



FILTRAGEM *ONLINE* DE EVENTOS RAROS E/OU EXÓTICOS BASEADA EM  
DADOS DE CALORIMETRIA DE ALTA ENERGIA EM UM DETECTOR  
FINAMENTE SEGMENTADO

Danilo Lima de Souza

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.

Orientador: José Manoel de Seixas

Rio de Janeiro  
Março de 2012

FILTRAGEM *ONLINE* DE EVENTOS RAROS E/OU EXÓTICOS BASEADA EM  
DADOS DE CALORIMETRIA DE ALTA ENERGIA EM UM DETECTOR  
FINAMENTE SEGMENTADO

Danilo Lima de Souza

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. Márcia Begalli, Dr.rer.nat.

---

Prof. Aline da Rocha Gesualdi, D.Sc.

---

Prof. Rui Seara, Dr.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2012

Lima de Souza, Danilo

Filtragem *online* de eventos raros e/ou exóticos baseada em dados de calorimetria de alta energia em um detector finamente segmentado/Danilo Lima de Souza. – Rio de Janeiro: UFRJ/COPPE, 2012.

XXX, 223 p.: il.; 29,7cm.

Orientador: José Manoel de Seixas

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2012.

Referências Bibliográficas: p. 197 – 210.

1. LHC. 2. ATLAS. 3. Redes Neurais. 4. Processamento de sinais. I. Manoel de Seixas, José. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

# Agradecimentos

À minha família.

Aos professores, alunos, ex-alunos e funcionários do Laboratório de Processamento de Sinais (LPS-COPPE-UFRJ).

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Aos pesquisadores do TDAQ e do grupo de estudos de eventos exóticos do ATLAS ligados direta ou indiretamente com o trabalho.

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

FILTRAGEM *ONLINE* DE EVENTOS RAROS E/OU EXÓTICOS BASEADA EM  
DADOS DE CALORIMETRIA DE ALTA ENERGIA EM UM DETECTOR  
FINAMENTE SEGMENTADO

Danilo Lima de Souza

Março/2012

Orientador: José Manoel de Seixas

Programa: Engenharia Elétrica

Do que é feita a matéria? Qual o candidato a matéria escura? Algumas das questões mais intrigantes da atualidade têm a oportunidade de serem respondidas em um futuro próximo e, para isso, contam com um ferramental tão grandioso quanto o desafio que é proposto, o acelerador de partículas LHC e o detector ATLAS. A cada colisão de partículas realizada, uma gigantesca quantidade de informação é produzida; contudo, apenas uma pequena parcela é importante para a caracterização dos fenômenos físicos de interesse. Para empreender essa separação, um eficiente sistema para detecção (filtragem) *online* de eventos é requerido. Os objetos de interesse, elétrons e *leptonjets*, são partículas extremamente importantes para a caracterização da física estudada e aparecem mascarados por um intenso ruído de fundo, que pode apresentar perfil de deposição de energia semelhante ao evento alvo, falseando-o. Os calorímetros são medidores de energia e, no ATLAS, são divididos em sete camadas finamente segmentadas, totalizando mais de 100.000 sensores. Para a identificação das partículas de interesse, os calorímetros são fundamentais. Neste trabalho, é proposta a utilização das análises de componentes principais e independentes no processo de extração de características, visando otimizar o desempenho do sistema de filtragem *online* no ATLAS. Para explorar toda a segmentação e granularidade disponíveis, ora a extração de características é efetuada separadamente, para cada camada do calorímetro, ora em conjunto. Através da abordagem proposta, é possível alcançar alta eficiência de discriminação (98% quando se objetiva identificar elétrons e, similarmente, 70% para *leptonjets*), 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.)

ONLINE FILTERING OF RARE AND/OR EXOTIC EVENTS BASED ON  
HIGH-ENERGY CALORIMETRY DATA IN A FINELY SEGMENTED DETECTOR

Danilo Lima de Souza

March/2012

Advisor: José Manoel de Seixas

Department: Electrical Engineering

What is matter made of? Which is the candidate for Dark Matter? Some of the most intriguing questions of our time have the opportunity to be answered in the near future, and therefore, they have an experiment as great as the challenge that is proposed, the LHC particle accelerator and the ATLAS detector. Each particle collision produces a tremendous amount of information; however, only a small part of that is important for the characterization of the physical phenomena of interest. To undertake this separation, an efficient system for detection (filtering) of online events is required. The objects of interest, electrons and leptonjets, are particles extremely important for the physical characterization of the theory, and appear masked by intense background noise, which can show an energy profile similar to the signal event, faking it. The calorimeters are energy big sensors and, in the ATLAS detector, are divided into seven fine segmented layers, totalizing more than 100,000 small sensors. These sensors are essential for the submitted task. In this work, we propose the use of principal and independent component analysis in the process of feature extraction to optimize the performance of the online ATLAS triggering system. To explore all available segmentation and granularity, the extraction has been performed in two ways: separately for each layer in the calorimeter or considering all layers together. Through the proposed approach, it has been possible to achieve high efficiency of discrimination (98% when electrons are the aim, and, similarly, 70% for leptonjets), resulting in purified data for the offline analysis.

# Sumário

<b>Lista de Figuras</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xxii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Contexto . . . . .	1
1.2 Motivação . . . . .	3
1.3 Objetivo . . . . .	4
1.4 Estado da Arte . . . . .	5
1.5 Abordagem proposta . . . . .	7
1.6 Organização do Trabalho . . . . .	8
<b>2 A Física de Partículas e o Experimento ATLAS</b>	<b>9</b>
2.1 Breve Histórico da Física de Partículas . . . . .	9
2.2 O Modelo Padrão . . . . .	11
2.3 Matéria Escura . . . . .	12
2.3.1 <i>Leptonjets</i> . . . . .	13
2.4 LHC . . . . .	15
2.4.1 Detector ATLAS . . . . .	18
<b>3 Filtragem <i>online</i> do detector ATLAS</b>	<b>26</b>
3.1 Aquisição e Filtragem <i>online</i> de Dados . . . . .	26
3.2 Plataforma Computacional do Ambiente de Filtragem . . . . .	31
3.3 Extração de Características e Teste de Hipótese . . . . .	31
3.4 Tempo de Processamento . . . . .	37
<b>4 Base de dados, Medidas de Desempenho, Metodologia Proposta e Algoritmos de Referência</b>	<b>39</b>
4.1 Caracterização do Sinal . . . . .	39
4.2 Filtragem de Nível 1 . . . . .	42
4.3 Normalização . . . . .	43
4.4 Avaliação de Desempenho . . . . .	44

4.5	Metodologia Proposta . . . . .	58
4.5.1	Metodologia Proposta para o Canal elétron/jato . . . . .	60
4.5.2	Metodologia proposta para o canal <i>leptonjets/jato</i> . . . . .	60
4.6	Desempenho dos algoritmos de referência . . . . .	64
<b>5</b>	<b>Resultados da Filtragem <i>online</i> de Nível 2 - canal elétron/jato</b>	<b>69</b>
5.1	<i>Menu e10</i> . . . . .	69
5.2	Anéis . . . . .	70
5.3	Pré-processamento não segmentado . . . . .	78
5.3.1	PCA . . . . .	78
5.3.2	ICA . . . . .	93
5.4	Pré-processamento Segmentado . . . . .	102
5.4.1	PCA . . . . .	105
5.4.2	ICA . . . . .	118
5.5	Comparação . . . . .	124
<b>6</b>	<b>Resultados da Filtragem <i>online</i> de Nível 2 - canal <i>leptonjets/jato</i></b>	<b>131</b>
6.1	<i>Menu 1j4e10</i> . . . . .	131
6.2	Anéis . . . . .	132
6.2.1	Regiões dentro da RoI . . . . .	138
6.3	Pré-processamento não Segmentado . . . . .	148
6.3.1	PCA . . . . .	148
6.3.2	ICA . . . . .	157
6.4	Pré-processamento Segmentado . . . . .	169
6.4.1	PCA . . . . .	171
6.4.2	ICA . . . . .	177
6.5	Comparação . . . . .	186
<b>7</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>192</b>
7.1	Trabalhos Futuros . . . . .	195
	<b>Referências Bibliográficas</b>	<b>197</b>
<b>A</b>	<b>Extração de Características e Teste de Hipótese</b>	<b>211</b>
A.1	Análise de Componentes Principais . . . . .	212
A.2	Análise de Componentes Principais Não-lineares . . . . .	214
A.2.1	Descorrelação não-linear . . . . .	216
A.3	Análise de Componentes Independentes . . . . .	217
A.4	Teste de Hipótese . . . . .	219
A.4.1	Classificador Linear de Fisher . . . . .	219
A.4.2	Classificador Neural . . . . .	220



# Lista de Figuras

1.1	Procedimento de detecção/decisão de eventos. . . . .	8
2.1	Cadeias típicas de decaimentos iniciando com um bóson de gauge escuro (rotulado $\gamma'$ ). A cadeia de decaimento pode ter diversas etapas e envolver outros estágios do setor escuro, como outros bósons de gauge escuros (rotulados $w'$ e $z'$ ), e Higgs escuros (rotulados $h'$ ). . . . .	14
2.2	Produção direta de um bóson de gauge escuro em um processo muito similar à produção de fótons no Modelo Padrão. . . . .	14
2.3	Cadeia de aceleração do LHC e seus detectores. . . . .	16
2.4	Esquema básico do detector ATLAS. . . . .	19
2.5	Sistema de coordenadas do ATLAS. . . . .	20
2.6	Disposição das camadas dos calorímetros Eletromagnético (EM) e Hadrônico (HAD) do detector ATLAS. . . . .	21
2.7	Granularidade e profundidade das camadas do calorímetro eletromagnético do detector ATLAS . . . . .	23
2.8	Imagem gerada em computador de um corte transversal no detector ATLAS para diferentes tipos de eventos. Comportamento de um múon, um fóton, um elétron, um próton, um nêutron e um neutrino em um sistema HEP. O duto de feixe é perpendicular ao plano da imagem. O múon passa completamente através dos calorímetros depositando apenas uma pequena quantidade de energia em cada seção até finalmente ser detectado pela câmara de múons. O fóton, mesmo sem massa, interage fortemente com o calorímetro eletromagnético e o elétron deposita toda a sua energia nesse mesmo calorímetro. Prótons depositam pouca energia no calorímetro eletromagnético e juntamente com nêutrons interagem fortemente com o calorímetro hadrônico. Jatos compostos de muitas partículas - majoritariamente píons - depositam energia tanto no calorímetro eletromagnético quanto no hadrônico e em uma região maior do que uma única partícula. O neutrino não interage com qualquer equipamento, ele passa indetectavelmente. . . . .	25

3.1	Sistema de filtragem do ATLAS. . . . .	27
3.2	Exemplo de RoI de um elétron por camada: <i>Pre Sampler</i> , canto superior esquerdo; <i>EM front layer</i> , canto superior direito; <i>EM middle layer</i> , segunda linha no lado esquerdo; <i>EM back layer</i> , segunda linha no lado direito; <i>HAD layer 0</i> , terceira linha no lado esquerdo; <i>HAD layer 1</i> , terceira linha no lado direito; <i>HAD layer 2</i> , canto inferior esquerdo. . . . .	28
3.3	Exemplo de RoI de um jato por camada: <i>Pre Sampler</i> , canto superior esquerdo; <i>EM front layer</i> , canto superior direito; <i>EM middle layer</i> , segunda linha no lado esquerdo; <i>EM back layer</i> , segunda linha no lado direito; <i>HAD layer 0</i> , terceira linha no lado esquerdo; <i>HAD layer 1</i> , terceira linha no lado direito; <i>HAD layer 2</i> , canto inferior esquerdo. . . . .	29
3.4	Exemplo de RoI de <i>leptonjets</i> por camada: <i>Pre Sampler</i> , canto superior esquerdo; <i>EM front layer</i> , canto superior direito; <i>EM middle layer</i> , segunda linha no lado esquerdo; <i>EM back layer</i> , segunda linha no lado direito; <i>HAD layer 0</i> , terceira linha no lado esquerdo; <i>HAD layer 1</i> , terceira linha no lado direito; <i>HAD layer 2</i> , canto inferior esquerdo. . . . .	30
3.5	Formação de anéis por camada dos calorímetros: (a) <i>Pre Sampler</i> ; (b) <i>EM middle layer</i> ; (c) <i>EM back layer</i> ; (d) <i>HAD layer 0</i> . . . . .	35
3.6	Eventos Formatados em Anéis. . . . .	36
4.1	Faixa de energia coberta pelos eventos. . . . .	40
4.2	Distribuição em $\eta$ e $\phi$ dos eventos simulados: (a) elétrons; (b) jatos; (c) <i>leptonjets</i> . . . . .	41
4.3	Elétron simulado normalizado por energia: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de elétrons: (c) Correlação linear; (d) Correlação não-linear; (e) Informação mútua. . . . .	45
4.4	Jato simulado normalizado por energia: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de jatos: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua. . . . .	46
4.5	<i>Leptonjets</i> simulado normalizado por energia: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de <i>leptonjets</i> : (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua. . . . .	47
4.6	Elétron simulado normalizado sequencialmente: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de elétrons: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua. . . . .	48

4.7	Jato simulado normalizado sequencialmente: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de jatos: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua. . . . .	49
4.8	<i>Leptonjets</i> simulado normalizado sequencialmente: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de <i>leptonjets</i> : (a) Correlação linear. (b) Correlação não-linear. (c) Informação mútua. . . . .	50
4.9	Elétron simulado normalizado por camada: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de elétrons: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua. . . . .	51
4.10	Jato simulado normalizado por camada: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de jatos: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua. . . . .	52
4.11	<i>Leptonjets</i> simulado normalizado por camada: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de <i>leptonjets</i> : (a) Correlação linear. (b) Correlação não-linear. (c) Informação mútua. . . . .	53
4.12	Curva ROC. . . . .	55
4.13	(a) Exemplo 1 de histograma de saída de classificador; (b) Exemplo 2 de histograma de saída de classificador; (c) Curvas ROC associadas aos histogramas mostrados em (a) e (b); (d) Produto SP associados aos histogramas mostrados nas Figuras (a) e (b). . . . .	56
4.14	(a) (b) (c) três diferentes exemplos de correlação entre a componente de máxima relevância de um processo decisório e as variáveis usadas na sua geração. (d) Acumulado de variáveis por limiar de correlação. . . . .	57
4.15	Procedimento de detecção/decisão de eventos. . . . .	58
4.16	Procedimento baseado em anéis para o canal elétron/jato: (a) Treinamento do algoritmo de classificação; (c) Teste de hipótese. . . . .	61
4.17	Procedimento não segmentado para o canal elétron/jato: (a) Treinamento do algoritmo de pré-processamento; (b) Treinamento do algoritmo de classificação; (c) Teste de hipótese. . . . .	62
4.18	Procedimento segmentado para o canal elétron/jato: (a) Treinamento dos algoritmos de pré-processamento; (b) Treinamento do algoritmo de classificação; (c) Teste de hipótese. . . . .	63

