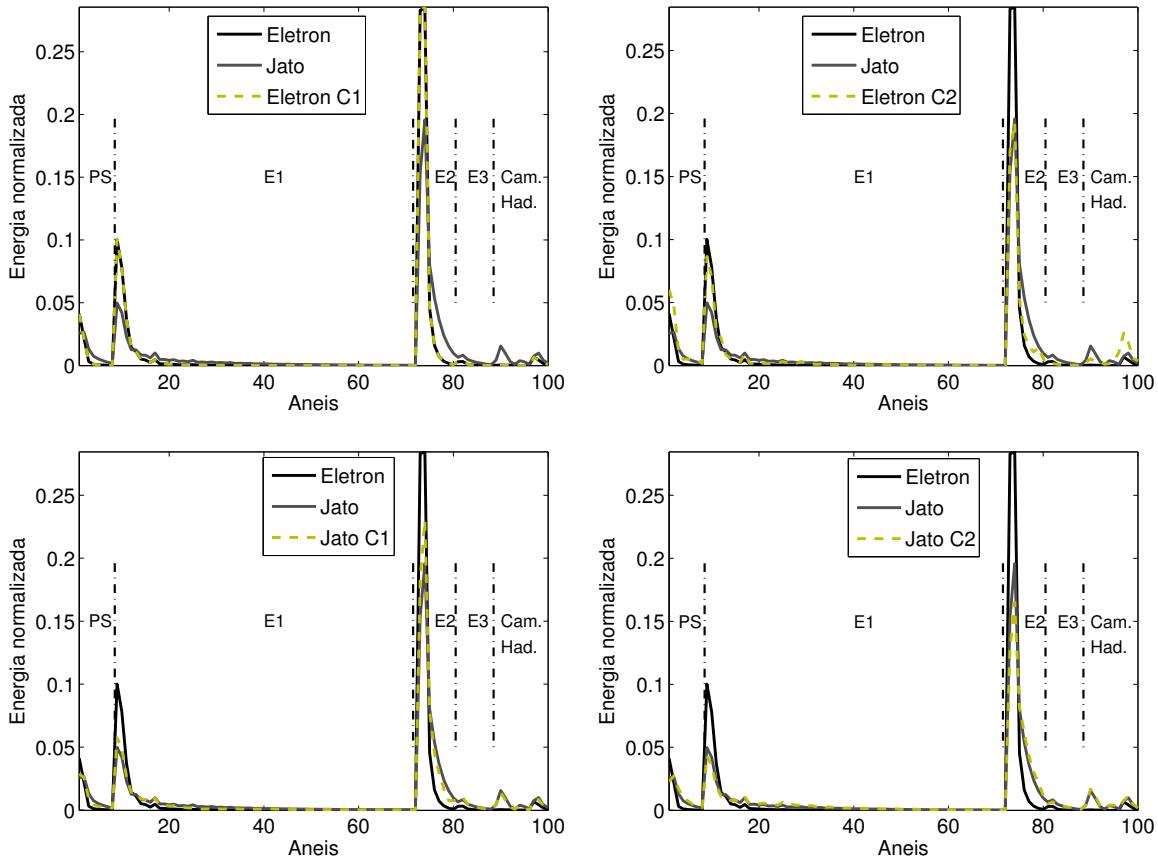


Figura 7.39: Eventos médios de elétrons e jatos nos *clusters* 1 e 2 comparados com os eventos médios considerando todo o conjunto de assinaturas E15i.

ponentes independentes. Para o ajuste dos patamares de decisão dos classificadores, novamente foi utilizado um algoritmo genético.

Para a escolha da topologia (número de neurônios ocultos) dos classificadores locais, foram realizadas diversas inicializações das redes neurais, variando-se (de 3 a 23) o número de neurônios na camada oculta. A Figura 7.40 mostra os resultados obtidos. Para o Discriminador do *cluster* 1, a máxima eficiência foi atingida para 13 neurônios ocultos, enquanto para o *cluster* 2, foram necessários 17 neurônios.

A Figura 7.41 mostra as curvas ROC para os discriminadores neurais operando sobre os anéis e sobre os componentes independentes e os pontos ótimos encontrados pelo AG para o pré-processamento através de ICA Local. Pode-se observar que, para um mesmo valor de eficiência (Ex. 96 %), o pré-processamento por ICA Local reduz a aceitação do falso alarme em aproximadamente 1,8 ponto percentual.

Um resumo dos resultados obtidos com a ICA Local é mostrado na Tabela 7.12.

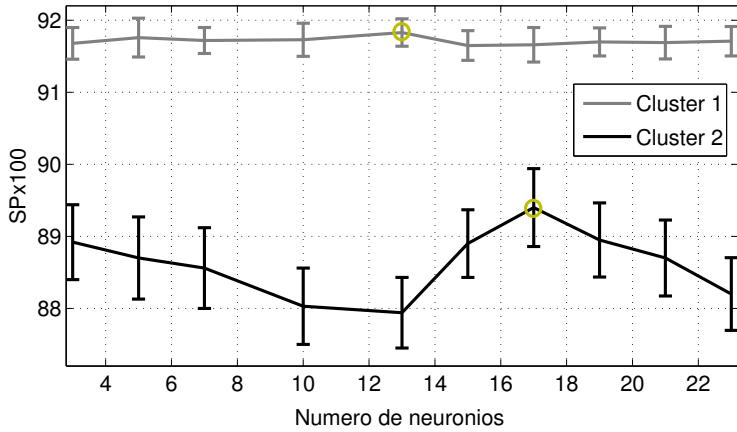


Figura 7.40: Variação do máximo SP com o número de neurônios.

Pode-se observar que o pré-processamento por ICA Local produz um aumento no índice SP, porém, o ponto (PD;PF) que produz esse máximo apresenta uma menor probabilidade de detecção do que para os demais discriminadores (esse fato é compensado por uma maior redução na aceitação de falso alarme).

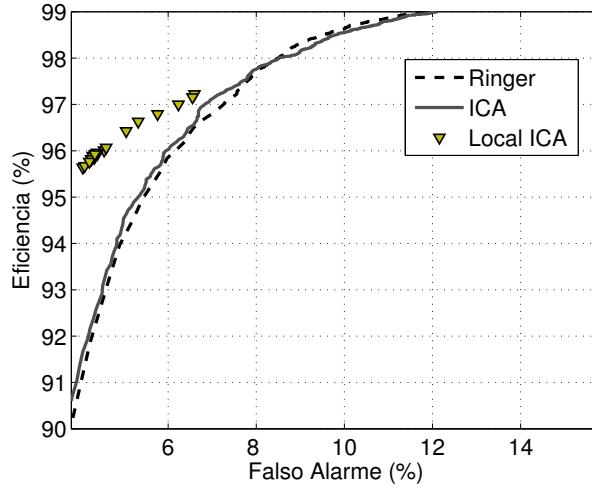


Figura 7.41: Comparativo de eficiência a partir das Curvas ROC.

Comparando o discriminador baseado na ICA Local com o *Neural Ringer* em termos da eficiência e do falso alarme para diferentes valores de η e energia (ver Figura 7.42), pode-se observar que, para os pontos de máximo SP, a eficiência obtida através do ICA Local é menor que a do *Neural Ringer* principalmente para energias menores do que 40 GeV e perto de $\eta \sim 0$ (interconexão entre os módulos do barril do calorímetro eletromagnético) e $|\eta| \sim 1,5$ (*crack*). Esta menor eficiência é compensada por uma redução na aceitação de jatos em quase todas as faixas de

Tabela 7.12: Comparação dos resultados obtidos com o pré-processamento por ICA local com os de outros discriminadores para o conjunto E15i. Foram considerados o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.

Discriminador	Máx. SP $\times 100$	PD (%)	PF (%)	PF _{PD=97,2%}
T2Calo	87,93	97,34	20,30	20,30
Ringer	94,94 \pm 0,12	96,69 \pm 0,27	6,79 \pm 0,30	7,31 \pm 0,25
ICA	95,17 \pm 0,10	96,86 \pm 0,52	6,50 \pm 0,39	6,81 \pm 0,31
Local	95,85 \pm 0,21	96,02 \pm 0,29	4,31 \pm 0,43	6,54 \pm 0,45

energia e η .

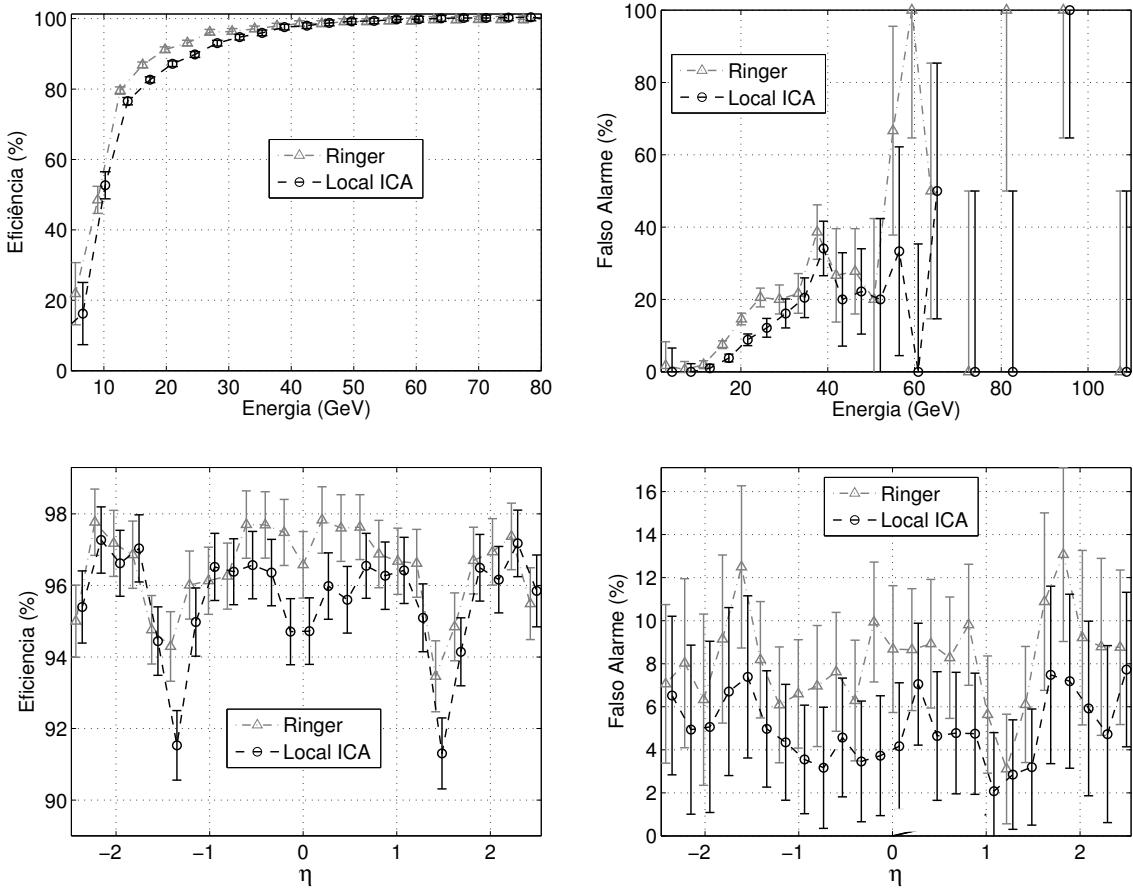


Figura 7.42: Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores *Neural Ringer* e Local ICA+MLP.

A Figura 7.43 possibilita uma comparação mais detalhada entre os discriminadores considerando as distribuições em η e energia dos eventos classificados corretamente.

mente através da ICA local e incorretamente pelos demais discriminadores (*Ringer* e ICA). De modo semelhante ao verificado para o conjunto E10, pode-se observar que a superioridade da ICA local é mais evidente nas regiões das interconexões do calorímetro ($|\eta| \sim 1,5$ e $|\eta| \sim 2,5$) e em baixas energias.

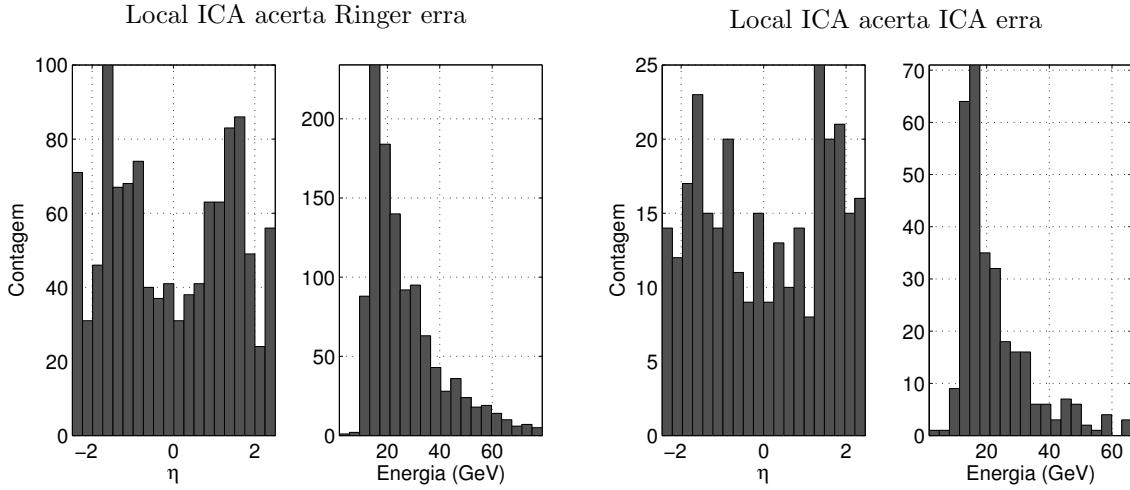


Figura 7.43: Comparativo de eficiência (esquerda) e falso alarme (direita) dos discriminadores *Neural Ringer* e Local ICA+MLP.

Mapas Auto-Organizáveis - SOM

Seguindo o procedimento descrito anteriormente para o conjunto E10, para estimar o tamanho ótimo dos mapas auto-organizáveis foram utilizados como critérios o erro médio de reconstrução (EMR) e o SP. A Figura 7.44 ilustra os resultados obtidos para os mapas não-segmentados (treinados concatenando os sinais de todas as camadas num único vetor de características) e segmentados (treinados separadamente para cada camada). Os valores do SP se referem à classificação obtida no conjunto de teste a partir da operação do SOM diretamente como classificador (associando, para o conjunto de treinamento, os neurônios de saída com a classe que mais os ativam). Novamente, foram utilizados dois tipos de mapas, quadrados ($N \times N$) e retangulares (com a razão entre os lados da grade igual a divisão entre os dois primeiros componentes principais do conjunto de treino). O tamanho ótimo dos mapas a ser utilizado foi escolhido com o propósito de obter um compromisso entre alto SP, baixo EMR e reduzido número de neurônios na grade.

A Tabela 7.13 mostra as dimensões escolhidas para os mapas e os respectivos

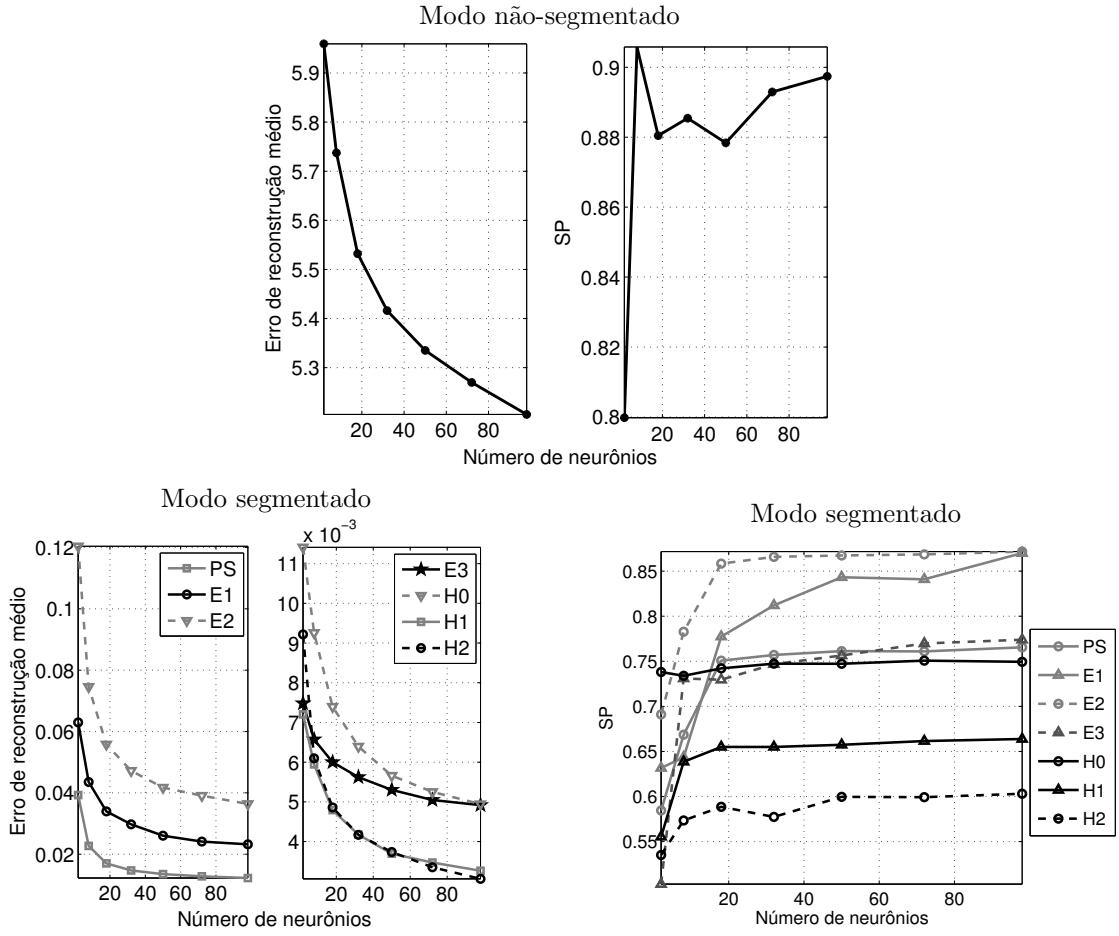


Figura 7.44: Erro quadrático médio (esquerda) e SP (direita) calculados variando-se as dimensões dos mapas, acima para o modo não-segmentado e abaixo para o segmentado.

valores do índice SP obtidos, considerando as diferentes abordagens para treinamento. Comparando com as dimensões ótimas escolhidas para o conjunto E10 (ver Tabela 7.4) observa-se que, para os mapas segmentados das camadas PS, E2, E3, H0 e H1 o número de neurônios dobrou, na camada H2 aumentou de 16 para 32. Apenas na camada E1 o tamanho permaneceu aproximadamente constante. Para o caso não segmentado, o tamanho do mapa permaneceu constante.

Considerando as diversas abordagens utilizadas (segmentado, não-segmentado, supervisionado e com ajuste por LVQ), classificadores neurais supervisionados foram treinados e a Figura 7.45 mostra as curvas ROC obtidas para cada caso e a Tabela 7.14 apresenta um resumo dos resultados. Percebe-se que, assim como no caso do conjunto E10, a maior eficiência (em termos do máximo SP) foi obtida para a abordagem supervisionada do SOM (treinada no modo não-segmentado).

Neste caso, o uso da abordagem segmentada produziu aumento da eficiência apenas para o SOM não-supervisionado. Com o treinamento supervisionado, a abordagem não-segmentada apresentou melhor desempenho.

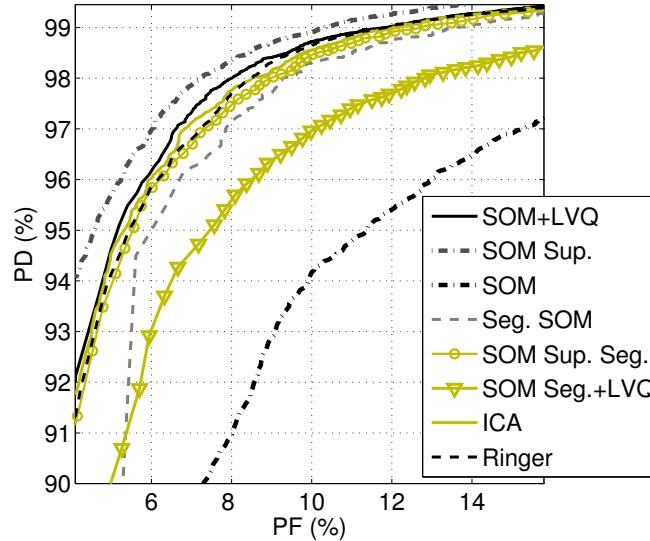


Figura 7.45: Curvas ROC para as diversas abordagens do SOM (conjunto E15i).

Um comparativo da eficiência e do falso alarme em função da energia e de η é mostrado na Figura 7.46 para os discriminadores SOM-Sup.+MLP e *Neural Ringer*. Pode-se observar que o pré-processamento através do SOM-Sup. é capaz de produzir ligeiro aumento na eficiência, principalmente próximo a $|\eta| \sim 1,5$ e também uma pequena redução do falso alarme.

A Figura 7.47 mostra as distribuições em η e energia dos eventos classificados corretamente pelo discriminador SOM+MLP e incorretamente pelo *Neural Ringer*.

Tabela 7.13: Máximo SP($\times 100$) para diferentes estratégias de treinamento dos mapas (aqui o SP é calculado considerando a operação do SOM diretamente como um classificador).

Camada	SOM Segmentado							SOM não-seg.
	PS	E1	E2	E3	H0	H1	H2	
Dimensões do mapa	10×5	10×5	10×5	12×6	12×6	6×6	6×6	12×6
Não-supervisionado	72.45	79.86	80.06	75.56	74.64	65.34	59.60	89.29
Ajustado por LVQ	76.13	84.31	86.73	76.93	75.07	66.15	59.97	90.58
Supervisionado	74.48	89.02	89.27	75.51	74.75	64.78	59.63	92.80

Tabela 7.14: Comparação dos resultados obtidos através do pré-processamento por SOM com os de outros discriminadores para o conjunto E15i. Foram considerados o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.

Discriminador	Máx. SP $\times 100$	PD (%)	PF (%)	PF _{PD=97,2%}
T2Calo	87,93	97,34	20,30	20,30
Ringer	94,94 \pm 0,12	96,69 \pm 0,27	6,79 \pm 0,30	7,31 \pm 0,25
ICA	95,17 \pm 0,10	96,86 \pm 0,52	6,50 \pm 0,39	6,81 \pm 0,31
SOM	91,49 \pm 0,15	96,42 \pm 0,45	13,31 \pm 0,38	14,95 \pm 0,30
SOM Seg.	94,70 \pm 0,18	96,22 \pm 0,49	6,81 \pm 0,41	7,94 \pm 0,33
SOM+LVQ	94,92 \pm 0,13	96,66 \pm 0,43	6,80 \pm 0,37	7,69 \pm 0,29
SOM Seg.+LVQ	93,74 \pm 0,17	96,51 \pm 0,51	8,98 \pm 0,44	9,97 \pm 0,41
SOM Sup.	95,54 \pm 0,09	97,20 \pm 0,33	6,10 \pm 0,39	5,96 \pm 0,23
SOM Sup. Seg.	95,22 \pm 0,11	96,87 \pm 0,41	6,41 \pm 0,31	6,61 \pm 0,27

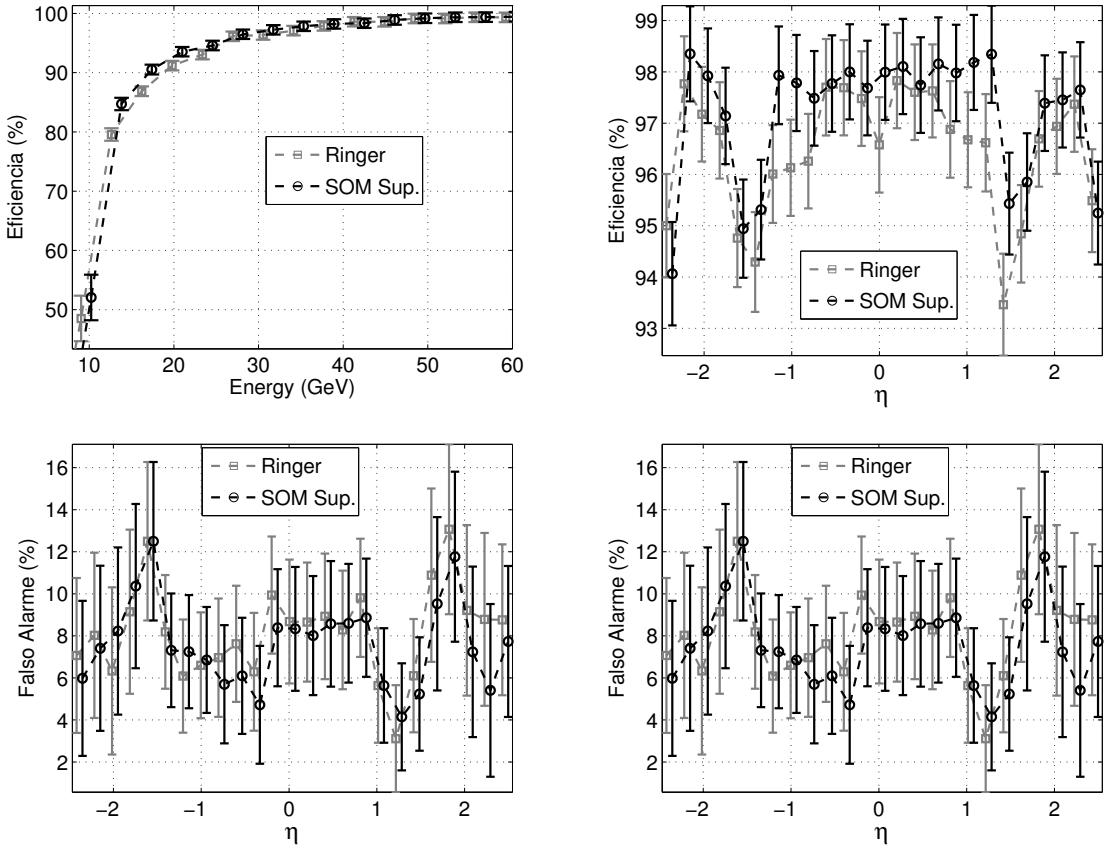


Figura 7.46: Eficiência (acima) e falso alarme (abaixo) em energia e η .

Percebe-se que, o pré-processamento por SOM é capaz de produzir aumento da eficiência e redução do falso alarme principalmente próximo ao *crack* e para eventos de energia menor que 40 GeV. Analisando-se os perfis de deposição de energia médios para esses eventos (ver Figura 7.48), percebe-se que o pré-processamento por SOM permite a identificação de eventos de elétrons que apresentam um aumento na parcela de energia hadrônica e de jatos que têm perfil muito semelhante a elétrons típicos (pouca energia hadrônica e concentração ao redor do centro).

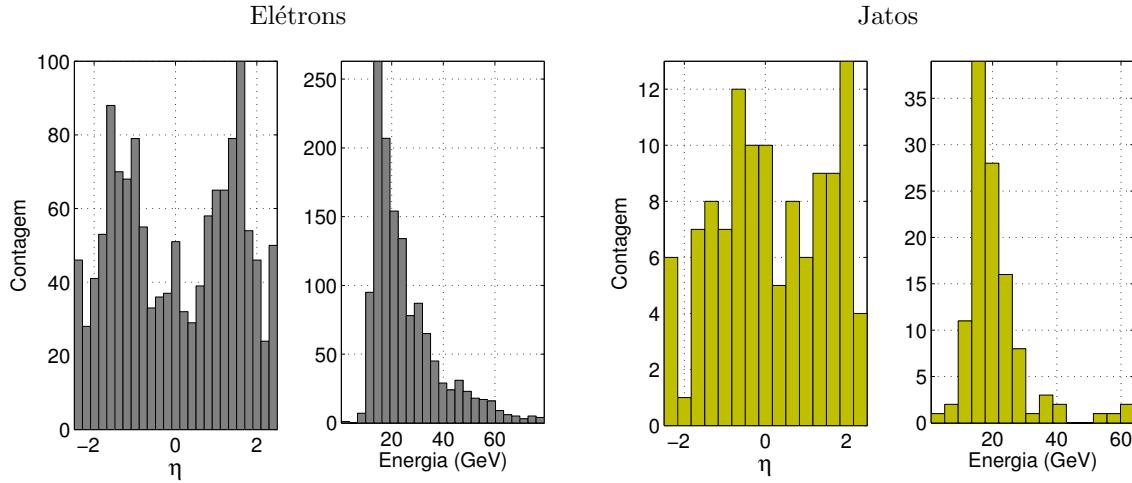


Figura 7.47: Distribuições em η e energia dos eventos de elétrons (esquerda) e jatos (direita) classificados corretamente pelo discriminador SOM+MLP e incorretamente pelo *Neural Ringer*.

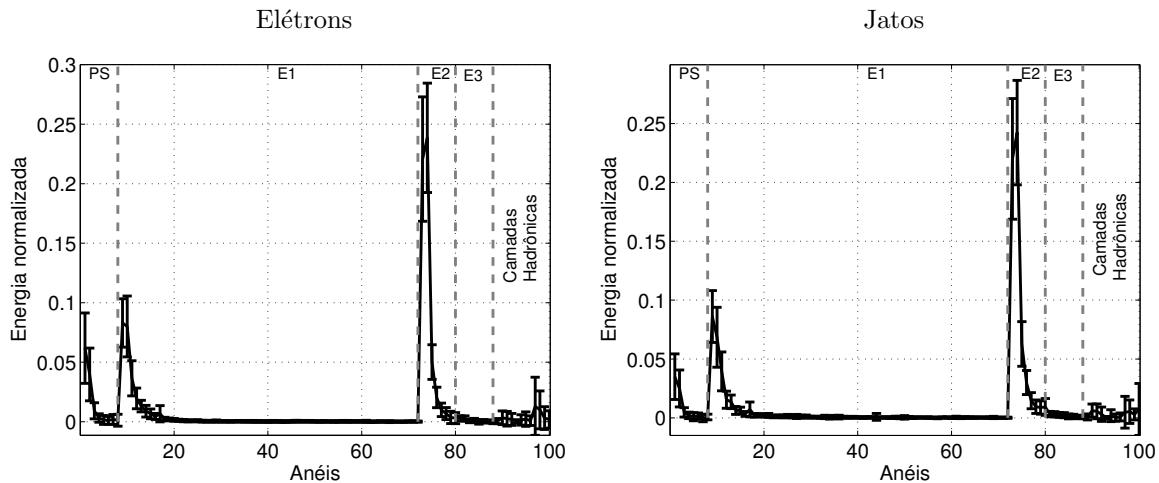


Figura 7.48: Eventos médios (e desvio padrão) de elétrons (esquerda) e jatos (direita) classificados corretamente pelo discriminador SOM+MLP e incorretamente pelo *Neural Ringer*.

Modelo com Restrições Estruturais - PNL

Para o modelo PNL, os algoritmos utilizados apresentam limitações quanto à dimensionalidade do problema (número de componentes a serem estimados), de modo que, para um número elevado de componentes, a estimação se torna lenta e com baixa precisão (muitas vezes divergindo, ou seja, não consegue finalizar a estimação do modelo). Assim, conforme feito anteriormente para os sinais do conjunto E10, a estimação do modelo PNL foi realizada apenas no modo segmentado.

Um dos parâmetros a ser estimado no modelo PNL utilizado é o grau de não-linearidade (que é representado pelo número de neurônios utilizados para a estimação de cada função não-linear). Com o objetivo de determinar o grau de não-linearidade mais adequado para o problema, o número de neurônios foi variado de dois a quatro, (realizando-se 10 inicializações para cada caso) e os componentes independentes estimados foram utilizados como entrada para classificadores neurais supervisionados. Pode-se observar que o melhor desempenho foi obtido para dois neurônios. Aumentando-se o grau da não-linearidade a eficiência de discriminação cai. Este resultado é idêntico ao encontrado para os sinais do conjunto E10 e de certa forma já era esperado pois, devido às características construtivas dos calorímetros não esperava-se fenômenos fortemente não-lineares. Um resumo dos resultados é apresentado na Tabela 7.15.

Tabela 7.15: Comparação entre discriminadores baseados no pré-processamento através do modelo PNL.

Discriminador	Ringer	PNL (2 neur.)	PNL (3 neur.)	PNL (4 neur.)
Máx. SP×100	94.35	94.70	93.70	90.83
P_F p/ P_D = 97%	8.67±0.20	7.69±0.35	9.67±0.38	17.39±0.40

Na Figura 7.49 são mostradas as curvas ROC obtidas para as diversas abordagens de treinamento do modelo PNL. Pode-se observar que o pré-processamento a partir do modelo PNL modificado produziu um aumento na eficiência em relação ao *Ringer* e ao discriminador neural com pré-processamento por ICA. O PNL+PCD também foi capaz de superar estes dois discriminadores apenas em uma pequena região da ROC, enquanto que o PNL+PCA apresentou desempenho inferior ao *Neural Ringer*. Os resultados estão resumidos na Tabela 7.16. Percebe-se que, em comparação com

o *Neural Ringer*, o pré-processamento através do modelo PNL modificado é capaz de reduzir em aproximadamente 1,3 ponto percentual o falso alarme para PD=97,2 %.