4.19	Desempenho do algoritmo T2Ca1o. (a) Histograma de saída para elétrons. (b) Probabilidades de detecção e falso alarme em função da $E_T$ . (c) Histograma de saída para jatos. (d) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (e) Histograma de saída para <i>leptonjets</i> . (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	66
4.20	Desempenho independente dos cortes do algoritmo T2Ca1o. (a) Histograma de saída para elétrons. (b) Histograma de saída para jatos. (c) Histograma de saída para <i>leptonjets</i> . . . . .	68
5.1	Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis e normalizados sequencialmente. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	71
5.2	Relevância dos anéis normalizados sequencialmente para o processo decisório implementado pelo classificador linear de Fisher. . . . .	74
5.3	Variação de desempenho do RingerNeuralHypo operando sobre dados formatados em anéis e normalizados sequencialmente em função do número de neurônios na camada intermediária da rede: (a) Erro quadrático médio final de treinamento para diferentes condições de inicialização do algoritmo neural; (b) Produto $SP_{e/j}$ associado à melhor rede; (c) Probabilidade de detecção $PD_e$ associada à melhor rede; (d) Probabilidade de falso alarme $PF_j$ associada à melhor rede; (e) Produto $SP_{e/j/l_j}$ associado à melhor rede; (f) Probabilidade de detecção $PD_{l_j}$ associada à melhor rede. . . . .	76
5.4	Desempenho do RingerNeuralHypo operando sobre dados formatados em anéis e normalizados sequencialmente. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	77
5.5	(a) Erro quadrático médio de treinamento do algoritmo neural. (b) Relevância dos anéis normalizados sequencialmente para o processo decisório implementado pelo classificador neural. . . . .	79
5.6	Curvas de carga: (a) eventos normalizados por energia; (b) eventos normalizados sequencialmente; (c) eventos normalizados por camada. . . . .	80

5.7	Variaco de desempenho do classificador linear de Fisher operando sobre componentes principais extradas de dados formatados em anis e normalizados por camada em funo da energia retida do sinal: (a) Produto $SP_{e/j}$ ; (b) Probabilidade de deteco $PD_e$ ; (c) Probabilidade de falso alarme $PF_j$ ; (d) Produto $SP_{e/j/l_j}$ ; (e) Probabilidade de deteco $PD_{lj}$ . . . . .	84
5.8	Desempenho do classificador linear de Fisher operando sobre componentes principais extradas de dados formatados em anis e normalizados por camada. (a) Histogramas de sada. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de deteco de <i>leptonjets</i> . (d) Probabilidades de deteco e falso alarme em funo da $E_T$ . (e) Probabilidades de deteco e falso alarme em funo da coordenada $\eta$ do detector. (f) Probabilidades de deteco e falso alarme em funo da coordenada $\phi$ do detector. . . . .	85
5.9	Medidas qualitativas das variveis de entrada do classificador linear de Fisher operando sobre componentes principais extradas de dados formatados em anis e normalizados por camada: (a) Relevncia das componentes principais no processo decisrio implementado; (b) Correlao entre as componentes principais extradas; (c) Correlao entre a componente de mxima relevncia no processo decisrio e os anis usados na extrao das componentes principais. . . . .	86
5.10	Variaco de desempenho do classificador neural operando sobre componentes principais extradas de dados formatados em anis normalizados por camada em funo do nmero de neurnios na camada intermediria da rede e da energia retida do sinal: (a) Erro quadrtico mdio (MSE) final de treinamento; (b) Produto $SP_{e/j}$ associado  melhor rede; (c) Probabilidade de deteco $PD_e$ associada  melhor rede; (d) Probabilidade de falso alarme $PF_j$ associada  melhor rede; (e) Produto $SP_{e/j/l_j}$ associado  melhor rede; (f) Probabilidade de deteco $PD_{lj}$ associada  melhor rede. . . . .	89
5.11	Desempenho do classificador neural operando sobre componentes principais extradas de dados formatados em anis e normalizados por camada. (a) Histogramas de sada. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de deteco de <i>leptonjets</i> . (d) Probabilidades de deteco e falso alarme em funo da $E_T$ . (e) Probabilidades de deteco e falso alarme em funo da coordenada $\eta$ do detector. (f) Probabilidades de deteco e falso alarme em funo da coordenada $\phi$ do detector. . . . .	91

5.12	(a) Erro quadrático médio de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por camada; (b) Relevância das componentes principais no processo decisório implementado; (c) Correlação entre as componentes principais extraídas; (d) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes principais. . . . .	92
5.13	Variação de desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por camada. Resultados em função da energia retida do sinal: (a) Produto $SP_{e/j}$ ; (b) Probabilidade de detecção $PD_e$ ; (c) Probabilidade de falso alarme $PF_j$ ; (d) Produto $SP_{e/j/l_j}$ ; (e) Probabilidade de detecção $PD_{l_j}$ . . . . .	96
5.14	Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por camada. (a) Histograma de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da energia transversa ( $E_t$ ). (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	97
5.15	Medidas qualitativas das variáveis de entrada do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por camada: (a) Relevância das componentes independentes no processo decisório implementado; (b) Informação mútua entre as componentes independentes extraídas; (c) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes. . . . .	99
5.16	Variação de desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados sequencialmente em função do número de neurônios na camada intermediária da rede e da energia retida do sinal: (a) Erro quadrático médio (MSE) final de treinamento; (b) Produto $SP_{e/j}$ associado à melhor rede; (c) Probabilidade de detecção $PD_e$ associada à melhor rede; (d) Probabilidade de falso alarme $PF_j$ associada à melhor rede; (e) Produto $SP_{e/j/l_j}$ associado à melhor rede; (f) Probabilidade de detecção $PD_{l_j}$ associada à melhor rede. . . . .	101

5.17	Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados sequencialmente. (a) Histograma de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	103
5.18	(a) Erro quadrático médio de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados sequencialmente: (b) Relevância das componentes independentes no processo decisório implementado; (c) Informação Mútua entre as componentes independentes extraídas; (d) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes. . . . .	104
5.19	Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) <i>Pre Sampler</i> ; (b) <i>EM front layer</i> . Normalização sequencial: (c) <i>Pre Sampler</i> ; (d) <i>EM front layer</i> . Normalização por camada: (e) <i>Pre Sampler</i> ; (f) <i>EM front layer</i> . . . . .	106
5.20	Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) <i>EM middle layer</i> ; (b) <i>EM back layer</i> . Normalização sequencial: (c) <i>EM middle layer</i> ; (d) <i>EM back layer</i> . Normalização por camada: (e) <i>EM middle layer</i> ; (f) <i>EM back layer</i> . . . . .	107
5.21	Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) <i>HAD layer 0</i> ; (b) <i>HAD layer 1</i> . Normalização sequencial: (a) <i>HAD layer 0</i> ; (b) <i>HAD layer 1</i> . Normalização por camada: (a) <i>HAD layer 0</i> ; (b) <i>HAD layer 1</i> . . . . .	108
5.22	Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) <i>HAD layer 2</i> . Normalização sequencial: (b) <i>HAD layer 2</i> . Normalização por camada: (c) <i>HAD layer 2</i> . . . . .	109
5.23	Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	114



5.24	Desempenho do classificador neural operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	117
5.25	Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	122
5.26	Desempenho do classificador neural operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{e/j}$ e probabilidade de detecção de <i>leptonjets</i> . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	125
5.27	Curvas ROC para o canal e/j obtidas com cada uma das abordagens propostas. . . . .	129
5.28	Curvas de eficiência na detecção de <i>leptonjets</i> obtidas com cada uma das abordagens propostas. . . . .	130
6.1	Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis e normalizados por camada. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{lj/j}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	134
6.2	Relevância dos anéis normalizados por camada para o processo decisório implementado pelo classificador linear de Fisher. . . . .	135



6.3	Varição de desempenho do classificador neural operando sobre dados formatados em anéis e normalizados por energia: (a) Erro quadrático médio final de treinamento para diferentes inicializações do algoritmo neural; (b) Produto $SP_{ij}$ associado à melhor rede; (c) Probabilidade de detecção $PD_{ij}$ associada à melhor rede; (d) Probabilidade de falso alarme $PF_j$ associada à melhor rede. . . . .	137
6.4	Desempenho do classificador neural operando sobre dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	139
6.5	(a) Erro quadrático médio de treinamento do algoritmo neural. (b) Relevância dos anéis normalizados por energia para o processo decisório implementado pelo classificador neural. . . . .	140
6.6	Formação de regiões de análise dentro da RoI dado as diferentes resoluções espaciais por camada dos calorímetros. . . . .	141
6.7	Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis pertencentes a região $\Delta R = 0,3$ e normalizados pela energia de RoI. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	144
6.8	Relevância dos anéis normalizados por energia para o processo decisório implementado pelo classificador linear de Fisher. . . . .	145
6.9	Desempenho do classificador neural operando sobre dados formatados em anéis pertencentes a região $\Delta R = 0,3$ e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	147
6.10	(a) Erro quadrático médio de treinamento do algoritmo neural. (b) Relevância dos anéis normalizados por energia para o processo decisório implementado pelo classificador neural. . . . .	148

6.11	Variação de desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia: (a) Produto $SP_{ij}$ ; (b) Probabilidade de detecção $PD_{ij}$ ; (c) Probabilidade de falso alarme $PF_j$ . . . . .	151
6.12	Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	153
6.13	Medidas qualitativas das variáveis de entrada do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia: (a) Relevância das componentes principais no processo decisório implementado; (b) Correlação entre as componentes principais extraídas; (c) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes principais. . . . .	154
6.14	Variação de desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia, em função do número de neurônios na camada intermediária da rede e da energia retida do sinal: (a) Erro quadrático médio (MSE) final de treinamento; (b) Produto $SP_{ij}$ associado à melhor rede; (c) Probabilidade de detecção $PD_{ij}$ associada à melhor rede; (d) Probabilidade de falso alarme $PF_j$ associada à melhor rede. . . . .	156
6.15	Desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia: (a) Histogramas de saída; (b) Curva ROC; (c) Produto $SP_{ij}$ ; (d) Probabilidades de detecção e falso alarme em função da $E_T$ ; (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector; (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	158

6.16	(a) Erro quadrático médio de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia; (b) Relevância das componentes principais no processo decisório implementado; (c) Correlação entre as componentes principais extraídas; (d) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes principais. . . . .	159
6.17	Variação de desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia. Resultados em função da energia retida do sinal: (a) Produto $SP_{lj/j}$ ; (b) Probabilidade de detecção $PD_{lj}$ ; (c) Probabilidade de falso alarme $PF_j$ . . . . .	162
6.18	Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{lj/j}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	163
6.19	Medidas qualitativas das variáveis de entrada do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia: (a) Relevância das componentes independentes no processo decisório implementado; (b) Informação mútua entre as componentes independentes extraídas; (c) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes. . . . .	165
6.20	Variação de desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis normalizados por energia em função do número de neurônios na camada intermediária da rede e da energia retida do sinal: (a) Erro quadrático médio (MSE) final de treinamento; (b) Produto $SP_{lj/j}$ associado à melhor rede; (c) Probabilidade de detecção $PD_{lj}$ associada à melhor rede; (d) Probabilidade de falso alarme $PF_j$ associada à melhor rede. . . . .	167

6.21	Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	168
6.22	(a) Erro médio quadrático de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia; (b) Relevância das componentes independentes no processo decisório implementado; (c) Informação Mútua entre as componentes independentes extraídas; (d) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes. . . . .	170
6.23	Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	174
6.24	Desempenho do classificador neural operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	178
6.25	Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	182

6.26	Desempenho do classificador neural operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por camada. (a) Histogramas de saída. (b) Curva ROC. (c) Produto $SP_{lj}$ . (d) Probabilidades de detecção e falso alarme em função da $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada $\eta$ do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada $\phi$ do detector. . . . .	185
6.27	Curvas ROC para o canal $lj$ obtidas com cada uma das abordagens propostas. . . . .	190
6.28	Produto $SP_{lj}$ obtido com cada uma das abordagens propostas. . . . .	191
A.1	Minimização do desvio ortogonal. . . . .	213
A.2	Componentes de um problema clássico de detecção de sinais. . . . .	219
A.3	Aprendizado supervisionado. . . . .	221
A.4	Rede Percéptron de Múltiplas Camadas. . . . .	223

# Lista de Tabelas

2.1	Uma visão esquemática da constituição da matéria segundo o Modelo Padrão. . . . .	12
2.2	Região de cobertura em $\eta$ e granularidade das camadas dos calorímetros. . . . .	22
3.1	Número de anéis por camada do calorímetro. . . . .	36
3.2	Tempo gasto nas diferentes etapas do processo filtragem implementado pelo RingerNeuralHypo. . . . .	38
4.1	Número de RoI obtidas para cada padrão após a filtragem de nível um com o uso do menu e10. . . . .	42
4.2	Desempenho do algoritmo T2Calo. . . . .	64
5.1	Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis e normalizados de diferentes maneiras. . . . .	70
5.2	Desempenho do classificador RingerNeuralHypo operando sobre dados formatados em anéis e normalizados de diferentes maneiras. . . . .	73
5.3	Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	82
5.4	Desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	88
5.5	Desempenho do classificador linear operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	94
5.6	Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	98
5.7	Melhor desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	111

5.8	Desempenho do classificador linear de Fisher operando sobre as componentes principais extraídas de forma segmentada mostradas na Tabela 5.7.	113
5.9	Desempenho do classificador neural operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	115
5.10	Desempenho do classificador neural operando sobre as componentes principais extraída de forma segmentada mostradas na Tabela 5.9. . . . .	116
5.11	Melhor desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	118
5.12	Desempenho do classificador linear de Fisher operando sobre as componentes independentes extraídas de forma segmentada mostradas na Tabela 5.11. . . . .	120
5.13	Desempenho do classificador neural operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	121
5.14	Desempenho do classificador neural operando sobre as componentes independentes extraída de forma segmentada mostradas na Tabela 5.13. . .	123
5.15	Comparação dos resultados entre as diferentes abordagens propostas. . . .	126
5.16	Tempo requerido, em ms, para selecionar região de interesse e construir topologicamente os sinais em anéis para cada camada dos calorímetros. .	129
6.1	Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis e normalizados de diferentes maneiras. . . . .	133
6.2	Desempenho dos classificadores neural sobre dados formatados em anéis e normalizados de diferentes maneiras. . . . .	136
6.3	Anéis dentro da região $\Delta R$ por camada do calorímetro. . . . .	142
6.4	Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis, delimitados pela região $\Delta R$ , e normalizados de diferentes maneiras. . . . .	143
6.5	Desempenho do classificador neural operando sobre dados formatados em anéis, delimitados pela região $\Delta R$ e normalizados de diferentes maneiras. .	146
6.6	Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	150
6.7	Desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	155