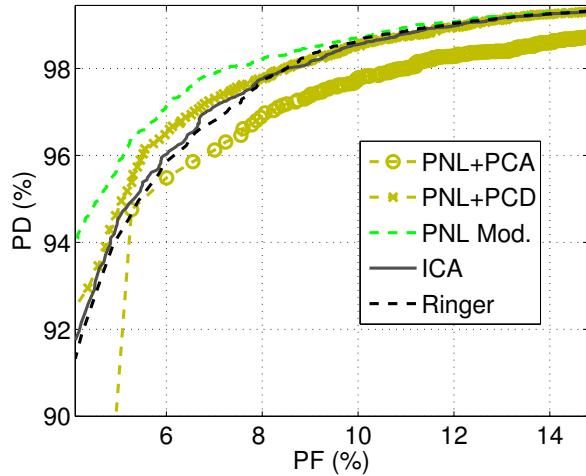


Figura 7.49: Curvas ROC obtidas variando-se o número de neurônios utilizados para estimar as funções não-lineares do modelo PNL (conjunto E15i).

Tabela 7.16: Comparação dos resultados obtidos através do pré-processamento pelo modelo PNL com os de outros discriminadores para o conjunto E15i. Foram considerados o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.

Discriminador	Máx. SP $\times 100$	PD (%)	PF (%)	PF _{PD=97,2%}
T2Calo	87,93	97,34	20,30	20,30
Ringer	94,94 \pm 0,12	96,69 \pm 0,27	6,79 \pm 0,30	7,31 \pm 0,25
ICA	95,17 \pm 0,10	96,86 \pm 0,52	6,50 \pm 0,39	6,81 \pm 0,31
PNL+PCA	94,55 \pm 0,17	96,13 \pm 0,32	7,01 \pm 0,37	8,42 \pm 0,33
PNL+PCD	95,20 \pm 0,13	96,88 \pm 0,25	6,46 \pm 0,41	6,67 \pm 0,31
PNL Mod.	95,54 \pm 0,19	97,57 \pm 0,43	6,47 \pm 0,29	5,98 \pm 0,35

Na Figura 7.50, é apresentada uma comparação de desempenho entre os discriminadores PNL+MLP e *Neural Ringer*. Pode-se observar que, em termos da eficiência, os dois discriminadores apresentam desempenho semelhante, embora o pré-processamento através do modelo PNL produza uma queda no desempenho próximo ao *crack* ($|\eta| \sim 1,5$). Considerando o falso alarme, o uso da NLICA produz redução aproximadamente uniforme em quase todas as faixas de energia e η .

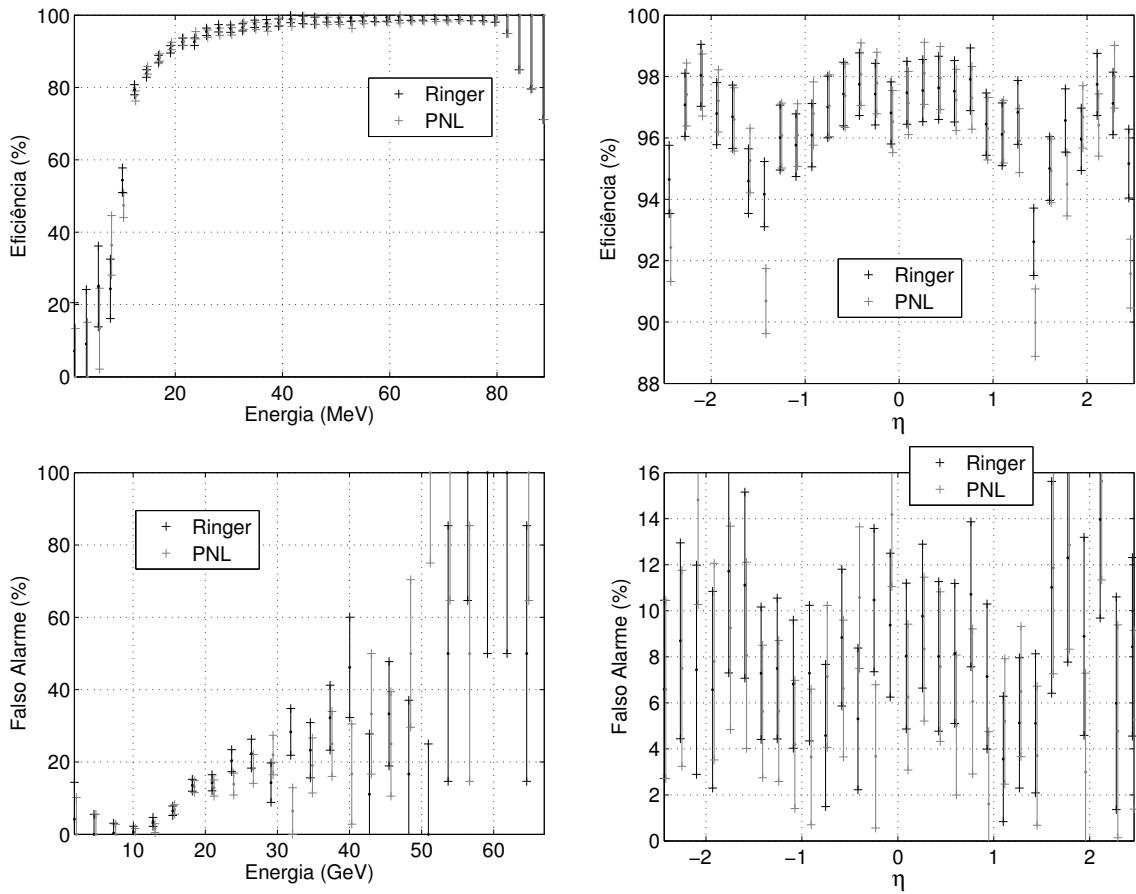


Figura 7.50: Eficiência (acima) e falso alarme (abaixo) em energia e η .

7.2.3 Estudo da Relevância por Camada

Para avaliar a relevância de cada uma das camadas no processo de discriminação, foram treinados classificadores neurais retirando-se parte da informação (referente a uma ou mais camadas do calorímetro). A Figura 7.51 mostra o máximo SP obtido em função da parcela da informação retirada. Pode-se observar que, o discriminador treinado sem as camadas H1 e H2 apresenta maior eficiência do que aquele treinado com toda as sete camadas (de modo semelhante ao observado no conjunto simulado E10). Isto indica que a informação destas camadas (que são bastante influenciadas por ruído de medição, uma vez que a maior parte dos eventos deposita uma pequena quantidade de energia nestas seções) confunde o discriminador, dificultando o processo de discriminação. Foi verificado ainda que, retirando-se a informação de todas as camadas hadrônicas o desempenho obtido é equivalente ao do discriminador treinado com toda a informação.

De modo semelhante ao realizado para o conjunto de dados E10, foram treinados

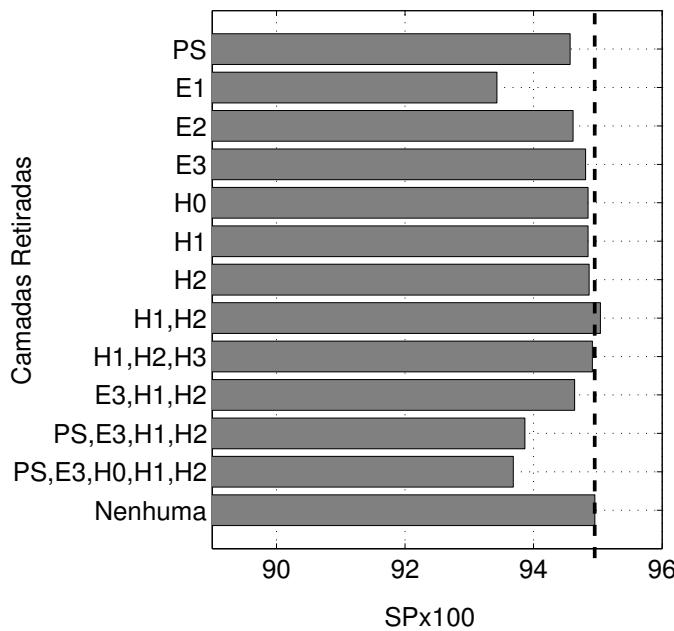


Figura 7.51: Máximo SP obtido após treinamento das redes neurais retirando (de modo segmentado) parte da informação.

classificadores especialistas nas informações de cada uma destas camadas e uma rede combinadora foi utilizada para melhor explorar toda a segmentação da informação (entre os métodos de combinação de classificadores estudados para o conjunto E10, a rede combinadora apresentou melhor resultado e, portanto, será utilizada para esta tarefa no restante desta tese). Para encontrar o número de neurônios ótimo para a camada oculta da rede combinadora, foram realizadas diversas inicializações, variando-se de 1 a 25 o número de neurônios nesta camada. Conforme mostrado na Figura 7.52, de modo semelhante ao observado anteriormente (para os sinais do conjunto E10), um pequeno número de neurônios (no caso 3 neurônios) é suficiente para resolver o problema.

Com a rede combinadora, foi repetido o procedimento de retirada da informação de modo segmentado, realizado anteriormente nesta Seção para a rede neural de duas camadas. A Figura 7.53 mostra os valores máximos do SP e as curvas ROC obtidas em cada um dos casos. Pode-se observar que, usando a rede combinadora, é possível retirar a informação de até quatro camadas (PS, E3, H1 e H2 ou PS, H0, H1 e H2) e ainda assim é obtido um desempenho superior ao do discriminador neural convencional operando sobre todas as camadas.

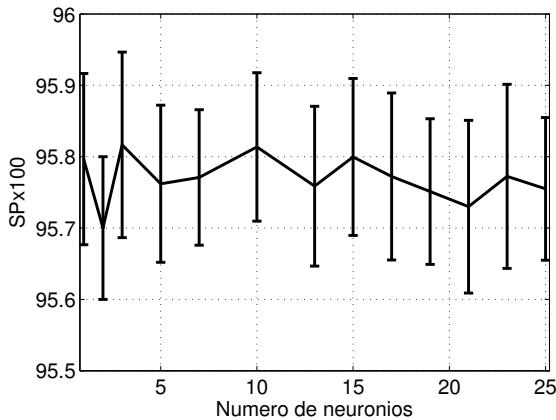


Figura 7.52: Variação do desempenho da rede neural combinadora variando-se o número de neurônios da camada oculta.

7.2.4 Comentários e Discussão

Na Tabela 7.17 é apresentada uma comparação entre os diversos discriminadores propostos até aqui para o conjunto E15i. Considerando os algoritmos de pré-processamento para o discriminador neural, o melhor desempenho (em termos do máximo SP), foi, assim como para o conjunto E10, obtido para o discriminador baseado em ICA Local. O uso do SOM supervisionado e do modelo PNL modificado acabou produzindo valores semelhantes do SP. Uma vantagem do SOM é a possibilidade de associar, também, as ferramentas de visualização do evento inerentes ao modelo. Em geral, pode-se verificar que o uso da informação supervisionada, como era esperado, sempre produziu aumento na eficiência de discriminação.

Considerando o uso dos classificadores segmentados, observa-se que a exploração da informação segmentada através da rede combinadora foi bastante eficiente na separação elétron/jato, sendo este o modelo de discriminador que produziu maior eficiência de discriminação. É importante notar ainda que, utilizando apenas a informação de duas camadas (E1 e E2), é possível obter desempenho equivalente (ligeiramente inferior) ao *Neural Ringer*. Com essa última configuração, é possível reduzir o tempo total de processamento em aproximadamente 30 %.

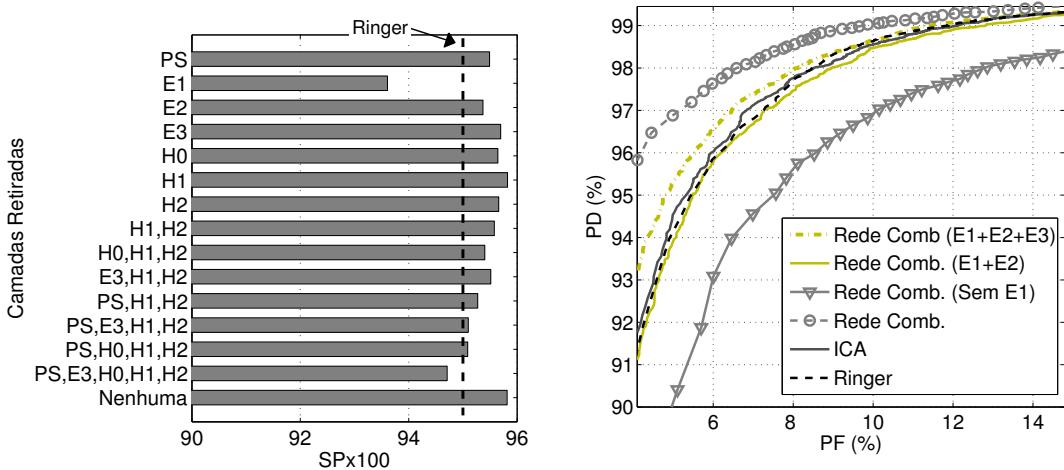


Figura 7.53: Máximo SP (esquerda) e curvas ROC (direita) obtidas para a rede combinadora retirando-se a informação (de modo segmentado) de uma ou mais camadas do calorímetro, a linha tracejada vertical indica o desempenho do *Neural Ringer*.

Tabela 7.17: Comparaçāo dos resultados obtidos para diferentes discriminadores no conjunto E15i considerando o māximo SP($\times 100$), as probabilidades de decisāo (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=97,2%.

Discriminador	Máx. SP $\times 100$	PD (%)	PF (%)	PF _{PD=97,2%}
T2Calo	87,93	97,34	20,30	20,30
Ringer	$94,94 \pm 0,12$	$96,69 \pm 0,27$	$6,79 \pm 0,30$	$7,31 \pm 0,25$
ICA	$95,17 \pm 0,10$	$96,86 \pm 0,52$	$6,50 \pm 0,39$	$6,81 \pm 0,31$
SOM	$91,49 \pm 0,15$	$96,42 \pm 0,45$	$13,31 \pm 0,38$	$14,95 \pm 0,30$
SOM Seg.	$94,70 \pm 0,18$	$96,22 \pm 0,49$	$6,81 \pm 0,41$	$7,94 \pm 0,33$
SOM+LVQ	$94,92 \pm 0,13$	$96,66 \pm 0,43$	$6,80 \pm 0,37$	$7,69 \pm 0,29$
SOM Seg.+LVQ	$93,74 \pm 0,17$	$96,51 \pm 0,51$	$8,98 \pm 0,44$	$9,97 \pm 0,41$
SOM Sup.	$95,54 \pm 0,09$	$97,20 \pm 0,33$	$6,10 \pm 0,39$	$5,96 \pm 0,23$
SOM Sup. Seg.	$95,22 \pm 0,11$	$96,87 \pm 0,41$	$6,41 \pm 0,31$	$6,61 \pm 0,27$
Local	$95,85 \pm 0,21$	$96,02 \pm 0,29$	$4,31 \pm 0,43$	$6,54 \pm 0,45$
PNL+PCA	$94,55 \pm 0,17$	$96,13 \pm 0,32$	$7,01 \pm 0,37$	$8,42 \pm 0,33$
PNL+PCD	$95,20 \pm 0,13$	$96,88 \pm 0,25$	$6,46 \pm 0,41$	$6,67 \pm 0,31$
PNL Mod.	$95,54 \pm 0,19$	$97,57 \pm 0,43$	$6,47 \pm 0,29$	$5,98 \pm 0,35$
Rede Comb.	$96,01 \pm 0,20$	$96,47 \pm 0,37$	$4,45 \pm 0,37$	$5,42 \pm 0,30$
Rede Comb. E1+E2	$94,84 \pm 0,23$	$96,65 \pm 0,31$	$6,95 \pm 0,44$	$7,32 \pm 0,23$

Capítulo 8

Resultados - Dados Experimentais

Neste capítulo serão apresentados os resultados obtidos através da aplicação da metodologia proposta para sinais medidos no detector ATLAS. Os sinais utilizados são divididos em dois tipos. Inicialmente, será verificada a capacidade dos algoritmos propostos para identificar (e consequentemente eliminar) os sinais induzidos no detector por raios cósmicos. A seguir, serão utilizados dados medidos no detector na fase inicial de operação do LHC, onde a energia total dos feixes está limitada a 7 TeV (funcionando em plena capacidade, o LHC atingirá energias da ordem de 14 TeV).

8.1 Eficiência dos Sistemas de Classificação Propostos na Rejeição de Raios Cósmicos

Para avaliar a eficiência de classificação dos raios cósmicos, foi realizado um estudo comparativo entre diferentes discriminadores. Para o T2Calo e o *Neural Ringer* foram utilizadas versões destes algoritmos implementadas no sistema de *software* do detector (Athena). Considerando os discriminadores baseados em componentes independentes (ICA e NLICA), foram empregadas as versões de cada modelo de estimativa que apresentaram melhor desempenho para o conjunto simulado de assinaturas E15.

A Tabela 8.1 mostra os resultados obtidos para diferentes discriminadores. Pode-se observar que os classificadores que utilizam os sinais em anéis apresentam desempenho superior ao T2Calo em pelo menos 1,1 ponto percentual. Com o pré-

processamento por ICA e Local ICA, foi possível melhorar ligeiramente a taxa de rejeição obtida pelo *Neural Ringer*.

Quando a informação segmentada dos sinais em anéis é explorada na classificação através da rede combinadora, é possível aumentar a rejeição dos raios cósmicos em relação ao *Neural Ringer*. Considerando a rede combinadora utilizando apenas as informações das camadas E1 e E2 observa-se que o desempenho diminui um pouco, mas ainda fica superior ao obtido pelo T2Calo.

Tabela 8.1: Comparaçāo do desempenho de diferentes discriminadores na rejeição de raios cósmicos.

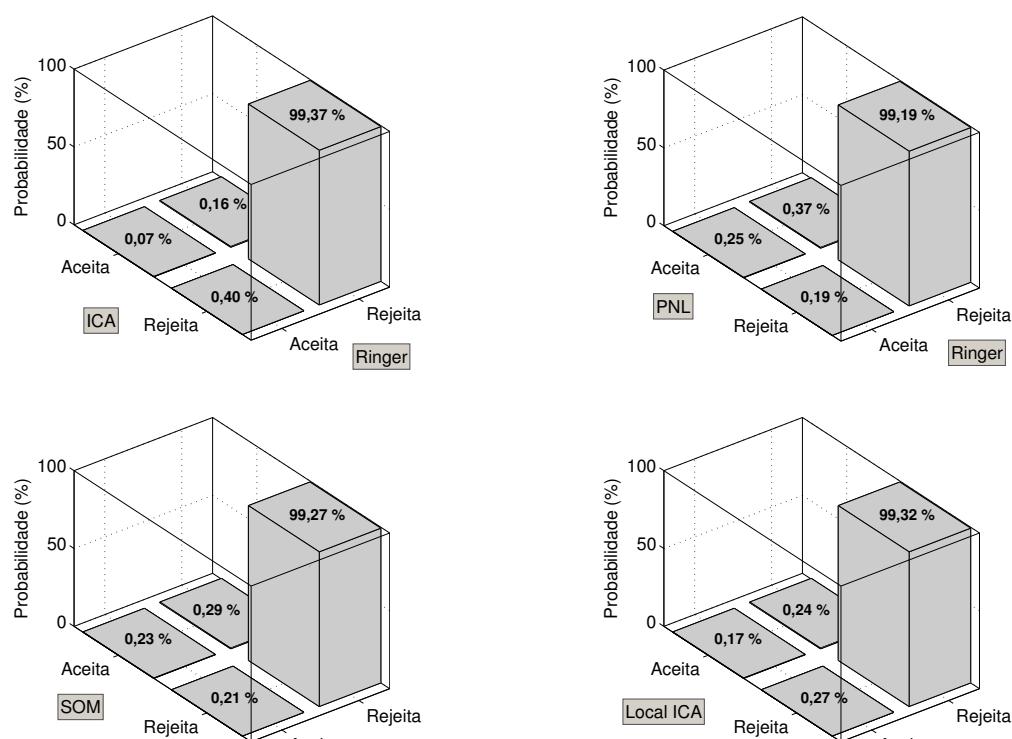
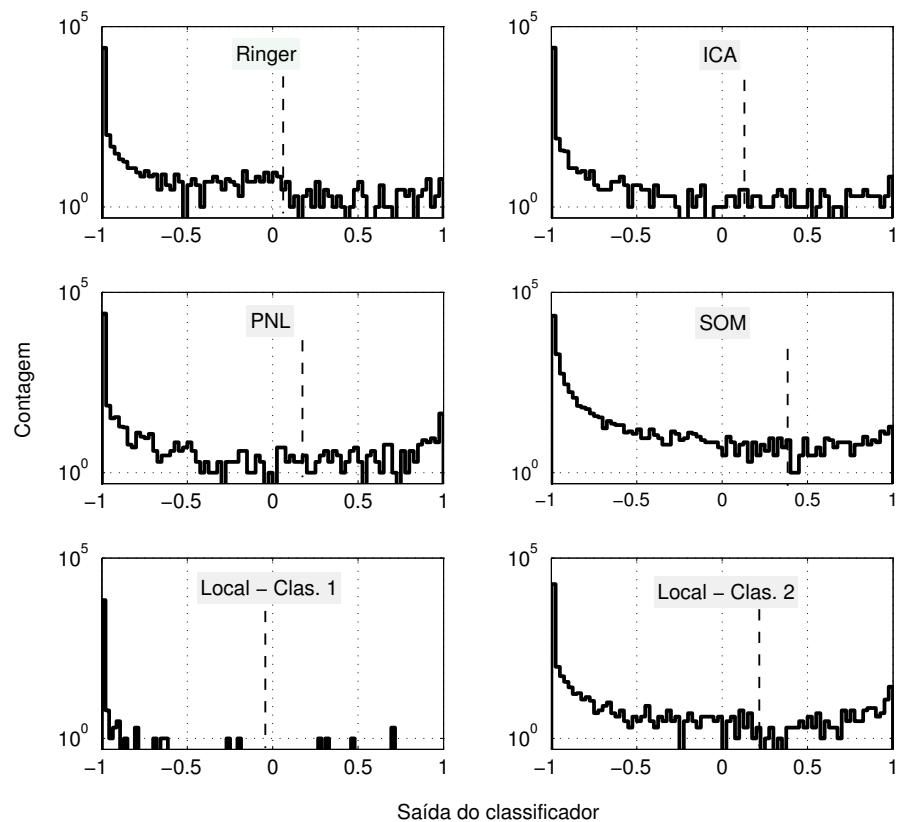
Discriminador	T2Calo	Ringer	ICA	Local	SOM	PNL	R. Comb.	R. Comb._(E1+E2)
Rejeição (%)	98,21	99,56	99,77	99,59	99,48	99,38	99,70	99,03

As distribuições dos valores obtidos nas saídas das redes neurais para os discriminadores com pré-processamento por ICA, SOM, PNL e ICA Local são mostradas na Figura 8.1 (no treinamento foi utilizada a saída alvo igual a 1 para elétrons). Os patamares de corte estão indicados em linhas tracejadas verticais. Percebe-se que, para todos os casos houve uma alta concentração perto de -1, indicando que a maioria das assinaturas de raios cósmicos é facilmente classificada como ruído de fundo.

Ainda na Figura 8.1, pode-se observar um comparativo de desempenho entre os diferentes discriminadores que utilizam o pré-processamento por ICA/NLICA e o *Neural Ringer* a partir da parcela das assinaturas que eles acertam e erram em conjunto. Nesta análise é possível observar que há uma grande concordância na parcela de assinaturas rejeitadas.

A Figura 8.2 mostra o perfil médio dos eventos de raios cósmicos que são aprovados (aceitos incorretamente como elétrons) conjuntamente pelo *Ringer* e pelo discriminador baseado em ICA Local. Pode-se observar que, os eventos aceitos tem, em média, um perfil que se assemelha ao perfil médio de elétrons.

Por fim, são apresentadas na Figura 8.3, as eficiências na rejeição dos raios cósmicos em função da energia e η para diferentes discriminadores. Observa-se que o T2Calo apresenta uma flutuação muito maior que os outros discriminadores (tanto em energia como em η) e uma tendência para menores valores de rejeição. Comparando os discriminadores neurais, observa-se que, o pré-processamento por



Comparativo de rejeição

Figura 8.1: Saída dos classificadores neurais (acima) e comparativo de rejeição (abaixo) para eventos de raios cósmicos.

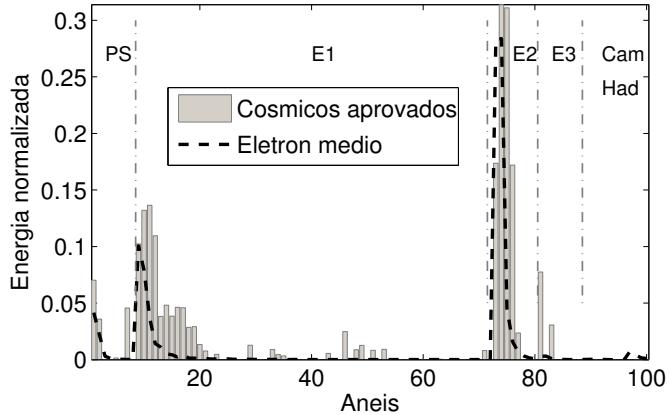


Figura 8.2: Perfil médio dos eventos de raios cósmicos aprovados conjuntamente pelo discriminador baseado em ICA Local e pelo *Neural Ringer*. O perfil médio dos elétrons do conjunto E15 é mostrado para comparação.

ICA/NLICA produz uma rejeição aproximadamente constante (com uma pequena queda na região do *crack* $|\eta| \sim 1,5$), atingindo resultados bastante semelhantes aos do *Neural Ringer*.

8.2 Colisões do LHC

Conforme mostrado no Capítulo 6, os sinais experimentais disponíveis até o momento tem características físicas bem distintas dos sinais simulados utilizados para desenvolvimento dos discriminadores, estando concentrados em uma faixa de energia muito baixa.

Conforme mostrado na Seção 4.2, as características do perfil de deposição nos calorímetros depende da energia da partícula. Sabendo que a faixa de energia dos sinais experimentais não estava presente nos conjuntos simulados utilizado¹, é provável que os discriminadores treinados para os sinais simulados não operem de modo adequado nestes sinais experimentais.

Deste modo, aplicando-se diretamente aos sinais experimentais a versão do *Neural Ringer* treinada para o conjunto simulado E10 é obtida uma eficiência de apenas 31,83 %, para um falso alarme igual a 19,50 %. A Figura 8.4 mostra as saídas do

¹Os sinais simulados estavam concentrados em energias maiores que 10 GeV (assinaturas de segundo nível do tipo E10 e E15), enquanto que, os experimentais estão numa faixa inferior (assinaturas tipo E3)

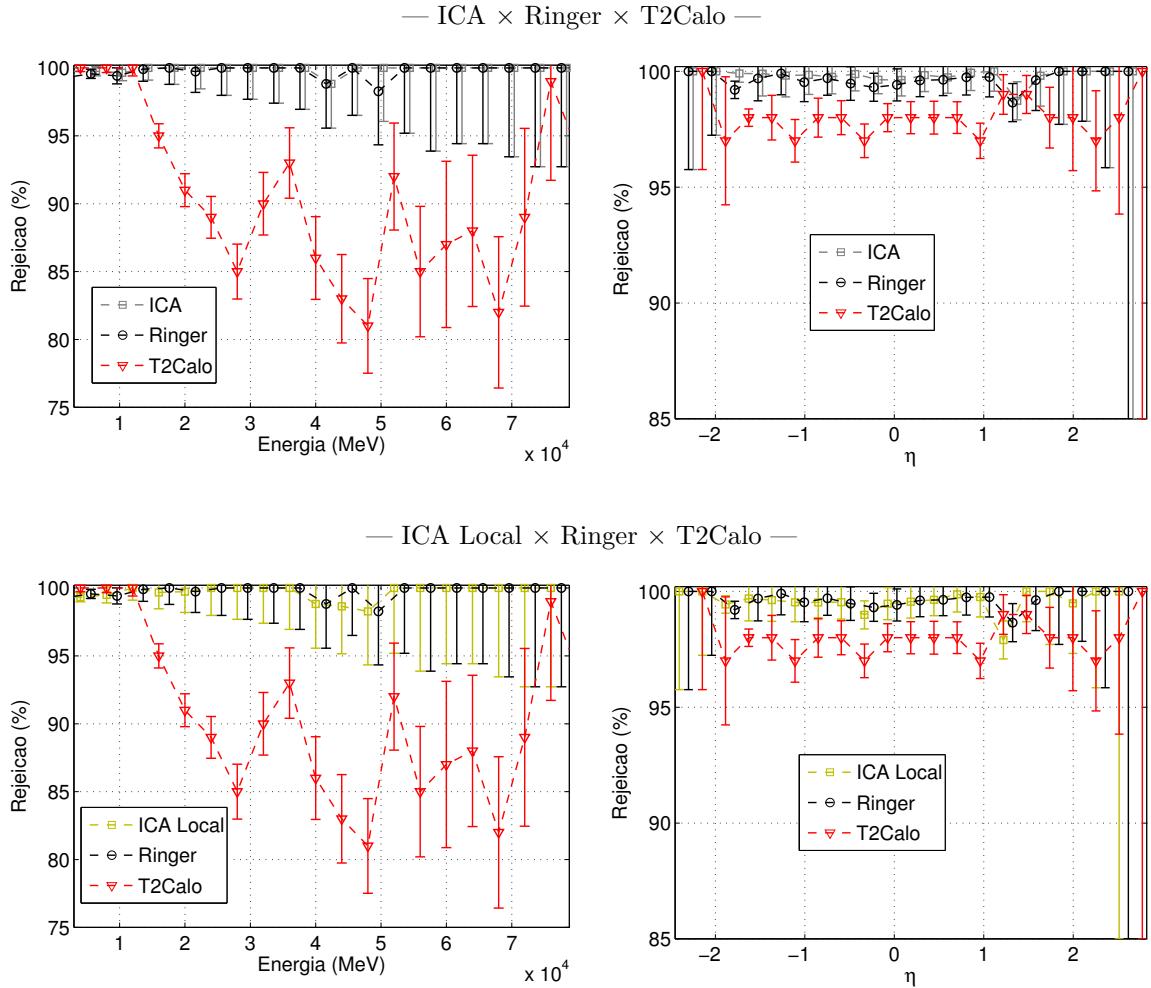


Figura 8.3: Comparativo de desempenho em energia e η de diferentes discriminadores na rejeição de eventos de raios cósmicos.

classificador neural. Pode-se observar que, as distribuições dos eventos de elétrons e jatos é bem semelhante, com uma maior concentração perto de -1 (ou seja, a rede neural não foi treinada para identificar o perfil dos elétrons de baixa energia adquiridos experimentalmente).

Neste contexto, para melhor avaliar o desempenho das técnicas de pré-processamento propostas neste trabalho, o processo de treinamento será repetido (seguindo a mesma metodologia utilizada para os sinais simulados, porém com um maior foco nas técnicas que apresentaram melhor desempenho para aqueles conjuntos), utilizando agora os sinais experimentais.

A informação da reconstrução *offline* será utilizada como verdade para o treinamento. Conforme exposto anteriormente no Capítulo 6, os eventos aprovados no

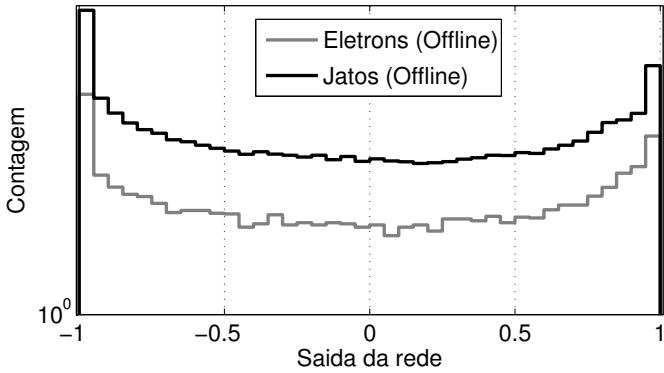


Figura 8.4: Saídas do *Neural Ringer* (treinado para os sinais simulados E10) para os sinais experimentais.

critério *electron tight* do *offline* serão considerados como “elétrons” (as aspas serão utilizadas no texto para diferenciar das assinaturas que são efetivamente elétrons, pois sabe-se que mesmo o critério mais seletivo da análise *offline* pode aceitar também partículas hadrônicas). O conjunto de “jatos” é composto pelas assinaturas que não foram aceitas no critério *electron loose*.

Considerando as assinaturas de segundo nível do tipo E3 (com energia maior que 3 GeV) o discriminador de elétrons padrão do ATLAS (T2Calo) pode ser configurado de duas formas diferentes (de acordo com a física de interesse):

- **T2EM3_Loose:** este corte apresenta maior eficiência na identificação de elétrons e, consequentemente maior aceitação de falso alarme. Para os sinais disponíveis, este corte identificou corretamente 98,10 % dos “elétrons” para um falso alarme igual a 64,07 %.
- **T2EM3_Medium:** num ajuste diferente, o T2Calo é capaz de eliminar maior parcela das assinaturas hadrônicas em troca de uma pequena redução na eficiência. Para o conjunto utilizado, este critério identificou 96,50 % dos “elétrons”, aceitando como falso alarme 55,73 % dos “jatos”.

8.2.1 Resultados - Anéis

Considerando o classificador neural operando diretamente sobre os anéis (*Neural Ringer*), inicialmente foi realizado um estudo para determinar se, no contexto dos sinais experimentais, o número de neurônios ocultos utilizado como padrão para os

sinais simulados (10 neurônios) também se mantém como uma boa opção. Foram treinadas várias redes, variando-se o número de neurônios ocultos de 1 a 18. Para cada nova configuração, foram realizadas 10 inicializações distintas, selecionado-se aleatoriamente a composição dos conjuntos de treino, teste e validação (conforme descrito no Capítulo 6).

A Figura 8.5 mostra os valores obtidos para o máximo SP (média e desvio padrão) variando-se o número de neurônios ocultos. Pode-se observar que o melhor SP foi obtido para uma rede de 5 neurônios ocultos. Este número será usado como referência para os discriminadores treinados com os sinais experimentais. Analisando os valores obtidos para o máximo SP, fica evidente que a eficiência para os sinais experimentais diminuiu em comparação aos simulados (onde o SP máximo para o discriminador neural operando sobre os anéis estava na faixa de 0,98 e 0,95, respectivamente para sinais E10 e E15i). Esta queda no desempenho pode ser associada a fatores como:

- Maior sobreposição entre as características de elétrons e jatos medidas no calorímetro quando a energia é mais baixa;
- Maior influencia de fenômenos inerentes dos sinais experimentais como ruído de medição (que é agravado para baixa energia, pois a relação sinal-interferência é menor) e empilhamento (sobreposição de eventos);
- Incerteza na determinação do tipo de partícula, pois sabe-se que, o critério de seleção da análise *offline* aceita assinaturas hadrônicas como sendo elétrons.

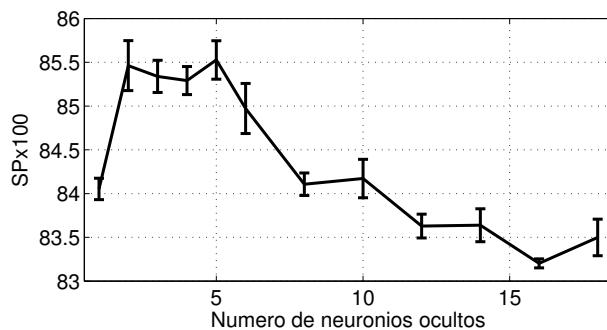


Figura 8.5: Máximo SP (média e desvio padrão) obtidos variando-se o número de neurônios ocultos.

Na Figura 8.6, são mostradas as distribuições da saída da rede neural (com 5 neurônios ocultos) e da projeção de um discriminador linear de Fisher. Observa-se que, a rede neural produziu maior separação entre as classes. Ainda na mesma Figura, pode-se visualizar as curvas ROC obtidas pelos dois discriminadores em comparação com os valores das probabilidades de detecção (PD) e falso alarme (PF) para os cortes do T2Calo. É possível visualizar que, a rede neural proporciona melhor desempenho que o T2Calo (para os dois cortes), porém, o discriminante de Fisher perde para o corte T2ME3_loose e apresenta desempenho aproximadamente igual ao T2EM3_tight.

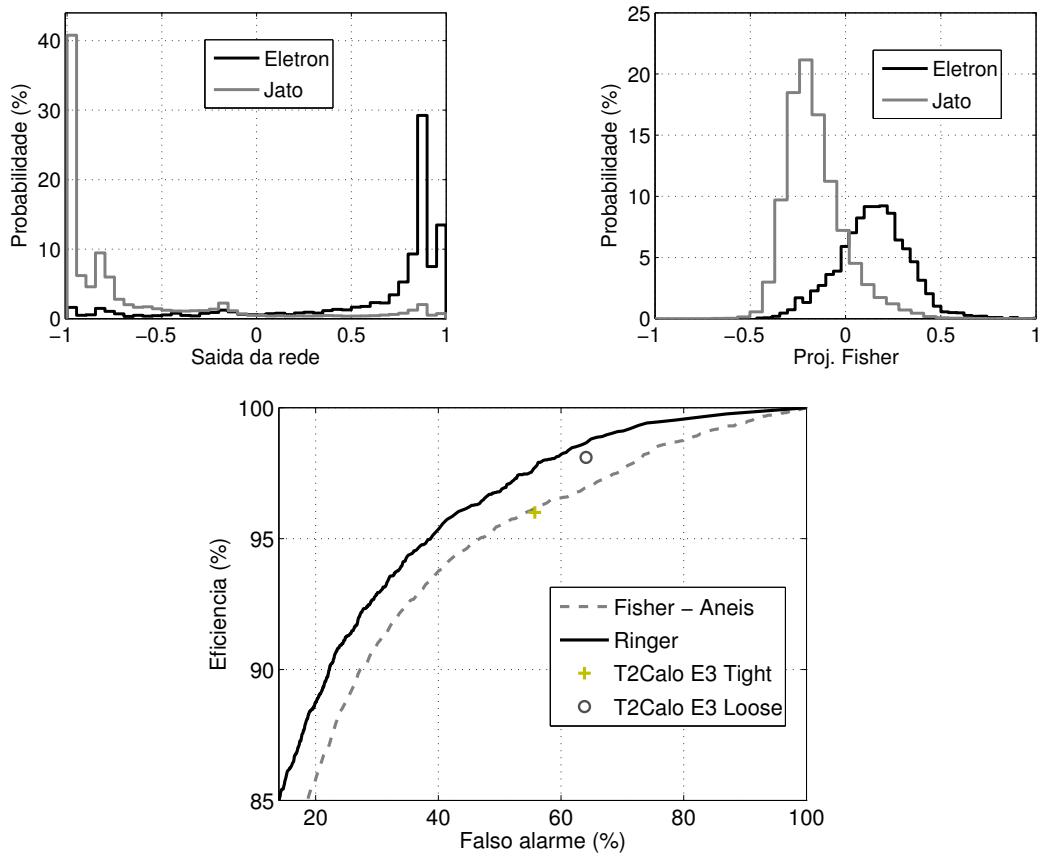


Figura 8.6: Distribuição das saídas para os discriminadores neural (esquerda e acima) e de Fisher (direita e acima) e as curvas ROC (abaixo).

A Tabela 8.2 compara os resultados obtidos até aqui. Pode-se observar que, o ponto de máximo SP para o *Neural Ringer* ($SP_{max} \times 100 = 85,55$; $PD \sim 84\%$ e $PF \sim 13\%$) apresenta eficiência menor que os cortes do T2Calo, dificultando a comparação entre os discriminadores. Para contornar esse problema, é preciso modificar o patamar de corte do classificador. Na Figura 8.7, é mostrada a variação do SP

Tabela 8.2: Comparaço dos resultados obtidos para o *Neural Ringer* e o T2Calo considerando o mximo SP($\times 100$), as probabilidades de deciso (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=96% e PD=98%.

Discriminador	$SP_{Max.} \times 100$	PD (%)	PF (%)	$PF_{PD=96\%}$	$PF_{PD=98\%}$
T2EM3_loose	63,36	98,10	64,07	-	64,07
T2EM3_tight	68,01	96,30	55,73	55,73	-
Ringer	$85,55 \pm 0,21$	$84,00 \pm 0,19$	$12,95 \pm 0,11$	$43,16 \pm 0,16$	$57,63 \pm 0,17$

e das probabilidades de deteco e falso alarme para diferentes valores do patamar de deciso. Pode-se observar que, h uma regio central onde os valores do SP se mantêm perto do mximo. Assim, ajustando-se o patamar para -0,70, é possível obter eficincia de 93,0 % para um falso alarme de 30,4 % ($SP \times 100 = 80,9$). Reduzindo ainda mais o patamar para -0,81, chega-se à mesma eficincia do corte T2E3_medium (~96 %), para um falso alarme igual a 43,2 % (porm, neste caso o SP caiu para 75,1).

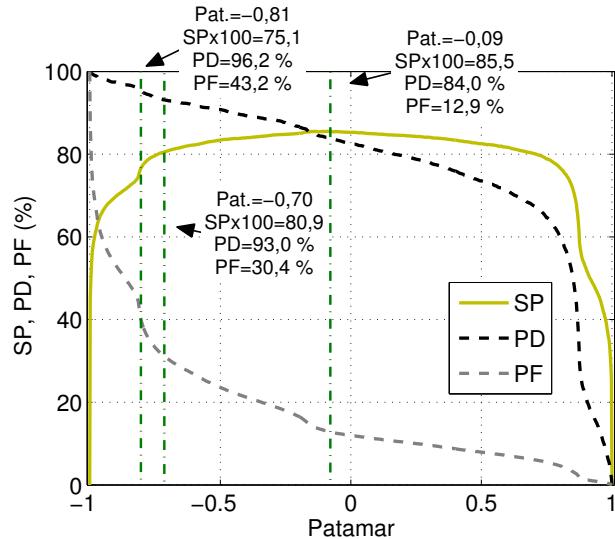


Figura 8.7: SP.

Na Figura 8.8 é apresentada uma comparação entre o *Neural Ringer* e os dois cortes do T2Calo em termos da eficincia e do falso alarme. Para o corte T2EM3_Loose, o *Neural Ringer* foi ajustado para a mesma probabilidade de deteco (~98 %). Pode-se verificar que, o *Ringer* é superior para altos valores de $|\eta|$ e pximo a $\eta = 0$ e para baixa energia, porm o discriminador neural apresentou eficincia inferior na regio do crack. Observando o falso alarme, fica evidente a superioridade do *Ringer*.

para quase todas as faixas de energia e pseudorapidez (η).

Considerando agora o corte T2EM3_ *Medium*, mostra que, de modo semelhante ao verificado no critério T2EM3_ *Loose*, para uma mesma probabilidade de detecção ($\sim 96\%$), o *Neural Ringer* reduz significativamente o falso alarme de modo aproximadamente uniforme em energia. Analisando as distribuições em η , observa-se novamente que há uma inversão nas curvas de falso alarme dos dois discriminadores. É interessante notar que, o *Ringer* consegue reduzir bastante a aceitação de falso alarme na região do *crack*.

Embora o *Neural Ringer* tenha produzido um desempenho de discriminação superior aos cortes do T2Calo, há uma grande discrepância entre os resultados obtidos para os sinais simulados. Neste contexto, visando uma melhor caracterização das assinaturas que o *Ringer* classifica incorretamente, foram selecionados os eventos onde a saída do classificador neural é muito próxima ao valor alvo atribuído à classe oposta (assinaturas de “elétrons” com saída da rede neural próxima a -1 e assinaturas de “jatos” com saída próxima a 1). As características destas assinaturas (em termos dos perfis de deposição médios e das distribuições em energia e η) serão comparadas a seguir.

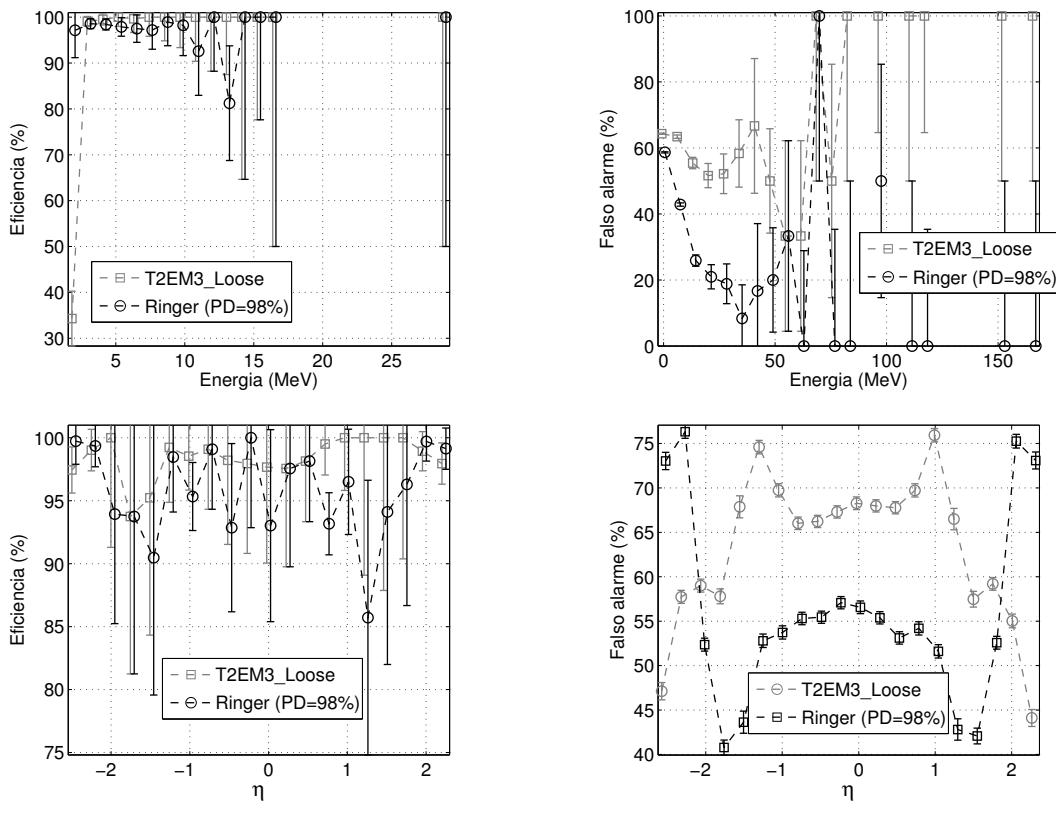
Os perfis médios dos eventos classificados incorretamente pelo *Neural Ringer* (para as duas classes) são mostrados na Figura 8.9 e comparados com os eventos médios de “elétrons” e “jatos”. É possível verificar que, os eventos classificados incorretamente pelo *Ringer* apresentam perfis médios bastante semelhantes ao perfil da outra classe.

Analizando a energia e a posição de interação (em função da coordenada η) dos eventos classificados incorretamente pelo *Neural Ringer* (ver Figura 8.10), observa-se que, os erros na identificação de elétrons estão concentrados nas regiões onde a seleção *offline* apresenta baixa eficiência (baixas energias, $\eta \sim 0$, $|\eta| \sim 1,5$ e $|\eta| \sim 2,5$), conforme mostrado anteriormente na Figura 6.19.

A partir do que foi observado nas Figuras 8.9 e 8.10, há uma forte indicação que uma parcela dos eventos que o *Ringer* classifica incorretamente seja, na verdade, proveniente de erro na identificação pela análise *offline*. Como, de certo modo, os cortes do T2Calo são um sub-conjunto dos critérios do *offline*² este discriminador

²As quatro variáveis calculadas pelo T2Calo são estimadas novamente na seleção *offline*, porém,

— Ringer \times T2EM3_Loose —



— Ringer \times T2EM3_Medium —

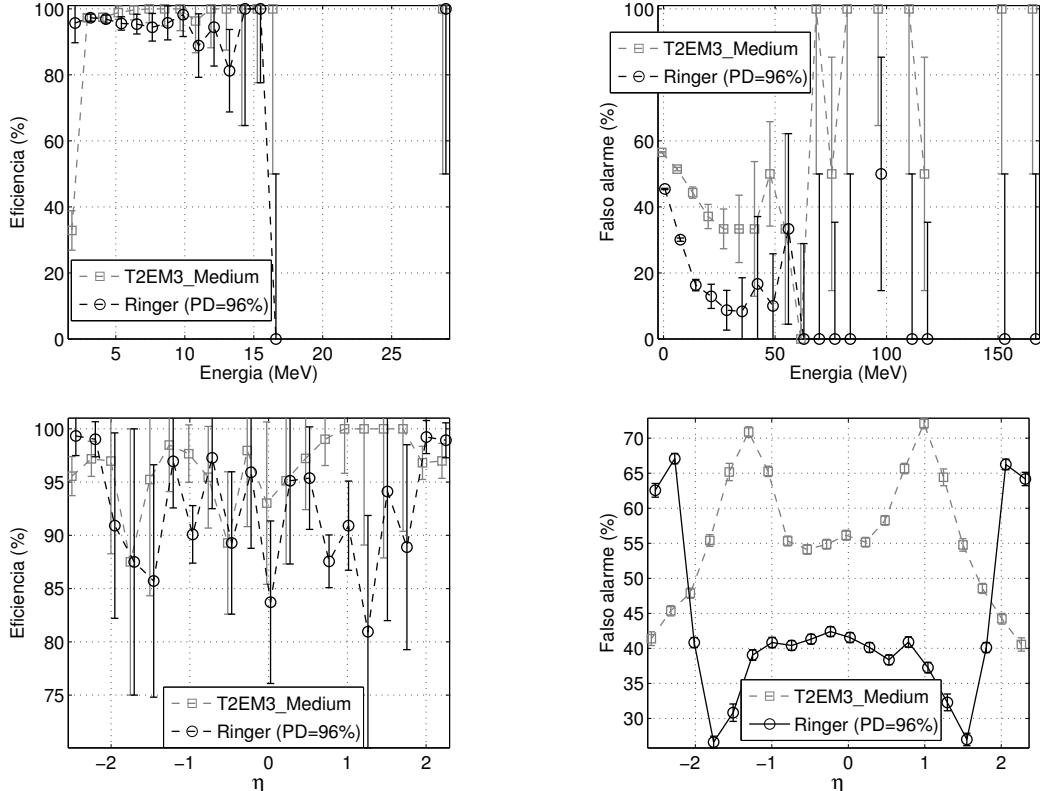


Figura 8.8: Comparativo de desempenho entre o *Neural Ringer* e os dois cortes do T2Calo.

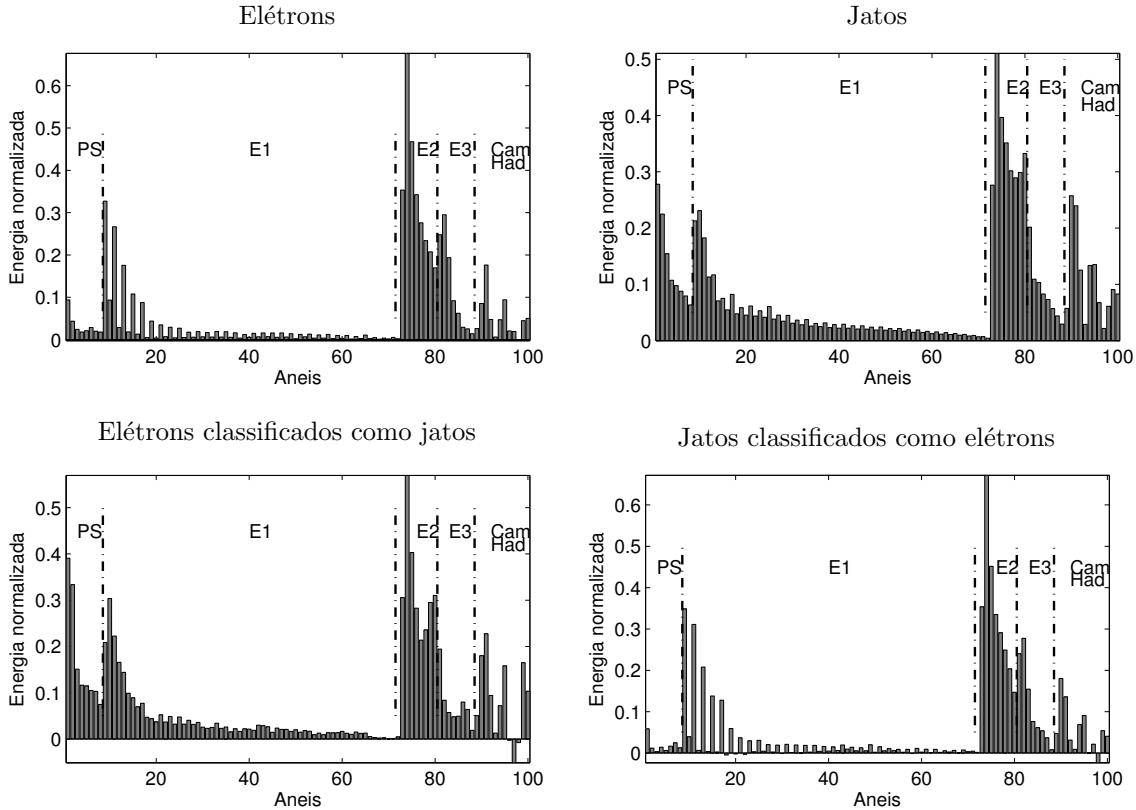


Figura 8.9: Perfil médio dos eventos classificados incorretamente pelo *Neural Ringer*. Os sinais médios de “elétrons” e “jatos” são mostrados para comparação.

tem seu desempenho menos influenciado pelos erros da seleção *offline*.

8.2.2 Resultados - ICA e NLICA

Considerando os discriminadores baseados em pré-processamento por ICA/NLICA, para os sinais experimentais, os algoritmos de extração de características foram re-treinados apenas nas configurações que obtiveram melhor desempenho para os sinais simulados (em geral, foram sempre as que utilizam, de algum modo, informação supervisão para estimar os componentes independentes). Assim, os componentes independentes foram estimados utilizando os modelos/abordagens de treinamento a seguir:

- ICA (linear e segmentado) com compactação por PCD;

sem restrições de tempo de processamento, o processo de cálculo é mais preciso e a elas são adicionados outros parâmetros de análise, totalizando aproximadamente 20 variáveis, conforme descrito anteriormente na Seção 6.5.2.

Elétrons classificados incorretamente

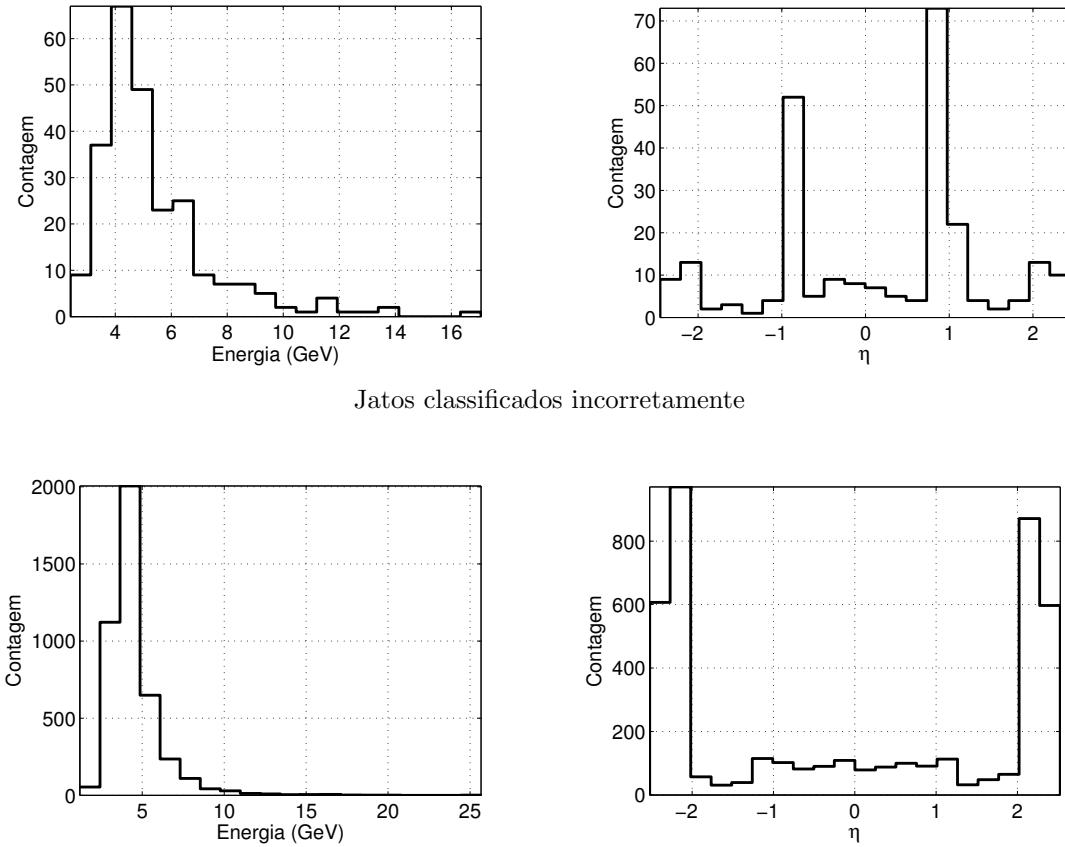


Figura 8.10: Distribuições em energia e η dos eventos classificados incorretamente pelo *Neural Ringer*.

- ICA Local;
- SOM Supervisionado (não-segmentado);
- PNL Modificado (segmentado).

Os parâmetros inerentes a cada modelo foram calculados novamente para o conjunto de sinais experimentais seguindo os mesmos procedimentos descritos nos sinais experimentais.

O modelo linear da ICA foi estimado utilizando o algoritmo JADE de modo segmentado. A compactação por PCD reduziu o número de sinais referentes a cada camada conforme mostrado na Tabela 8.3. A PCD produziu uma compactação de 70 % no número de componentes.

Para a ICA Local, os sinais foram inicialmente mapeados em dois agrupamentos através do algoritmo *k-means*. A distribuição das assinaturas de cada classe nos

Tabela 8.3: Número de componentes estimados a partir da compactação por PCD.

Layer	PS	E1	E2	E3	H0	H1	H0	Total
Rings	8	64	8	8	4	4	4	100
PCD	6	5	5	6	3	3	2	30

grupos é mostrada na Figura 8.11. A maior parcela dos elétrons ($\sim 89\%$) ficaram concentrados no primeiro grupo, juntamente com grande parte dos jatos ($\sim 72\%$), indicando que há uma grande semelhança entre grande parte das assinaturas das duas classes.

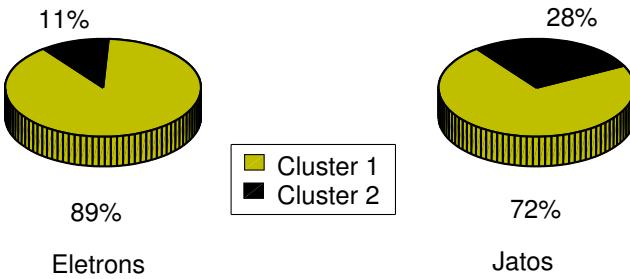


Figura 8.11: Distribuição das assinaturas de cada classe nos agrupamento.

A seguir, os componentes independentes foram estimados localmente e utilizados para alimentar os classificadores neurais. Variando-se o número de neurônios em cada rede neural (de 1 a 20), chegou-se a valores ótimos de eficiência para 11 neurônios no agrupamento 1 e 3 no 2. A seguir, foi utilizado um algoritmo genético para realizar a busca pelos patamares de corte dos classificadores locais que maximiza o SP.

Para o SOM, foi necessário repetir a busca pelo tamanho ótimo do mapa, considerando como parâmetros a serem analisados o erro de reconstrução médio e o SP. Para o mapa supervisionado não-segmentado, melhor compromisso entre os dois parâmetros foi obtido para um mapa com uma grade de neurônios retangular de 14×7 neurônios.

No modelo PNL, novamente foi identificado que o grau de não-linearidade ótimo para o problema era baixo, pois melhor eficiência de discriminação foi obtida quando as funções não-lineares foram estimadas por redes neurais de apenas dois neurônios

ocultos.

A Figura 8.12 mostra as curvas ROC obtidas a partir dos discriminadores treinados com os componentes independentes. Pode-se observar que, a ICA linear é capaz de melhorar o desempenho em comparação com o *Neural Ringer*. O pré-processamento a partir do SOM e do modelo PNL foi capaz de melhorar ligeiramente o desempenho em relação a ICA linear. Por fim, a ICA Local produziu a maior eficiência (em termos do máximo SP).

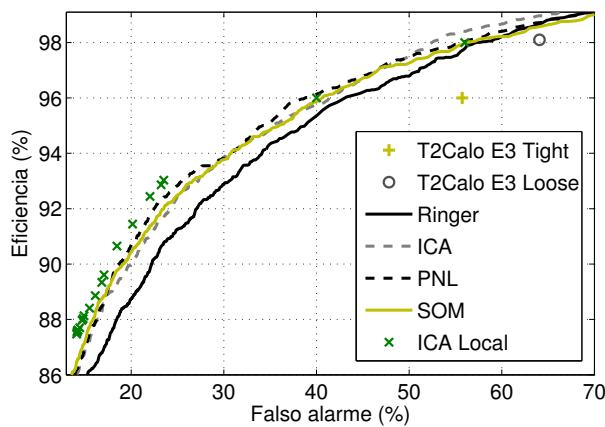


Figura 8.12: Curvas ROC para os discriminadores baseados em ICA/NLICA para os sinais experimentais.

Um resumo dos resultados obtidos com o pré-processamento por ICA/NLICA é mostrado na tabela 8.4. Os discriminadores são comparados em termos do máximo SP, das probabilidades de detecção e falso alarme que produziram o máximo SP e das probabilidades de falso alarme para diferentes valores de PD. Percebe-se que, o pré-processamento por ICA Local maximiza o SP e minimiza a probabilidade de falso alarme para PD=93 %, porém, quando a PD aumenta, o desempenho piora. Para PD=96 % e PD=98 % (correspondentes aos cortes do T2Calo), o discriminador baseado nos componentes independentes estimados pelo modelo PNL apresentou maior redução do falso alarme.

Na Figura 8.13, os discriminadores baseados na ICA linear e na ICA Local foram comparados com o *Neural Ringer* (todos ajustado para uma probabilidade de detecção de $\sim 93\%$). Pode-se observar que, o *Ringer* apresenta eficiência ligeiramente mais alta próximo a 10 GeV e na região do *crack*. Essa diferença é compensada por uma redução no falso alarme em quase todas as faixas de energia e η . No gráfico do

Tabela 8.4: Comparaçao dos resultados obtidos para diferentes discriminadores baseados no pré-processamento por ICA/NLICA considerando o máxmo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=93%, PD=96% e PD=98%.

Discriminador	SP _{Max.} $\times 100$	PD (%)	PF (%)	PF _{PD=93%}	PF _{PD=96%}	PF _{PD=98%}
T2EM3_loose	63,36	98,10	64,07	-	-	64,07
T2EM3_tight	68,01	96,30	55,73	-	55,73	-
Ringer	85,55 \pm 0,21	84,00 \pm 0,19	12,95 \pm 0,11	30,40 \pm 0,13	43,16 \pm 0,16	57,63 \pm 0,17
ICA	85,96 \pm 0,24	86,36 \pm 0,22	14,48 \pm 0,13	27,29 \pm 0,12	41,44 \pm 0,20	53,16 \pm 0,29
SOM Sup.	86,26 \pm 0,20	85,34 \pm 0,29	12,81 \pm 0,11	26,81 \pm 0,17	40,65 \pm 0,25	55,89 \pm 0,27
PNL Mod.	86,06 \pm 0,18	86,10 \pm 0,15	13,98 \pm 0,12	25,60 \pm 0,21	39,01 \pm 0,22	56,01 \pm 0,26
Local	86,67 \pm 0,27	87,61 \pm 0,30	14,22 \pm 0,17	23,52 \pm 0,10	42,02 \pm 0,26	56,01 \pm 0,29

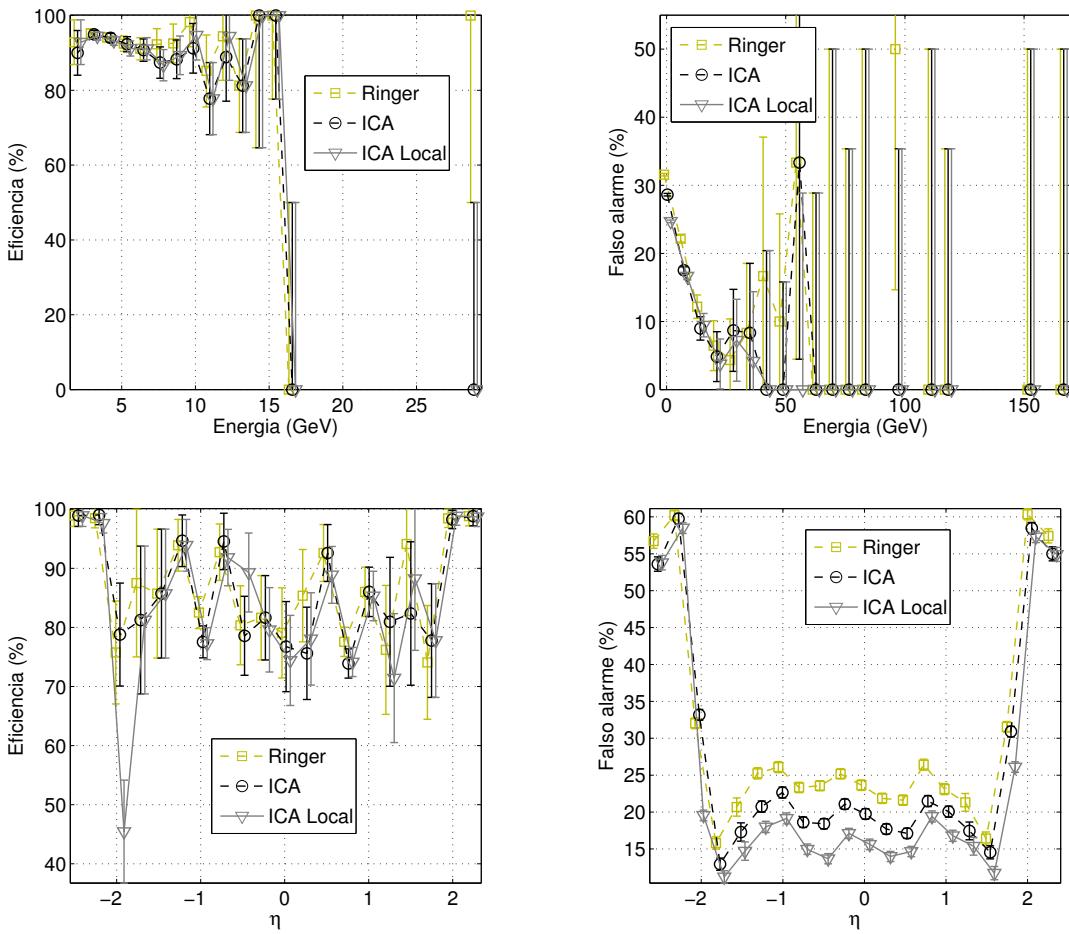


Figura 8.13: Comparativo de desempenho entre o *Neural Ringer* e os discriminadores neurais com pré-processamento por ICA (linear) e ICA Local.

falso alarme em η é possível visualizar claramente que há uma gradual redução do *Ringer* para a ICA linear e depois para a ICA Local.