6.8	Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis normalizados de diferentes maneiras. . . . .	161
6.9	Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	164
6.10	Melhor desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	171
6.11	Desempenho do classificador linear de Fisher operando sobre as componentes principais extraídas de forma segmentada mostradas na Tabela 6.10.	173
6.12	Desempenho do classificador neural operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	175
6.13	Desempenho do classificador neural operando sobre as componentes principais extraída de forma segmentada mostradas na Tabela 6.12. . . . .	176
6.14	Melhor desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	179
6.15	Desempenho do classificador linear de Fisher operando sobre as componentes independentes extraídas de forma segmentada apresentadas na Tabela 6.14. . . . .	180
6.16	Desempenho do classificador neural operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras. . . . .	181
6.17	Desempenho do classificador neural operando sobre as componentes independentes extraída de forma segmentada apresentadas na Tabela 6.16. .	184
6.18	Comparação dos resultados das diferentes abordagens propostas. . . . .	187



# Nomenclatura

$E$  Energia, p. 17

$E_T$  Energia Transversa, p. 32

$E_m$  *Missing Energy* - Energia Perdida, p. 21

$I_{X;Y}$  Informação mútua entre  $X$  e  $Y$ , p. 44

$L$  Luminosidade, p. 16

$P_T$  Momento transverso, p. 5

$R_i$  Relevância da  $i$ -ésima variável, p. 57

$\Delta R$  Raio do cone de ação da região analisada em uma RoI, p. 138

$\Delta\eta$  Variação espacial da coordenada  $\eta$ , p. 138

$\Delta\phi$  Variação espacial da coordenada  $\phi$ , p. 138

$\eta$  Coordenada do detector ATLAS, p. 19

$\mathbf{A}$  Matriz de mistura, p. 211

$\mathbf{C}_x$  Matriz de covariância da amostra  $\mathbf{X}$ , p. 213

$\mathbf{S}_B$  Matriz de espalhamento entre classes, p. 220

$\mathbf{S}_W$  Matriz de espalhamento intraclasses, p. 220

$\mathbf{V}$  Matriz de branqueamento, p. 217

$\mathbf{W}$  Matriz de separação, p. 212

$\mathbf{X}$  Amostra de Dados, p. 211

$\mathbf{m}_x$  Vetor média da amostra  $\mathbf{X}$ , p. 213

$\mathbf{s}$  vetor de sinais fonte, p. 211

$\mathbf{v}$  vetor de observação branqueado, p. 217  
 $\mathbf{x}$  vetor de observação, p. 211  
 $\mu$  múon, p. 11  
 $\phi$  Coordenada do detector ATLAS, p. 19  
 $\pi^{(+)(-)(0)}$  Mésons, p. 10  
 $\rho_{x,y}$  Correlação entre as variáveis  $x$  e  $y$ , p. 44  
 $\tau$  tau, p. 11  
 $PD_e$  Probabilidade de detecção de elétrons, p. 54  
 $PD_{lj}$  Probabilidade de detecção de *leptonjets*, p. 54  
 $PF_j$  Probabilidade de falso alarme de jatos, p. 54  
 $SP_{e/lj}$  Produto SP para canal elétron/jato considerando a eficiência na detecção de *leptonjets*, p. 55  
 $SP_{e/j}$  Produto SP para canal elétron/jato, p. 55  
 $SP_{lj/j}$  Produto SP para canal *leptonjets*/jato, p. 55  
 $\upsilon_{x,y}$  Correlação não-linear entre as variáveis  $x$  e  $y$ , p. 44  
 $a_{ij}$  coeficientes de mistura da matriz  $\mathbf{A}$ , p. 211  
 $b$  quark *bottom*, p. 11  
 $c$  quark *charm*, p. 11  
 $d$  quark *down*, p. 11  
 $e$  elétron, p. 11  
 $f(\cdot)$  Função não-linear, p. 44  
 $g(\cdot)$  Função não-linear, p. 44  
 $p(x)$  Função densidade de probabilidade de  $x$ , p. 44  
 $p(x,y)$  Função densidade de probabilidade conjunta de  $x$  e  $y$ , p. 44  
 $s$  quark *strange*, p. 11  
 $t$  quark *top*, p. 11

$u$  quark  $up$ , p. 11

$\nu_e$  neutrino do elétron, p. 11

$\nu_\mu$  neutrino do múon, p. 11

$\nu_\tau$  neutrino do tau, p. 11

$x_i$  variável de observação  $i$ , p. 211

SP Produto SP, p. 55

# Nomenclatura

- EM  $E_T$  Variável de corte do algoritmo T2Calo, p. 34
- Geant Simulador de interação entre partículas e partículas/detector, p. 31
- HAD  $E_T$  Variável de corte do algoritmo T2Calo, p. 34
- Pythia Simulador de Monte Carlo para colisões de partículas, p. 31
- RingerNeuralHypo Algoritmo de filtragem de nível 2 do ATLAS, p. 34
- Ringer Algoritmo de extração de características do classificador RingerNeuralHypo, p. 34
- Rshape Variável de corte do algoritmo T2Calo, p. 34
- Rstrip Variável de corte do algoritmo T2Calo, p. 34
- T2Calo Algoritmo de filtragem de nível 2 do ATLAS, p. 6
- e10 Menu de seleção de eventos, p. 5
- 1j4e10 Menu de seleção de eventos, p. 5
- ALICE *A Large Ion Collider Experiment*, p. 2
- ATLAS *A Toroidal LHC Apparatus*, p. 2
- BSS *Blind Source Separation* - Separação Cega de Fontes, p. 211
- CDF *Collider Detector at Fermilab*, p. 5
- CERN *Conséil Européen pour la Recherche Nucléaire*, p. 2
- CMS *Compact Muon Solenoid*, p. 2
- COPPE Instituto Alberto Luiz Coimbra de Pós-graduação e Pesquisa de Engenharia, p. 1
- CP *Charge Parity*, p. 17

D0 Detector do Fermilab, p. 5

DELPHI *DEtector with Lepton, Photon and Hadron Identification*, p. 5

DESY *Deutsches Elektronen-SynchrotronDie*, p. 5

DM *Dark Matter* - Matéria Escura, p. 12

ED *Eigenvalue Decomposition* - Decomposição em Autovalores, p. 214

EF Filtro de Eventos, p. 26

EM1 Primeira camada eletromagnética - EM *front layer*, p. 21

EM2 Segunda camada eletromagnética - EM *middle layer*, p. 21

EM3 Terceira camada eletromagnética - EM *back layer*, p. 21

EM Eletromagnético, p. 21

Fermilab *Fermi National Accelerator Laboratory*, p. 5

H1 Experimento do acelerador HERA, p. 5

HAD0 Primeira camada hadrônica - HAD *layer 0*, p. 21

HAD1 Segunda camada hadrônica - HAD *layer 1*, p. 21

HAD2 Terceira camada hadrônica - HAD *layer 2*, p. 21

HAD Hadrônico, p. 21

HEP *High Energy Physic* - Física de Alta Energia, p. 6

HERA *Hadron Electron Ring Accelerator*, p. 5

HLT *High-level Trigger*, p. 26

ICA *Independent Component Analysis* - Análise de Componentes Independentes, p. 3

LEP *Large Electron-Positron collider*, p. 5

LHC *Large Hadron Collider*, p. 2

LHCb *LHC-beauty*, p. 2

LVL1 Filtragem de nível 1, p. 26

LVL2 Filtragem de nível 2, p. 26

MLP *Multilayer Perceptron* - Percéptron de Múltiplas Camadas, p. 222

NLICA *Nonlinear Independent Component Analysis* - Análise de Componentes Independentes Não-Lineares, p. 196

NLPCA *Nonlinear Principal Component Analysis* - Análise de Componentes Principais Não-Lineares, p. 3

PCA *Principal Component Analysis* - Análise de Componentes Principais, p. 3

PCD *Principal Components for Discrimination* - Componentes Principais Discriminantes, p. 7

PS *Pre Sampler*, p. 21

PS *Proton Synchrotron*, p. 16

QCD *Quantum Chromodynamics* - Cromodinâmica Quântica, p. 14

RLS *Recursive Least Squares*, p. 214

RNA Rede Neural Artificial, p. 3

ROC Receiver Operating Characteristic, p. 54

RoI *Region of Interest* - Região de Interesse, p. 26

SM *Standard Model* - Modelo Padrão, p. 2

SPS *Super Proton Synchrotron*, p. 16

SUSY *Supersymmetry* - Supersimetria, p. 4

SVD *Singular Value Decomposition* - Decomposição em Valor Singular, p. 214

SVM *Support Vector Machine* - Máquina de Vetor de Suporte, p. 196

TDAQ *Trigger Data Acquisition* - Filtragem e Aquisição de Dados, p. 26

Zeus Detector do acelerador HERA, p. 6

# Capítulo 1

## Introdução

As diferentes técnicas de processamento digital de sinais e de análise de componentes, sejam elas: lineares, não-lineares e independentes; tiveram crescimento significativo na pesquisa científica básica e aplicada nas últimas décadas. Como resultado, temos a aparição freqüente de produtos industriais, das mais diferentes áreas e segmentos tecnológicos, que de alguma maneira se beneficiaram desse avanço.

A solução de um dado problema depende, muitas vezes, não apenas do desempenho final; é freqüente encontrarmos aplicações com requisitos de outra natureza, como: velocidade, custo computacional, portabilidade etc. Na escolha de cada uma das técnicas empregadas pesam características positivas e negativas e, o crescente avanço computacional permite a muitas delas, que em um período anterior eram requeridas apenas em situações específicas de abundância tecnológica, se aproximarem das ditas clássicas ou de mais fácil aplicação computacional. Como o avanço tecnológico traz consigo o desejo de solucionar problemas cada vez maiores e mais complexos, seja teórica ou experimentalmente, crescem também os recursos exigidos em suas soluções. Esse ciclo encontra na física de partículas mais uma aplicação.

### 1.1 Contexto

Em algumas instâncias de aplicação, a detecção de padrões requerida deve ser feita *online*, em janelas de tempo bastante limitadas. Aumentar a velocidade de processamento em um ambiente de alta granularidade (dimensão elevada) exige a utilização de técnicas de pré-processamento sobre um grande número de sensores (muitas vezes heterogêneos), o que pode implicar processamento segmentado e até distribuído (múltiplos processadores que trabalham paralelamente para alcançar o objetivo final). Não só isso, a complexidade pode crescer caso a informação procurada esteja inserida em uma quantidade de dados muito grande. No caso de eventos raros serem o alvo da detecção, a eficiência é crucial no sistema, uma vez que a perda da informação de interesse pode comprometer severamente

a funcionalidade do sistema. É com todos esses ingredientes que trabalha o sistema de filtragem de um detector de partículas.

O detector ATLAS [1] - *A Toroidal LHC Apparatus* - é um dos 4 detectores construídos ao redor do colisionador de partículas LHC [2]- *Large Hadron Collider* - no CERN, em Genebra, Suíça. O CERN - Centro Europeu de Pesquisa Nuclear - é o maior laboratório para a Física de altas energias na atualidade, conduzindo vários experimentos em regime de colaboração internacional. O projeto e a montagem do ATLAS foram conduzidos por uma colaboração de 37 países e contaram com mais de mil pesquisadores de aproximadamente 150 universidades e centros de pesquisa. O detector ATLAS, já em operação, é o maior detector de propósito geral do mundo, atualmente. Quando se atingir a operação em capacidade máxima, o LHC gerará aproximadamente  $10^9$  interações por segundo no detector. Uma das principais tarefas do detector será selecionar, dentro desse conjunto enorme de dados, os raros eventos de interesse. Espera-se uma frequência menor do que 1 Hz para esses eventos.

Os primeiros resultados do LHC são apresentados e pesquisadores dos quatro maiores experimentos - ALICE [3], ATLAS CMS [4] e LHCb [5] - começam a revelar conquistas dos primeiros meses de operação do LHC com energia em torno de 3,5TeV por feixe, uma energia três vezes e meia maior do que a alcançada por um acelerador de partículas anteriormente. Com essas primeiras medidas, os experimentos estão redescobrimdo as partículas que estão no coração do Modelo Padrão - SM (*Standard Model*) -, o bojo teórico que contém o entendimento atual das partículas da matéria e as forças que agem entre elas. Esse é um passo essencial antes do trabalho sobre novas descobertas. Dentre as bilhões de colisões já armazenadas, “redescobrir” velhas partículas mostra que o Modelo Padrão está trabalhando como esperado nos experimentos do LHC e esses estão bem preparados para entrar em um novo território de pesquisa. O LHC, que ainda evoluciona em seus testes operacionais, está progredindo continuamente para condições mais ambiciosas de operação. A luminosidade - uma medida da taxa de colisão - já aumentou em mais de mil vezes desde março de 2010. Esse rápido progresso tem permitido a geração de uma massa de dados que começa a ser processada ao redor do mundo nos diferentes centros de processamento computacional da colaboração.

Uma parcela importante das informações necessárias para a caracterização dos eventos é obtida do sistema de calorimetria, subdividido em partes eletromagnética e hadrônica, cada uma composta por diferentes camadas. Os calorímetros são medidores de energia formados a partir de um enorme número de sensores (também chamados de células). Esses sensores quantificam a energia perdida (depositada) pelas partículas incidentes, sendo os perfis dessa deposição de energia usados para a identificação de partículas e remoção de ruído de fundo do experimento. Com uma alta taxa de eventos e dimensão elevada de dados (da ordem de 1000 sensores de calorimetria por região de interesse), a tarefa de processamento de sinais ganha destaque na cadeia de operação implementada,



o que sugere o uso de técnicas de compactação de dados, além do uso de técnicas de processamento inteligente de sinais.

Dentre as principais técnicas de compactação, a Análise de Componentes Principais [6] - PCA (*Principal Component Analysis*) - é encontrada com grande frequência na literatura. Baseada em estatística de segunda ordem, obtém uma base de transformação em que os dados transformados estão dispostos hierarquicamente nas direções de máxima variância. Por sua vez, a Análise de Componentes Principais Não-linear [7] - NLPCA (*Nonlinear Principal Component Analysis*) - baseada em estatística de ordem superior, generaliza a PCA, obtendo uma base de transformação em que os dados transformados estão dispostos de maneira a minimizar a correlação não-linear entre si. Além delas, a Análise de Componentes Independentes [7] - ICA (*Independent Component Analysis*) busca a independência estatística nas variáveis transformadas. A ICA pode ser implementada usando algoritmos de NLPCA.