8.2.3 Estudo da relevância por camada

A partir das análises realizadas nos sinais simulados, observou-se que, o modo mais eficiente para realizar o estudo da relevância por camada é treinar classificadores especialistas na informação de cada camada do calorímetro, e utilizar uma rede neural combinadora para aproveitar ao máximo toda a segmentação da informação.

Assim, este procedimento foi repetido para os sinais experimentais. Inicialmente foram treinados classificadores especialistas. A Figura 8.14 mostra os valores do máximo SP obtidos por cada discriminador especialista. Pode-se observar que, de modo semelhante ao obtido para os sinais simulados, a camada mais importante para a discriminação é a E1, seguida da E2 e PS. As camadas hadrônicas e a E3 são as menos discriminantes (este resultado é bastante semelhante ao observado nos sinais simulados).

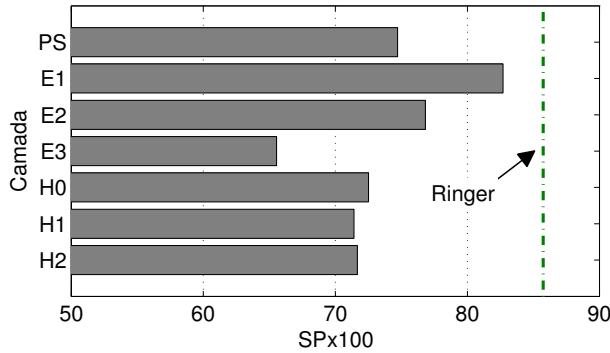


Figura 8.14: SP máximo×100 obtido por cada classificador especialista (segmentado).

Em sequencia, a rede combinadora foi treinada utilizando a informação dos classificadores especialistas. O processo de treinamento foi re-iniciado retirando-se, por camada, a informação dos classificadores especialistas. A Figura 8.15 ilustra os resultados obtidos em termos do máximo SP e das curvas ROC. Pode-se ver que, mesmo retirando a informação de cinco (das sete) camadas do calorímetro, obtém-se eficiência maior que a do *Neural Ringer*.

Os resultados obtidos a partir da utilização da rede combinadora para explorar a informação dos classificadores especialistas são comparados com os de outros discriminadores na Tabela 8.5. Pode-se observar que as duas configurações da rede combinadora são capazes de obter melhor desempenho que os outros discrimina-

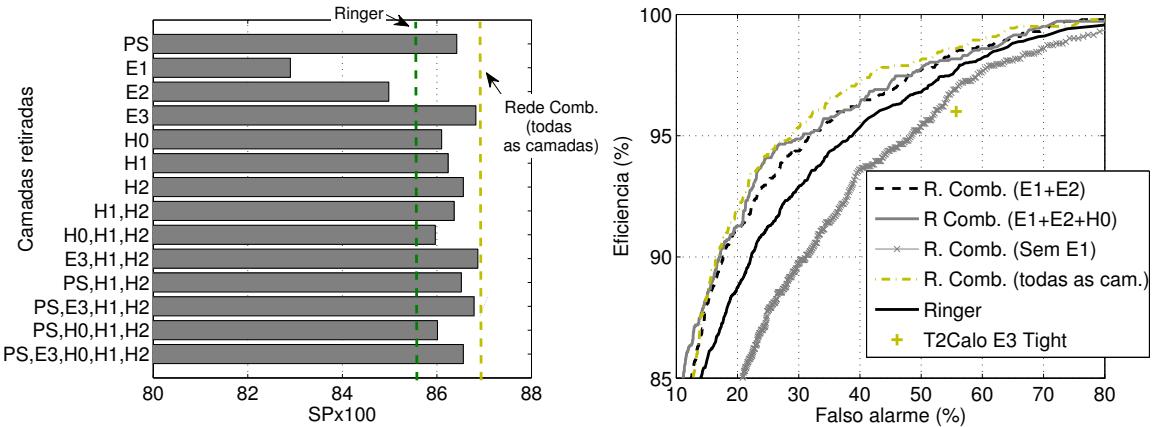


Figura 8.15: Desempenho obtido, em termos do máximo SP (esquerda) e das curvas ROC (direita), através da rede combinadora.

dores. Considerando os valores fixados para a probabilidade de detecção, a rede combinadora (utilizando a informação de todas as camadas) produz uma redução do falso alarme sempre maior que 8 pontos percentuais em relação ao *Neural Ringer*.

Tabela 8.5: Comparaçao dos resultados obtidos para diferentes discriminadores considerando o máximo SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=93%, PD=96% e PD=98%.

Discriminador	SP _{Max.} × 100	PD (%)	PF (%)	PF _{PD=93%}	PF _{PD=96%}	PF _{PD=98%}
T2EM3_loose	63,36	98,10	64,07	-	-	64,07
T2EM3_tight	68,01	96,30	55,73	-	55,73	-
Ringer	85,55 ± 0,21	84,00 ± 0,19	12,95 ± 0,11	30,40 ± 0,13	43,16 ± 0,16	57,63 ± 0,17
Rede Comb.	86,80 ± 0,23	89,44 ± 0,18	15,80 ± 0,10	21,95 ± 0,09	33,76 ± 0,21	48,76 ± 0,23
R. Comb(E1+E2)	86,26 ± 0,21	87,50 ± 0,17	14,91 ± 0,14	25,21 ± 0,12	39,48 ± 0,23	51,92 ± 0,29

Os discriminadores T2Calo (corte EM3-medium), *Neural Ringer* e Rede Combinadora (utilizando a informação de todas as camadas) são comparados na Figura 8.16 em função do desempenho em energia e η . Pode-se observar que, embora o T2Calo apresente maior eficiência para algumas regiões em η , a rede combinadora é capaz de reduzir o falso alarme para toda a faixa de energia e especialmente na região onde $|\eta| < 1$.

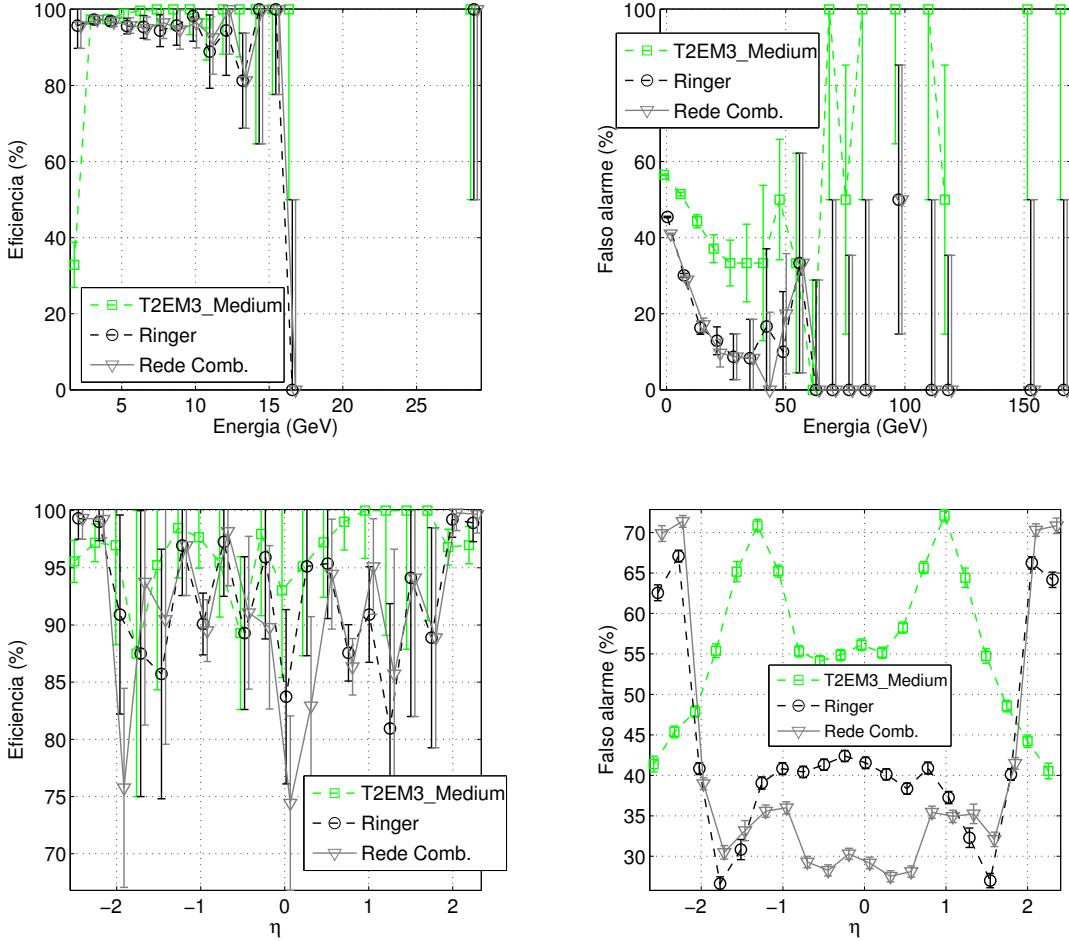


Figura 8.16: Comparativo de desempenho entre o *Neural Ringer* e os discriminadores neurais com pré-processamento por ICA (linear) e ICA Local.

8.2.4 Comentários e Discussão

Considerando o conjunto de sinais experimentais produzidos em colisões do LHC, os discriminadores propostos são comparados com outros classificadores na Figura 8.17. Percebe-se que, o pré-processamento por ICA e ICA Local é capaz de revelar características discriminantes, tornando mais simples a tarefa de classificação. É possível observar também que, a exploração de toda a segmentação inerente aos sinais dos calorímetros (através dos classificadores especialistas associados pela rede combinadora) produz a maior eficiência entre todos os discriminadores.

A Tabela 8.6 resume os resultados obtidos para o conjunto de sinais experimentais. De modo semelhante ao observado nos sinais simulados, o pré-processamento por ICA Local produz o melhor desempenho (em termos do máximo SP) entre todos os modelos de ICA/NLICA utilizados. Uma particularidade dos sinais adquiridos

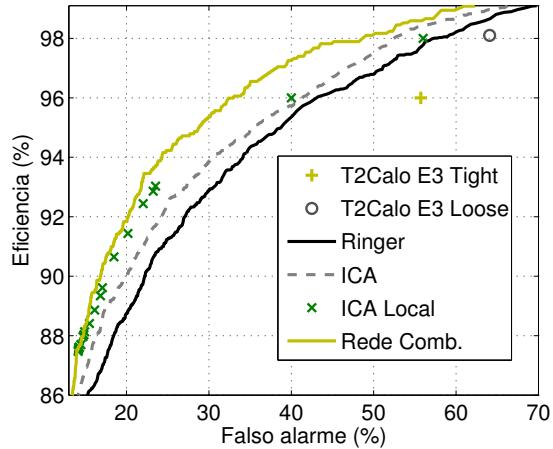


Figura 8.17: Comparativo de desempenho entre diferentes discriminadores em termos das curvas ROC.

nas colisões é que, devido possivelmente a fatores como a diferente faixa de energia (em baixa energia os perfis de elétrons e jatos apresentam mais semelhanças e associado a isso, o ruído de medição também é mais influente), e os erros inerentes da identificação *offline*, a eficiência dos métodos propostos é mais baixa do que a obtida para os sinais simulados em abordagens equivalentes.

Tabela 8.6: Comparaçāo dos resultados obtidos para diferentes discriminadores considerando o máxim SP($\times 100$), as probabilidades de decisão (PD) e falso alarme (PF) para o melhor SP e a probabilidade de falso alarme para PD=93%, PD=96% e PD=98%.

Discriminador	$SP_{Max.} \times 100$	PD (%)	PF (%)	PF _{PD=93%}	PF _{PD=96%}	PF _{PD=98%}
T2EM3_loose	63,36	98,10	64,07	-	-	64,07
T2EM3_tight	68,01	96,30	55,73	-	55,73	-
Ringer	$85,55 \pm 0,21$	$84,00 \pm 0,19$	$12,95 \pm 0,11$	$30,40 \pm 0,13$	$43,16 \pm 0,16$	$57,63 \pm 0,17$
ICA	$85,96 \pm 0,24$	$86,36 \pm 0,22$	$14,48 \pm 0,13$	$27,29 \pm 0,12$	$41,44 \pm 0,20$	$53,16 \pm 0,29$
SOM Sup.	$86,26 \pm 0,20$	$85,34 \pm 0,29$	$12,81 \pm 0,11$	$26,81 \pm 0,17$	$40,65 \pm 0,25$	$55,89 \pm 0,27$
PNL Mod.	$86,06 \pm 0,18$	$86,10 \pm 0,15$	$13,98 \pm 0,12$	$25,60 \pm 0,21$	$39,01 \pm 0,22$	$56,01 \pm 0,26$
Local	$86,67 \pm 0,27$	$87,61 \pm 0,30$	$14,22 \pm 0,17$	$23,52 \pm 0,10$	$42,02 \pm 0,26$	$56,01 \pm 0,29$
Rede Comb.	$86,80 \pm 0,23$	$89,44 \pm 0,18$	$15,80 \pm 0,10$	$21,95 \pm 0,09$	$33,76 \pm 0,21$	$48,76 \pm 0,23$
R. Comb(E1+E2)	$86,26 \pm 0,21$	$87,50 \pm 0,17$	$14,91 \pm 0,14$	$25,21 \pm 0,12$	$39,48 \pm 0,23$	$51,92 \pm 0,29$

8.3 Estimativa do custo computacional dos algoritmos propostos

Conforme detalhado na seção 4.3.4, o *Neural Ringer* está implementado na plataforma de *software* do ATLAS como uma sub-rotina do T2Calo e seu fluxo de processamento envolve as etapas a seguir:

1. **Seleção de região** (T2Calo) - são selecionados os dados utilizados pelo T2Calo ($0,4927 \pm 0,0787$) ms;
2. **Pré-processamento** (T2Calo) - as variáveis de decisão do T2Calo são calculadas ($0,1408 \pm 0,0148$) ms;
3. **Seleção de região (Ringer)** - são selecionados os dados utilizados pelo *Neural Ringer* ($0,4375 \pm 0,0996$) ms;
4. **Anelamento** - construção dos anéis ($0,0986 \pm 0,0165$) ms;
5. **Normalização** - o vetor com os anéis é normalizado ($0,0026 \pm 0,0015$) ms;
6. **Classificação Neural** - a rede neural produz a decisão de aceitação/rejeição do evento ($0,0387 \pm 0,0018$) ms.

Em conjunto, as seis etapas totalizam ($1,2109 \pm 0,1288$) ms.

Com o uso de uma etapa de pré-processamento para a classificação neural (através da NLICA), é necessário verificar o efeito desta modificação no tempo total para tomada de decisão.

As rotinas para operação da ICA Local e do modelo PNL foram implementadas no sistema de filtragem do ATLAS e seu custo computacional avaliado a partir da propagação de um conjunto de eventos através das duas cadeias de processamento. Nesta avaliação, as etapas de 1 a 5 do fluxo do *Neural Ringer* não são modificadas. Apenas na etapa 6, é adicionado o pré-processamento para a rede neural através dos algoritmos de NLICA.

No caso da ICA Local, a etapa 6 é constituída de 3 procedimentos:

1. **Seleção do cluster** - as assinaturas que chegam são divididas nos dois clusters a partir do cálculo da distância (euclidiana) para o centro de cada cluster;

2. **ICA+Rede Neural** - dentro de cada cluster, as assinaturas são projetadas nos componentes independentes e propagadas através da rede classificadora;
3. **Comparação com o patamar** - as saídas das redes são comparadas com os patamares de decisão, produzindo a classificação propriamente dita.

O tempo médio gasto em cada uma das etapas foi respectivamente $(0,0183 \pm 0,0037)$ ms, $(0,0474 \pm 0,0059)$ ms e $(0,0081 \pm 0,0045)$ ms, totalizando $(0,0738 \pm 0,0083)$.

Para os modelos PNL e SOM, dois procedimentos substituem a etapa 6:

1. **PNL-ICA+Rede Neural** - os sinais são mapeados nos componentes independentes e propagados pela rede classificadora;
2. **Comparação com o patamar** - a saída da rede é comparada com o patamar de decisão.

Para o modelo PNL, as etapas acima foram completadas respectivamente em $(0,0747 \pm 0,0074)$ ms e $(0,0067 \pm 0,0043)$ ms, totalizando $(0,0814 \pm 0,0085)$ ms.

O SOM, embora esteja diretamente relacionado com a NLICA, conforme descrito anteriormente, foi utilizado no contexto deste trabalho como uma transformação linear (projeção nas direções \mathbf{d}_i) dos sinais de entrada \mathbf{x} (composto pelos 100 anéis). Deste modo, numa implementação *online*, a matriz de projeção estimada pelo método ($\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K]^T$) pode ser incluída juntamente com os pesos da camada de entrada da rede neural. Para o classificador SOM+MLP, a entrada da rede neural é definida por:

$$\mathbf{z} = \mathbf{D}\mathbf{x}. \quad (8.1)$$

Este sinal é projetado na matriz de pesos \mathbf{W} da camada de entrada, gerando o sinal \mathbf{WDx} que é utilizado como entrada para os neurônios da primeira camada. Para operação do classificador pode-se definir uma nova rede que utiliza $\mathbf{WD} = \mathbf{W} \cdot \mathbf{D}$ como matriz de pesos da camada de entrada. Deste modo, o discriminador SOM+MLP pode ser simplificado para apenas uma rede MLP, que tem como entrada os sinais em anéis \mathbf{x} . Considerando que este discriminador tem o mesmo número de neurônios ocultos que o *Neural Ringer*, o custo computacional das duas abordagens é aproximadamente o mesmo.

A Tabela 8.7 resume os resultados mostrados acima, incorporando também o tempo total dos discriminadores e uma comparação (percentual) com o tempo gasto pelo *Neural Ringer*. Observa-se que, o uso do pré-processamento por NLICA contribui pouco para o aumento do custo computacional, implicando em apenas 2,90% e 3,53% de aumento em relação ao *Neural Ringer*, respectivamente para a ICA Local e a PNL-ICA.

Tabela 8.7: Comparaçāo dos tempos de processamento (em ms) para diferentes discriminadores. O tempo gasto na transformaçāo através da NLICA, quando for o caso, estā incluído no tempo do classificador neural.

Discriminador	Pre-process.	Class. (e NLICA)	Total	% do Ringer
T2Calo	$0,6335 \pm 0,0801$	$0,0134 \pm 0,0013$	$0,6469 \pm 0,0802$	53,42
<i>Neural Ringer</i>	$1,1722 \pm 0,1287$	$0,0387 \pm 0,0018$	$1,2109 \pm 0,1288$	100,00
SOM	$1,1722 \pm 0,1287$	$\sim 0,0387 \pm 0,0018$	$\sim 1,2109 \pm 0,1288$	$\sim 100,00$
ICA Local	$1,1722 \pm 0,1287$	$0,0738 \pm 0,0083$	$1,2460 \pm 0,1520$	102,90
PNL	$1,1722 \pm 0,1287$	$0,0814 \pm 0,0085$	$1,2536 \pm 0,1526$	103,53

Considerando o tempo total gasto pelo discriminador PNL+MLP (pior caso para as abordagens baseadas em NLICA), um incremento de aproximadamente 94 % é produzido em comparaçāo ao T2Calo. Considerando a cadeia completa de processamento para identificaçāo de elétrons no L2 do ATLAS (que pode incluir, além da calorimetria, a análise da trajetória da partícula no detector de traços), o tempo total gasto para a decisão é $7,88 \pm 2,04$ ms. Este valor corresponde a apenas ~ 20 % do tempo médio por evento especificado para o L2 (que é de 40 ms).

8.3.1 Abordagens segmentadas

Considerando que, em todos os conjuntos de dados utilizados, foram obtidos bons resultados realizando os processos de extraçāo de características e classificação de modo segmentado, é importante avaliar o impacto desta abordagem no tempo de processamento.

Na Tabela 8.8 são mostrados os tempos gastos nos processos de seleção da informação da região de interesse, produção dos sinais em anéis (anelamento) e normalização para cada camada do calorímetro. Pode-se observar que, a primeira camada eletromagnética é a que demanda mais tempo de processamento (por apresentar

maior quantidade de sensores), seguida das camadas hadrônicas (para as camadas hadrônicas o processo é um pouco mais demorado pois o calorímetro hadrônico é segmentado em um maior número de módulos e, dependendo da posição de interação, a RoI pode conter informação de vários destes módulos).

A Figura 8.18 mostra que as etapas de seleção da região de interesse, anelamento e normalização, em conjunto, são responsáveis por aproximadamente 44 % do tempo total de processamento. Pode-se observar também, a distribuição do tempo gasto nas etapas listadas acima entre as sete camadas do calorímetro.

Uma característica particular das camadas hadrônicas é que as células do barril somente podem ser selecionadas de uma única vez para todas as três camadas (o tempo de seleção é $0,1132 \pm 0,0773$). Ou seja, se apenas uma camada for utilizada, é necessário solicitar os dados de todas as camadas, porém, na utilização de mais de uma camada este tempo é contado apenas uma vez. Tanto na Tabela 8.8 como na Figura 8.18 o tempo de solicitação das informações do barril foi contabilizado para todas as camadas hadrônicas, pois o objetivo é comparar os tempos gastos no processamento em cada segmento.

Tabela 8.8: Tempo gasto para seleção da região de interesse e construção dos sinais em anéis para cada camada do calorímetro, os valores percentuais são referentes apenas ao total destas duas etapas.

Camada	Seleção da RoI	Anelamento	Total da camada	%do total
PS	$0,0071 \pm 0,0018$	$0,0091 \pm 0,0044$	$0,0162 \pm 0,0047$	2,2
E1	$0,1264 \pm 0,0309$	$0,0311 \pm 0,0077$	$0,1575 \pm 0,0318$	21,8
E2	$0,0794 \pm 0,0091$	$0,0203 \pm 0,0021$	$0,0997 \pm 0,0093$	13,8
E3	$0,0425 \pm 0,0056$	$0,0111 \pm 0,0018$	$0,0536 \pm 0,0059$	7,4
H0	$0,1230 \pm 0,0775$	$0,0080 \pm 0,0029$	$0,1309 \pm 0,0776$	18,1
H1	$0,1214 \pm 0,0775$	$0,0071 \pm 0,0027$	$0,1285 \pm 0,0776$	17,8
H2	$0,1292 \pm 0,0778$	$0,0071 \pm 0,0018$	$0,1364 \pm 0,0778$	18,9

Considerando o uso dos classificadores segmentados, foi mostrado que é possível obter eficiência de discriminação equivalente a do *Neural Ringer* (em alguns casos superior), mesmo sem utilizar a informação de algumas das camadas. A maior redução que ainda mantém bom desempenho para todos os conjuntos de dados utilizados é obtida quando são mantidas apenas as informações das camadas E1

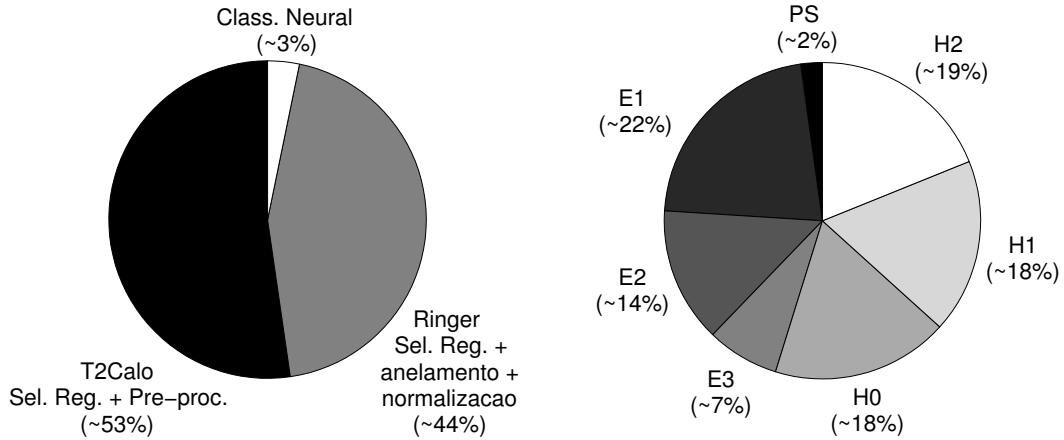


Figura 8.18: Parcela do tempo total gasta em cada etapa de processamento do *Neural Ringer* (esquerda) e parcela do tempo específico das etapas de seleção de região, anelamento e normalização gasto para cada camada do calorímetro.

e E2. A Tabela 8.9 compara o *Neural Ringer* com a rede combinadora operando apenas sobre as informações das camadas E1 e E2 em termos do desempenho nos diversos conjuntos apresentados e do tempo de processamento. É importante notar que, para um desempenho equivalente, a rede combinadora necessita de apenas 80,2 % do tempo de processamento utilizado pelo *Ringer*.

Tabela 8.9: Comparativo entre o *Neural Ringer* e a rede combinadora que utiliza apenas as informações das camadas E1 e E2.

Discriminador	Sim. E10 SP×100	Sim. E15i SP×100	Exp. E3 SP×100	Cósmicos Rej.(%)	Tempo proces. (ms)
Ringer	$97,99 \pm 0,18$	$94,94 \pm 0,12$	$85,55 \pm 0,21$	99,56	$1,2109 \pm 0,1288$
R. Comb _(E1+E2)	$98,08 \pm 0,26$	$94,84 \pm 0,23$	$86,26 \pm 0,21$	99,03	$0,9717 \pm 0,0867$

Capítulo 9

Conclusões

A física de partículas ou física de altas energias (*HEP-High-Energy Physics*) se propõe a estudar as características fundamentais da matéria. Para isso, ao longo dos anos, vêm sendo construídos experimentos, cada vez mais complexos, que envolvem profissionais de diversas áreas como engenharia, computação, matemática e física nas etapas de projeto, montagem e operação.

O conhecimento adquirido em experimentos de HEP está bastante difundido na sociedade em aplicações como a rede mundial de computadores (*WWW-World-Wide Web*), que foi idealizada inicialmente para troca de informação entre os diversos institutos participantes da colaboração do CERN ao redor do mundo, e na medicina, em tratamentos como a rádio-terapia e exames tipo a cintilografia e o PET (*positron emission tomography* ou tomografia por emissão de pósitrons), que produz uma imagem em três dimensões das características funcionais do corpo humano.

O Grande Colisionador Hadrônico (*LHC-Large Hadron Collider*), que produziu suas primeiras colisões no final de 2009, no CERN, iniciou uma nova era de descobertas, estabelecendo recordes em energia e luminosidade. Diversos avanços tecnológicos foram necessários desde o projeto até a conclusão da montagem do LHC e seus diversos detectores.

Para lidar com a enorme quantidade de informação produzida nas colisões do LHC, os detectores são dotados de um sistema de filtragem (*trigger*) *online*, que deve apresentar, entre outras características:

- Alta eficiência na detecção da física de interesse;
- Baixa aceitação do ruído de fundo;

- Rapidez na tomada de decisão.

O bom funcionamento do sistema de filtragem é fundamental para o desempenho do detector, porém, atender a todas as demandas inerentes à aplicação é uma tarefa de difícil solução. Em geral, os sistemas de filtragem são divididos em estágios (níveis) sequenciais de seleção, onde o ruído de fundo (informação não relevante) é gradualmente reduzido.

No detector ATLAS (o maior detector de propósito geral do LHC), o sistema de filtragem é sub-dividido em três níveis. O discriminador padrão utilizado no experimento para seleção de elétrons no segundo nível de filtragem (T2Calo) opera através de cortes lineares em parâmetros calculados a partir de informações do perfil de deposição de energia medido nos calorímetros. Os elétrons são extremamente importantes para o experimento, pois estão presentes em decaimentos de diversos fenômenos físicos de interesse.

Num trabalho anterior [9], foi proposto um classificador alternativo ao T2Calo que combina o pré-processamento da informação medida nos calorímetros em anéis concêntricos com a utilização de classificadores neurais supervisionados (arquitetura Perceptron de Múltiplas Camadas). Este discriminador (*Neural Ringer*) está implementado na plataforma de *software* do detector e opera em paralelo com o T2Calo, apresentando maior eficiência de discriminação. Em [10] foi proposta uma extensão ao *Neural Ringer* incorporando uma etapa de pré-processamento linear (através da análise de componentes independentes - ICA) ao classificador neural. Foi observado que a extração de características através da ICA produziu um aumento na eficiência de discriminação em comparação ao discriminador operando diretamente sobre os sinais em anéis.

Neste contexto, este trabalho propôs a utilização do modelo não-linear da análise de componentes independentes (NLICA) para a extração de características dos sinais em anéis, como uma etapa de pré-processamento para o discriminador neural. A estimativa da NLICA foi realizada a partir de três estratégias distintas: pelo modelo pós não-linear (PNL), por mapas auto-organizáveis (para estimar um modelo não-linear sem restrições estruturais) e pela ICA local. Com os algoritmos utilizados, foi possível obter maior eficiência de discriminação para regiões do detector que apresentam menor número de células sensoras (*cracks* e interconexões dos módulos

do calorímetro) e para eventos de menor energia (que apresentam menor relação sinal-interferência).

Comparando os três métodos, observou-se que o pré-processamento por ICA local apresentou maior eficiência (em termos do máximo SP) para os dois conjuntos simulados utilizados e também para os sinais experimentais. O uso do pré-processamento por mapas auto-organizáveis possibilita, além de um discriminador *online* com alta eficiência, a visualização dos eventos em gráficos de duas ou três dimensões. Foi mostrado também que, o uso da informação das classes na estimação dos componentes independentes produziu, em todos os casos, aumento na eficiência de discriminação em comparação com as versões não-supervisionadas.

Considerando o tempo de processamento, percebeu-se, que a adição do pré-processamento por NLICA não contribui significativamente para o aumento no custo computacional. Na verdade, a maior parte do tempo é gasta para solicitação da informação e formação dos sinais em anéis, sendo que, estas duas etapas são realizadas em bloco, por cada camada do calorímetro.

Numa tentativa de explorar ao máximo a segmentação disponível nos calorímetros do ATLAS, foram treinados classificadores especialistas nas informações de cada camada. Deste modo, foi possível verificar a relevância dos diversos segmentos na identificação de elétrons. Para combinar a informação dos especialistas foram empregados diversos métodos de combinação de classificadores como média, votação e uma rede neural combinadora. Observou-se que, a rede combinadora (que na operação *online* pode ser implementada através de uma rede neural de três camadas com a camada de entrada não completamente conectada) apresentou maior eficiência entre os métodos de combinação de classificadores utilizados.

Através da rede combinadora, verificou-se que a retirada de 5 das 7 camadas influencia pouco na eficiência de discriminação (em comparação com o *Neural Ringer*). Por outro lado, com o uso apenas desta parcela da informação (referente apenas às camadas E1 e E2) é possível reduzir o tempo total de processamento em aproximadamente 20 %.

9.1 Trabalhos Futuros

À medida que o conhecimento a respeito do detector e dos fenômenos físicos envolvidos aumenta, os algoritmos de filtragem evoluem continuamente, incorporando novas informações. Recentemente, o discriminador padrão do ATLAS para o canal elétron/jato (T2Calo) foi modificado, e agora opera com patamares de corte variáveis em função de η .

Esta estratégia tem como objetivo melhorar a eficiência em áreas como o *crack* e as interconexões entre módulos, pois com a menor quantidade de sensores, as características dos eventos destas regiões são modificadas e os cortes utilizados no restante do detector não produzem bom desempenho de discriminação.

Considerando o bom desempenho obtido com o pré-processamento por ICA/NLICA, outra opção seria a utilização de algoritmos de extração de características que estão diretamente relacionados com a análise de componentes independentes como a fatorização de matrizes não-negativas (*non-negative matrix factorization*) [210] ou a análise de componentes esparsos (*sparse component analysis*) [211, 212]. Para o uso da NMF, o procedimento de normalização precisa ser modificado para produzir sempre sinais não-negativos. A aplicação da SCA é mais direta, pois, em cada assinatura, apenas uma pequena parte dos 100 anéis tem considerável nível de energia, caracterizando a esparsidade dos sinais.

Pode-se ainda testar outras arquiteturas de classificadores neurais como as funções de base radial (*radial basis functions*) e máquinas de vetor de suporte (*support vector machine*) [5], visando otimizar o processo de classificação. Sendo que, neste caso, seria necessário um trabalho adicional para a implementação do discriminador no sistema de filtragem do detector.

Embora os discriminadores propostos neste trabalho tenham sido desenvolvidos para o sistema de filtragem *online*, os altos índices de eficiência obtidos indicam que é possível também pensar numa aplicação à filtragem *offline*, especialmente para os discriminadores baseados no pré-processamento por mapas auto-organizáveis, devido às ferramentas de visualização inerentes ao método.

Referências Bibliográficas

- [1] ATLAS COLABORATION, “ATLAS Experiment at CERN Large Hadron Collider”, *Journal of Instrumentation*, v. 3, n. S08003, pp. 1–407, 2008.
- [2] EVANS, L., BRYANT, P., “LHC Machine”, *Journal of Instrumentation*, v. 3, n. S08001, pp. 1–158, 2008.
- [3] JOLLIFFE, I. T., *Principal Component Analysis*. 2nd ed. Springer: Nova York, Estados Unidos, 2002.
- [4] HYVARINEN, A., KARHUNEN, J., OJA, E., *Independent Component Analysis*. Wiley: Nova York, Estados Unidos, 2001.
- [5] HAYKIN, S., *Neural Networks and Learning Machines*. 3rd ed. Prentice Hall: Nova Jersey, Estados Unidos, 2008.
- [6] COMMITTEE ON ELEMENTARY PARTICLE PHYSICS IN THE 21ST CENTURY, N. R. C., *Revealing the Hidden Nature of Space and Time: Charting the Course for Elementary Particle Physics*. National Academic Press: Washington-DC, Estados Unidos, 2006.
- [7] ATLAS-COLLABORATION, *Letter of Intents for a General Purpose pp Experiment at the Large Hadron Collider at CERN*, Tech. rep., CERN, Genebra, Suíça, Outubro 1992.
- [8] SEIXAS, J. M., L. CALÔBA, M. S., “Neural Second-Level trigger system based on calorimetry”, *Computer Physics Communications*, v. 95, n. 2-3, pp. 143–157, Junho 1996.
- [9] DOS ANJOS, A. R., *Sistema de filtragem online aplicada a um ambiente com alta taxa de eventos*, Ph.D. Thesis, COPPE/UFRJ, Dezembro 2006.

- [10] TORRES, R. C., *Sistema Online de Filtragem em um Ambiente com Alta Taxa de Eventos e Fina Granularidade*, Ph.D. Thesis, COPPE/UFRJ, Março 2010.
- [11] CALOBA, L., SEIXAS, J., PEREIRA, F., “Neural Discriminating Analysis for a Second-Level Trigger System”. In: *Proceedings of the International Conference on Computing in High Energy Physics (CHEP95)*, Rio de Janeiro, Brasil, Setembro 1995.
- [12] WIGMANS, R., *Calorimetry: Energy Measurement in Particle Physics*. Clarendon Press: Gloucestershire, Reino Unido, 2000.
- [13] ALMEIDA, L. B., *Nonlinear Source Separation*. Morgan and Claypool, 2006.
- [14] ROBERT, C., CASELLA, G., *Monte Carlo Statistical Methods*. Springer-Verlag: Nova York, Estados Unidos, 2004.
- [15] ATLAS-COLLABORATION, “Experiment Overview”, Disponível em :<<http://www.atlas.ch/>>. Acesso em: 27 de agosto de 2010.
- [16] FERNOW, R. C., *Introduction to Experimental Particle Physics*. Cambridge University Press: Cambridge-Reino Unido, 1986.
- [17] MARTIN, B. R., *Nuclear and Particle Physics, An Introduction*. Wiley: West Sussex, Reino Unido, 2006.
- [18] CHUNG, K. C., *Introdução à Física Nuclear*. EdUERJ: Rio de Janeiro, 2001.
- [19] VELTMAN, M. J. G., *Facts and Mysteries in Elementary Particle Physics*. World Scientific: Nova Jersey, Estados Unidos, 2003.
- [20] ELLIS, R. K., STIRLING, W. J., WEBBER, B. R., *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press: Cambridge, Reino Unido, 2003.
- [21] PRICE, M., “The LHC project”, *Nuclear Instruments and Methods in Physics Research A*, v. 478, pp. 46–61, 2002.

- [22] CERN-COLABORATION, “CERN”, Disponível em :<<http://www.cern.ch/>>. Acesso em: 20 de março de 2007.
- [23] BERTONE, G., HOOPER, D., SILK, J., “Particle dark matter, evidence, candidates and constraints”, *Physics Reports*, v. 405, pp. 279–390, 2005.
- [24] LYKKEN, J. D., “Introduction to Supersymmetry”, Lectures at the Theoretical Advanced Study Institute (TASI 96):Fields, Strings, and Duality, Junho 1996.
- [25] BIGI, I., SANDA, A., *CP violation*. Cambridge University Press: Cambridge, Reino Unido, 1999.
- [26] DESY, “Deutsches Elektronen-SynchrotronDie”, Disponível em :<<http://www.desy.de/>>. Acesso em: 02 de julho de 2009.
- [27] KEK, “KEK High-Energy Accelerator”, Disponível em: <<http://www.kek.jp/>>. Acesso em: 02 de julho de 2009.
- [28] FERMILAB, “Fermi National Accelerator Laboratory”, Disponível em :<<http://www.fnal.gov/>>. Acesso em: 02 de julho de 2009.
- [29] SLAC, “SLAC Laboratory”, Disponível em: <<http://www.slac.stanford.edu/>>. Acesso em: 02 de julho de 2009.
- [30] BROOKHAVEN, “Brookhaven National Laboratory”, Disponível em: <<http://www.bnl.gov/>>. Acesso em: 02 de julho de 2009, 2009.
- [31] LHC-COLLABORATION, “Large Hadron Collider”, Disponível em: <<http://lhcb.web.cern.ch/lhc/>>. Acesso em: 28 de março de 2007.
- [32] CMS COLABORATION, “CMS Experiment at the CERN LHC”, *Journal of Instrumentation*, v. 3, n. S08004, pp. 1–334, 2008.
- [33] LHCb COLABORATION, “LHCb Detector at the LHC”, *Journal of Instrumentation*, v. 3, n. S08005, pp. 1–205, 2008.
- [34] LHCf COLABORATION, “LHCf Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation*, v. 3, n. S08006, pp. 1–36, 2008.

- [35] ALICE COLABORATION, “ALICE Experiment at the CERN LHC”, *Journal of Instrumentation*, v. 3, n. S08002, pp. 1–245, 2008.
- [36] TOTEM COLABORATION, “TOTEM Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation*, v. 3, n. S08007, pp. 1–107, 2008.
- [37] ROSSI, L., “The ATLAS tracking and vertexing detector”, *Nuclear Instruments and Methods in Physics Research Section A*, v. 580, n. 2, 2007.
- [38] PUZO, P., “ATLAS Calorimetry”, *Nuclear Instruments and Methods in Physics Research A*, v. 494, n. 1, pp. 340–345, 2002.
- [39] PONTECORVO, L., “The ATLAS muon spectrometer”, *The European Physical Journal C - Particles and Fields*, v. 34, n. 1, 2004.
- [40] PADILLA, C., “The ATLAS Trigger System”. In: *Proceedings of the 16th IEEE-NPSS Real Time Conference*, v. 1, pp. 326–333, Pequim, China, 2009.
- [41] DREES, J., “Review of final LEP results or a tribute to LEP”. In: *Proceedings of the International Symposium on Lepton and Photon Interactions at High Energies*, pp. 349–373, Roma, Itália, Julho 2001.
- [42] KANE, G., PIERCE, A., (eds), *Perspectives on LHC Physics*. World Scientific: Singapura, 2008.
- [43] ATLAS-COLLABORATION, *ATLAS Detector and Physics Performance Technical Design Report, Volume 2*, Tech. rep., CERN, Genebra, Suiça, Maio 1999.
- [44] HAUSER, R., “The ATLAS High Level Trigger System”. In: *Proceedings of the U.S. Atlas Computing Meeting*, v. 1, pp. 1–5, Upton, EUA, 2003.
- [45] ATLAS-COLLABORATION, *ATLAS Detector and Physics Performance Technical Design Report, Volume 1*, Tech. rep., CERN, Genebra, Suiça, Maio 1999.

- [46]ADRAGNA, ET AL., P., “The ATLAS Hadronic Tile Calorimeter: From Construction Toward Physics”, *IEEE Transactions on Nuclear Science*, v. 53, n. 5, pp. 1275–1281, Junho 2006.
- [47]WATTS, G., “Review of Triggering”. In: *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference*, v. 1, pp. 282–287, Portland, EUA, Outubro 2003.
- [48]VOLKER LINDENSTRUTH, I. K., “Overview of trigger systems”, *Nuclear Instruments and Methods in Physics Research A*, v. 535, n. 1, pp. 48–56, 2004.
- [49]DA COSTA, C., *Projeto de Circuitos Digitais com FPGA*. Érica, 2009.
- [50]DINIZ, P. S. R., DA SILVA, E. A. B., NETO, S. L., *Processamento Digital de Sinais*. Bookman, 2004.
- [51]SMITH, M. J. S., *Application Specific Integrated Circuits*. Addison-Wesley Professional: Nova York, Estados Unidos, 1997.
- [52]CDF-COLLABORATION, “The CDF Experiment”, Disponível em: <<http://www-cdf.fnal.gov/>>. Acesso em: 21 de abril de 2010.
- [53]D0-COLLABORATION, “The D0 Experiment”, Disponível em: <<http://www-cdf.fnal.gov/>>. Acesso em: 21 de abril de 2010.
- [54]RIU, I., ABOLINS, M., ET. AL, P. A., “Integration of the Trigger and Data Acquisition Systems in ATLAS”, *IEEE Transactions on Nuclear Science*, v. 55, n. 1, pp. 106–112, 2008.
- [55]SEEZ, C., “The CMS trigger system”, *The European Physical Journal C - Particles and Fields*, v. 34, n. 1, pp. 151–159, 2004.
- [56]KIESLING, C., JANAUSCHEK, L., PLACAKYTE, R., et al., “The H1 neural network trigger”. In: *Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing Conference*, v. 1, pp. 319–324, Melbourne, Austrália, 2004.

- [57] LINDSEY, C., DENBY, B., HAGGERTY, H., et al., “Real Time Track Finding with a VLSI Neural Network”, *Nuclear Instruments and Methods in Physics Research A*, v. A, n. 317, pp. 346–356, 1992.
- [58] LINDSEY, C. S., DENBY, B., HAGGERTY, H., “Drift chamber tracking with neural networks”, *IEEE Transactions on Nuclear Science*, v. 40, n. 4, pp. 607–614, 2002.
- [59] CUTTS, D., HOFTUN, J. S., SORNBORGER, A., “The use of neural networks in the D0 data aquisition system”, *IEEE Transaction on Nuclear Science*, v. 36, n. 5, pp. 1490–1493, 1989.
- [60] DENBY, B., “Neural Networks in high-energy physics, a ten year perspective”, *Computer Physics Communications*, v. 1, n. 119, pp. 219–231, 1999.
- [61] NGUYET, T. T., “Searches for new physics by the H1 experiment at HERA”, *Acta Physica Polonica B Proceedings Supplement*, v. 1, n. 2, pp. 407–409, 2008.
- [62] ZIMMERMANN, J., KIESLING, C., “Neural networks for the H1 experiment”. In: *Proceedings of the IEEE Nuclear Science Symposium*, v. 3, pp. 1869–1872, Roma, Itália, 2004.
- [63] KOLANOSKI, H., “Application of artificial neural networks in particle physics”, *Nuclear Instruments and Methods in Physics Research A*, v. 1, n. 367, pp. 14–20, 1995.
- [64] SPENCER, J. E., “Real-Time Applications of Neural Nets”, *IEEE Transactions on Nuclear Science*, v. 36, n. 5, pp. 1485–1489, 1989.
- [65] DENBY, B., GARDA, P., GRANADO, B., et al., “Fast triggering in high-energy physics experiments using hardware neural networks”, *IEEE Transactions on Neural Networks*, v. 14, n. 5, pp. 1010–1027, 2003.
- [66] WON, E., “A hardware implementation of artificial neural networks using field programmable gate arrays”, *Nuclear Instruments and Metohds in Physics Research A*, v. 581, n. 1, pp. 816–820, 2007.

- [67] WESTERHOFF, S., MEYER, H., “Neural nets as a tool for separating gamma and hadron induced Air Showers”. In: *Proceedings of the AIHENP*, v. 1, n. 1, pp. 1–8, Pisa, Itália, 1995.
- [68] DOS ANJOS, A., TORRES, R., SEIXAS, J., “Neural triggering system operating on high resolution calorimetry information”, *Nuclear Instruments and Methods in Physics Research A*, v. 559, n. 1, pp. 134–138, 2006.
- [69] PRORIOL, J., “Selection of variables for neural network analysis. Comparison of several methods with high energy physics analysis”, *Nuclear Instruments and Methods in Research A*, v. 361, n. 3, pp. 581–585, 1995.
- [70] WHITESON, S., WHITESON, D., “Machine learning for event selection in high energy physics”, *Engineering Applications of Artificial Intelligence*, v. 22, n. 8, pp. 1203–1217, 2009.
- [71] RIGGI, S., CARUSO, R., INSOLIA, A., et al., “A neural network approach to event-by-event cosmic ray primary mass identification”, *Proceedings of Science*, v. ACAT07, n. 035, pp. 1–13, 2007.
- [72] MARONE, A., ANS F. CADINI, C. F., ZIO, E., et al., “Employing Neural Networks to Determine the Position of Interaction of Medium-High Energy Gamma Rays”. In: *Proceedings of the IEEE Nuclear Science Symposium*, v. 1, pp. 645–649, Orlando, Estados Unidos, 2009.
- [73] CHOULAKIAN, V., “The optimality of the centroid method”, *Psychometrika Journal*, v. 68, n. 3, pp. 473–475, 2006.
- [74] ALDRICH, J., “R. A. Fisher and the making of maximum likelihood 1912–1922”, *Statistical Science*, v. 12, n. 3, pp. 162–176, 1997.
- [75] ATLAS-COLLABORATION, *First-Level Trigger Technical Design Report*, Tech. rep., CERN, Genebra, Suiça, Junho 1998.
- [76] ATLAS-COLLABORATION, *ATLAS High-Level Trigger, Data Acquisition and Controls Technical Design Report*, Tech. rep., CERN, Genebra, Suiça, Maio 2003.

- [77] CERQUEIRA, A. S., CALÔBA, L. P., SEIXAS, J. M., “Analog System for Building the First-Level Triggering Signal Provided By the Hadronic Calorimeter of ATLAS Detector”, *Nuclear Instruments and Methods in Physics Research - A*, v. 570, n. 1, pp. 117–125, 2007.
- [78] GARVEY, J., HILLIER, S., MAHOUT, G., et al., “Use of an FPGA to identify electromagnetic clusters and isolated hadrons in the ATLAS level-1 calorimeter trigger”, *Nuclear Instruments and Methods in Physics Research A*, v. 512, n. 1, pp. 400–432, 2003.
- [79] DOS ANJOS, A., ABOLINS, M., ARMSTRONG, S., “The Second Level Trigger of the ATLAS Experiment at CERN LHC”, *IEEE Transactions on Nuclear Science*, v. 51, n. 3, pp. 909–914, 2004.
- [80] LANE, W. G., CUNNINGHAM, D., *Gigabit Ethernet Networking*. Sams: Nova York, Estados Unidos, 1999.
- [81] SHILDT, H., *C++: The Complete Reference*. McGraw-Hill: Nova York, Estados Unidos, 2003.
- [82] ATLAS-COLLABORATION, *Athena Developer Guide*, Tech. rep., CERN, Genebra, Suiça, Agosto 2001.
- [83] CORTI, G., CATTANEO, M., CHARPENTIER, M., “Software for the LHCb Experiment”, *IEEE Transactions on Nuclear Science*, v. 53, n. 3, pp. 1323–1328, 2006.
- [84] SCHIAVI, C., “Implementation and Performance of the High Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC”, *IEEE Transactions on Nuclear Science*, v. 53, n. 5, pp. 1424–1429, 2006.
- [85] MELLO, A. G., DOS ANJOS, A., ARMSTRONG, S., et al., “Overview of the High-Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC”, *IEEE Transactions on Nuclear Science*, v. 53, n. 5, pp. 2839–2843, 2006.
- [86] BISHOP, C. M., *Neural Networks for Pattern Recognition*. Clarendon Press: Gloucestershire, Reino Unido, 1995.

- [87] COVER, T. M., THOMAS, J. A., *Elements of Information Theory*. Wiley: Nova Jersey, Estados Unidos, 2006.
- [88] LIMA JR., H., SEIXAS, J., “A segmented principal component analysis applied to calorimetry information at ATLAS”, *Nuclear Instruments and Methods in Physics Research A*, v. 559, n. 1, pp. 129–133, Abril 2006.
- [89] SIMAS FILHO, E., SEIXAS, J. M., CALOBA, L. P., “Nonlinear Independent Component Analysis, Theoretical Review and Applications”, *Learning and Nonlinear Models*, v. 5, n. 2, pp. 99–120, 2007.
- [90] HARMAN, H. H., *Modern Factor Analysis*. 2nd ed. University of Chicago Press: Chicago, Estados Unidos, 1967.
- [91] PARK, H. M., OH, S. H., LEE, S. Y., “Adaptive noise cancelling based on independent component analysis”, *Electronics Letters*, v. 38, n. 15, pp. 832–833, 2002.
- [92] MOURA, N. N., FILHO, E. F. S., SEIXAS, J. M., “Advances in Sonar Technology”, chap. Independent Component Analysis for Passive Sonar Signal Processing, pp. 91–110, In-Tech, 2009.
- [93] SARPERI, L., ZHU, X., NANDI, A. K., “Blind OFDM receiver based on independent component analysis for MIMO systems”, *IEEE Transactions on Wireless Communications*, v. 6, n. 11, pp. 4079–4089, 2007.
- [94] KWAK, K. C., PEDRYCZ, W., “Face recognition using an enhanced independent component analysis approach”, *IEEE Transactions on Neural Networks*, v. 18, n. 2, pp. 530–541, 2007.
- [95] ESCUDERO, J., HORNERO, R., ABASOLO, D., et al., “Artifact removal in magneto-encephalogram background activity with independent component analysis”, *IEEE Transactions on Biomedical Engineering*, v. 54, n. 11, pp. 1965–1973, 2007.
- [96] PAPOULIS, A., *Probability, Random Variables, and Stochastic Processes*. 3rd ed. McGraw-Hill: Nova York, Estados Unidos, 1991.

- [97] CICHOCKI, A., AMARI, S., *Adaptive Blind Signal and Image Processing*. John Wiley and Sons: West Sussex, Reino Unido, 2002.
- [98] HYVÄRINEN, A., OJA, E., “Independent Component Analysis: Algorithms and Applications”, *Neural Networks*, v. 13, n. 4-5, pp. 411–430, 2000.
- [99] JUTTEN, C., KARHUNEN, J., “Advances in Nonlinear Blind Source Separation”. In: *Proceedings of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 245–256, Nara, Japão, 2003.
- [100] HYVÄRINEN, A., PAJUNEN, P., “Nonlinear Independent Component Analysis: Existence and Uniqueness results”, *Neural Networks*, v. 12, n. 3, pp. 429–439, 1999.
- [101] ROJAS, F., PUNTONET, C. G., ROJAS, I., “Independent component analysis evolution based method for nonlinear speech processing”, *Artificial Neural Nets Problem Solving Methods, PT II*, v. 2687, pp. 679–686, 2003.
- [102] MIYABE, S., JUANG, B.-H., SARUWATARI, H., et al., “Kernel-based nonlinear independent component analysis for underdetermined blind source separation”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, v. 1, pp. 1641–1644, Taipei, Taiwan, 2009.
- [103] HARITOPOULOS, M., YIN, H., ALLINSON, N. M., “Image denoising using self-organizing map-based nonlinear independent component analysis”, *Neural Networks*, v. 15, n. 8-9, pp. 1085–1098, 2002.
- [104] ALMEIDA, M. S., ALMEIDA, L. B., “Wavelet-based separation of nonlinear show-through and bleed-through image mixtures”, *Neurocomputing*, v. 72, n. 1-3, pp. 57–70, 2008.
- [105] LU, C.-J., WU, J.-Y., FAN, C.-R., et al., “Forecasting stock price using Nonlinear independent component analysis and support vector regression”. In: *Proceedings of the IEEE International Conference on Industrial Engi-*

neering and Engineering Management, v. 1, pp. 2370–2374, Hong Kong, 2009.

- [106] DUARTE, L. T., JUTTEN, C., “A Mutual Information Minimization Approach for a Class of Nonlinear Recurrent Separating Systems”. In: *Proceedings of the International Workshop on Machine Learning for Signal Processing*, v. 1, pp. 122–127, Thessaloniki, Greece, 2007.
- [107] DUARTE, L. T., JUTTEN, C., MOUSSAOUI, S., “Ion Selective Electrode Array Based on a Bayesian Nonlinear Source Separation Method”. In: *Independent Component Analysis And Signal Separation, 8th International Conference, Lecture Notes In Computer Science*, pp. 662–669, Springer: Paraty, Brasil, Março 2009.
- [108] TALEB, A., JUTTEN, C., “Source Separation in post-nonlinear mixtures”, *IEEE Transactions on Signal Processing*, v. 47, n. 10, pp. 2807–2820, 1999.
- [109] ROJAS, F., PUNTONET, C. G., RODRÍGUEZ-ÁLVAREZ, M., et al., “Blind Source Separation in Post-Nonlinear Mixtures Using Competitive Learning, Simulated Annealing, and a Genetic Algorithm”, *IEEE Transactions on Systems, Man and Cybernetics-Part C*, v. 34, n. 4, pp. 407–416, 2004.
- [110] NGUYEN, T. V., PATRA, J. C., DAS, A., “A Geometric Approach to Post-Nonlinear Mixture in Blind Source Separation”. In: *Proceedings of the International Conference on Communications Systems*, pp. 260–264, Cin-gapura, Setembro 2004.
- [111] WEI, C., KHOR, L., WOO, W., et al., “A Novel Iterative Conditional Maximization Method for Post-Nonlinear Underdetermined Blind Source Separation”. In: *Proceedings of the International Conference on Digital Signal Processing*, pp. 551–554, Cardiff, Wales, Inglaterra, Julho 2007.
- [112] SOLAZZI, M., UNCINI, A., “Spline neural networks for blind separation of post-nonlinear-linear mixtures”, *IEEE Transactions on Circuits and Systems - I: Regular Papers*, v. 51, n. 4, pp. 817–829, 2004.

- [113] WANG, J., WANG, M., FANG, Y., “Post-nonlinear blind image separation algorithm with SOM initialization method”. In: *Proceedings of the IET Conference on Wireless, Mobile and Sensor Networks*, v. 1, pp. 978–982, Xangai, China, 2007.
- [114] KAI, S., QI, W., MINGLI, D., “Approach to nonlinear blind source separation based on niche genetic algorithm”. In: *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications*, v. 1, pp. 441–445, Jinan, China, 2006.
- [115] TAN, Y., WANG, J., “Nonlinear Blind Source Separation using Higher Order Statistics and a Genetic Algorithm”, *IEEE Transactions on Evolutionary Computation*, v. 5, n. 6, pp. 600–612, 2001.
- [116] WOO, W. L., DLAY, S. S., “Nonlinear blind source separation using a hybrid RBF-FMLP network”, *IEE Proceedings on Vision, Image and Signal Processing*, v. 8, n. 2, pp. 173–183, 2005.
- [117] GAO, P., WOO, W., DLAY, S. S., “Neural network approaches to nonlinear blind source separation”. In: *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*, v. 1, pp. 78–81, Sydney, Australia, 2005.
- [118] GRIFFEL, D. H., *Applied Functional Analysis*. Dover: West Sussex, Reino Unido, 1985.
- [119] PAJUNEN, P., HYVARINEN, A., KARHUNEN, J., “Nonlinear blind source separation by self-organizing maps”. In: *Proceedings of the International Conference on Neural Information Processing*, v. 1, pp. 1207–1210, Hong Kong, Setembro 1996.
- [120] WANG, J., WANG, M. Y. F., “Post-nonlinear blind image separation algorithm with SOM initialization method”. In: *Proceedings of the IET Conference on Wireless, Mobile and Sensor Networks*, v. 1, pp. 978–982, Xangai, China, 2007.

- [121] YIN, H., “Intelligent Data Engineering and Automated Learning”, v. 2690, chap. Nonlinear Multidimensional Data Projection and Visualisation, pp. 377–388, *Lecture Notes in Computer Science*, Springer, 2003.
- [122] BISHOP, C. M., SVENSEN, M., WILLIANS, C. K. I., “GTM: The generative topographic mapping”, *Neural Computation*, v. 10, pp. 215–234, 1998.
- [123] PAJUNEN, P., KARHUNEN, J., “A maximum likelihood approach to nonlinear blind source separation”. In: *Proceedings of the International Conference on Artificial Neural Networks*, v. 1, pp. 541–546, Lausanne, Suíça, 1997.
- [124] LAPPALAINEN, H., HONKELA, A., “Advances in Independent Component Analysis”, chap. Bayesian nonlinear independent component analysis by multi-layer perceptrons, pp. 93–121, Springer, 2000.
- [125] LAPPALAINEN, H., “Ensemble learning for independent component analysis”. In: *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 7–12, Aussois, France, 1999.
- [126] LAPPALAINEN, H., MISKIN, J. W., “Advances in Independent Component Analysis”, chap. Ensemble learning, pp. 75–92, Springer, 2000.
- [127] VALPOLA, H., “Nonlinear independent component analysis using ensemble learning: Theory”. In: *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 351–356, Helsinki, Finlândia, 2000.
- [128] HONKELA, A., “Approximating nonlinear transformations of probability distributions for nonlinear independent component analysis”. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, v. 3, pp. 2169–2174, Budapest, Hungria, 2004.
- [129] HONKELA, A., VALPOLA, H., ILIN, A., et al., “Blind separation of nonlinear mixtures by variational bayesian learning”, *Digital Signal Processing*, v. 17, n. 5, pp. 914–934, 2007.

- [130] ILIN, A., ACHARD, S., JUTTEN, C., “Bayesian versus constrained structure approaches for source separation in post-nonlinear mixtures”. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, v. 3, pp. 2181–2186, Budapest, Hungria, Julho 2004.
- [131] ALMEIDA, L. B., “Linear and Nonlinear ICA based on Mutual Information”, *The Journal of Machine Learning Research*, v. 4, n. 7, pp. 1297–1318, 2004.
- [132] ALMEIDA, L. B., HYVÄRINEN, A., “Separating a real-life nonlinear image mixture”, *Journal of Machine Learning Research*, v. 6, pp. 1199–1229, 2005.
- [133] ZHENG, C.-H., HUANG, D.-S., LI, K., et al., “MISEP Method for Post-nonlinear Blind Source Separation”, *Neural Computation*, v. 19, n. 9, pp. 2557–2578, 2007.
- [134] SUN, Z.-L., “An extension of MISEP for post-nonlinear-linear mixture separation”, *IEEE Transactions on Circuits and Systems*, v. 56, n. 8, pp. 654–658, 2009.
- [135] KARHUNEN, J., MALAROIU, S., “Local independent component analysis using clustering”. In: *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 1–6, Aussois, France, 1999.
- [136] DUDA, R. O., HART, P. E., STORK, D. G., *Pattern Classification*. 2nd ed. Wiley: Nova York, Estados Unidos, 2000.
- [137] KOHONEN, T., OJA, E., SIMULA, O., et al., “Engineering Applications of the Self-Organizing Map”, *Proceedings of the IEEE*, v. 84, n. 10, pp. 1358–1384, Outubro 1996.
- [138] KOTHARI, R., PITTS, D., “On finding the number of clusters”, *Pattern Recognition Letters*, v. 20, n. 1, pp. 405–416, 1999.
- [139] TAKASU, A., “Lecture Notes In Computer Science”, v. 1532, chap. On the Number of Clusters in Cluster Analysis, pp. 62–63, Springer, 1998.

- [140] KARHUNEN, J., MALAROIU, S., ILMONIEMI, M., “Local linear independent component analysis based on clustering”, *International Journal of Neural Systems*, v. 10, pp. 439–451, 2000.
- [141] HONDA, K., ICHIHASHI, H., OHUE, M., et al., “Extraction of local independent components using fuzzy clustering”. In: *Proceedings of 6th International Conference on Neural Networks and Soft Computing*, v. 1, pp. 837–842, Zakopane, Polônia, 2000.
- [142] MAENAKA, T., HONDA, K., ICHIHASHI, H., “Local independent component analysis with fuzzy clustering and regression-principal component analysis”. In: *Proceedings of IEEE International Conference on Fuzzy Systems*, v. 1, pp. 857–862, Vancouver, Canada, 2006.
- [143] PALMIERI, F., BUDILLON, A., “Advances in Independent Component Analysis”, chap. Multi-class independent component analysis for rank deficient distributions, pp. 145–160, Springer-Verlag, 2000.
- [144] LAN, T., ERDOGMUS, D., “Local Linear ICA for Mutual Information Estimation in Feature Selection”. In: *Proceedings of the IEEE Workshop on Machine Learning for Signal Processing*, v. 1, n. 1, pp. 3–8, Connecticut, Estados Unidos, Setembro 2005.
- [145] LAN, T., HUANG, Y., ERDOGMUS, D., “Independent Component Analysis and Blind Signal Separation”, v. 3889, chap. A Comparison of Linear ICA and Local Linear ICA for Mutual Information Based Feature Ranking, pp. 823–830, *Lecture Notes in Computer Science*, Springer, 2006.
- [146] GRUBER, P., THEIS, F., STADLTHANNER, K., et al., “Denoising using Local ICA and kernel-PCA”. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, Julho 2004.
- [147] SANCHEZ-POBLADOR, V., MONTE-MORENO, E., SOLÉ-CASALS, J., “Independent Component Analysis and Blind Signal Separation”, v. 3195, chap. ICA as a Preprocessing Technique for Classification, pp. 1165–1172, *Lecture Notes in Computer Science*, Springer, 2004.

- [148] BLAKE, C. L., MERZ, C. J., “UCI Repository of Machine Learning Databases”, Disponível em: <<http://www.ics.uci.edu/mlearn/MLRepository.html>>. Acesso em: 09 de maio de 2010.
- [149] CARDOSO, J.-F., SOULOUMIAC, A., “Blind Beamforming for Non-gaussian Signals”, *IEE Proceedings-F*, v. 140, n. 6, pp. 362–370, Novembro 1993.
- [150] CABALLERO, R. G., ORELLANA, C. J. G., MACÍAS, M. M., et al., “Independent Component Analysis Applied to Breast Cancer Detection on Digitized Mammograms”, *International Congress Series*, v. 1281, n. 1, pp. 1052–1057, 2005.
- [151] HYVÄRINEN, A., “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”, *IEEE Transactions on Neural Networks*, v. 10, n. 3, pp. 626–634, 1999.
- [152] ZHENG, C. H., HUANG, D. S., SHANG, L., “Feature Selection in Independent Component Subspace for Microarray Data Classification”, *Neurocomputing*, v. 69, n. 1, pp. 2407–2410, 2006.
- [153] APOLLONI, B., BASSIS, S., BREGA, A., “Feature Selection via Boolean Independent Component Analysis”, *Information Sciences*, v. 179, pp. 3815–3831, 2009.
- [154] GOLDBAUM, M. H., “Unsupervised Learning with Independent Component Analysis can Identify Patterns of Glaucomatous Visual Field Defects”, *Transactions of the American Ophthalmology Society*, v. 103, n. 1, pp. 270–280, 2005.
- [155] BONNET, N., NUZILLARD, D., “Independent Component Analysis, a New Possibility for Analysing Series of Electron Energy Loss Spectra”, *Ultramicroscopy*, v. 102, n. 1, pp. 327–337, 2005.
- [156] EGERTON, R. F., *Electron Energy Loss Spectroscopy in the Electron Microscope*. Plenum: Nova York, Estados Unidos, 1996.

- [157] BELOUCHRANI, A., MERAIN, K. A., CARDOSO, J. F., et al., “A Blind Source Separation Technique based on Second Order Statistics”, *IEEE Transactions on Signal Processing*, v. 45, n. 2, pp. 434–444, 1997.
- [158] OVEISI, F., “EEG Signal Classification Using Nonlinear Independent Component Analysis”. In: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, v. 1, pp. 361–364, Taipei, Taiwan, 2009.
- [159] GOLDBERG, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley: Boston, Estados Unidos, 1989.
- [160] LU, C. J., FAN, C. R., CHIU, C. C., “Forecasting Stock Price Using Nonlinear Independent Component Analysis and Support Vector Machine”. In: *Proceedings of the International Conference on Industrial Engineering and Engineering Management*, v. 1, pp. 2370–2374, Hong Kong, 2009.
- [161] LANG, M. J., “Application of Kohonen network classifier in TeV gamma-ray astronomy”, *Journal of Physics G: Nuclear Particle Physics*, v. 1, n. 24, pp. 2279–2287, 1998.
- [162] BECKS, K. H., DRESS, J., FLAGMEYER, U., et al., “Separation of hadronic W-decays from QCD-background with self-organizing maps”, *Nuclear Instruments and Methods in Physics Research A*, v. 426, n. 2, pp. 599–604, 1999.
- [163] LANGE, J. S., “Transputer self-organizing map algorithm for beam background rejection at the BELLE silicon vertex detector”, *Nuclear Instruments and Methods in Physics Research A*, v. 420, n. 1, pp. 288–309, 1999.
- [164] WYZYKOWSKI, L., BELOKUROV, V., “Self-organizing maps, an application to the OGLE data and the Gaia Science Alerts”. In: *Proceedings of the International Conference on Classification and discovery in large astronomical surveys, AIP Conference Proceedings*, v. 1083, n. 1, pp. 201–206, Rinberg Castle, Alemania, 2008.