A extração de característica é um importante bloco de processamento em um ambiente com as características acima mencionadas. Um processamento inteligente nessa etapa do processo pode influenciar positivamente outros blocos de processamento, como o teste de hipótese, tendo em vista que este usa diretamente a informação gerada neste ponto. No que diz respeito ao teste de hipótese, diferentes algoritmos são encontrados na literatura: dos clássicos lineares como o de Fisher [8] às Redes Neurais Artificiais [9] - RNA - usadas como classificadores de padrões. As redes neurais fazem uso da estatística dos dados na solução de problemas de reconhecimento de padrões, classificação e filtragem, o que as credenciam cada vez mais na solução de problemas desse tipo, onde a tarefa é dita complexa.

## 1.2 Motivação

A física experimental de altas energias, ao longo do tempo, mostrou-se como um dos ramos da ciência que mais se relaciona e demanda recursos da área de tecnologia e de sistemas de processamento. Dada a natureza peculiar dos experimentos realizados e sua busca por confirmar, experimentalmente, teorias propostas pela sua contra-parte teórica, muitas vezes esses recursos são tão complexos que a sua aplicação se torna uma atividade pioneira.

Ao buscarmos (pontualmente) o desconhecido é primordial a construção de sistemas de filtragem robustos, uma vez que o evento de interesse é (normalmente) raro. Dentre os eventos gerados nas colisões do LHC, apenas uma pequena parcela será útil para a caracterização dos processos físicos de interesse. Por exemplo, as assinaturas do instável bóson de Higgs podem ser encontradas a partir de seu decaimento (em um estágio final) em elétrons ou múons. Os taus também podem levar aos modelos de Higgs estendidos. A identificação desses e outros objetos é de interesse do sistema de filtragem.

Neste trabalho, o objeto de maior interesse na filtragem do detector ATLAS será o elétron. Os elétrons estão imersos em um intenso ruído de fundo composto de jatos hadrônicos. Em termos da calorimetria (equipamento usado para caracterizar um evento através da deposição energética desse em seus elementos passíveis - sensores), as assinaturas dos jatos podem ser confundidas com as dos elétrons quando apresentam perfis de deposição concentrados nas camadas eletromagnéticas e pouca energia nas camadas hadrônicas, estas últimas projetadas especificamente para absorver interações hadrônicas. O canal elétron/jato (e/j) é fundamental para o desempenho do detector.

Outro ponto de interesse crescente é a explicação para o que os físicos e astrofísicos chamam de Matéria Escura. Ela tem sido um problema a ser desvendado por pelo menos 75 anos. Dentre as diferentes demandas sobre o assunto destaca-se: do que ela é feita? Qual é a sua natureza? Nenhuma das partículas do Modelo Padrão (que encapsula as leis da Física que conhecemos hoje) têm as propriedades para preencher os requisitos da matéria escura fria, requerendo uma revisão no conhecimento das leis atuais da Física. Das várias teorias existentes encontra-se a que atribui um anômalo excesso de léptons no setor visível da matéria como resultado da aniquilação da matéria escura em uma escala elevada de energia. Esses jato de léptons - *leptonjets* - seria um dos objetos de interesse do sistema de filtragem. Contudo, ele está imerso em um ruído de fundo composto de partículas hadrônicas, provenientes ou não desse mesmo decaimento com modos hadrônicos.

Revelar a natureza da matéria escura fria no universo é um dos ramos de pesquisa mais excitantes hoje em dia. Felizmente, um conjunto de experimentos em operação, sendo desenvolvidos ou em planejamento, prometem um rápido progresso na descoberta das propriedades da matéria escura fria nos próximos anos [10–14].

### 1.3 Objetivo

A volta das operações do LHC no fim de 2009, após o incidente de 2008 - uma conexão elétrica defeituosa entre dois dos ímãs do acelerador resultou em danos mecânicos e liberação de hélio (a partir dos ímãs) para dentro do túnel do LHC - permitiram abrir uma nova janela de pesquisa física na escala de TeV e permitirão a descoberta do bóson de Higgs, se ele existir, ou de possíveis esquemas alternativos para mecanismos da quebra espontânea da simetria. Além disso, o LHC possibilitará a resposta de uma das questões mais instigantes dos dias atuais advinda de observações astronômicas, astrofísicas e de experimentos com raios cósmicos: qual é a partícula física candidata a componente da Matéria Escura do Universo? Supersimetria - *Supersymmetry* (SUSY) - é um cenário teoricamente atrativo para a física além do Modelo Padrão que pode fornecer um adequado candidato à Matéria Escura no caso da paridade R ser conservável. Se a SUSY for acessível na escala de TeV, como defendidos por alguns estudiosos, ela será acessível no LHC dentro dos primeiros anos de coleta de dados.

Dentre os diferentes canais de interesse contidos no ATLAS, este trabalho se encontra focado na discriminação de eventos exóticos dentro do canal elétron/jato. Os elétrons com alto momento transversal ( $P_T$ ) são componentes de muitas das prováveis assinaturas dos bósons de Higgs. Porém, é esperado um grande ruído de fundo de jatos hadrônicos que tem perfil de decaimento próximo ao dos elétrons. A separação eficiente entre esses padrões é crucial para o bom funcionamento do detector.

Além disso, será analisado o caso dos chamados *leptonjets*, que representam o decaimento em estágio final de múltiplos léptons dentro de uma mesma região de análise. Almeja-se maximizar a detecção de *leptonjets* em detrimento ao ruído de fundo também composto por jatos hadrônicos, o que aumentaria o entendimento de possíveis candidatos à Matéria Escura, considerando a literatura abordada no trabalho.

Para alcançar estes objetivos serão seguidas duas linhas de pesquisa. A primeira considera a infraestrutura existente do detector ATLAS, incluindo o uso de canal e *menu* existentes, como, por exemplo, o canal elétron/jato e seu *menu* de seleção  $e10$ , responsável por filtrar elétrons com determinada especificidade em cada um dos níveis de filtragem *online*. Nessa abordagem a separação entre elétrons e jatos é maximizada e um ponto de operação é obtido. Sobre esse ponto de operação é verificada a eficiência na detecção de *leptonjets*. O objetivo é fazer uso de uma estrutura consolidada no experimento para estudar os eventos exóticos, sem perda de eficiência para a primeira.

A segunda linha de pesquisa propõe um novo canal *leptonjets*/jato de seleção e um novo *menu* de seleção, por exemplo,  $1j4e10$ . Esse *menu* seria responsável por filtrar *leptonjets* com as especificidades desejadas mesmo permeado pelo ruído de fundo característico. Essa segunda linha de pesquisa apresenta uma alternativa futura, dado que a inserção de um novo canal de seleção é uma tarefa complexa, principalmente, em um ambiente de colaboração tão distribuído e segmentado, em que os requisitos de eficiência são tão elevados.

## 1.4 Estado da Arte

Os problemas relacionados ao projeto de filtragem de eventos (velocidade, dimensão de dados elevada e níveis de ruído) tendem a tornar o processo de classificação uma tarefa bastante complexa.

Dentre as técnicas de processamento inteligente de sinais, as redes neurais artificiais têm sido usadas em aplicações científicas e comerciais devido a sua adequada capacidade de generalização e fácil aplicação comparadas às técnicas tradicionais estatísticas. O uso de algoritmos neurais na solução de problemas experimentais da física de partículas vem ganhando espaço, de maneira que é possível encontrar publicações cuja a aplicação de redes neurais soluciona a busca por determinadas assinaturas de partículas do Modelo Padrão para diferentes colisionadores e detectores já construídos.

No domínio da física de altas energias, as redes neurais são geralmente usadas em problemas de reconhecimento de padrões: identificação de jatos de quark e glúon [15–17]; *top* quarks [18]; bóson de Higgs [19–21] seja no detector DELPHI [22] do acelerador LEP [23] seja no detector CDF II [24] do Tevatron [25] no Fermilab [26]; buscas por bósons de Higgs do Modelo Padrão [27, 28] tanto no CDF II quanto para o D0 [29], ambos detectores do Tevatron, assim como algumas previsões para a sua implementação no LHC [30]; detecção de traço [31–33]; filtragem [34, 35]; mineração de dados e tarefas de classificação em geral [36].

Em [37], os autores apresentam um algoritmo de identificação de elétrons baseado em redes neurais artificiais alimentado por dados de calorimetria do detector Zeus [38] acoplado ao acelerador/colisionador elétron-próton HERA [39] do laboratório DESY [40] em Hamburgo, Alemanha. No mesmo laboratório, o Experimento H1[41] do acelerador HERA, faz uso em seu segundo nível de filtragem *online* de eventos de um algoritmo neural [42, 43], enquanto os autores de [44] apresentam uma alternativa de filtragem de primeiro nível do experimento L3 [45] do acelerador/colisionador LEP baseada em algoritmos neurais.

Um resumo dos primeiros trabalhos envolvendo o uso de algoritmos neurais em física de altas energia, assim como algumas perspectivas futuras, é apresentado em [46, 47]. A primeira menção de redes neurais no contexto de física de altas energia - *High Energy Physics* (HEP) - data de 1988 [31]. Nos anos seguintes, pesquisadores usaram técnicas neurais para classificação de padrão, análise de evento e aplicações de filtragem. Nos dias de hoje, uma busca na literatura revela um grande número de publicações em que técnicas neurais têm sido, majoritariamente, testadas e avaliadas em cada uma dessas aplicações, em detrimento dos algoritmos atuais de operação.

Como uma colisão entre partículas é, no Modelo Padrão, uma interação entre os constituintes elementares que elas contém, a análise do conjunto de dados que incide sobre os detectores permite uma determinação experimental das propriedades físicas dos constituintes em uma avaliação da validade do modelo. Reconstruir um evento em HEP envolve tanto reconhecimento de padrão em baixo nível, tal como encontrar trajetórias ou encontrar e classificar os léptons candidatos (elétrons, múon, tau) quanto determinação de processos físicos, em que propriedades mais sofisticadas do evento, tais como distribuição angular dos jatos, presença de léptons mais jatos, etc., são realizadas. Outro ponto a se ressaltar diz respeito ao reconhecimento de padrão aplicado: *online* ou *offline*.

No caso específico do detector ATLAS, onde o trabalho está inserido, algoritmos e abordagens com desempenho superior ao T2Ca1o (software de referência do segundo nível de filtragem do ATLAS) vêm sendo propostos na literatura, contudo, a estratégia a ser adotada ainda não está definida, sendo essa uma das frentes de pesquisa mais importantes nessa fase do experimento.

Os trabalhos [36, 48] propuseram o uso de um classificador neural supervisionado

(rede percéptron de múltiplas camadas) para o segundo nível de filtragem do canal elétron/jato do detector ATLAS. Utilizando informação especialista a respeito do problema, os sinais medidos no sistema de calorimetria foram formatados topologicamente em anéis concêntricos a partir do centro de deposição de energia. Essa informação foi usada para alimentar um discriminador neural. Em [49], o sistema neural de filtragem - RingerNeuralHypo - foi incorporado ao ATLAS. Foi demonstrado que o algoritmo em questão supera o software de referência - T2Calo - em termos de eficiência com um aumento no tempo de processamento da aproximadamente 90%, mas ainda dentro do tempo permitido de operação do segundo nível de filtragem.

No que diz respeito ao uso de técnicas de análise de componentes, o seu uso seguiu praticamente o mesmo caminho trilhado pelas redes neurais. Quando experimentos eletrônicos entraram em operação nos anos 1970, o número de trajetórias candidatas falsas produzidas pelo algoritmo de busca de trajetórias ao analisar os equipamentos responsáveis era tão grande que um passo adicional entre a busca de trajetórias e a identificação de trajetórias fora necessário para a rápida eliminação da maior parte das trajetórias falsas. Um dos métodos aplicados foi a Análise de Componentes Principais (PCA). A PCA usava um conjunto de trajetórias válidas para determinar uma região de um espaço característico multidimensional em que as trajetórias válidas deveriam estar. Ao mapear o conjunto de medidas que constituem uma trajetória candidata para um ponto nesse espaço, uma decisão rápida pode ser tomada quer a trajetória candidata seja aceita ou não [50].

De volta ao ATLAS, os autores de [51] propõem um processamento segmentado da informação de calorimetria através da PCA. O objetivo é aumentar a velocidade de processamento com a redução da dimensionalidade da informação usada no segundo nível de filtragem do detector ATLAS. Devido a fatores como alta taxa de eventos e dimensão elevada dos dados, no trabalho [52] métodos de compactação da informação como PCA e PCD - *Principal Components for Discrimination* - foram aplicados em conjunto com o modelo linear da ICA. Já em [53], a extensão dos métodos de análise de componentes alcança a abordagem não-linear da ICA.

## 1.5 Abordagem proposta

A abordagem usada no procedimento de detecção/decisão de eventos é mostrada na Figura 1.1.

Os dados de calorimetria providos pelo primeiro nível de filtragem são condicionados na etapa de formatação do sinal de maneira a preservar as características físicas de deposição de energia do processo, facilitando a operação sobre os mesmos pela etapa seguinte. A cada evento, regiões diferentes do calorímetro são sensibilizadas e usadas pelos algoritmos de extração de características nas mais diferentes abordagens.

O pré-processamento envolve diferentes técnicas de processamento digital de sinais,

o resultado é o fornecimento de um conjunto de dados com maior informação agregada à etapa de teste de hipótese e consequente decisão elétron/jato com verificação da aceitação de *leptonjets* ou, em um segundo cenário, *leptonjets/jato*.

O algoritmo de classificação apresenta uma estimativa do evento pertencer a classe de interesse ou não. No caso dos classificadores neurais, essa estimativa é baseada em um treinamento prévio, já para o algoritmo de referência do sistema, a inferência é realizada através de cortes sobre variáveis previamente estabelecidas, baseadas em características físicas do processo de colisão e/ou do detector.

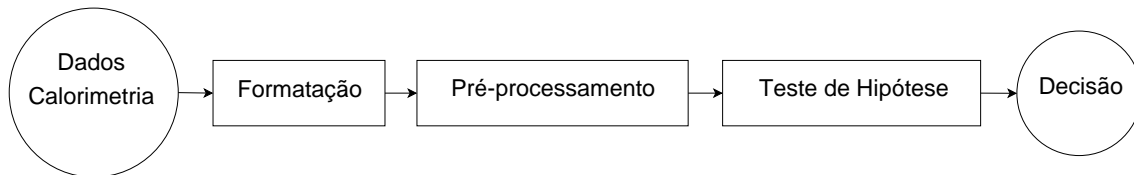


Figura 1.1: Procedimento de detecção/decisão de eventos.

## 1.6 Organização do Trabalho

O trabalho está organizado da seguinte maneira: o Capítulo 2 apresenta o ambiente no qual o experimento está submetido, assim como uma breve introdução à física de partículas. O Capítulo 3 descreve o ambiente de filtragem do detector ATLAS. No Capítulo 4, dados, medidas de desempenho e metodologia usados no trabalho são explicados, assim como o desempenho dos algoritmos de referência atuais. Resultados são apresentados nos Capítulos 5 e 6. As conclusões e sugestões para trabalhos futuros são destacados no Capítulo 7.

## Capítulo 2

# A Física de Partículas e o Experimento ATLAS

O histórico de descobertas na área de física de partículas remete ao começo do século XIX com pequenos e grandes avanços nos últimos dois séculos. Essas descobertas sempre tiveram uma relação muito íntima com o ferramental tecnológico contemporâneo. Dessa maneira, as lacunas presentes nas teorias atuais requerem equipamentos tão sofisticados para a sua descoberta que encontram no LHC e no ATLAS esperança para algumas de suas perguntas.

### 2.1 Breve Histórico da Física de Partículas

No começo dos anos trinta do século passado, a estrutura do átomo estava bem estabelecida e a estrutura do núcleo era uma das áreas de pesquisa mais desenvolvida. Acreditava-se que os componentes básicos da matéria seriam elétrons, prótons, nêutrons e neutrinos - postulados por Wolfgang Pauli, em 1931, para explicar uma perda anômala de energia no decaimentos de nêutrons, detectados diretamente apenas em 1956. O nêutron, detectado em 1932, havia sido sugerido um pouco antes para explicar a massa nuclear. Antes, pensava-se que o núcleo poderia ser constituído de prótons e elétrons, com excesso de prótons para explicar sua carga positiva. Contudo, medições do *spin* nuclear descartaram essa possibilidade [54].

A hipótese e a detecção do nêutron colocaram o problema da estabilidade do núcleo em evidência e com ela algumas questões. Sendo o núcleo composto de prótons e nêutrons, como conciliar a existência de um grande número de prótons, particularmente nos elementos pesados, em um espaço tão pequeno? A repulsão elétrica entre eles seria tão grande que levaria o núcleo a explodir.

Em 1935, o pesquisador Hideki Yukawa propôs a existência de uma nova partícula que seria a mediadora da interação que manteria nêutrons e prótons coesos no núcleo.



A interação entre prótons e nêutrons deveria ser mediada por alguma partícula, ou seja, prótons e nêutrons interagiriam trocando uma partícula. Essa partícula foi denominada méson  $\pi$ , ou pión. Um pión poderia ser emitido por um nêutron e absorvido por um próton, ou vice-versa, fazendo com que o nêutron e o próton exercessem uma força um sobre o outro. Essa outra força foi chamada de força nuclear e a correspondente interação de forte [54].

A partícula proposta por Yukawa, o méson  $\pi$ , foi finalmente detectada, em 1947, com a massa por ele prevista, em um laboratório na Universidade de Bristol, em emulsões fotográficas sobre as quais incidiam partículas cósmicas. Em 1948, mésons  $\pi^+$  e  $\pi^-$  foram produzidos em aceleradores de partículas, na Universidade de Berkeley, e em 1950 foi produzido o méson  $\pi^0$ , também em colisões provocadas em aceleradores. A curiosidade nesse registro histórico fica por conta da contribuição brasileira [55].

Mais tarde, em 1953, o pesquisador Murray Gell-Mann, sugeriu que certas partículas subatômicas teriam uma propriedade chamada estranheza. A estranheza é uma propriedade da matéria, análoga à carga elétrica, e governa a velocidade com que as partículas decaem. Mas admitindo que existe tal propriedade, é possível explicar, modelar, prever vários processos físicos. Analogamente, há outras propriedades da matéria que não sabemos exatamente o que são, mas que admitindo sua existência os físicos podem, por exemplo, prever o resultado de certos processos. A suposição da estranheza permitiu aos físicos prever, com sucesso, que determinadas partículas seriam produzidas em certas reações, se decaíssem em determinado tempo.

Murray Gell-Mann e outro físico chamado George Zweig concluíram, independentemente, que padrões na constituição de bárions e mésons resultariam naturalmente se algumas das partículas fundamentais do átomo fossem formadas por partículas ainda mais fundamentais que ficaram conhecidas como quarks. Hoje aceita-se que os quarks, assim como os elétrons, são as partículas verdadeiramente elementares da matéria, uma espécie de tijolo básico para a construção de toda a matéria, inclusive dos nêutrons e prótons [54].

O problema com a teoria dos quarks era que tais partículas tinham propriedades muito peculiares, para não dizer misteriosas: sua carga elétrica seria fracionária ( $1/3e$ ,  $2/3e$ ), não existiriam como partículas livres e constituiriam os hádrons sempre em pares quarks-antiquarks ( $q\bar{q}$ , mésons) ou em tríades de quarks ( $qqq$ , bárions) [54].

Por outro lado, comparando o mundo dos hádrons e o dos léptons, notava-se que havia apenas seis léptons e muitos hádrons. Isso reforçava a hipótese de que esses seriam partículas compostas de outras mais elementares.

A evidência experimental dos quarks foi considerada convincente apenas a partir da década de 1970, a chamada década de ouro da física de partículas, através de reações de altas energias em aceleradores/colisionadores de partículas e mais tarde refinadas como o acelerador Linear de Stanford, o Tevatron do Fermilab, em Batavia, Illinois e o Grande Colisor Elétron-Pósitron do CERN.



Desde a década de 1970, os físicos conseguiram um ferramental teórico para explicar, em parte, diversos fenômenos. Porém, há evidências de que essa teoria não seja suficiente para explicar algumas outras questões. No mínimo, o LHC deverá determinar a exatidão dessa teoria, o que já será uma grande contribuição dessa máquina. Não só isso, é possível que o LHC, ao investigar a questão, encontre novos fenômenos que ajudem no entendimento da matéria.

## 2.2 O Modelo Padrão

A teoria que tenta descrever a natureza da matéria ou do que é feito o universo e como se aglutinam suas partes em termos de quatro forças, quatro partículas mediadoras e doze partículas fundamentais, é chamada de Modelo Padrão. As doze partículas fundamentais são os seis léptons e os seis quarks; as quatro partículas mediadoras são os fótons, os glúons, as partículas  $W$  e  $Z$  e os grávitons; as quatro forças são a eletromagnética, a forte, a fraca e a gravitacional [54, 56].

As partículas fundamentais, ou partículas de matéria, são chamadas de férmions. Léptons e quarks são, portanto, subclasses de férmions. Léptons não são influenciados pela força nuclear forte, não estão encerrados dentro de partículas maiores e podem viajar por conta própria. Elétrons, múons e neutrinos são léptons. Quarks sofrem influência da força forte e estão sempre confinados em partículas maiores (hádrons) [54, 56].

As partículas mediadoras (com *spin* inteiro) que transmitem as quatro forças da natureza são chamadas bósons. Enquanto os férmions são partículas de matéria, os bósons são partículas de força. A Tabela 2.1 esquematiza a constituição da matéria segundo o Modelo Padrão. Cada partícula de matéria possui a sua partícula de antimatéria.

O Modelo Padrão das partículas elementares não é um simples modelo físico, é um referencial teórico que incorpora a cromodinâmica quântica (a teoria da interação forte) e a teoria eletrofraca (a teoria da interação eletrofraca que unifica as interações eletromagnética e fraca). Aqui aparece uma grande dificuldade do Modelo Padrão, talvez a maior: ele não consegue incluir a gravidade, porque a força gravitacional não tem a mesma estrutura das três outras forças, não se adapta à teoria quântica, a partícula mediadora hipotética - o gráviton - não foi ainda detectada [54, 56]. Nesse contexto se insere o bóson de Higgs. Pelo modelo, a interação com o campo de Higgs (ao qual esta partícula estaria associada) daria massa às partículas. Acredita-se que todo o espaço seja preenchido com esse campo e interagindo com ele, as partículas adquirem suas massas. Partículas que interajam fortemente com o campo são pesadas, enquanto aquelas que interajam fracamente tornam-se mais leves. Porém, o modelo não explica completamente essas interações e o bóson de Higgs está ainda por ser detectado. É um paradoxo que a massa, uma propriedade tão familiar à matéria, seja um dos assuntos mais nebulosos e instigantes na pesquisa de física de partículas. Um dos principais objetivos de projeto do ATLAS é detectar essa partícula,

Tabela 2.1: Uma visão esquemática da constituição da matéria segundo o Modelo Padrão.

Matéria			
Partículas Matéria			
Léptons (Férmions)		Quarks (Férmions)	
elétron ( $e$ )		quark <i>up</i> ( $u$ )	
neutrino do elétron ( $\nu_e$ )		quark <i>down</i> ( $d$ )	
múon ( $\mu$ )		quark <i>charm</i> ( $c$ )	
neutrino do múon ( $\nu_\mu$ )		quark <i>estranho</i> ( $s$ )	
tau ( $\tau$ )		quark <i>bottom</i> ( $b$ )	
neutrino do tau ( $\nu_\tau$ )		quark <i>top</i> ( $t$ )	
Hádrons			
Bárions (três quarks)		Mésons (pares quarks-antiquarks)	
Forças (interações) fundamentais			
Eletromagnética	Fraca	Forte	Gravitacional
Eletrofraca			
Partículas de força (Bósons)			
Fótons	$W$ & $Z$	Glúons	Grávitons (não detectados)

se existir, o que seria uma das maiores descobertas físicas de todos os tempos. O bóson de Higgs foi previsto teoricamente em 1964 pelo físico escocês Peter Higgs e usado posteriormente por Steven Winberg (1967) e Abdus Salam (1968) para explicar porque outras partículas, os bósons  $W$  e  $Z$ , têm massa.

Outra dificuldade do Modelo Padrão é explicar a assimetria matéria-antimatéria. Por que, atualmente, praticamente tudo é feito de matéria? Por que a antimatéria é raramente encontrada na natureza? Há, ainda, diversas outras dificuldades. Algumas são resultantes das limitações do modelo. Como toda teoria física, esse modelo não pode explicar tudo. Outras, como a do bóson de Higgs, podem levar a modificações na teoria. Se a partícula, prevista teoricamente pelo modelo para explicar a massa das partículas, não for detectada a teoria teria que ser modificada. O importante aqui é dar-se conta que o Modelo Padrão da física de partículas é a melhor teoria sobre a natureza jamais elaborada pelo homem, com muitas confirmações experimentais. Por exemplo, o modelo previu a existência das partículas  $Z$  e  $W$ , dos quarks *charme* e *top* que foram todas posteriormente detectadas com as propriedades previstas. Mas, nem por isso é uma teoria definitiva. As teorias físicas não são definitivas, ainda que sejam tão bem sucedidas como o Modelo Padrão [54, 56].

## 2.3 Matéria Escura

A Matéria Escura - *Dark Matter* (DM) - tem sido um problema astrofísico por, pelo menos, 75 anos - desde as divulgações das observações de Fritz Zwicky de uma grande velocidade de dispersão dos membros da galáxia Coma [57]. O problema de rotação das

galáxias - as estrelas giram “rápido demais” para pertencerem à gravidade Newtoniana (caso toda a matéria presente seja visível) - que remete às medidas de Baabcock sobre a galáxia Andrômeda em 1939 [58] é similar. Esse cenário levou, contudo, várias décadas para ser reconhecido como um problema real e em sua forma moderna remete ao fim dos anos 1970 e começo dos anos 1980 quando o então chamado paradigma da matéria escura fria apareceu (nesse contexto, fria significa matéria se movendo com velocidade não relativística). Hoje em dia a abundância de dados de estudos da radiação de fundo, medidas de distância de supernovas e uma grande compreensão das galáxias têm, conjuntamente, solidificado o chamado "Modelo Padrão" da cosmologia [59, 60].

Revelar a natureza da matéria escura no Universo é um dos ramos de pesquisa mais excitantes hoje em dia. Felizmente, um conjunto de experimentos em operação [10–14], sendo desenvolvidos ou em planejamento prometem um rápido progresso na descoberta das propriedades da matéria escura nos próximos anos. Existem diferentes teorias sobre o assunto, entre elas as que produzem no estado final *leptonjets*.

### 2.3.1 *Leptonjets*

À luz das recentes observações astrofísicas, os autores de [61] propuseram uma ampla classe de teorias em que a aniquilação da matéria escura (da escala de TeV) na borda galáctica contribui para o excesso anômalo de raios cósmicos de léptons. Embora essa matéria escura seja, provavelmente, inacessível para aceleradores/colisionadores, ela está acompanhada por um setor escuro (da escala de GeV) que se acopla, embora muito fracamente, ao Modelo Padrão. Os estados do setor escuro são relativamente leves e podem ser produzidos em aceleradores de alta energia, criando decaimentos em cascata a partir de si e, finalmente, regressando para o setor visível como elétrons, múons e, possivelmente, píons. Os léptons obtidos emergem no detector como jatos de léptons, chamados *leptonjets* [62], que são múltiplos múons ou elétrons altamente colimados resultantes do decaimento desses estados do setor escuro altamente "impulsionados". Como observado em [62] e posteriormente desenvolvido em [63], a adição da supersimetria também carrega consigo um número interessante e novo de assinaturas de colisão com eventos mais espetaculares. Parte da fenomenologia é também relevante para cenários com vales escondidos [64] ou onde a matéria escura decai em um setor da escala de GeV [65] e, como tais, podem descrever uma grande classe de modelos. Recentemente, esforços experimentais na busca de tais objetos têm sido relatados em [66].

A natureza do setor escuro implica a presença de uma complicada cadeia de decaimentos. Alguns dos típicos decaimentos são mostrados na Figura 2.1. Para um setor escuro com mistura cinética com o Modelo Padrão, o bóson de gauge ( $\gamma'$ ) decai para um di-lépton como estado final (primeiro esquema no lado esquerdo da Figura 2.1). Por outro lado, um setor escuro não-abeliano, resulta em complicadas cadeias de decaimentos

(esquemas restantes da Figura 2.1). O Higgs do setor escuro ( $h'$ ), necessário para quebrar o grupo não-abeliano, pode também participar de tais cascatas e, inevitavelmente, produzir múltiplos léptons no estado final (facilmente  $> 2$  e possivelmente 8) que provêm uma única assinatura do setor escuro. Modelos simplificados não invocam a SUSY [67] e produzem um *leptonjet* com 4 léptons.

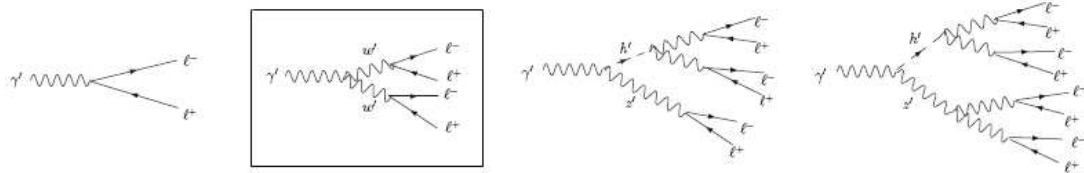


Figura 2.1: Cadeias típicas de decaimentos iniciando com um bóson de gauge escuro (rotulado  $\gamma'$ ). A cadeia de decaimento pode ter diversas etapas e envolver outros estágios do setor escuro, como outros bósons de gauge escuros (rotulados  $w'$  e  $z'$ ), e Higgs escuros (rotulados  $h'$ ) [63].