- [165] HEIKKINEN, A., “Separation of Higgs boson signal from Drell-Yan background with self-organizing maps”, *Proceedings of Science*, v. ACAT07, n. 065, pp. 1–8, 2007.
- [166] PRORIOL, J., “Selection of variables for neural network analysis. Comparison of several methods with high-energy physics data”, *Nuclear Instruments and Methods in Physics Research A*, v. 361, n. 1, pp. 581–585, 1995.
- [167] MONTGOMERY, D. C., RUNGER, G. C., *Estatística Aplicada e Probabilidade para Engenheiros*. LTC: Rio de Janeiro, Brasil, 2003.
- [168] WOLTER, M., “Multivariate Analysis Methods in Physics”, *Physics of Particles and Nuclei*, v. 38, n. 2, pp. 255–268, 2007.
- [169] AKRAS, S., BOUMIS, P., “A principal component analysis approach to the morphology of planetary nebulae”, *Astrophysics Journal*, v. 1, n. 1, pp. 1–8, 2007.
- [170] CADAVID, A. C., LAWRENCE, J. K., RUZMAIKIN, A., “Principal Components and Independent Component Analysis of Solar and Space Data”, *Solar Physics Journal*, v. 248, n. 2, pp. 247–261, 2008.
- [171] HUANG, X., LEE, S. Y., PREBYS, E., et al., “Application of independent component analysis to Fermilab BOOSTER”, *Physical Review Special Topics, Accelerators and Beams*, v. 8, n. 6, pp. 1–14, 2005.
- [172] FERNANDEZ, C., NAYERI, S., “Independent Component Analysis Applications in Physics”. In: *Proceedings of the Int. Joint Conference on Neural Networks*, v. 1, n. 1, pp. 2213–2216, Montreal, Canada, 2005.
- [173] COSTAGLIA, M., KURUOGLU, E. E., AHMED, A., “Astrophysical Source Separation using Particle Filters”, *Lecture Notes in Computer Science - Independent Component Analysis and Blind Signal Separation*, v. 3195, pp. 930–937, 2004.
- [174] IGUAL, J., LINHARES, R., SALAZAR, A., “Source Separation of Astrophysical Ice Mixtures”, *Lecture Notes in Computer Science - Independent*

Component Analysis and Blind Signal Separation, v. 3889, pp. 368–375, 2006.

- [175] MOUDDEN, Y., ABRIAL, P., MELIN, J. B., et al., “Independent Component Separation from Incomplete Spherical Data using Wavelets. Application to CMB Data Analysis”. In: *Proceedings of the International Conference on Physics in Signal and Image Processing*, v. 1, pp. 1–6, Toulouse, France, 2005.
- [176] VIOO, R., ANDREANI, P., “A Modified ICA Approach for Signal Separation in CMB maps”, *Astronomy and Astrophysics Journal*, v. 802, pp. 1–12, Fevereiro 2008.
- [177] NOJUN KWAK, M., CHOI, C.-H., “Feature Extraction Based on ICA for Binary Classification Problems”, *IEEE Transactions on Knowledge and Data Engineering*, v. 15, n. 6, pp. 1374–1388, 2003.
- [178] CANER, E. S., SEIXAS, J., “Neural Discriminating Analysis on Pre-processed Data”. In: *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 415–418, Pafos, Cyprus, Setembro 1999.
- [179] LU, H., SETIONO, R., LIU, H., “Effective Data Mining Using Neural Networks”, *IEEE Transactions on Knowledge and Data Engineering*, v. 8, n. 6, pp. 957–961, 1996.
- [180] STEPPE, J. M., ANS S. K. ROGERS, K. W. B. J., “Integrated Feature Architecture Selection”, *IEEE Transactions on Neural Networks*, v. 7, n. 4, pp. 1007–1014, 1996.
- [181] MCGARRY, K. J., WERMTER, S., MACINTYRE, J., “Knowledge Extraction from Radial Basis Functions Networks and Multi-Layer Perceptrons”. In: *Proceedings of the International Joint Conference on Neural Networks*, v. 4, pp. 2494–2497, Washington-DC, Estados Unidos, 1999.

- [182] KWAK, N., CHOI, C. H., CHOI, J. Y., “Lecture Notes In Computer Science (ICANN)”, v. 2130, chap. Feature Extraction Using ICA, pp. 568–573, Springer, 2001.
- [183] SIMAS FILHO, E. F., SEIXAS, J. M., CALOBA, L. P., “Modified Post-Nonlinear ICA Model for Online Neural Discrimination”, *Neurocomputing*, v. 73, n. 16-18, pp. 2820–2828, 2010.
- [184] GEORGIEVA, P., RALESCU, A., RALESCU, D., “Cross-cumulants measure for independence”, *Journal of Statistical Planning and Inference*, v. 137, n. 1, pp. 1085–1098, 2006.
- [185] HAUPT, R. L., HAUPT, S. E., *Practical Genetic Algorithms*. Wiley-Interscience: Nova York, Estados Unidos, 2004.
- [186] TREES, H. L. V., *Detection, Estimation, and Modulation Theory, Part I*. Wiley: Nova York, Estados Unidos, 2001.
- [187] MOURA, N. N., SIMAS FILHO, E. F., SEIXAS, J. M., “Advances in Sonar Signal Processing”, chap. Independent Component Analysis for Passive Sonar Signal Processing, pp. 91–110, In-Tech: Vienna, Austria, 2009.
- [188] TORRES, R. C., LIMA, D. E. F., SIMAS FILHO, E. F., et al., “Neural online filtering based on preprocessed calorimeter data”. In: *IEEE Nuclear Science Symposium Conference Record*, pp. 530–536, Orlando, Estados Unidos, Maio 2009.
- [189] KANUNGO, T., MOUNT, D. M., NETANYAHU, N. S., et al., “An Efficient k-Means Clustering Algorithm, Analysis and Implementation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, pp. 881–892, 2002.
- [190] ALSABTI, K., “An efficient k-means clustering algorithm”. In: *Proceedings of IPPS/SPDP Workshop on High Performance Data Mining*, v. 1, pp. 1–6, Genebra, Suiça, 1998.
- [191] DING, C., HE, X., “K-means Clustering via Principal Component Analysis”. In: *Proc. of Int'l Conf. Machine Learning*, v. 1, pp. 225–232, 2004.

- [192] GRAY, R. M., “Vector Quantization”, *IEEE ASSP Magazine*, v. Part 1, pp. 4–29, Abril 1984.
- [193] FLEXER, A., “On the use of self-organizing maps for clustering and visualization”, *Intelligent Data Analysis*, v. 5, pp. 373–384, October 2001.
- [194] KOHONEN, T., *Self Organizing Maps*. 3rd ed. Springer: Berlin, Alemanha, 2001.
- [195] KUNCHEVA, L., *Combining Pattern Classifiers, Methods and Algorithms*. Wiley: Nova York, Estados Unidos, 2004.
- [196] RIEDMILLER, M., BRAUN, H., “A direct adaptive method for faster back-propagation learning, the RPROP algorithm”, *Proc. of Int. Conf. on Neural Networks*, pp. 586–591, 1993.
- [197] NGUYEN, D., WIDROW, B., “Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights”. In: *Proceedings of the International Joint Conference on Neural Networks*, v. 3, pp. 21–26, São Diego, Estados Unidos, Junho 1990.
- [198] SCOTT, D. W., *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley: Nova York, Estados Unidos, 1992.
- [199] JOE, H., “Relative Entropy Measures of Multivariate Dependence”, *Journal of the American Statistical Association*, v. 84, n. 405, pp. 586–591, 1989.
- [200] SCHLICKEISER, R., *Cosmic Ray Astrophysics*. Springer: Nova York, Estados Unidos, 2002.
- [201] GAISSER, T. K., *Cosmic Rays and Particle Physics*. Cambridge University Press: Cambridge, Reino Unido, 1990.
- [202] ATLAS-COLLABORATION, *Electron and photon reconstruction and identification in ATLAS: expected performance at high energy and results at 900 GeV*, Tech. rep., CERN, Genebra, Suíça, Março 2010.
- [203] FLOWERDEW, M. J., *Standard Model Z/gamma to ee production in early data at ATLAS*, Ph.D. Thesis, University of Liverpool, Dezembro 2009.

- [204] KERSCHEN, N., “Electron performance in the ATLAS experiment”. In: *Proceedings of Physics at LHC Conference*, pp. 1–5, Hamburg, Alemanha, Junho 2010.
- [205] KUNA, M., “Electron and Photon Reconstruction and Identification with the ATLAS Detector”. In: *Proceedings of the Topical Seminar on Innovative Particle and Radiation Detectors*, pp. 1–5, Siena, Itália, Junho 2010.
- [206] TORRES, R. C., SIMAS FILHO, E. F., LIMA, D. E. F., et al., “Signal Processing”, chap. Segmented Online Neural Filtering based on Independent Components of Pre-Processed Information, pp. 337–358, In-Tech: India, 2010.
- [207] KNOLL, G. F., *Radiation Detection and Measurement*. 3rd ed. Wiley: Nova York, Estados Unidos, 2000.
- [208] KASKI, S., VENNA, J., T.KOHONEN, “Coloring that reveals high-dimensional structures in data”. In: *Proceedings of the International Conference on Neural Information Processing*, novembro 1999.
- [209] HIMBERG, J., “A SOM based cluster visualization and its application for false coloring”. In: *Proceedings of the Int. Joint Conference on Neural Networks*, Julho 2000.
- [210] CICHOCKI, A., ZDUNEK, R., PHAN, A. H., et al., *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley: West Sussex, Reino Unido, 2009.
- [211] SPARSE SIGNAL REPRESENTATIONS, B. I. R., “A Mutual Information Minimization Approach for a Class of Nonlinear Recurrent Separating Systems”. In: *Proceedings of 2004 IEEE International Symposium on Circuits and Systems*, v. 5, pp. 684–687, Vancouver, Canadá, 2004.
- [212] GEORGIEV, P. G., THEIS, F., CICHOCKI, A., “Sparse component analysis and blind source separation of underdetermined mixtures”, *IEEE Transactions on Neural Networks*, v. 16, n. 4, pp. 992–996, 2005.

- [213] BASE, A. M., GRUBER, P., THEISA, F., et al., “Blind source separation based on self-organizing neural network”, *Engineering Applications of Artificial Intelligence*, v. 19, n. 3, pp. 305–311, 2006.
- [214] GERSHO, A., “On the Structure of Vector Quantizers”, *IEEE Transactions on Information Theory*, v. 28, n. 2, pp. 157–166, Março 1982.
- [215] KOHONEN, T., “Improved versions of learning vector quantization”. In: *Proceedings of the International Joint Conference on Neural Networks*, v. 1, pp. 545–550, San Diego, Estados Unidos, 1990.
- [216] BARAS, J. S., DEY, S., “Combined Compression and Classification with Learning Vector Quantization”, *IEEE Transactions on Information Theory*, v. 45, n. 6, pp. 1911–1920, 1999.
- [217] FREEMAN, J. A., SKAPURA, D. M., *Neural Networks Algorithms, Applications, and Programming Techniques*. Addison Wesley, 1991.
- [218] OJA, E., “Neural Networks, Principal Components, and Subspaces”, *International Journal of Neural Systems*, v. 1, n. 1, pp. 61–68, 1989.
- [219] DE OLIVEIRA E SOUZA FILHO, J. B., *Classificação Neural de Sinais Passivos*, Ph.D. Thesis, COPPE/UFRJ, Julho 2007.
- [220] PEEBLES JR., P. Z., *Probability, Random Variables and Random Signal Principles*. Mc Graw Hill: Nova York, Estados Unidos, 2001.
- [221] SPIEGEL, M. R., SCHILLER, J. J., SRINIVASAN, R. A., *Probability and Statistics*. 2nd ed. McGraw-Hill: Nova York, Estados Unidos, 2000.
- [222] SHANNON, C. E., “A Mathematical Theory of Communication”, *The Bell System Technical Journal*, v. 27, n. 6, pp. 379–423, Julho 1948.
- [223] HYVARINEN, A., “New approximations of differential entropy for independent component analysis and projection pursuit”, *Advances in Neural Information Processing*, v. 10, pp. 273–279, 1998.
- [224] CICHOCKI, A., UNBEHAUEN, R., “Robust Neural Networks with On-Line Learning for Blind Identification and Blind Separation of Sources”, *IEEE*

Transactions on Circuits and Systems-I: Fundamental Theory and Applications, v. 43, n. 11, Novembro 1996.

- [225] LUENBERGER, D. G., *Linear and Nonlinear Programming*. Addison-Wesley: Boston, Estados Unidos, 1984.
- [226] MICHAL, A. D., *Matrix and Tensor Calculus*. 1st ed. Dover: West Sussex, Reino Unido, 2008.
- [227] CARDOSO, J. F., “Source separation using higher order moments”. In: *Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, v. 4, pp. 2109–2112, Glasgow, Reino Unido, 1989.
- [228] AKUZAWA, T., MURATA, N., “Multiplicative Nonholonomic Newton-like Algorithm”, *Chaos, Solitons and Fractals*, v. 12, n. 4, pp. 785–793, 2001.
- [229] AKUZAWA, T., “Extended Quasi-Newton Method for the ICA”. In: *Proceedings of the Int. Workshop Independent Component Anal. Blind Signal Separation*, pp. 521–525, Helsinki, Finlândia, 2000.
- [230] CUENCA, W. M., LEVY, A. F., SEIXAS, J. M., et al., “Análise de Componentes Independentes para Extração Cega de Ruídos Gaussianos dos Sinais de Descargas Parcias em Equipamentos de Alta Tensão”. In: *Anais do Congresso Brasileiro de Redes Neurais*, v. 1, pp. 133–138, São Paulo, 2003.
- [231] CICHOCKI, A., DOUGLAS, S., AMARI, S., “Robust techniques for independent component analysis with noisy data”, *Neurocomputing*, v. 22, n. 1-3, pp. 113–129, 1998.
- [232] HYVÄRINEN, A., “Independent Component Analysis in the Presence of Gaussian Noise by Maximizing Joint Likelihood”, *Neurocomputing*, v. 22, n. 1-3, pp. 49–67, 1998.
- [233] HYVÄRINEN, A., “Sparse Code Shrinkage: Denoising of Nongaussian Data by Maximum Likelihood Estimation”, *Neural Computation*, v. 11, n. 7, pp. 1739–1768, 1999.

- [234] PARASCHIV-IONESCU, A., JUTTEN, C., AMINIAN, K., et al., “Source separation in strong noisy mixtures a study of wavelet de-noising pre-processing”. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 2, pp. 1681–1684, Orlando, Estados Unidos, 2002.
- [235] TAN, Y., WANG, J., “Nonlinear blind separation using an RBF network model”. In: *Proceedings of the IEEE International Symposium on Circuits and Systems*, v. 3, pp. 634–637, Genebra, Suiça, 2000.
- [236] SHANMUGAN, K. S., BREIPOHL, A. M., *Random Signals, Detection, Estimation and Data Analysis*. Wiley: Nova York, Estados Unidos, 1988.
- [237] FISHER, R. A., “The Use of Multiple Measurements in Taxonomic Problems”, *Annals of Eugenics*, v. 7, pp. 179–188, 1936.
- [238] SIMAS FILHO, E. F., ALMEIDA, L. A. L., LIMA, A. C. C., “Vibration Monitoring of On-Load Tap Changers Using a Genetic Algorithm”. In: *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, v. 3, pp. 2288–2293, Ottawa, Canada, 2005.
- [239] SIMAS FILHO, E. F., ALMEIDA, L. A. L., “Filtragem inversa de medições de histerese térmica utilizando Redes Neurais”. In: *Anais do XVI Congresso Brasileiro de Automática*, v. 1, pp. 1–6, Salvador, Brasil, Outubro 2006.
- [240] ZOPPOLI, R., PARISINI, T., SANGUINETI, M., “Neural Aproximators for Function Optimization”, *Proceedings of the 35th conference on decision and control*, v. 3, pp. 3290–3293, 1996.
- [241] ICHIKAWA, Y., SAWA, T., “Neural Networks Application for Direct Feedback Controllers”, *IEEE Transactions on Neural Networks*, v. 3, n. 2, pp. 224–231, 1992.
- [242] WASSERMAN, P., *Neural Computing, Theory and Practice*. Van Nostrand Reinhold: Nova York, Estados Unidos, 1989.
- [243] WHITLEY, D., “An Overview of Evolutionary Algorithms: Practical Issues and Common Pitfalls”, *Colorado Advanced Software Institute Press*, 2002.

- [244] HOLLAND, J., “Adaptation in Natural and Artificial Systems”, *University of Michigan Press*, 1975.
- [245] MITCHEL, M., *An Introduction to Genetic Algorithms*. The MIT Press, 1996.
- [246] ACKLEY, D., “A Connectionist Machine for Genetic Hillclibimg”, *Kluwer Academic Publishers*, 1987.
- [247] SYSWERDA, G., “Uniform Crossover in Genetic Algorithms”, *Proceedings of the 3rd International Conference on Genetic Algorithms*, pp. 2–9, 1989.
- [248] TANOMARU, J., “Motivação, Fundamentos e Aplicações de Algoritmos Genéticos”, *II Congresso Brasileiro de Redes Neurais, III Escola de Redes Neurais*, 1995.

Apêndice A

Aspectos Teóricos das Técnicas de Extração de Características

Neste apêndice serão fornecidos os detalhes da teoria envolvida nos diversos métodos de extração de características utilizados neste trabalho.

A.1 Mapas Auto-Organizáveis

O mapa auto-organizável (SOM - *Self Organizing Map*) é uma rede neural com treinamento não-supervisionado, baseado na aprendizagem competitiva, que é capaz de realizar uma organização topológica das entradas. O SOM foi proposto por Teuvo Kohonen em 1982 [194], sendo capaz de realizar um mapeamento não-linear dos sinais de um espaço de entrada contínuo de dimensão k para um espaço de características discreto que, em geral, é bidimensional. Cada neurônio da grade está diretamente conectado a todos os nós de entrada. Na Figura A.1 pode-se visualizar o diagrama de um mapa auto-organizável bi-dimensional.

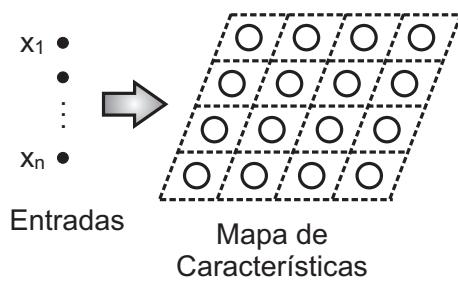


Figura A.1: Diagrama de um mapa auto-organizável

O mapa auto-organizável compacta a informação e preserva relações topológicas ou métricas do conjunto de sinais. Os SOM estão ligados à ICA por conseguirem extrair informações ocultas dos sinais de forma não supervisionada [213]. Uma aproximação das componentes independentes não-lineares pode ser obtida utilizando mapas auto-organizáveis [4].

Três processos estão envolvidos na formação do mapa auto-organizável: a **competição**, onde, para cada vetor de entrada, há apenas um neurônio vencedor; a **cooperação**, quando o neurônio vencedor determina uma vizinhança topológica de neurônios excitados; e a **adaptação**, que procede ao ajuste dos pesos sinápticos para reforçar a resposta do neurônio vencedor, e de seus vizinhos, ao padrão de entrada.

Considerando vetores de entrada $\mathbf{x} = [x_1, x_2, \dots, x_k]^T$, como os neurônios são totalmente conectados às entradas, o vetor de pesos sinápticos do neurônio j pode ser definido por: $\mathbf{w}_j = [w_{1j}, w_{2j}, \dots, w_{kj}]^T$. A atualização do vetor de pesos é feita através da equação:

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n)h_{ij}(n)(\mathbf{x}(n) - \mathbf{w}_j(n)), \quad (\text{A.1})$$

sendo $\eta(n)$ a taxa de aprendizagem, um tipo de função de vizinhança $h_{ij}(n)$ usualmente utilizada é definida por:

$$h_{ij}(n) = \exp(-d_{ij}^2/2\sigma^2(n)) \quad (\text{A.2})$$

onde d_{ij} é a distância do neurônio j para o neurônio vencedor i e $\sigma(n)$ é a largura da função vizinhança na n -ésima iteração.

... falar de outras funções de vizinhança utilizadas ...

O mapa de características possui algumas propriedades, listadas a seguir [5]:

1. é formado pelo conjunto de vetores de pesos sinápticos \mathbf{w}_i no espaço de saída discreto e fornece uma boa aproximação para o espaço de entrada;
2. é ordenado de modo topológico. Padrões de entrada semelhantes são mapeados para regiões adjacentes no mapa de características;
3. regiões do espaço de entrada que possuem alta probabilidade de ocorrência são mapeadas para domínios maiores do espaço de saída;

4. a matriz de pesos sinápticos pode ser definida por:

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_P], \quad (\text{A.3})$$

onde P é o número de neurônios do mapa.

No mapa de características, o neurônio que apresentar maior saída é considerado o vencedor, ou seja a saída do SOM é do tipo “vencedor leva tudo”(WTA - *winner takes all*). O neurônio ativado é escolhido a partir de sua semelhança com a entrada \mathbf{x}_A apresentada. É comum a utilização da distância euclidiana como métrica da proximidade entre dois vetores; nesse caso, o neurônio vencedor é aquele que minimiza $i(\mathbf{x}) = \|\mathbf{x}_A - \mathbf{w}_j\|$.

Uma outra forma de operar um mapa auto-organizável é utilizar as projeções dos sinais de entrada no mapa de características, ou seja as saídas u_j de cada neurônio j que podem ser calculadas por:

$$u_j = \mathbf{x}^T \mathbf{w}_j \quad (\text{A.4})$$

O vetor $\mathbf{u} = [u_1, \dots, u_K]^T$ pode ser considerado como a projeção de \mathbf{x} no mapa de características.

Os mapas auto-organizáveis pertencem à classe de algoritmos de codificação vetorial, sendo capazes de encontrar, de forma otimizada, um número fixo de vetores ou palavras de código que melhor representem o conjunto de sinais.

A.1.1 Quantização Vetorial por Aprendizado

A quantização vetorial (VQ - *Vector Quantization*) é uma técnica de codificação onde um espaço de entrada é mapeado em um grupo finito de vetores representativos (*codebook*) [214]. A codificação é definida como um particionamento do espaço de entrada em um número finito de regiões. O quantizador realiza um mapeamento do espaço \mathbb{R}^k , em um subconjunto finito Y de \mathbb{R}^k :

$$Q : \mathbb{R}^k \rightarrow \mathbf{Y} \quad (\text{A.5})$$

sendo $\mathbf{Y} = \{y_1, y_2, \dots, y_k\}$ o livro de código (*codebook*). Para cada palavra de código

y_i existe uma partição R_i do espaço de entrada que satisfaz:

$$R_i = Q^{-1}(\mathbf{y}_i) = \{\mathbf{x} \in \mathbb{R}^k : Q(\mathbf{x}) = \mathbf{y}_i\} \quad (\text{A.6})$$

$$\bigcup_{i=1}^N R_i = \mathbb{R}^k, \quad R_i \cap R_j = \emptyset, \quad i \neq j \quad (\text{A.7})$$

Quando um quantizador vetorial possui mínima distorção é denominado **quantizador de Voronoi**. Neste caso, diz-se que o espaço de entrada está particionado de acordo com a regra do vizinho mais próximo, e as partições criadas são chamadas de células de Voronoi [192]. Usando-se a distância euclidiana como parâmetro de distorção, o quantizador Q^* é dito ótimo se, para qualquer outro quantizador Q , com o mesmo número de pontos, a condição abaixo é satisfeita:

$$E\|\mathbf{x} - Q^*(\mathbf{x})\|^2 \leq E\|\mathbf{x} - Q(\mathbf{x})\|^2 \quad (\text{A.8})$$

As palavras de código ou os vetores de Voronoi podem ser calculados de modo aproximado pelo algoritmo SOM. O *codebook* é formado a partir dos pesos sinápticos dos neurônios do mapa. As células de Voronoi são compostas pelos pontos do espaço de entrada que estão mais próximos do vetor de código correspondente.

Em um problema de classificação, pode-se empregar a quantização vetorial por aprendizado (*Learning Vector Quantization*) [215], que utiliza informações sobre as classes para mover ligeiramente os vetores de Voronoi, visando a uma melhora no desempenho de decisão do classificador.

Na sua forma básica, o algoritmo LVQ escolhe aleatoriamente um vetor de entrada \mathbf{x} ; quando seu rótulo de classe $\mathcal{C}_{\mathbf{x}_i}$ e o de um vetor de Voronoi \mathbf{w}_c concordam, então, \mathbf{w}_c é movido na direção de \mathbf{x} :

$$\mathcal{C}_{\mathbf{w}_c} = \mathcal{C}_{\mathbf{x}_i} \rightarrow \mathbf{w}_c(n+1) = \mathbf{w}_c(n) + \alpha[\mathbf{x} - \mathbf{w}_c(n)] \quad (\text{A.9})$$

onde α é a taxa de aprendizagem ($0 < \alpha < 1$). Em caso contrário, \mathbf{w} é afastado de

\mathbf{x} :

$$\mathcal{C}_{\mathbf{w}_c} \neq \mathcal{C}_{\mathbf{x}_i} \rightarrow \mathbf{w}_c(n+1) = \mathbf{w}_c(n) - \alpha[\mathbf{x} - \mathbf{w}_c(n)] \quad (\text{A.10})$$

Conforme proposto em [215], podem ser implementadas algumas modificações na forma básica do algoritmo de LVQ, visando a melhorar o desempenho do método. Chega-se, então, aos algoritmos LVQ-2 e LVQ-2.1, que ajustam dois vetores de código simultaneamente.

Alguns exemplos da aplicação da quantização vetorial por aprendizado para compressão de sinais e classificação podem ser encontrados em [215] e [216].

A.1.2 Classificação a Partir do Mapa de Características

Considerando um problema de classificação, o mapeamento auto-organizável consegue transformar o conjunto de sinais, revelando características ocultas. A nova organização do conjunto de entrada pode ser utilizada para guiar o processo de discriminação. Em [217] é proposta uma estratégia de classificação a partir do mapa de características onde uma rede neural MLP é conectada às saídas do SOM (ver Figura A.2). A MLP é treinada com supervisão usando informações a respeito das classes de sinais.

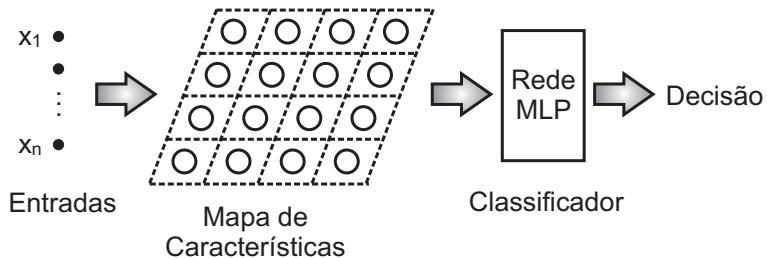


Figura A.2: Diagrama da classificação a partir do mapa de características

A.2 Análise de Componentes Principais

A.3 Técnicas de Pré-Processamento - Compactação

No processamento de sinais multi-dimensionais, é comum a utilização de técnicas de processamento de sinais que visam a redução da dimensionalidade do problema. O objetivo é projetar os sinais N-dimensionais observados em

A.3.1 Análise de Componentes Principais

A análise de componentes principais (PCA - *Principal Component Analysis*) é uma técnica estatística de processamento de sinais diretamente ligada à transformação de *Karhunen-Loëve* [3]. O objetivo da PCA é encontrar uma transformação linear onde os sinais projetados sejam não-correlacionados e grande parcela da energia (variância) esteja concentrada num pequeno número de componentes. Para isso, são exploradas informações da estatística de segunda ordem.

A análise de componentes principais é bastante usada para compactação de informação. Como a PCA projeta os sinais em componentes ordenados por energia, uma métrica geralmente utilizada para reduzir a dimensão dos dados consiste na seleção apenas dos componentes de maior energia, de modo que o sinal recuperado a partir da informação compactada tenha pequeno erro médio quadrático se comparado ao original. A seguir serão desenvolvidos, de forma resumida, os fundamentos matemáticos da PCA.

Considerando-se um vetor $\mathbf{x} = [x_1, \dots, x_N]^T$ aleatório com N elementos, assume-se que ele tenha média zero:

$$\mathcal{E}\{\mathbf{x}\} = 0 \quad (\text{A.11})$$

onde $\mathcal{E}\{\cdot\}$ é o operador esperança. Se \mathbf{x} tem média não nula faz-se $\mathbf{x} \leftarrow \mathbf{x} - \mathcal{E}\{\mathbf{x}\}$.

A projeção z_i de \mathbf{x} na direção de \mathbf{v}_i pode ser expressa por:

$$z_i = \mathbf{v}_i^T \mathbf{x} = \sum_{k=1}^N v_{ki} x_k \quad (\text{A.12})$$

Na transformação por PCA, os componentes z_i ($i = 1, \dots, N$) devem ser ortogonais e ordenados (de modo decrescente) pela variância das projeções, sendo, então, z_1 a projeção de máxima variância. Para tornar a variância independente da norma de \mathbf{v}_i , faz-se:

$$\mathbf{v}_i \leftarrow \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} \quad (\text{A.13})$$

Fazendo-se com que $\|\mathbf{v}_i\| = 1$, torna-se a variância função apenas da direção das projeções.

Como $\mathcal{E}\{\mathbf{x}\} = 0$, então $\mathcal{E}\{z_i\} = 0$, logo a variância da projeção z_i é calculada por $\mathcal{E}\{z_i^2\}$. Seguindo a definição da PCA, z_1 tem máxima variância; logo, \mathbf{v}_1 pode ser encontrado pela maximização de [4]:

$$J_1^{PCA}(\mathbf{v}_1) = \mathcal{E}\{z_1^2\} = \mathcal{E}\{(\mathbf{v}_1^T \mathbf{x})^2\} = \mathbf{v}_1^T \mathcal{E}\{\mathbf{x}\mathbf{x}^T\} \mathbf{v}_1 = \mathbf{v}_1^T \mathbf{C}_x \mathbf{v}_1, \quad (\text{A.14})$$

onde \mathbf{C}_x é a matriz de covariância de \mathbf{x} .

A solução para o problema de maximização da equação (A.14) pode ser encontrada na álgebra linear, em função dos autovetores $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N$ da matriz \mathbf{C}_x . A ordem dos autovetores é tal que os autovalores associados satisfazem $d_1 > d_2 > \dots > d_N$. Desta forma, tem-se:

$$\mathbf{v}_i = \mathbf{e}_i, \quad 1 \leq i \leq N \quad (\text{A.15})$$

Percebe-se que a PCA de \mathbf{x} e a decomposição por autovalores da matriz \mathbf{C}_x (de dimensão $N \times N$) são equivalentes. Limitações computacionais na extração das componentes principais utilizando as equações (A.12) e (A.15) aparecem quando a dimensão N do vetor \mathbf{x} aumenta, pois o processo de obtenção dos autovetores se torna proibitivamente lento. Nesse caso, uma solução é utilizar métodos iterativos de extração das componentes principais, através de redes neurais [218, 219].

A.3.2 Redução de Dimensão

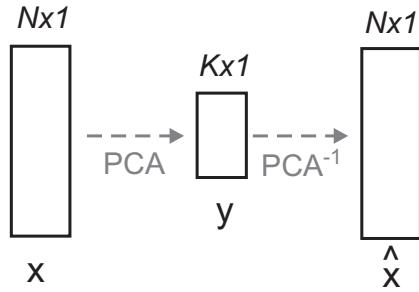


Figura A.3: Compressão e recuperação do sinal \mathbf{x} utilizando a transformação por PCA.

A principal aplicação da PCA é a compactação da informação. A redução de dimensão é obtida utilizando-se para a reconstrução do sinal original \mathbf{x} um número K de componentes principais sendo $K < N$. Na Figura A.3 é ilustrado o processo de redução de dimensão utilizando análise de componentes principais. Em geral, o número de componentes é escolhido visando a preservar uma parcela V_e da energia total, de modo que $\hat{\mathbf{x}} \approx \mathbf{x}$. A variância explicada V_e de um conjunto de componentes pode ser calculada usando-se:

$$V_e(K) = \frac{\sum_{i=1}^K d_i}{\sum_{i=1}^N d_i}, \quad (\text{A.16})$$

sendo d_i o autovalor da matriz \mathbf{C}_x de covariância do processo correspondente à componente i .

A transformação por PCA é ótima no sentido de representação do sinal nas primeiras componentes, mas não há garantia de que a compactação facilite o processo de classificação. Quando as direções de maior variância coincidem com as de melhor discriminação das classes, então a PCA é também útil para o reconhecimento de padrões, em caso contrário, a redução de dimensão pode dificultar a separação. Entretanto, em problemas de classificação onde a dimensão da entrada é excessivamente grande o pré-processamento por PCA reduz o custo computacional e consequentemente o tempo de processamento.