A mistura cinética entre o portador de força escura e o fóton do Modelo Padrão induz uma pequena carga nos campos do Modelo Padrão eletromagneticamente carregados. Conseqüentemente, os bósons de gauge escuros podem ser diretamente produzidos nos colisionadores via um processo análogo à produção de fótons no Modelo Padrão, como mostra a Figura 2.2. Após os bósons serem produzidos eles tipicamente decaem em múltiplos léptons, formando os já mencionados *leptonjets*.

O desafio encontrado na filtragem dos *leptonjets* se refere à característica de deposição de energia desses eventos nos calorímetros. Eles possuem semelhanças eletromagnéticas com outros objetos de interesse ( $e$ ), porém, a multiplicidade de elétrons dentro de um mesmo cone de decaimento acaba os aproximando do ruído de fundo (jatos hadrônicos) que também podem conter partículas eletromagnéticas.

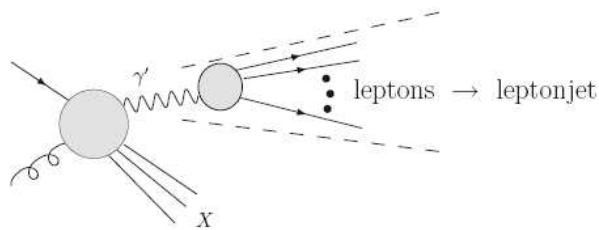


Figura 2.2: Produção direta de um bóson de gauge escuro em um processo muito similar à produção de fótons no Modelo Padrão [63].

*Leptonjets* podem surgir se existir um setor leve oculto composto de partículas instáveis com massas no intervalo de MeV a GeV. Uma classe desses modelos contém uma partícula (um fóton oculto) que tem uma pequena mistura cinética com o fóton do Modelo Padrão. Devido a essa mistura, o fóton oculto pode decair para partículas mais leves

com carga elétrica. Por exemplo, um fóton oculto com massa de 100 MeV decai exclusivamente para elétrons, enquanto um com massa de 1 GeV decai para elétrons, múons e píons. No Tevatron e no LHC, fótons ocultos e outras partículas leves ocultas são produzidos com grande impulso, provocando em seus decaimentos visíveis estruturas assemelhadas a jatos hadrônicos (jatos de QCD - *Quantum Chromodynamics*), o que obriga o desenvolvimento de técnicas que eficientemente isolem o sinal da física de interesse do ruído hadrônico. Tais objetos podem ter alguns estados hadrônicos em si, por exemplo, se a assinatura contiver píons carregados ou se Higgs escuros aparecerem com modos hadrônicos de decaimento. Ainda assim, espera-se que o conteúdo eletromagnético dos léptons seja mais rico e permita sua distinção dos jatos hadrônicos.

Outro motivador do estudo de *leptonjets* diz respeito às recentes associações entre estes e a produção de Higgs do Modelo Padrão. Os autores de [68, 69] cogitam, inclusive, a possibilidade da produção abundante dessas partículas, tanto no LEP quanto no Tevatron, sem que tenham sido detectadas. Nesse cenário, Higgs decairiam, predominantemente, em setores escondidos seja diretamente seja através de estados supersimétricos leves. Cascatas subsequentes aumentam a multiplicidade de partículas no setor escondido que, após decaírem novamente para o Modelo Padrão, aparecem no detector como *clusters* de léptons colimados. Os autores acreditam que um Higgs decaindo em *leptonjets* pode não ser observado quando a topologia do evento mimetiza a do ruído hadrônico. Até os dias de hoje, Higgs decaindo para *leptonjets* não têm sido alvo de qualquer análise experimental e a eficiência das buscas existentes para esse tipo de sinal é baixa. A única exceção é a busca apresentada pelo D0 [70].

## 2.4 LHC

Em 2005 começou a funcionar o maior acelerador de partículas do mundo. Conhecido pela sigla LHC (*Large Hadron Collider*), o equipamento pode ser considerado o maior empreendimento científico da atualidade, seja por seu tamanho, tecnologia, número de cientistas envolvidos ou por seus objetivos. A máquina fará partículas se chocarem com velocidades e energias impressionantes, ajudando na busca de respostas sobre como a natureza se comporta [71, 72].

Com 27 km de circunferência, o LHC (Figura 2.3) está encravado a 100m de profundidade na fronteira da Suíça com a França, onde se encontra a Organização Européia para a Pesquisa Nuclear (CERN), que o abriga.

O objetivo do LHC é ajudar a compreender a estrutura da matéria em dimensões inferiores ao tamanho dos prótons:  $10^{-18}$ m. As motivações são várias. Acredita-se que a origem da quebra da simetria eletrofraca será elucidada. Isso pode acarretar a descoberta de um ou mais bósons de Higgs. Outro campo de atividade será a busca por novos tipos de simetria e partículas, especificamente as que compreendem a teoria da Supersimetria -

*Supersymmetry* (SUSY) [67]. SUSY é o mais proeminente e cuidadoso modelo de estudo de todas as extensões do Modelo Padrão propostas. Ele postula a simetria entre férmions e bósons e introduz um novo espectro de partículas bastante rico. Além disso, algumas implementações da SUSY prevêm um excelente candidato para a matéria escura em nosso Universo. Sua primeira missão será estudar elementos previstos ou mal compreendidos na teoria atual, o chamado Modelo Padrão. A segunda missão do acelerador é buscar novos fenômenos físicos na altíssima escala de energia que será atingida pelo LHC em pequeníssimos volumes de espaço [72]. O número de colisões por centímetro quadrado produzida por segundo é chamado de luminosidade ( $L$ ), dado por

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

onde  $n$  é o número de feixes,  $N_i$ , o número de partículas por feixe,  $A$ , a área transversal do feixe e  $f$ , a frequência de colisão. Quanto maior a luminosidade do experimento maior a quantidade de partículas geradas, ou seja, mais informação para ser analisada. A luminosidade é uma importante medida para caracterizar o desempenho de um acelerador.

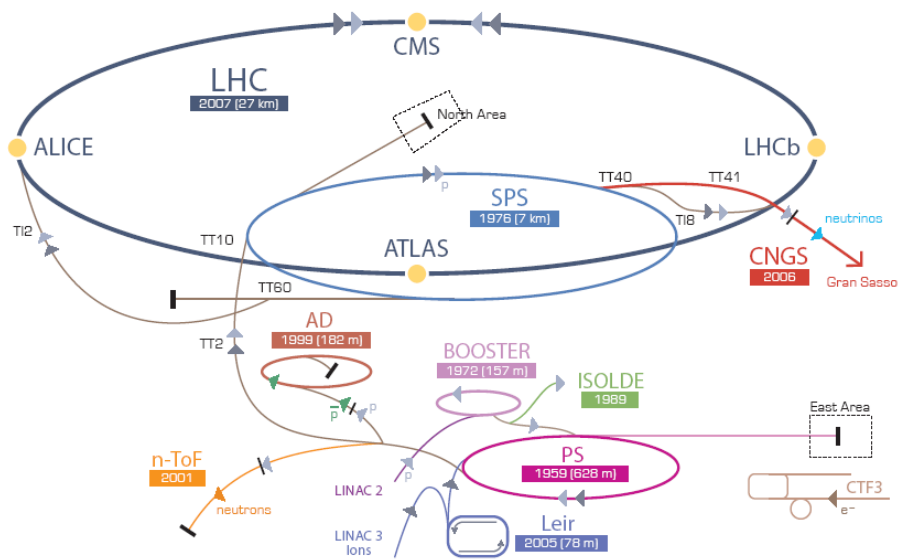


Figura 2.3: Cadeia de aceleração do LHC e seus detectores [73].

Na escala de energia, os fenômenos produzidos, quando a energia é menor do que 20MeV, são chamados de física de baixas energias. A faixa entre 20MeV e 400MeV corresponde a física de energias intermediárias e, finalmente, fenômenos com energia superior a 400MeV são estudados na física de altas energias [72]. Os experimentos no LHC atingirão níveis de energia de aproximadamente 14 TeV. Para alcançar a luminosidade desejada, o LHC tem de fazer uso dos antigos aceleradores PS - *Proton Synchrotron* - e SPS - *Super Proton Synchrotron* - como seus injetores, que foram reequipados para prover exatamente a densidade de feixe desejada.



Em modernos aceleradores, partículas podem chegar a 99,999999986% da velocidade da luz. Os feixes de partículas são então colididos contra alvos estacionários ou outros feixes de partículas de mesma velocidade que se movem em sentido contrário. A maneira como as partículas se dispersam frente ao alvo pode ser usada para revelar detalhes da estrutura interna das partículas que o atingem. Quando o movimento rápido das partículas cessa ou é diminuído drasticamente nas colisões, sua energia de movimento (energia cinética) é convertida em "chuveiros" de novas partículas, coerentemente com a famosa equação:  $E = mc^2$  (nesse caso  $m = E/c^2$ ). Os chuveiros de partículas produzidos dessa maneira podem ser usados para estudar como o mundo quântico funciona. É importante entender que, em muitos casos, não há sentido em que as partículas no interior do chuveiro estivessem, originalmente, no interior das partículas que colidiram. Elas foram, na verdade, criadas puramente pela energia gerada e não existiam antes da colisão ocorrer [74].

No LHC, quando operar em luminosidade nominal ( $10^{34} \text{cm}^{-2}\text{s}^{-1}$ ), a cada segundo, um pacote com cerca de 3 trilhões de prótons, viajando com velocidade próxima a da luz no vácuo (300 mil km/s), irá atravessar outro com características idênticas. Choques ocorrerão à taxa de 40 milhões por segundo. No momento da colisão de um próton com outro, a energia liberada será suficiente para se transformar em centenas de outras partículas, incluindo outros prótons. A reconstrução dos subprodutos oriundos de uma colisão ajuda a entender qual foi o mecanismo (ou força) que participou na transformação da energia em matéria. Reconstruí-los é o trabalho dos detectores, máquinas tão sofisticadas e gigantescas quanto o próprio acelerador. O LHC tem quatro detectores principais e todos eles contam com a participação de pesquisadores brasileiros [74].

- ALICE [3] - *A Large Ion Collider Experiment*: é um detector de íons pesados dedicado a explorar a potencial física de interação núcleo-núcleo sob a elevada energia alcançada pelo LHC. O objetivo é estudar a física da matéria na interação em condições de densidade e energia extrema, na qual a formação de uma nova fase da matéria, o plasma de quarks-glúons, é esperado. A existência dessa fase e suas propriedades são questões fundamentais na Cromodinâmica Quântica - *Quantum Chromodynamics* (QCD) - para a compreensão de confinamento e simetria. Para esse efeito, pretende-se realizar um estudo abrangente dos hádrons, elétrons, fótons e múons produzidos na colisão de núcleos pesados. ALICE também vai estudar as colisões próton-próton, assim como realizar uma comparação com colisões chumbo-chumbo em áreas da física em que o ALICE é competitivo com outros experimentos do LHC.
- LHCb [5] - *LHC-beauty*: Quatorze bilhões de anos atrás, o Universo começou com o que foi chamado Big Bang. Dentro de um espaço infinitamente pequeno, a energia confinada formava quantidades iguais de matéria e antimatéria. Mas como

o Universo esfriou e expandiu, sua composição foi alterada. Apenas um segundo após o Big Bang, a antimatéria tinha desaparecido, deixando a matéria para dar forma a tudo o que vemos ao nosso redor - das estrelas e galáxias, a Terra e toda vida que ele suporta. LHCb é um experimento criado para explorar o que aconteceu depois do Big Bang, que permitiu a matéria sobreviver e criar o Universo em que vivemos hoje. Ele é especializado em realizar medidas precisas da violação CP e decaimentos raros de hádrons B (compostos com o quark *beauty* ou b-quark).

- CMS [4] - *Compact Muon Solenoid*: É um dos dois detectores acoplados ao LHC com propósitos múltiplos. Entre as suas atribuições estão o estudo dos aspectos das colisões de íons pesados, buscar evidências que comprovariam uma física além do Modelo Padrão (Supersimetria e as dimensões espaciais extras), detecção do bóson de Higgs e explorar a física na faixa de TeV.
- ATLAS [75] - *A Toroidal LHC Apparatus*: Além do CMS, é o outro detector com propósitos múltiplos, incluindo um programa para a detecção do bóson de Higgs. O detector é complexo e pode ser reconfigurado para estudo de nova física na remota possibilidade de não serem encontradas evidências do bóson de Higgs. A montagem do ATLAS está concluída. Como um dos detectores de propósito geral do LHC, ele possibilitará estudos sobre forças básicas que formaram o nosso universo, desde o início dos tempos. Dentre os estudos de interesse, destacam-se: a origem da massa, espaço de dimensão extra, buracos negros microscópicos, evidências para candidatos a matéria escura no universo e estudo sobre a não similaridade entre matéria e antimatéria [76].

### 2.4.1 Detector ATLAS

O ATLAS é um dos detectores de propósito geral acoplados ao acelerador/colisionador de partículas LHC. Em funcionamento desde 2009, o ATLAS busca experimentalmente respostas para vários questionamentos que a teoria de física de partículas apresenta. Além disso, permitirá descobertas possíveis apenas com os níveis de energia fornecidos pelo LHC. Suas dimensões são surpreendentes: tem aproximadamente 45 metros de comprimento, mais de 25 metros de altura e pesa aproximadamente 7 mil toneladas. Para isso, conta com um sistema de filtragem de eventos que separa a física ordinária dos eventos que possam representar o objeto de interesse. Esse sistema tem papel fundamental na eficiência final do detector.

O detector ATLAS está dividido em um característico sistema dispostos em camadas, conforme mostrado na Figura 2.4. A função do Detector de traço (*inner detector*) é identificar as partículas com carga elétrica e medir seu momento a partir da curvatura de sua trajetória. Na sequência estão os calorímetros Eletromagnético e Hadrônico que



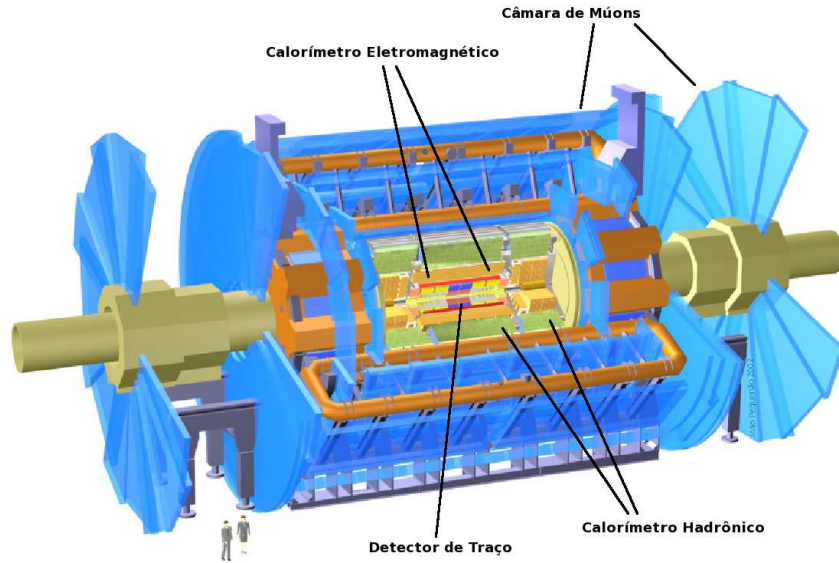


Figura 2.4: Esquema básico do detector ATLAS (adaptado de [77]).

medem a energia depositada pelas partículas. Na interação com as células do calorímetro são produzidos perfis de deposição. Em um último estágio está o sistema de detecção de múons (*Muon detectors*) [78]. Outra característica do detector é o seu sistema de coordenadas. Dado o formato cilíndrico dos elementos detectores dispostos ao redor do ponto de impacto, o sistema de coordenadas  $xyz$  acompanha a direção dos feixes de partículas provenientes da colisão (Figura 2.5).

O eixo  $z$  segue a direção do feixe de partículas, enquanto os eixos  $x$  e  $y$  formam um plano transversal ao feixe. A direção positiva de  $x$  aponta para o centro do anel LHC, enquanto  $y$  positivo aponta para cima. O ângulo obtido:

$$\phi = \arctan\left(\frac{x}{y}\right) \quad (2.2)$$

representa a rotação em torno do eixo de colisão. O ângulo polar  $\theta$  em relação ao eixo de espalhamento em  $z$  é usado no cálculo da chamada pseudo-rapidez  $\eta$ , que representa a direção de projeção das partículas após a colisão, definida por

$$\eta = -\log\left(\tan\frac{\theta}{2}\right) \quad (2.3)$$

Grandes valores de pseudo-rapidez ( $\eta \rightarrow \infty$ ) indicam que a colisão ocorrida entre as partículas não foi frontal. Nesse tipo de colisão o resultado é pouco interessante para o estudo do detector, que foi projetado para um melhor desempenho com valores baixos,  $\eta < 3$ .

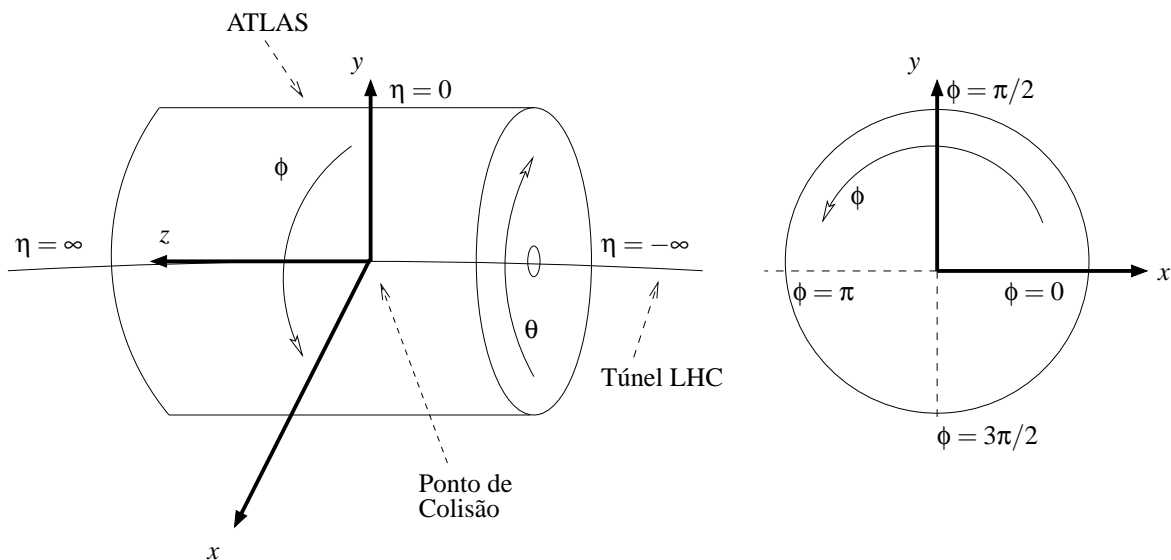


Figura 2.5: Sistema de coordenadas do ATLAS.

### Detector de Traço

O detector de traços se localiza diretamente em torno da região de colisão das partículas. É formado por detectores discretos de alta resolução, fabricados em silício. Eles têm capacidade para reconhecimento de padrões, medidas de momento e de vertex, definindo trajetórias de partículas carregadas com grande precisão [75, 79]. O número total de camadas de precisão deve ser limitado por causa do excesso de matéria que estes detectores introduzem, o que provocaria absorção de parte da energia. Segundo o projeto do detector, a trajetória das partículas cruzará entre e três e quatro camadas de sensores.

### O Sistema de Calorimetria

Calorímetros granularmente estruturados representam um papel central em todos os experimentos modernos de física de partículas. Eles são assim construídos para otimizar o processo de detecção, seja abordando aspectos de resolução espacial ou de energia depositada pela partícula. Resolução superior em energia é importante, principalmente, para medir com precisão a energia de partículas alvo na presença de um intenso ruído de fundo. O enorme campo de energia coberto pelos sensores, exige uma representação em palavras de 16 *bits*, em que o limite inferior corresponde ao típico ruído eletrônico e o limite superior, ao máximo valor de energia depositada por partículas provenientes de diferentes decaimentos.

A alta granularidade é outro diferencial desses equipamentos. A maioria dos alvos físicos, assim como a minimização do ruído eletrônico e de *pile-up*, apresentam forte resposta a essa característica. Além disso, os calorímetros são longitudinalmente seccionados em 2 ou 3 seções para melhorar a identificação da partícula.

Para minimizar o efeito de *pile-up* no detector, a resposta dos sensores usados precisa

ser extremamente rápida, dada a frequência de colisões esperada. Para uma independência satisfatória entre amostras sucessivas é necessário a garantia de um percentual máximo de energia residual dentro do tempo de medição do próximo evento.

Outras características requeridas nos calorímetros atuais são a hermeticidade e a resistência à radiação. A hermeticidade garante baixos níveis de energia perdida - *missing energy*  $E_m$  - resultante de vários decaimentos. A resistência a radiação é obrigatória para os equipamentos que serão submetidos a doses extremamente altas ao longo do tempo de operação. Acredita-se que as doses de radiação acumuladas durante um ano de operação podem chegar a 500Mrad [80].

O sistema de calorimetria do ATLAS vem operando continuamente desde agosto de 2006. Ele é subdividido em eletromagnético e hadrônico, que por sua vez são seccionados em 7 camadas: 4 eletromagnéticas e 3 hadrônicas (Figura 2.6). Atualmente todas as células dos calorímetros estão conectadas ao sistema de leitura de dados do ATLAS e desde 2008 testes de desempenho dos calorímetros são realizados. Ao todo 182.468 sensores (células), ou 97,2% dos total de sensores do ATLAS estão nessa região [81].

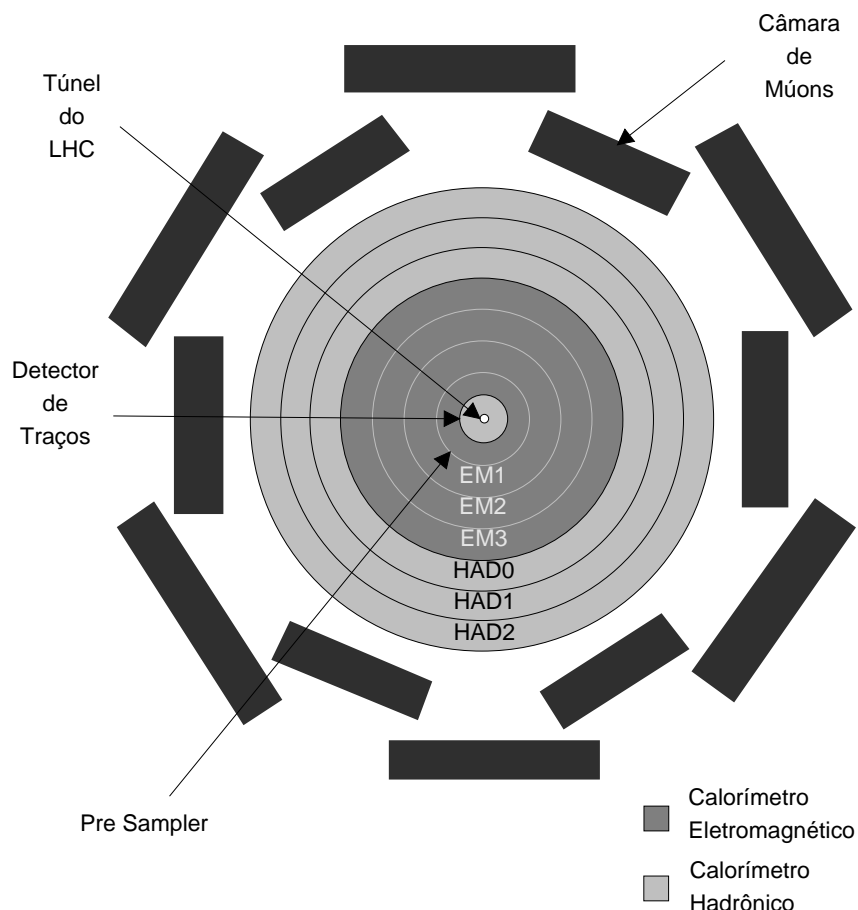


Figura 2.6: Disposição das camadas dos calorímetros Eletromagnético (EM) e Hadrônico (HAD) do detector ATLAS.

O perfil de energia das partículas, gerado na interação delas com as camadas dos

Tabela 2.2: Região de cobertura em  $\eta$  e granularidade das camadas dos calorímetros [75].

<i>Pre Sampler</i>	Barril	Tampa
Cobertura	$ \eta  < 1,52$	$1,5 <  \eta  < 1,8$
Granularidade ( $\Delta\eta \times \Delta\phi$ )	$0,025 \times 0,1$	$0,025 \times 0,1$
Eletromagnético	Barril	Tampa
Cobertura	$ \eta  < 1,52$	$1,375 <  \eta  < 3,2$
Granularidade 1 ( $\Delta\eta \times \Delta\phi$ )	$0,003 \times 0,1$	$0,003 \times 0,1$ a $0,1 \times 0,1$
Granularidade 2 ( $\Delta\eta \times \Delta\phi$ )	$0,025 \times 0,025$	$0,025 \times 0,025$ a $0,1 \times 0,1$
Granularidade 3 ( $\Delta\eta \times \Delta\phi$ )	$0,05 \times 0,025$	$0,05 \times 0,025$
Hadrônico	Barril	Barril estendido
Cobertura	$ \eta  < 1$	$0,8 <  \eta  < 1,7$
Granularidade 1 e 2 ( $\Delta\eta \times \Delta\phi$ )	$0,1 \times 0,1$	$0,1 \times 0,1$
Granularidade 3 ( $\Delta\eta \times \Delta\phi$ )	$0,2 \times 0,1$	$0,2 \times 0,1$

calorímetros, é fundamental na discriminação do evento de interesse. Após um evento ser aceito pelo sistema de filtragem, essas informações são armazenadas em mídia permanente para posterior análise *off-line*. A granularidade ou a quantidade de células por unidade de área varia entre as camadas dos calorímetros. Cada camada do calorímetro eletromagnético é dividida em barril e tampa e do hadrônico, em barril e barril estendido. Essas partes fecham quase que hermeticamente o espaço ao redor da colisão. O barril é responsável pela região de baixa pseudo-rapidez,  $|\eta| < 1,5$  para as camadas eletromagnéticas e  $|\eta| < 1,0$  para as hadrônicas. A Tabela 2.2 traz informações sobre a região de cobertura em  $\eta$  e a granularidade das camadas dos calorímetros, incluindo informações sobre o pré-irradiador e as tampas [75].

O calorímetro eletromagnético é construído de argônio líquido como material absorvedor e eletrodos de chumbo que interagem com partículas no desenvolvimento das cascatas eletromagnéticas, resultantes da colisão, o qual está situado logo após o detector de traços. O calorímetro eletromagnético também inclui um pré-irradiador (*Pre Sampler*), que funciona, praticamente, como um calorímetro muito fino, posicionado antes dos calorímetros de argônio líquido, com a função de recuperar a informação perdida no material morto da seção eletromagnética. O calorímetro eletromagnético mede a energia total de pósitrons, elétrons e fótons. Essas partículas produzem chuveiros de pares pósitron/elétron no material. Os elétrons ou pósitrons, são defletidos pelos campos elétricos dos átomos, provocando a irradiação de fótons. Os fótons então produzem pares pósitron/elétron, que por sua vez irradiam mais fótons e assim por diante. Essa sequência, em cascata, é chamada de chuveiro eletromagnético. O número final de pares é proporcional à energia da partícula original [74, 75]. A Figura 2.7 ilustra a granularidade e a profundidade do calorímetro eletromagnético.