A.4 Análise de Componentes Independentes

A.4.1 Princípios de Estimação dos Componentes Independentes

No modelo básico da ICA (ver equação (5.4)) assume-se que a matriz \mathbf{A} é quadrada e não são considerados os atrasos temporais nem a existência de ruído aditivo. O princípio básico para a extração das componentes independentes é obtido do teorema do limite central. Como a soma de duas variáveis aleatórias independentes é sempre mais próxima de uma distribuição normal do que as variáveis originais, os sinais misturados x_i , que são gerados a partir do somatório ponderado das fontes s_i , têm distribuições de probabilidade mais semelhantes à Gaussiana quando comparadas aos sinais originais. As fontes podem ser obtidas então pela maximização da não-Gaussianidade.

Maximização da não-Gaussianidade

A **curtose** é o cumulante de quarta ordem, e para uma variável y de média zero e variância unitária é definida por [220]:

$$kurt(y) = \mathcal{E}\{y^4\} - 3(\mathcal{E}\{y^2\})^2. \quad (\text{A.17})$$

Variando no intervalo $[-2, \infty)$, a curtose é igual a zero para uma variável Gaussiana, os valores negativos indicam sub-Gaussianidade e os positivos super-Gaussianidade.

A curtose é um parâmetro estatístico facilmente calculado a partir das realizações da variável aleatória, porém seu valor pode ser bastante influenciado por um pequeno conjunto de pontos na cauda da distribuição [221], sendo, nesse caso, pouco robusta para a estimativa da não-Gaussianidade. Conhecidos como *intrusos* (ou *outliers*) esses pontos podem realmente pertencer à variável aleatória, ou ter sido artificialmente introduzidos por algum fenômeno desconhecido, como erro de medida ou de digitação.

Uma estimação alternativa da não-Gaussianidade pode ser obtida a partir da **negentropia**, que é calculada por [87]:

$$J(y) = H(y_{gauss}) - H(y), \quad (\text{A.18})$$

onde $H(\cdot)$ é a entropia, e y_{gauss} é uma variável aleatória Gaussiana com a mesma média e variância de y . A entropia é um dos conceitos básicos da teoria da informação e pode ser interpretada como o grau de informação contido em uma variável. Para uma variável aleatória discreta a entropia é definida como [222]:

$$H(Y) = - \sum_i P(Y = a_i) \log P(Y = a_i), \quad (\text{A.19})$$

onde os a_i são os possíveis valores da variável Y , e $P(Y = a_i)$ é a probabilidade de Y ser igual a a_i .

Um resultado importante obtido a partir da teoria da informação é que uma variável Gaussiana tem a máxima entropia entre todas as variáveis de mesma variância. Considerando a equação (A.18), a negentropia é sempre não negativa e zero quando a variável é Gaussiana, servindo como uma medição da não-Gaussianidade. O grande problema no cálculo de $J(\cdot)$ é a necessidade de se estimar as probabilidades da equação (A.19). Para evitar esse cálculo, utilizam-se aproximações da negentropia. Conforme descrito em [4], existem duas aproximações mais utilizadas para a negentropia, uma faz uso de cumulantes de ordem superior:

$$J(Y) \approx \frac{1}{12} E\{Y^3\}^2 + \frac{1}{48} kurt(Y)^2, \quad (\text{A.20})$$

e outra utiliza funções não-polinomiais [223]:

$$J(Y) \approx [k_1(E\{G_1(Y)\})^2 + k_2(E\{G_2(Y)\} - E\{G_2(\nu)\})^2], \quad (\text{A.21})$$

onde ν é uma variável aleatória Gaussiana de média zero e variância unitária. As funções não-lineares recomendadas em [223] são $G_1(y) = y \exp(-y^2/2)$ e $G_2(y) = |y|$ ou $G_2(y) = \exp(-y^2/2)$.

O uso de cumulantes traz de volta o problema da pouca robustez a *outliers*. É mostrado em [223] que o uso das funções não-polinomiais leva ao método da máxima entropia [4].

Minimização da Informação Mútua

Um outro método de estimação de ICA, também derivado da teoria da informação, é obtido pela minimização da informação mútua. A informação mútua $I(\cdot)$ entre m

variáveis aleatórias escalares y_i é definida como [98]:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(\mathbf{y}) \quad (\text{A.22})$$

A entropia $H(y_i)$ pode ser interpretada como o comprimento de código (ou a quantidade de informação) necessário para representar a variável y_i . Conforme a equação (A.22), a informação mútua é a diferença entre o somatório das entropias de cada uma das m variáveis y_i e a entropia do vetor aleatório $\mathbf{y} = [y_1, y_2, \dots, y_m]$. Pode-se provar que a codificação mais eficiente é obtida quando se utiliza o conjunto de variáveis \mathbf{y} . Utilizar as variáveis isoladamente sempre gera um maior código, menos quando as y_i são independentes, pois desta forma uma variável não carrega informação sobre as demais, sendo a informação mútua igual a zero. Desta forma, $I(y_1, y_2, \dots, y_m)$ pode ser utilizada como uma medida da dependência entre as variáveis. A matriz \mathbf{W} de transformação inversa da ICA, conforme equação ??, pode ser estimada através da minimização da informação mútua dos sinais s_i recuperados.

ICA através da Descorrelação Não-Linear

A igualdade da equação:

$$\mathcal{E}\{g(x)h(y)\} = \mathcal{E}\{g(x)\}\mathcal{E}\{h(y)\} \quad (\text{A.23})$$

repetida aqui para comodidade do leitor, garante que as variáveis x e y são independentes quando todas funções $g(\cdot)$ e $h(\cdot)$, integráveis em x e y são descorrelacionadas. Portanto, a extração das ICs pode ser obtida testando-se a correlação entre todas as funções não-lineares $g(\cdot)$ e $h(\cdot)$.

Existem alguns algoritmos propostos na literatura para o problema da decorrelação não-linear, como o *Hérault-Jutten* [4] e o *Chichocki-Unbehauen* [224], mas como não é possível testar a descorrelação entre todas as funções não-lineares, escolhem-se $f(\cdot)$ e $g(\cdot)$ visando-se a obter boas aproximações das componentes independentes. O algoritmo *Hérault-Jutten*, por exemplo, aconselha o uso de $f(y) = y^3$ e $g(y) = \operatorname{arctg}(y)$, já o *Chichocki-Unbehauen* sugere uma função polinomial e a tangente hiperbólica.

A PCA não-linear (NLPCA - *Non-linear Principal Component Analysis*) pode ser vista como uma extensão não linear da PCA, e é capaz de encontrar projeções

descorrelacionadas não-linearmente. Enquanto o objetivo da PCA é minimizar o erro médio quadrático de reconstrução do sinal projetando as componentes numa base ortonormal, a NLPCA pode ser definida de modo simples através da função-objetivo a ser minimizada:

$$J(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n) = \mathcal{E}\left\{\|\mathbf{x} - \sum_{i=1}^n g_i(\mathbf{w}_i^T \mathbf{x}) \mathbf{w}_i\|^2\right\}, \quad (\text{A.24})$$

onde $g_1(\cdot), g_2(\cdot), \dots, g_n(\cdot)$ é um conjunto de funções escalares e não-lineares, e os vetores \mathbf{w}_i formam a base do sub-espacô onde serão projetadas as entradas \mathbf{x} . Quando o mínimo de $J(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n)$ for encontrado, o produto $\mathbf{w}_i^T \mathbf{x}$ dará as componentes principais não-lineares. Se $g_i(y) = y$ para todo i , então equação (A.24) se reduz à função objetivo da PCA. Quando os sinais satisfazem ao modelo da ICA, mostrado na equação (5.4), a NLPCA obtém uma aproximação das componentes independentes.

A.4.2 Pré-Processamento dos Sinais para ICA

Em geral, os algoritmos de extração das componentes independentes têm seu trabalho simplificado quando os sinais são centralizados, ou seja, têm sua média removida fazendo-se:

$$\mathbf{x} \leftarrow \mathbf{x} - \mathcal{E}\{\mathbf{x}\} \quad (\text{A.25})$$

Outra transformação importante é o branqueamento. Um vetor $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ é dito branco quando os elementos z_i são descorrelacionados e têm variância unitária. O branqueamento pode ser realizado por uma transformação linear:

$$\mathbf{z} = \mathbf{V}\mathbf{x} \quad (\text{A.26})$$

O branqueamento, que é apenas a descorrelação seguida de uma normalização, pode ser realizado por uma transformação através de PCA. Com as variáveis branqueadas a extração da ICA é facilitada, pois os sinais já estão descorrelacionados.

Em problemas com vetores de entrada de alta dimensão, é importante a compactação da informação através de PCA ou Análise de Relevância para facilitar o processo de extração das componentes independentes.

A.4.3 Principais Algoritmos para ICA

Diversos algoritmos vêm sendo propostos para a extração das componentes independentes. Essas rotinas diferem basicamente no princípio teórico no qual fundamentam a obtenção das componentes independentes (não-Gaussianidade, informação mútua, descorrelação não-linear, etc) e na forma fazem a otimização da função objetivo escolhida. Os principais parâmetros para avaliação de desempenho são o tempo de processamento (complexidade computacional) e a precisão na extração das componentes.

Um estudo comparativo entre diversos métodos de estimativa das componentes independentes foi realizado em [4]. O algoritmo **FastICA**, descrito com detalhes em [4] e [98], é o que apresenta menor custo computacional. Algoritmos que utilizam descorrelação não linear e NLPCA têm desempenho semelhante ao FastICA em termos da precisão na obtenção da matriz \mathbf{W} , porém exigem maior esforço de computação. O algoritmo **JADE** (*Joint Approximate Diagonalization of Eigenmatrices*) proposto em [149] também é muito utilizado em ICA, mostrando bons resultados.

Algoritmo FastICA

Considerando as aproximações da negrentropia mostradas nas Equações (A.20) e (A.21), e o fato de que a minimização da negrentropia leva à independência estatística, no trabalho [151] foram propostos algoritmos de ponto fixo para ICA (chamados FastICA), que utilizam iterações semelhantes às de Newton [225]. Entre as vantagens deste algoritmo pode-se citar simplicidade computacional, baixa utilização de memória e boas características de convergência [98].

A partir de algumas manipulações da equação (A.21), o algoritmo FastICA para estimativa de uma componente independente é formulado a seguir para sinais pré-branqueados:

1. Escolha um vetor de pesos inicial \mathbf{w} de modo aleatório;
2. Faça $\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - E\{g'(\mathbf{w}^T \mathbf{x})\}\mathbf{w}$;
3. $\mathbf{w} = \mathbf{w}^+ / \| \mathbf{w}^+ \|$;

4. Se o algoritmo não tiver收敛ido voltar para o passo 2.

Os autores sugerem o uso de uma das funções $g(\cdot)$ a seguir:

$$g_1(x) = \tgh(a_1x), \quad (\text{A.27})$$

$$g_2(x) = x \exp(-a_2 u^2/2), \quad (\text{A.28})$$

$$g_3(x) = x^3, \quad (\text{A.29})$$

onde $1 \leq a_1 \leq 2$ e $a_2 \approx 1$. A escolha da função não-linear pode ser guiada pelas características a seguir [151]: a função $g_1(\cdot)$ é indicada quando não há informação a respeito da estatística das componentes independentes, pois o algoritmo apresenta resultados satisfatórios para qualquer tipo de distribuição; o uso de $g_2(\cdot)$ é indicado quando as componentes independentes são super-Gaussianas e o $g_3(\cdot)$ deve ser utilizada para estimar componentes sub-Gaussianas.

Para estimar mais de uma componente independente pode-se utilizar métodos de ortogonalização deflacionária como o de Gram-Schmidt [4].

Algoritmo JADE

No algoritmo JADE (*Joint Approximate Diagonalization of Eigenmatrices*), as informações estatísticas de segunda e quarta ordem são utilizadas a partir de uma abordagem tensorial. Tensores [226] são generalizações de alta-dimensão das matrizes. O tensor cumulante de quarta ordem \mathbf{T}_4 é uma “matriz” de quatro dimensões onde cada elemento é definido por $q_{ijkl} = \text{cum}(x_i, x_j, x_k, x_l)$, os índices i,j,k e l variaram de 1 até N (onde N é o número de sinais) e $\text{cum}(x_i, x_j, x_k, x_l)$ é o cumulante de quarta ordem:

$$\begin{aligned} \text{cum}(x_i, x_j, x_k, x_l) &= E\{x_i, x_j, x_k, x_l\} - E\{x_i, x_j\}E\{x_k, x_l\} \\ &\quad - E\{x_i, x_k\}E\{x_j, x_l\} - E\{x_k, x_j\}E\{x_i, x_l\} \end{aligned} \quad (\text{A.30})$$

Sabe-se que a diagonalização da matriz de correlação (\mathbf{C}_y) produz a descorrelação entre os componentes de \mathbf{y} [4]. Para sinais independentes, apenas quando $i=k=j=l$ os cumulantes de quarta-ordem são diferentes de zero. Considerando isso, os métodos Tensoriais de ICA propõe a diagonalização de \mathbf{T}_4 para alcançar a independência estatística [149].

Embora teoricamente simples, a utilização de métodos tensoriais de ICA exigem uma grande quantidade de recursos computacionais para a decomposição em

auto-valores de matrizes de quarta-ordem. O algoritmo JADE propõe um método aproximado para a diagonalização de \mathbf{T}_4 , se tornando mais leve computacionalmente.

Considerando que os dados satisfazem o modelo da ICA para dados pré-branqueados, pode-se escrever:

$$\mathbf{z} = \mathbf{V}\mathbf{A}\mathbf{s} = \mathbf{W}^T\mathbf{s} \quad (\text{A.31})$$

onde $\mathbf{x} = \mathbf{As}$ são os sinais observados, \mathbf{V} é a matriz de branqueamento e $\mathbf{W}^T = VA$ é a matriz de misturas branqueada. Neste caso, pode-se provar (ver [4] que o tensor cumulante de \mathbf{z} tem uma estrutura especial e suas auto-matrizes são descritas por:

$$\mathbf{M} = \mathbf{w}_m \mathbf{w}_m^T \quad (\text{A.32})$$

onde $m=1,\dots,N$ e w_n são as colunas da matriz \mathbf{W}^T

O algoritmo JADE utiliza a transformação linear F_{ij} da matriz \mathbf{M} definida por:

$$F_{i,j}(\mathbf{M}) = \sum m_{kl} \text{cum}(x_i, x_j, x_k, x_l) \quad (\text{A.33})$$

onde m_{kl} é um elemento da matriz \mathbf{M} .

A decomposição em autovalores é vista como um processo de diagonalização, então busca-se a matriz \mathbf{W} que diagonaliza $F(\mathbf{M})$ para qualquer \mathbf{M} (ou seja, $Q = \mathbf{W}F(\mathbf{M}_i)\mathbf{W}^T$ é uma matriz diagonal).

A função custo do método JADE busca a diagonalização de \mathbf{Q} pela maximização da soma dos elementos de sua diagonal. As matrizes \mathbf{M}_i utilizadas são as automatrizes do tensor cumulante dos dados, pois assim tem-se um conjunto de N matrizes que contém toda a informação relevante a respeito dos cumulantes.

Os métodos tensoriais [227, 149] foram, provavelmente, a primeira classe de algoritmos capazes de executar a ICA de modo realmente eficiente [4]. Atualmente, estes métodos são mais utilizados para sinais de baixa dimensão, pois o custo computacional aumenta rapidamente com o número de componentes a serem estimados.

Algoritmo Multiplicativo com Iteração de Newton

Um algoritmo multiplicativo para ICA foi proposto por Akuzawa e Murata em [228]. Usando a curtose como função para avaliar a independência, esse algoritmo utiliza

estratégia de otimização de segunda ordem, através do método de Newton [225], para estimar as componentes independentes.

O algoritmo proposto por Akuzawa não requer pré-branqueamento, operando diretamente sobre os dados medidos. Resultados experimentais obtidos em [229, 230] indicam que o algoritmo de Akuzawa apresenta melhor desempenho que FastICA e JADE quando os sinais estão contaminados por ruído Gaussiano.

Considerando a transformação linear $\mathbf{Y} = \mathbf{C}\mathbf{X}$, o objetivo do algoritmo de Akuzawa é encontrar a matriz \mathbf{C} que maximiza a independência entre as componentes de \mathbf{y} . Os passos a seguir são executados durante as iterações:

1. Escolher C_0 (a matriz de separação inicial) e a matriz Δ_0 ($N \times N$);
2. Calcular a iteração $C_t = \exp(\Delta_t - 1)C_{t-1}$;
3. Avalie a função custo em C_t usando uma expansão de segunda ordem em torno de C_{t-1} ;
4. Δ_t é escolhido como o ponto de sela da função custo;
5. Retorne para o passo 2 até convergir.

Mais detalhes a respeito da execução do passo 4 podem ser encontradas em [228]. Modificações no método de Akuzawa foram propostos em [229] com o objetivo de reduzir o custo computacional pela substituição das iterações de Newton pelo método quasi-Newton [225].

Comparado com FastICA e JADE, o algoritmo multiplicativo de Akuzawa é mais lento (mesmo em sua versão modificada que utiliza o método de otimização quasi-Newton), suas vantagens aparecem quando o nível de ruído aumenta, neste caso o método de Akuzawa apresenta melhores resultados.

A.4.4 Extensões ao Modelo Básico de ICA

A análise de componentes independentes em sua formulação básica mostrada nas equações (5.4) e (??) é empregada com sucesso em uma grande variedade de aplicações. O modelo básico de ICA não considera que os sinais podem estar contaminados por ruído aditivo, ou que o sinal misturado seja gerado de forma não-linear.

No ambiente do calorímetro do ATLAS, o ruído estará presente nas células absorvedoras de energia, e por características de construção do sistema, transformações não-lineares dos sinais poderão estar presentes.

ICA para Sinais Ruidosos

Sabe-se que os sinais reais no ambiente do calorímetro do ATLAS estarão contaminados por ruído. A formulação básica da ICA, mostrada na equação (5.4), não considera a presença de ruído. Um modelo mais realista permite a obtenção do modelo da ICA para sinais ruidosos (*Noisy ICA*) [4]:

$$\mathbf{x} = \mathbf{As} + \mathbf{n}, \quad (\text{A.34})$$

onde $\mathbf{n} = (n_1, n_2, \dots)$ é o vetor de ruído. Em geral, considera-se que o ruído é gaussiano e independente das componentes independentes. Assume-se também que a matriz de covariância do ruído $\Sigma = \sigma^2 I$ é conhecida.

A equação (A.34) pode ser associada a ruído de sensores, uma vez que o ruído n_i é adicionado separadamente a cada fonte x_i . O ruído de fonte, que é adicionado às componentes independentes s_i é representado pela equação a seguir:

$$\mathbf{x} = \mathbf{A}(s + \mathbf{n}) \quad (\text{A.35})$$

Um caso especial que pode simplificar bastante o problema de estimação das componentes independentes em ambiente ruidoso acontece quando existem poucas fontes de ruído. Se o número total de componentes de ruído é menor que o de componentes independentes, o modelo básico da ICA pode ser adaptado. Define-se o vetor $\tilde{\mathbf{s}} = (s_1, \dots, s_k, n_1, \dots, n_l)^T$, onde s_i ($i = 1, \dots, k$) são as componentes independentes e n_j ($j = 1, \dots, l$) são as fontes de ruído. Assumindo-se que o número de misturas é $k + l$, então o modelo básico da ICA pode ser aplicado a $x = A\tilde{s}$ e usando-se um algoritmo deflacionário consegue-se estimar as k componentes menos gaussianas, que são as próprias ICs.

Na maioria das vezes, deseja-se considerar que o ruído foi adicionado a cada uma das misturas; assim, $k + l$ é maior que o número de misturas, e o modelo básico da ICA não pode ser aplicado ao vetor \tilde{s} .

O processo de estimação das componentes independentes já é bastante difícil na formulação básica, e quando o ruído é considerado o cenário piora consideravelmente. Nos modelos das equações (A.34) e (A.35) pode-se verificar que as ICs não são obtidas apenas com a inversão da matriz de mistura \mathbf{W} :

$$\mathbf{Wx} = \mathbf{s} + \mathbf{Wn} \quad (\text{A.36})$$

A solução do problema da equação (A.36) envolve duas etapas: o uso de métodos de otimização numérica para obter uma aproximação do ruído e, a partir de então, a estimação das componentes independentes. Em uma abordagem mais simples e que parece ser a mais promissora para o *noisy ICA*, são utilizadas técnicas de remoção de tendência para reduzir os efeitos do ruído, adaptando-se os métodos básicos de ICA para o caso ruidoso [231].

O algoritmo FastICA pode ser adaptado para sinais contaminados por ruído usando medidas não tendenciosas da Gaussianidade [232]:

$$\mathbf{w}^* = E\{\tilde{\mathbf{x}}g(\mathbf{w}^T)\tilde{\mathbf{x}}\} - (I + \tilde{\Sigma})\mathbf{w}E\{g'(\mathbf{w}^T\tilde{\mathbf{x}})\}, \quad (\text{A.37})$$

sendo que a norma do novo valor \mathbf{w}^* é tornada unitária após cada iteração, e $\tilde{\Sigma}$ é dado por:

$$\tilde{\Sigma} = E\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T\} = (\mathbf{C} - \Sigma)^{-1/2}\Sigma(\mathbf{C} - \Sigma)^{-1/2} \quad (\text{A.38})$$

onde $\mathbf{C} = E\{\mathbf{x}\mathbf{x}^T\}$ é a matriz de covariância dos sinais ruidosos observados. A função $g(.)$ pode ser escolhida entre as abaixo:

Outras técnicas como o “encolhimento de código esparso” (*Sparse Code Shrinkage*) [233] também são utilizadas para a extração das componentes independentes para sinais ruidosos. Em casos onde o nível de ruído é muito alto, pode-se tentar o uso de técnicas de processamento de sinais como *wavelets* ou filtragem adaptativa [234].

A.5 ICA Não-Linear

Conforme mostrado no Capítulo ??, o modelo da ICA não-linear (NLICA) apresenta uma formulação mais geral que o linear. A seguir será mostrado o desenvolvimento

teórico de um algoritmo para a estimação das componentes independentes no modelo pós não-linear.

A.5.1 Algoritmo Taleb-Jutten para o Modelo PNL

Um dos primeiros algoritmos para o modelo pós não-linear da ICA foi proposto por Taleb e Jutten no trabalho [108]. Este algoritmo é robusto a variações na distribuição de probabilidade das fontes, pois executa estimativa iterativa da estatística das componentes independentes estimadas através do cálculo da função escore:

$$\psi = p'_{Yi}(u)/p_{Yi}(u), \quad (\text{A.39})$$

conforme Figura A.4.

Cada função não-linear g_k ($k=1,\dots,N$) é modelado por redes MLP com um neurônio linear na saída:

$$g_k(u) = \sum_{h=1}^{N_H} \xi^h \sigma(\omega^h u - \eta^h), \quad (\text{A.40})$$

onde N_H é o número de neurônios ocultos. A divergência de Kullback-Lieber é utilizada para encontrar as regras de aprendizado para a estimação das funções não-lineares [108].

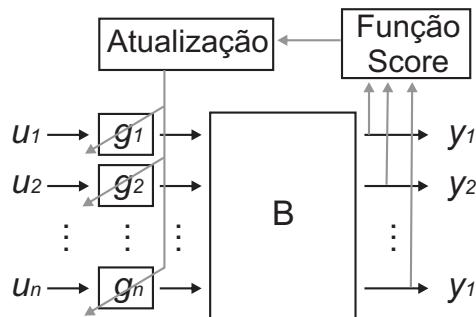


Figura A.4: Diagrama do algoritmo de Taleb-Jutten para o modelo PNL.

Como existem vários parâmetros a serem ajustados no modelo inverso proposto e a otimização envolve funções não-lineares, o algoritmo pode apresentar problemas de convergência para mínimos locais [99]. Diferentes procedimentos foram propostos na literatura para melhorar a eficiência de estimação em modelos PNL. Em [115, 114] um algoritmo genético [159] foi utilizado para executar uma busca global, evitando

o problema dos mínimos locais. O problema com esta abordagem é o aumento do custo computacional.

Redes neurais com arquiteturas alternativas também foram aplicadas com sucesso na separação de misturas PNL. Por exemplo, em [235] funções de base radial (RBF - *Radial Basis Function*). Em um outro trabalho [112], a separação foi realizada por redes neurais com funções de ativação do tipo *spline*.

Apêndice B

Conceitos Fundamentais em Classificação de Sinais

A seguir serão mostrados os fundamentos teóricos de algumas técnicas de classificação de padrões, iniciando-se com uma visão geral do problema de decisão binária. Serão apresentadas técnicas lineares como Filtros Casados e Análise de Discriminantes, e não-lineares, como Redes Neurais.

B.1 Teste de Hipóteses

Considerando-se inicialmente a discriminação entre duas hipóteses H_1 e H_0 , o problema de classificação pode ser resumido pelo esquema da Figura B.1. A fonte gera as saídas, que após passarem por um meio probabilístico, precisam ser detectadas a partir das observações do processo. As regras de decisão, que formam o sistema classificador, são projetadas para maximizar a probabilidade de detecção correta.

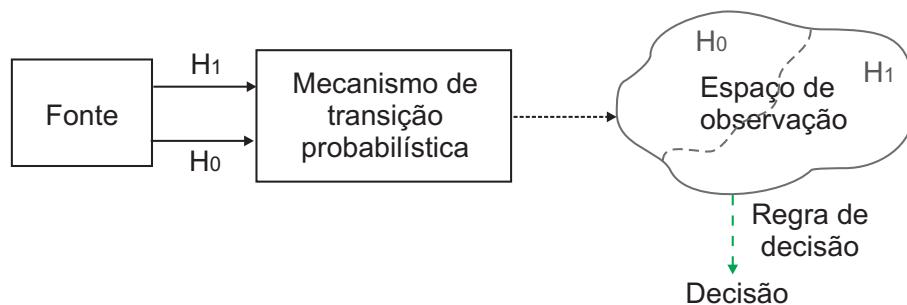


Figura B.1: Esquemático do problema de classificação binário.

No caso da decisão binária, cada vez que uma observação é efetuada 4 situações podem ocorrer:

decidir pela hipótese H_1 , sendo H_0 verdadeira;

decidir pela hipótese H_0 , sendo H_1 verdadeira;

decidir pela hipótese H_1 , sendo H_1 verdadeira;

decidir pela hipótese H_0 , sendo H_0 verdadeira.

As duas primeiras são erros de decisão e as duas últimas classificações corretas. Cada uma das hipóteses é associada a uma saída da fonte, que é mapeada em uma região do espaço de observação. Considerando um espaço de observação de dimensão N finita, um ponto neste espaço pode ser representado por um vetor:

$$\mathbf{r} = [r_1, r_2, \dots, r_N]. \quad (\text{B.1})$$

O mecanismo de transição probabilística gera pontos de acordo com as densidades de probabilidade condicionais $P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)$ e $P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)$. Quando essas probabilidades são conhecidas ou podem ser estimadas de alguma forma, o projeto do sistema classificador pode ser simplificado. Os critérios de *Bayes* e *Neyman-Pearson* são procedimentos clássicos utilizados para a escolha da regra de decisão.

B.2 Critério de Bayes

O critério de *Bayes* necessita do conhecimento das probabilidades a priori P_1 e P_0 de a fonte produzir H_1 ou H_0 , das probabilidades condicionais $P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)$ e $P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)$ e dos custos C_{ij} associados à escolha da hipótese i sendo j a verdadeira. O risco é, então, definido como [186]:

$$\begin{aligned} \mathfrak{R} = & C_{00}P_0 \int_{Z_0} P_{\mathbf{r}/H_0}(\mathbf{R}/H_0) d\mathbf{R} \\ & + C_{10}P_0 \int_{Z_1} P_{\mathbf{r}/H_0}(\mathbf{R}/H_0) d\mathbf{R} \\ & + C_{11}P_1 \int_{Z_1} P_{\mathbf{r}/H_1}(\mathbf{R}/H_1) d\mathbf{R} \\ & + C_{01}P_1 \int_{Z_0} P_{\mathbf{r}/H_1}(\mathbf{R}/H_1) d\mathbf{R} \end{aligned}, \quad (\text{B.2})$$

onde os elementos do espaço de observação que pertencem às partições Z_0 e Z_1 são associados, respectivamente, a H_0 e H_1 . As variáveis C_{ij} representam o custo da

escolha da hipótese i quando a hipótese verdadeira é a j . Em geral assume-se que o custo de uma decisão errada (C_{ij} sendo $i \neq j$) é maior do que o de um acerto (C_{ij} sendo $i = j$).

Minimizando o risco \mathfrak{R} da equação (B.2) chega-se a [186]:

$$\frac{P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)}{P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)} \gtrless_{H_0}^{H_1} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}; \quad (\text{B.3})$$

a expressão à esquerda é chamada razão de semelhança ($\Lambda(\mathbf{R})$) e a fração à direita é o valor limiar (patamar) do teste (κ). Com isso, a equação (B.3) se reduz a:

$$\Lambda(\mathbf{R}) \gtrless_{H_0}^{H_1} \kappa; \quad (\text{B.4})$$

então, se a razão de semelhança é maior que o patamar, decide-se por H_1 , caso contrário, escolhe-se H_0 .

Quando os custos não são conhecidos, pode-se adotar o critério *minimax*, que minimiza o risco máximo; após algumas considerações chega-se a:

$$\begin{aligned} C_{00} &= C_{11} = 0 \\ C_{01}P_M &= C_{10}P_F \end{aligned}, \quad (\text{B.5})$$

onde $P_F = \int_{Z_1} P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)d\mathbf{R}$ é a probabilidade de falso-alarme (terminologia usada em sistemas de radar, indicando que decidiu-se pela presença do alvo H_1 estando o mesmo ausente) e $P_M = \int_{Z_0} P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)d\mathbf{R}$ é a probabilidade de perda do alvo.

O teste de *Neyman-Pearson* [236] é utilizado quando não se tem informações sobre os custos ou as probabilidades a priori. Escolhe-se um valor limite para a probabilidade de falso-alarme e procura-se minimizar a probabilidade de perda do alvo para o valor escolhido. Como o critério utiliza P_F e P_M é preciso conhecer as probabilidades condicionais $P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)$ e $P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)$.

B.3 Filtros Casados

Quando o processo de detecção é realizado em ambiente ruidoso o desempenho dos classificadores decaiu bastante. Os filtros casados são utilizados em processos de

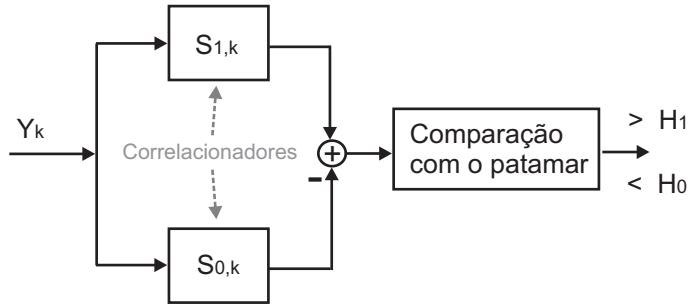


Figura B.2: Diagrama em blocos de um filtro casado.

detecção onde as formas de onda dos sinais corrompidos por ruído são conhecidas. Os filtros casados maximizam a razão sinal-ruído em ambiente de ruído branco. Um processo estocástico é dito branco quando sua densidade espectral de potência $S(\omega)$ é constante em todas as freqüências [220]:

$$S(\omega) = N_0/2 \quad (\text{B.6})$$

A autocorrelação do processo branco, calculada a partir da transformada de Fourier inversa de $S(\omega)$ é $R_{xx} = N_0/2\delta(t)$, ou seja, o ruído branco tem correlação nula com qualquer cópia deslocada dele mesmo.

No caso de sinais digitais, conforme mostrado na Figura B.2, o filtro casado é implementado através do cálculo da correlação do sinal recebido $y[k]$ com as formas de ondas esperadas $s_1[k]$ e $s_0[k]$. As saídas dos correlacionadores são subtraídas e comparadas com um patamar de decisão; se o valor for maior que o patamar decide-se por H_1 , caso contrário por H_0 .

Se o ruído de fundo não for branco, o desempenho dos filtros cai bastante. Esse problema é minimizado se o ruído é conhecido. Pode-se, então, projetar um filtro branqueador. A desvantagem do branqueador é que ele atua também sobre o sinal, alterando suas características estatísticas.

Embora muito utilizadas, por sua formulação matemática relativamente simples e bom desempenho em diversas aplicações, as técnicas mostradas até aqui apresentam as desvantagens de necessitarem conhecimento prévio a respeito das distribuições de probabilidade, custos ou forma de onda dos sinais de entrada. Em muitos casos práticos essas informações não estão disponíveis, sendo necessária a utilização de outros métodos de classificação. Além disso, o desempenho dessas técnicas piora

bastante em ambiente de ruído colorido e também quando as classes estão muito sobrepostas.

B.4 Discriminante Linear de Fisher

A análise de discriminantes busca a direção \mathbf{w} onde as projeções \mathbf{y} dos sinais de entrada \mathbf{x} sejam maximamente separáveis. A análise por discriminante de Fisher (FDA - *Fisher Discriminant Analysis*) busca a direção ótima de discriminação utilizando 2 parâmetros: a distância inter-classes, e a distância intra-classes [237].

Numa formulação matricial o objetivo é encontrar a direção \mathbf{w}_0 que maximiza a expressão:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \quad (\text{B.7})$$

onde \mathbf{m}_i é a média da classe i , $\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$ é a matriz de separação inter-classes e $\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2$ é a matriz de separação intra-classes, sendo:

$$\mathbf{S}_i = \sum_{\mathbf{x} \in \mathcal{D}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \quad (\text{B.8})$$

Pode-se provar que a direção ótima que maximiza (B.7) é dada por [136]:

$$\mathbf{w} = \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \quad (\text{B.9})$$

O discriminante de Fisher é capaz de encontrar a transformação linear ótima dos sinais de entrada de modo que os sinais projetados $\mathbf{y} = \mathbf{w}^T \mathbf{x}$ tenham máxima separação. Pode-se realizar a análise por discriminante de Fisher de modo analítico usando-se as equações (B.7), (B.8) e (B.9), ou de modo iterativo a partir de uma rede neural de uma camada.

B.5 Classificadores Neurais

As redes neurais artificiais (RNA) [5] são modelos matemáticos que emulam algumas características do cérebro humano, sendo capazes de adquirir conhecimento (aprender) e generalizar (responder corretamente a estímulos novos). Devido ao poder computacional, obtido de sua estrutura paralelamente distribuída, e às habilidades

de aprender e generalizar, as RNAs vêm sendo utilizadas em diversas aplicações como reconhecimento de padrões e classificação [238], processamento de sinais [239], aproximação de funções [240], controle e identificação de sistemas [241]. A seguir serão descritos os fundamentos matemáticos das redes neurais e sua aplicação como classificadores. Maior ênfase será dada às redes alimentadas adiante (*feedforward neural networks*) com treinamento supervisionado. Os livros [5] e [242] fornecem textos mais abrangentes sobre assunto.

Uma diferença fundamental entre os classificadores neurais e os métodos clássicos como os filtros casados é que nestes últimos é necessário formular um modelo matemático a partir dos sinais. Na abordagem neural, o classificador trabalha diretamente no conjunto de dados, ficando o modelo matemático implícito nos valores dos pesos sinápticos obtidos após o treinamento. Os classificadores neurais têm obtido resultados bastante animadores em problemas para os quais métodos clássicos não apresentam bom desempenho.

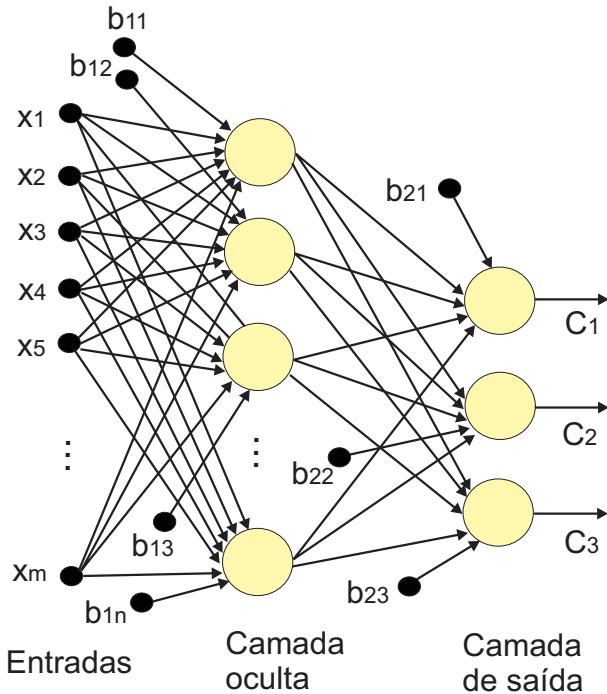


Figura B.3: Exemplo de uma rede neural utilizada para separação dos sinais de entrada em 3 classes.

Para o projeto do classificador neural dividem-se os pares entrada saída disponíveis nos conjuntos de treino, teste e validação. A divisão é importante para garantir que a rede treinada consiga generalizar bem o conhecimento adquirido. Apesar disso, é comum que a rede seja treinada com todos os dados disponíveis, e que o resultado seja validado com um conjunto de teste.

nas o conjunto de treino é considerado para o ajuste dos pesos sinápticos. O conjunto de validação é utilizado na parada do treinamento em caso de sobre-aprendizado. O sobre-aprendizado acontece quando a rede se ajusta demais ao conjunto de treino, perdendo capacidade de generalização. O resultado esperado do classificador é avaliado através do conjunto de teste, que não foi usado no ajuste dos pesos nem na parada do algoritmo.

Para decisão entre duas classes, pode-se usar na camada de saída um neurônio tipo tangente hiperbólica, associando-se $y = +1$ para uma classe e $y = -1$ para a outra. Para decisão entre N classes ($N \geq 3$), pode-se fazer com que a camada de saída tenha N neurônios. Associa-se um neurônio para cada classe. Quando a classe k estiver presente, a saída desejada no neurônio k associado deve ser $+1$ e nos demais -1 . Na fase de operação decide-se pela classe do neurônio com a maior saída. A rede da Figura B.3 é um exemplo de arquitetura que pode ser usada para a separação entre três classes a partir de um espaço de entrada de dimensão m .

Considerando uma rede de duas camadas (ver Figura B.3), a camada oculta é responsável por um mapeamento não-linear das entradas. Este mapeamento busca a máxima separação das classes diferentes, facilitando o processo de classificação realizado pela camada de saída.

Apêndice C

Algoritmos Genéticos

C.1 Algoritmo Genético como Método de Otimização

Os Algoritmos Genéticos (AG) pertencem à classe dos Algoritmos Evolutivos, que são ferramentas de busca e otimização bastante difundidas, utilizando operadores que simulam a evolução dos seres vivos para buscar a solução de problemas complexos [243].

Comparado com outros métodos de busca e otimização os AG's se destacam quando o problema tratado tem características altamente não lineares e o ambiente de busca é multi-modal (existem diversos máximos ou mínimos locais na função custo). Por ser um método de busca global, o AG tem menor probabilidade de convergir para soluções sub-ótimas (máximos/mínimos locais).

Outra vantagem é a facilidade de implementação, pois, em geral, os métodos numéricos precisam que a função objetivo $f(x)$ seja conhecida e diferenciável, pois se baseiam no cálculo das derivadas parciais. Já o AG utiliza apenas a função objetivo.

Idealizados inicialmente por John Holland e seus alunos nos anos 70 [244], na Universidade de Michigan, os Algoritmos Genéticos são métodos de busca baseados na teoria da evolução. O objetivo inicial de Holland era utilizar os mecanismos de evolução, adaptação e seleção natural dos seres vivos como modelos para sistemas computacionais.

Na natureza os indivíduos apresentam qualidades particulares que podem torná-los mais ou menos aptos para sobreviver às adversidades do meio ambiente. A seleção natural e os operadores genéticos, como a reprodução e a mutação, podem ocasionar o surgimento de exemplares com características mais favoráveis, melhorando as chances de sobrevivência da população.

Nos problemas computacionais, traçando um paralelo com a natureza, as possíveis soluções são chamadas cromossomos ou indivíduos, cada um possuindo características que o fazem melhor ou pior candidato à resolução do problema. Os melhores são selecionados para reproduzirem-se gerando uma nova geração. Espera-se que a cada iteração apareçam indivíduos mais aptos. A mutação, por sua vez, introduz aleatoriamente informações não existentes no conjunto original, contribuindo para a manutenção da diversidade populacional (explorando outras regiões do espaço de busca). Algumas características interessantes dos AG's esão listadas a seguir:

1. trabalham com os parâmetros codificados;
2. realizam a busca a partir de uma população de soluções candidatas (processamento em paralelo);
3. utilizam informações da função objetivo, e não de suas derivadas;
4. usam regras probabilísticas e não determinísticas.

C.2 Estrutura de um Algoritmo Genético

Não existe uma definição rigorosa para os Algoritmos Genéticos, entretanto, a maioria dos métodos assim denominados tem em comum as seguintes características: população composta de cromossomos, seleção de acordo com a função de adequabilidade, recombinação e mutação para a produzir uma nova geração [245]. A seguir tem-se uma breve descrição dos conceitos mais importantes na teoria dos Algoritmos Genéticos.

C.2.1 Conceitos Principais

- **Cromossomo** - Em geral, na teoria dos A.G.'s cada indivíduo é constituído de um único cromossomo, onde cada elemento equivale a um gene. Pode-se

utilizar codificação binária simples, códigos de gray, e até codificação decimal. A codificação binária é a mais comum, nela cada gene pode assumir apenas dois valores 0 ou 1. Os cromossomos podem ser vistos como um ponto do espaço de busca das prováveis soluções do problema.

- **Codificação** - As variáveis a serem otimizadas precisam ser codificadas nos cromossomos. A quantidade de bits destinada a cada variável deve ser escolhida de forma a minimizar o esforço computacional, mas sempre levando em conta a precisão desejada. Na figura C.1 tem-se a representação de um cromossomo binário, onde cada parâmetro é codificado por 4 bits distintos.

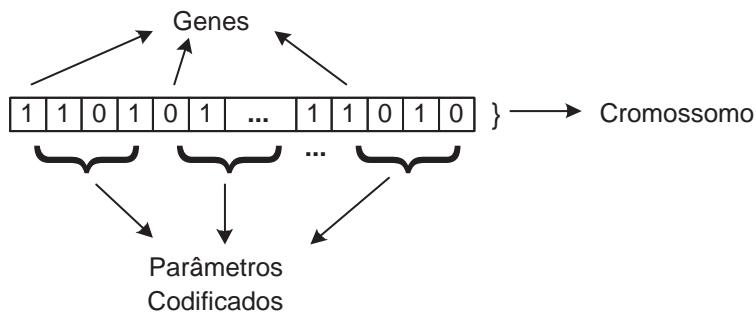


Figura C.1: Exemplo de um cromossomo binário

- **População** - Conjunto de cromossomos gerado inicialmente de forma aleatória, a população é atualizada a cada nova iteração do algoritmo, espera-se que a cada geração apareçam indivíduos mais aptos para a solução do problema. Em geral, usam-se populações de tamanho fixo.
- **Função Aptidão (*Fitness Function*)** - É na Função Aptidão ($fit(x)$) que são codificadas as informações sobre o sistema a ser otimizado. A cada geração são calculados os valores da Função Aptidão de todos os indivíduos. Esta informação será usada como base do processo de seleção.
- **Seleção** - Simula o processo de seleção natural, os indivíduos mais aptos, com maior valor da função aptidão, têm mais chances de serem escolhidos para o processo de recombinação.
- **Recombinação (*Crossover*)** - Promove a troca de informação genética. Os indivíduos são selecionados dois a dois, simulando um processo de reprodução

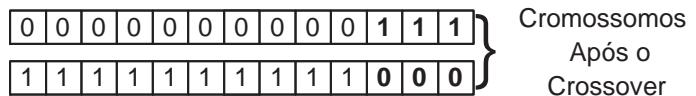
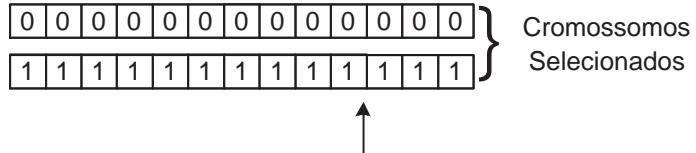
sexuada, e parte do material genético (bits) é trocado. Existem vários modos de proceder esta troca de bits, cada um dá origem a um tipo diferente de recombinação, a seguir temos a descrição dos tipos mais comuns de *crossover* em cromossomos binários:

- *Single-Point Crossover* - O modo mais simples e comum de recombinação, é escolhido aleatoriamente um ponto no cromossomo e são trocados os bits após este ponto.
- *Multi-Point Crossover* - Semelhante à recombinação de ponto único, pontos aleatórios são escolhidos, e trocados os bits dos segmentos que ficam entre estes pontos.
- *Uniform Crossover* - Alguns pontos são escolhidos de modo randômico e trocados os bits destes pontos. É gerado um cromossomo aleatório, as posições onde o valor do bit é igual a 1 serão trocadas. O *uniform crossover* foi proposto inicialmente em [246] e [247]. Na figura C.2 são ilustrados os processos de recombinação de ponto único e uniforme.
- **Mutação** - A mutação é responsável pela manutenção da diversidade genética, inserindo, de modo aleatório, novas informações nos indivíduos. No caso de codificação binária, o bit selecionado tem o seu valor invertido.
- **Fenótipo** - Em analogia com o termo biológico, o fenótipo de um indivíduo pode ser expresso pelo valor da função aptidão, ou pelos parâmetros decodificados.

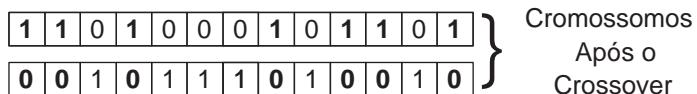
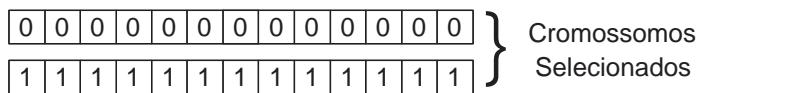
A recombinação e a mutação são controlados por parâmetros específicos, p_{rec} e p_{mut} , respectivamente. Estes parâmetros estabelecem a taxa de ocorrência dos processos. Exemplo: para $p_{rec} = 0.7$, a cada geração, 70% dos indivíduos são selecionados para participarem da recombinação, com $p_{mut} = 0.001$, a cada geração 0,1% dos bits sofrerão mutação.

C.2.2 Escalonamento de Aptidão

É importante salientar que problemas com a pressão seletiva (rigor com o qual é realizado o processo de seleção) do algoritmo podem levar à demora na convergência



(a)



(b)

Figura C.2: (a) Recombinação em ponto único e (b) recombinação uniforme.

ou à convergência prematura. A última se caracteriza quando o AG concentra a busca em uma pequena região do domínio, e acaba apontando para um mínimo local. E a convergência lenta, acontece se o algoritmo, embora próximo do mínimo global, não consegue atender às condições de parada. Estes problemas podem ser minimizados com a adoção de um procedimento simples chamado escalonamento de aptidão (*fitness scaling*). No início do processo de busca é interessante que a pressão seletiva seja pequena, permitindo que o AG explore a maior porção possível do espaço de domínio. Após muitas gerações, a população já se encontra em um estágio mais avançado, sendo conveniente aumentar a pressão seletiva, uma vez que os indivíduos tem valores de *fitness* muito próximos, dificultando o destaque dos melhores candidatos. O escalonamento linear foi proposto por Goldberg em [159] e o valor da função aptidão escalonada (fit_e) é dada por:

$$fit_e = a fit + b \quad (C.1)$$

os coeficientes a e b são calculados pela rotina `fitescala.m`, descrita no apêndice B, utilizando valores entre 1,2 e 2 para o fator multiplicador Cm .

C.2.3 Implementação de um Algoritmo Genético

Um AG com codificação binária pode ser facilmente implementado utilizando-se *strings* de bits para designar os cromossomos. Os demais operadores podem ser implementados com operações de manipulação de bits. A maior parte do processamento computacional fica no cálculo da função aptidão, nesta etapa os cromossomos devem ser decodificados de valores binários em reais para obtenção dos parâmetros (x_i), e com estes calcula-se os valores da função aptidão ($fit(x)$) para cada indivíduo.

Para que a busca com um AG se torne eficiente, é preciso escolher corretamente os parâmetros do algoritmo. O tamanho da população, o número de bits destinado a cada variável e as taxas de recombinação e mutação exercem papel fundamental na velocidade de convergência e na globalidade da busca efetuada.

O uso de populações pequenas não permite o mapeamento de todas as regiões do espaço de domínio, já populações muito grandes tornam o algoritmo muito lento. Baixas taxas de mutação podem fazer a busca perder a generalidade, ou seja, ficar concentrada em pequenas partes do espaço de busca, valores de p_{mut} muito altos podem gerar uma busca aleatória. Na escolha do p_{rec} deve-se permitir que uma parte da população permaneça inalterada, sob pena de gerar uma nova geração com indivíduos inferiores aos da anterior. No algoritmo desenvolvido, após uma série de testes práticos, onde diferentes valores destes parâmetros foram utilizados, decidiu-se pela utilização de: $p_{mut} = 5\%$, $p_{rec} = 80\%$ e uma população de 150 indivíduos.

C.2.4 O Algoritmo Genético Utilizado

O algoritmo Genético usado é semelhante ao proposto por Goldberg em [159], com modificações no processo de recombinação e a implementação de estratégias elitistas. Diferente do *single-point crossover*, proposto por Goldberg, onde é escolhido um ponto e trocados os bits que estão depois deste ponto, foi utilizada a recombinação ou *crossover uniforme*, que é feita escolhendo-se aleatoriamente o número e a posição dos bits a serem trocados. O elitismo foi implementado garantindo-se

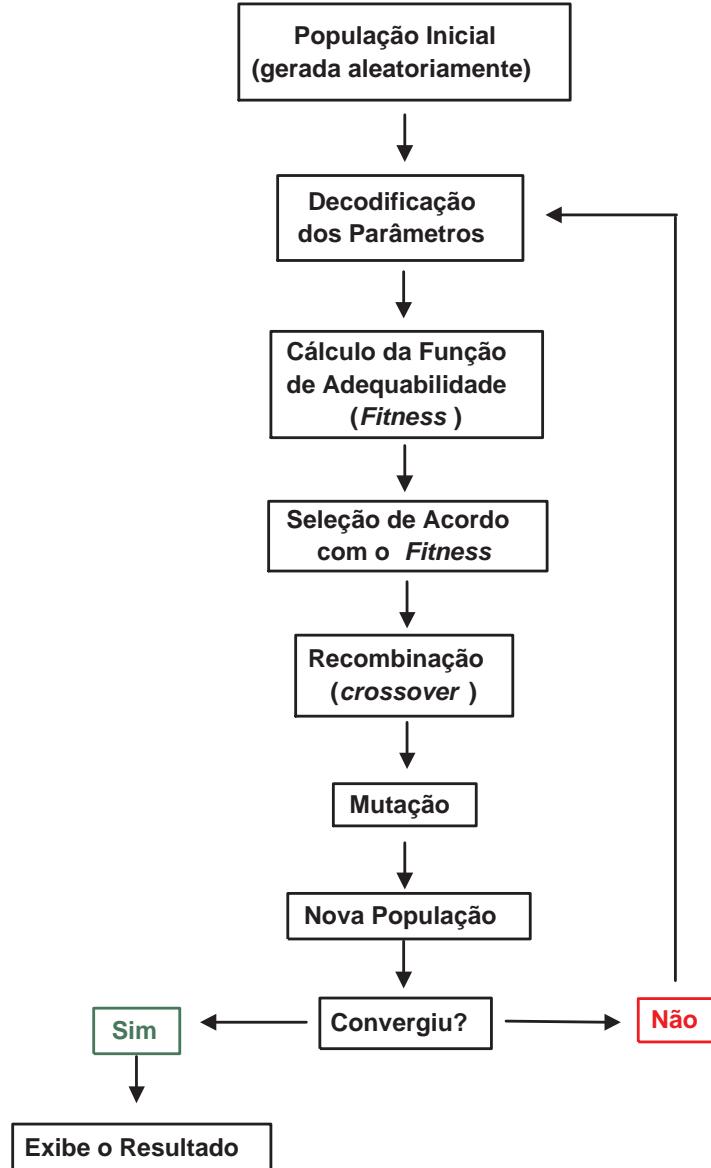


Figura C.3: Fluxo de um Algoritmo Genético

que cópias dos melhores indivíduos de cada geração estejam presentes na geração seguinte. Utilizou-se também, o escalonamento da aptidão, proposto em [159], este procedimento ameniza o problema da convergência prematura do algoritmo, regulando a pressão seletiva.

Escolheu-se como função objetivo a expressão do erro quadrático médio da equação ???. Como o AG foi desenvolvido para selecionar os indivíduos com maiores valores da função *fitness* ($fit(\mathbf{x})$), utilizou-se:

$$fit(\mathbf{x}) = \frac{1}{(f(\mathbf{x}, t) - f_e(\mathbf{x}, t))^2 + \zeta} \quad (C.2)$$

onde a constante ζ impede que $fit(\mathbf{x})$ assuma valores muito elevados à medida que o erro diminui.

Para uma boa amostragem do espaço de busca, é interessante usar populações grandes, mas, isso eleva muito o tempo de processamento. Na impossibilidade de fazer uso de um grande número de indivíduos, precisou-se implementar soluções alternativas. Na tentativa de corrigir problemas do algoritmo genético empregou-se alguns procedimentos:

Elitismo garante que os melhores indivíduos estejam presentes na geração seguinte, sem o risco de serem perdidos nos processos de *crossover* e mutação.

Recombinação uniforme para populações pequenas, age de modo mais eficiente que a de ponto único, facilitando a exploração das diversas regiões do espaço de busca.

Genocídio periódico Também visando evitar uma amostragem incompleta e a rápida perda de diversidade em populações pequenas, utilizou-se este procedimento de fácil implementação [248], que consiste em, a um número conveniente de gerações, aumentar o número de indivíduos de N para $\kappa \times N$, sendo κ um inteiro positivo, e depois, através de uma seleção rigorosa eliminar os $(\kappa - 1)N$ de menor aptidão.

Apêndice D

Produção Científica

A seguir, serão apresentados os resumos dos trabalhos publicados no desenvolvimento desta tese, com o propósito de facilitar a busca do leitor por mais referências a respeito da pesquisa desenvolvida. Os artigos estão em ordem cronológica (iniciando pelo mais recente) e divididos de acordo com o veículo de publicação.

D.1 Publicações em periódicos

1. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Modified Post-Nonlinear ICA for Online Neural Discrimination, Neurocomputing, vol. 73, no. 16-18, p. 2820-2828, 2010.*

Resumo:

Neste trabalho foi proposta uma modificação no modelo pós não-linear da ICA (*Independent Component Analysis*) que permite a estimativa de um conjunto de componentes que são ao mesmo tempo independentes e discriminantes. O algoritmo proposto foi aplicado para o problema da identificação de elétrons no segundo nível de filtragem do detector ATLAS. O método proposto produziu melhor desempenho se comparado ao modelo pós não-linear tradicional.

2. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. ATLAS Second-Level Electron/Jet Neural Discriminator based on Nonlinear Independent Components. Proceedings of Science (nova denominação do Journal of High Energy Physics/Online), v. ACAT10, p. 1-7, 2010.*

Resumo:

Um estudo detalhado a respeito da aplicação do modelo pós não-linear da análise de componentes independentes foi conduzido neste trabalho. Entre os tópicos estudados pode-se destacar a busca pelo grau de não-linearidade ótimo para o problema.

3. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Optimized Calorimeter Signal Compaction for an Independent Component based ATLAS Electron/Jet Second-Level Trigger. Proceedings of Science* (nova denominação do *Journal of High Energy Physics/Online*), v. ACAT08, p. 1-10, 2009.

Resumo:

Neste trabalho foram estudados alguns métodos de compactação de sinal para serem aplicados como pré-processamento à análise de componentes independentes. Entre os métodos testados pode-se destacar NLPCA (Análise de Componentes Principais Não-Lineares - *Nonlinear Principal Component Analysis*), PCA (Análise de Componentes Principais) e PCD (Componentes Principais de Discriminação - *Principal Components of Discrimination*). O processo de extração de características foi realizado de modo segmentado (a nível de cada camada do calorímetro).

4. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. High-Energy Particles Online Discriminators Based on Nonlinear Independent Components. Lecture Notes in Computer Science*, v. 5441, p. 718-725, 2009.

Resumo:

Um estudo comparativo entre três modelos distintos de estimação das componentes independentes não-lineares para o problema da filtragem *online* de elétrons no ATLAS foi conduzido neste trabalho. A NLICA foi estimada a partir: de um modelo sem restrições estruturais (utilizando mapas auto-organizáveis), de um modelo com restrições estruturais (utilizando um algoritmo para o modelo pós não-linear) e da ICA Local (que é uma abordagem diretamente ligada ao problema da ICA não-linear). Os resultados

obtidos foram comparados em termos do desempenho de discriminação.

5. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Self-organized mapping of calorimetry information for high efficient online electron/jet identification in ATLAS. Proceedings of Science* (nova denominação do *Journal of High Energy Physics/Online*), v. ACAT07, p. 1-13, 2007.

Resumo:

Um estudo da aplicação de mapas auto-organizáveis (SOM - *Self-organizing Maps*) para a extração de características na filtragem *online* de elétrons do ATLAS foi conduzido neste trabalho. Os SOM foram aplicados de modo não-segmentado (considerando como entrada um vetor composto pelos sinais em anéis gerados a partir de todas as camadas dos calorímetros).

6. *Eduardo F. Simas Filho, José M. de Seixas. Nonlinear Independent Component Analysis: Theoretical Review and Applications. Learning and Nonlinear Models*, v. 5, p. 99-120, 2007.

Resumo:

Um tutorial a respeito da teoria e dos diversos algoritmos para estimação do modelo não-linear da análise de componentes independentes (em suas diversas variações) foi apresentado neste trabalho. Foram realizados testes de desempenho com diversos algoritmos utilizando uma base de dados onde havia controle sobre o processo da mistura dos sinais (consequentemente os sinais a serem estimados eram conhecidos). Uma aplicação da NLICA para extração de características no segundo nível de filtragem do detector ATLAS também foi mostrada.

D.2 Capítulo de Livro

1. *Rodrigo C. Torres, Eduardo F. Simas Filho, Danilo E. F. Lima e J. M. Seixas. Segmented Online Neural Filtering based on Independent Components of Pre-Processed Information. Signal Processing*. Vienna: In-Tech, 2010, v.1 , p. 337-358, (aceito para publicação).

Resumo:

Neste trabalho foi realizado um estudo comparativo de diversas abordagens baseadas em análise de componentes independentes (ICA - *Independent Component Analysis*) para extração de características na filtragem online do ATLAS. Os componentes independentes foram estimados de ambos os modos, segmentado e não-segmentado. Para a classificação foram utilizados classificadores neurais e classificadores lineares (discriminante linear de Fisher).

D.3 Artigos em Conferências

1. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Análise de Componentes Independentes para uma Filtragem Online baseada em Calorimetria de Alta Energia e com Fina Segmentação.* In: Workshop de Teses e Dissertações em Inteligência Artificial (parte do Joint Conference SBIA/SBRN), p. 1-8, São Bernardo do Campo-SP, 2010.

Resumo:

Neste trabalho foi apresentado um resumo com os principais resultados obtidos até o momento da submissão, que incluiam a base de dados simulados com assinaturas E10.

2. *Rodrigo C. Torres, Danilo E. F. Lima, Eduardo F. Simas Filho e J. M. Seixas. Neural Online Filtering Based on Preprocessed Calorimeter Data.* In: IEEE Nuclear Science Symposium and Medical Imaging Conference, p. 530-536, Orlando, 2009.

Resumo:

Um novo conjunto de sinais simulados, com características mais próximas da operação real esperada para o detector, foi utilizado neste trabalho. Um estudo detalhado sobre diferentes métodos de compactação como Análise de Componentes Principais (PCA - *Principal Component Analysis*) e Componentes Principais de Discriminação (PCD - *Principal Components of Discrimination*) foi conduzido.

3. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Combinação de Classificadores Neurais Segmentados com Pré-processamento por Análise de Componentes Independentes para um Sistema Online de Filtragem.* In: Congresso Brasileiro de Redes Neurais e Inteligência Computacional, p. 1-5, Ouro Preto-MG, 2009.

Resumo:

Neste trabalho foi proposto o uso de classificadores neurais especialistas nas informações de cada camada do calorímetro. Alguns métodos para combinar as saídas dos classificadores segmentados foram testados. Com o uso dos classificadores especialistas foi possível aumentar a eficiência de discriminação e ainda identificar que existe redundância na informação disponível nas diversas camadas dos calorímetros. Neste caso, foi mostrado que, mesmo sem utilizar os sinais de algumas camadas o desempenho de discriminação se mantém quase inalterado. Como o processo de formação dos sinais em anéis é responsável pela maior parte do tempo de processamento, quanto menor a quantidade de informação necessária mais rápida a decisão é tomada.

4. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Local Independent Component Analysis Applied to Highly Segmented Detectors.* In: IEEE International Symposium on Circuits and Systems (ISCAS08), v. 1, p. 3005-3008, Seattle, 2008.

Resumo:

Neste trabalho foi proposta a utilização do modelo da ICA Local para a extração de características na filtragem online de segundo nível do ATLAS. Neste modelo um algoritmo não-supervisionado de agrupamento é utilizado para dividir o conjunto de sinais disponíveis em grupos de características semelhantes. A ICA é aplicada separadamente para os sinais de cada conjunto.

5. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Segmented Self-Organized Feature Extraction for Online Filtering in a High Event Rate Detector.* In: European Signal Processing Conference, p. 1-5,

Lausanne, Suiça, 2008.

Resumo:

Mapas auto-organizáveis foram aplicados de modo segmentado (a nível de cada camada do calorímetro) para extração de características sobre os sinais em anéis. Um estudo para busca do tamanho ótimo do mapa foi conduzido considerando o erro médio de representação obtido. Foi mostrado que com a abordagem segmentada obteve-se melhor desempenho de discriminação do que no modo não-segmentado.

6. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Online Neural Filtering Operating Over Segmented Discriminating Components.* In: IEEE International Conference on Electronics, Circuits and Systems, v. 1. p. 530-533, Malta, 2008.

Resumo:

Foi proposta a estimativa dos componentes principais de discriminação (PCD - *Principal Discriminating Components*) de modo segmentado (a nível de cada camada do calorímetro), aproveitando, desta forma, toda a segmentação e granularidade disponíveis ao detector. As PCDs (obtidas em cada camada) foram concatenadas e utilizadas para treinar um classificador neural.

7. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Segmented Overdetermined Nonlinear Independent Component Analysis for Online Neural Filtering.* In: *Simpósio Brasileiro de Redes Neurais*, p. 159-164, Salvador, 2008.

Resumo:

Uma modificação no modelo pós não-linear da ICA foi proposta neste trabalho visando à aplicação em misturas sobre-determinadas (onde o número de sinais observados é maior que o número de componentes independentes a serem estimadas). O modelo proposto parece ser mais adequado para o problema da extração de características no ATLAS pois espera-se que o número de componentes independentes seja menor que o de anéis (100).

8. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Análise Não-Linear de Componentes Independentes Segmentadas para Filtragem Online num Detector com Alta Taxa de Eventos.* In: *Congresso Brasileiro de Automática*, p.1-6, Juiz de Fora, 2008.

Resumo:

Neste trabalho o modelo pós não-linear da NLICA (*Nonlinear Independent Component Analysis*) foi utilizado para extração de características de modo segmentado (a nível de cada camada do calorímetro). Foi utilizado um algoritmo que não permite a compactação dos sinais, então, após a estimativa das componentes foi calculada a relevância e as menos relevantes foram descartadas. Uma limitação deste procedimento é que todas (100) componentes independentes não-lineares precisam ser calculadas e muitas delas não são utilizadas para a discriminação, neste caso o esforço computacional (e consequentemente o tempo necessário para tomada de decisão) aumenta consideravelmente.

9. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Segmented Independent Component Analysis for online filtering using highly segmented detectors.* In: *International Conference on Intelligent Systems Design and Applications*, Rio de Janeiro, v. 1. p. 659-664, 2007.

Resumo:

A análise de componentes independentes (ICA - *Independent Component Analysis*) foi aplicada de modo segmentado (a nível de cada camada do calorímetro) para extração de características relevantes na discriminação de elétrons no segundo nível de filtragem do ATLAS. Para compactação foi utilizada a análise de componentes principais (PCA - *Principal Component Analysis*)

10. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Análise de Componentes Independentes para Filtragem Online num Ambiente de Alta Taxa de Eventos e Informação Segmentada.* In: *VIII Congresso Brasileiro de Redes Neurais*, Florianópolis, p. 1-6, 2007.

Resumo:

A combinação PCA + ICA foi utilizada para compactação e extração de características no segundo nível de filtragem do ATLAS. Neste trabalho foi utilizada uma abordagem não-segmentada para extração de características, ou seja, os anéis produzidos em todas as camadas do calorímetro foram considerados como um único vetor de dados. Foi realizada também uma análise da relevância das componentes independentes estimadas, visando eliminar as componentes não-relevantes para a discriminação.

11. *Eduardo F. Simas Filho, Luiz P. Calôba e José M. de Seixas. Filtragem ótima para o Trigger de segundo nível do ATLAS baseado em calorimetria.* In: *XXVII Encontro Nacional de Física de Partículas e Campos*, Águas de Lindóia, p. 1-4, 2006.

Resumo:

Neste trabalho foram apresentados resultados iniciais da pesquisa onde os sinais em anéis foram identificados através de filtros casados. Visando a compactação dos sinais, testes foram conduzidos utilizando-se apenas os anéis mais energéticos de cada camada.

D.4 Resumos em Conferências

1. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. ATLAS Neural Second-Level Trigger based on Nonlinear Independent Components of Segmented Calorimeter Information.* In: *Experimental High-Energy Physics and Associated Technologies Workshop*, Rio de Janeiro, 2008.

Resumo:

Um estudo envolvendo diversos métodos de estimativa das componentes independentes no modelo não-linear foi conduzido neste trabalho.