A seção eletromagnética do calorímetro é dividida em 3 camadas, além do pré-irradiador, das quais a segunda é a mais profunda. Cada camada possui uma segmentação

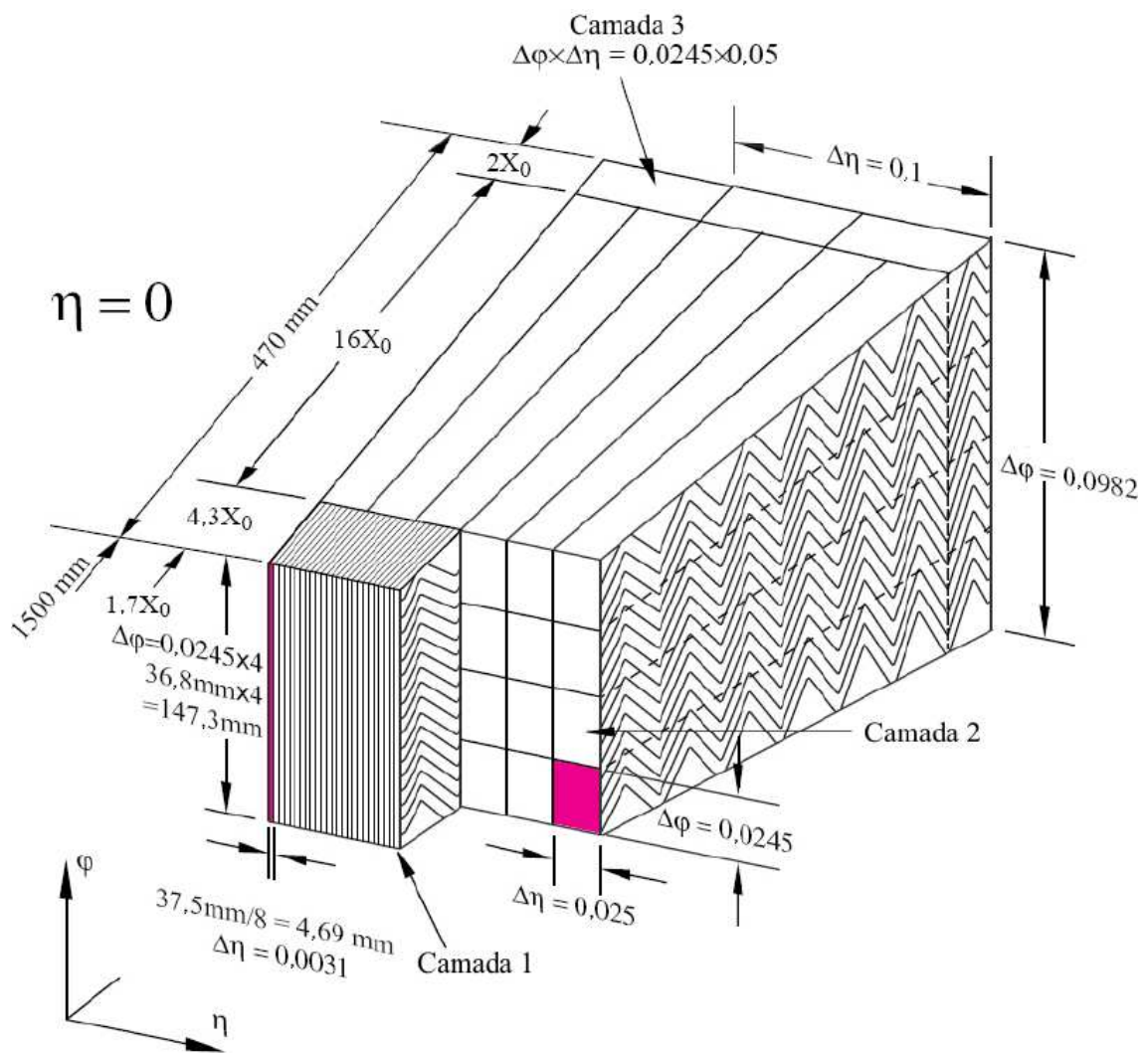


Figura 2.7: Granularidade e profundidade das camadas do calorímetro eletromagnético do detector ATLAS (adaptado de [75]).

específica que otimiza a relação custo-benefício do detector. As camadas mais internas são mais segmentadas, permitindo a localização precisa das partículas no plano  $\eta \times \phi$ , enquanto as mais externas são concebidas de forma menos segmentada e mais profunda, com o objetivo de absorver toda a energia da partícula incidente.

O calorímetro hadrônico envolve o calorímetro eletromagnético, sendo composto de ligas de aço como material absorvedor, separadas por telhas de material plástico cintilante. Quando as partículas atravessam as telhas elas emitem luz de intensidade proporcional a energia incidente. O calorímetro hadrônico é projetado para a detecção de partículas ou jatos que desenvolvem cascatas baseadas em hádrons, como nêutrons, prótons e píons [74, 75].

A Figura 2.8 mostra um corte transversal no detector ATLAS e o comportamento de diferentes partículas (léptons, fótons e hádrons) frente aos diferentes subsistemas.

Na Tabela 2.2, é possível perceber que o tamanho das células para o calorímetro hadrônico (em média) é maior do que o valor apresentado pelas células no calorímetro eletromagnético. A segmentação é também mais uniforme do que a da seção eletromagnética. Isso se deve ao fato dos chuveiros hadrônicos serem mais largos e profundos, provocando maiores flutuações nas medidas de energia e, portanto, não necessitando de uma segmentação tão fina.

## Detecção de Múons

O detector de múons envolve todo o sistema de calorimetria. Múons com elevado valor de momento transversal são assinaturas bastante promissoras e robustas da física de interesse no LHC. Para explorar esse potencial foi projetado um sistema de filtragem e medição sensível a esse tipo de partícula. A detecção é baseada na deflexão (magnética) de múons provida por um grande toróide com núcleo a ar e detectores de traço bastante precisos. Para  $|\eta| \leq 1$ , um grande ímã em formato de barril foi construído com oito solenóides circundando a seção hadrônica dos calorímetros. Para  $1,4 \leq |\eta| \leq 2,7$  a deflexão é garantida por ímãs na forma de tampas ao redor do barril. Essa configuração provê um campo magnético praticamente transversal à trajetória de eventuais múons. Na região do barril, as trajetórias são detectadas por câmaras organizadas em três camadas cilíndricas (estações) ao redor do eixo de colisão. Na região da tampa, as câmaras foram montadas verticalmente, também em número de três. As coordenadas dos traços são medidas precisamente por Câmaras de Traços na direção principal de deflexão do campo magnético, na maior parte do intervalo da pseudo-rapidez. Para grandes valores de pseudo-rapidez, câmaras com alta segmentação foram construídas para garantir a precisão da detecção nesta área bastante ruidosa. Tal como os calorímetros, esse detector está diretamente acoplado ao primeiro nível do sistema de filtragem de eventos do ATLAS [83, 84].

A Figura 2.8 mostra um corte transversal no detector ATLAS e o comportamento de um múon ao cruzar os diferentes subsistemas.



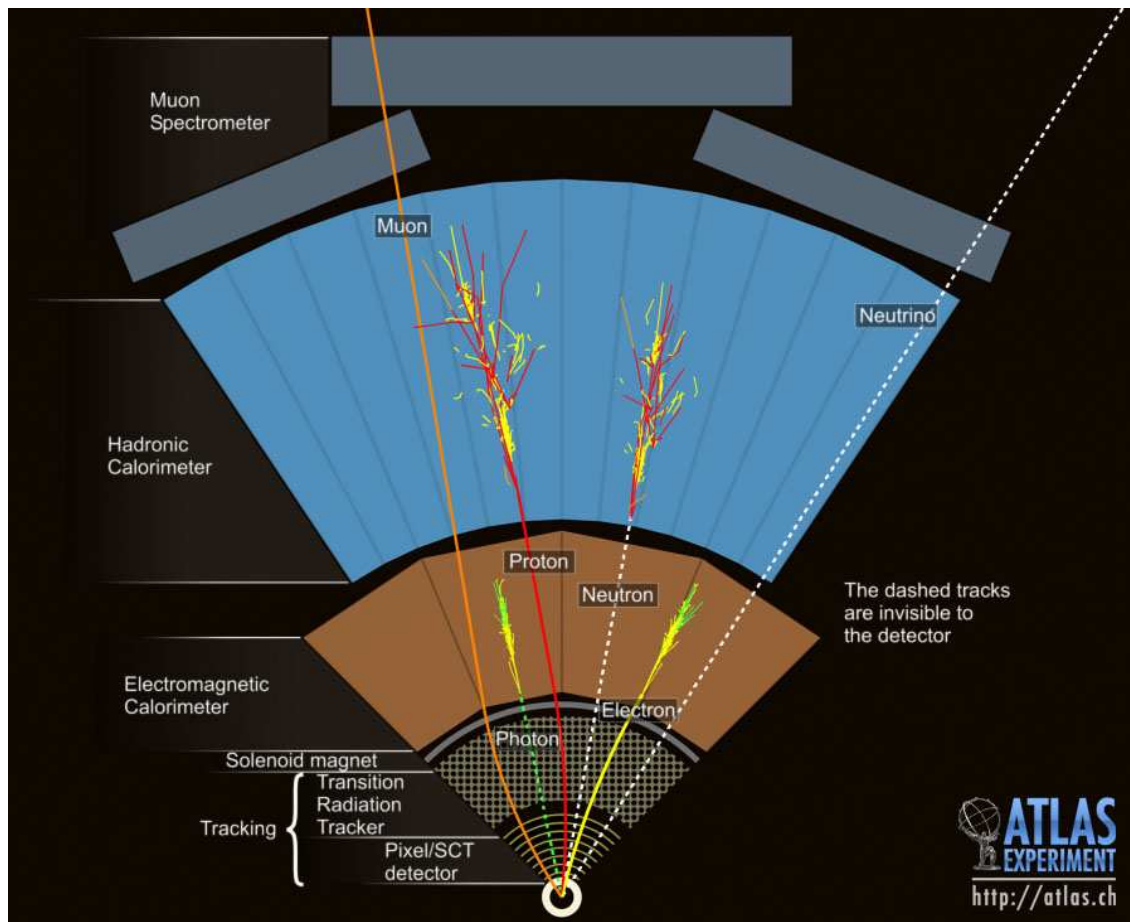


Figura 2.8: Imagem gerada em computador de um corte transversal no detector ATLAS para diferentes tipos de eventos. Comportamento de um múon, um fóton, um elétron, um próton, um nêutron e um neutrino em um sistema HEP. O duto de feixe é perpendicular ao plano da imagem. O múon passa completamente através dos calorímetros depositando apenas uma pequena quantidade de energia em cada seção até finalmente ser detectado pela câmara de múons. O fóton, mesmo sem massa, interage fortemente com o calorímetro eletromagnético e o elétron deposita toda a sua energia nesse mesmo calorímetro. Prótons depositam pouca energia no calorímetro eletromagnético e juntamente com nêutrons interagem fortemente com o calorímetro hadrônico. Jatos compostos de muitas partículas - majoritariamente píons - depositam energia tanto no calorímetro eletromagnético quanto no hadrônico e em uma região maior do que uma única partícula. O neutrino não interage com qualquer equipamento, ele passa indetectavelmente [82].

## Capítulo 3

# Filtragem *online* do detector ATLAS

No detector ATLAS a filtragem *online* empregada para experimentos físicos de altas energias - *High Energy Physics* (HEP) - está inserida no sistema de filtragem e aquisição de dados - *Triggering and Data Acquisition* (TDAQ) - com todas as características descritas no Capítulo 2.

### 3.1 Aquisição e Filtragem *online* de Dados

O sistema de filtragem de dados do ATLAS (Figura 3.1) está baseado em três níveis de seleção, iniciando com uma taxa de eventos de 40Mhz, uma colisão a cada 25ns (taxa de iteração de 1GHz a uma luminosidade de  $10^{34}\text{cm}^{-2}\text{s}^{-1}$ ) e finalizando com aproximadamente 100Hz de eventos selecionados [78]. Como os eventos de interesse, por exemplo, os bósons de Higgs, se existirem, conforme o Modelo Padrão ocorrem com frequência inferior a 0,001Hz e cada evento carrega aproximadamente 1 MByte de dados, o fluxo de dados será da ordem de 40 TBytes por segundo, impossibilitando o armazenamento completo desses eventos para análise *offline*. Assim, um sistema de filtragem *online* torna-se indispensável para o experimento.

O sistema de filtragem *online* do ATLAS é baseado em 3 níveis de seleção de eventos, o nível 1 (LVL1), o nível 2 (LVL2) e o filtro de eventos - *Event Filter* (EF) -, sendo que os dois últimos compõem a filtragem de alto nível - *High Level Trigger* (HLT).

O primeiro nível de filtragem do ATLAS realiza a seleção inicial dos eventos baseado na informação obtida com granularidade menos fina, gerada por um sub-conjunto de detectores (calorimetria e detecção rápida de múons). O tempo para a tomada de decisão é tão curto que torna-se necessário a redução da quantidade de informação a ser processada. O LVL1 então descarta eventos com características bem distintas dos canais de interesse desejados. A taxa máxima de saída do LVL1 está limitada em 75kHz (podendo ser expandida para 100kHz) [78].

É importante que se mantenha o tempo de latência (tempo para formar e informar



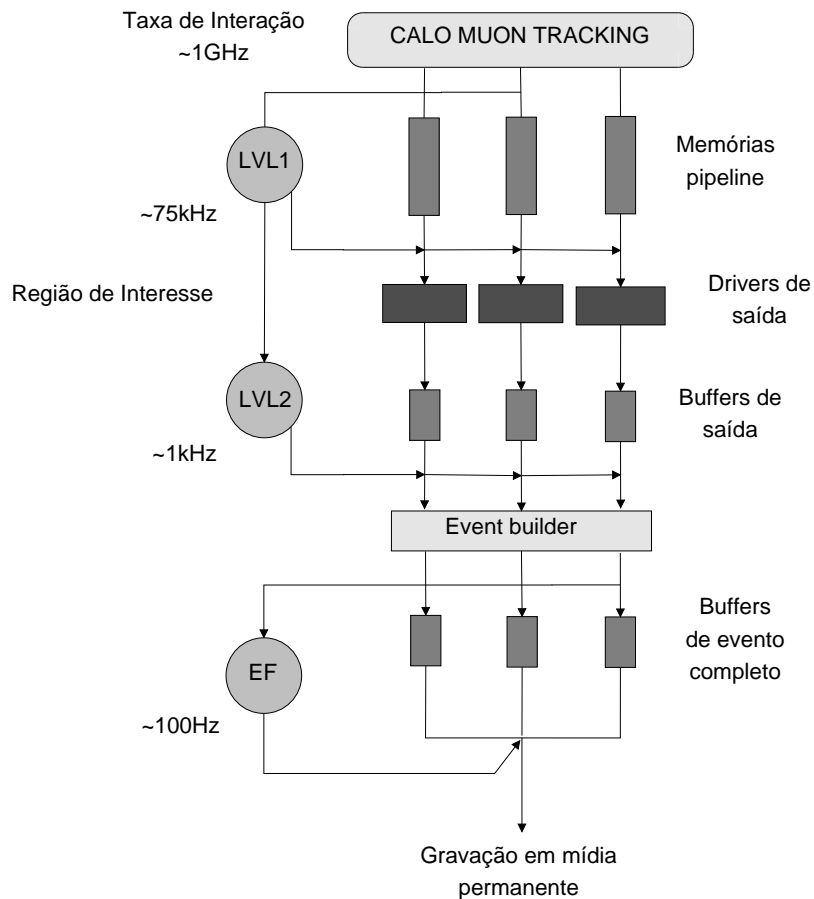


Figura 3.1: Sistema de filtragem do ATLAS (baseado em [75]).

a decisão do filtro) no valor mais baixo possível. Durante esse tempo, a informação de todos os canais do detector precisa ser retida em memórias do tipo *pipeline*, que estão posicionadas próximas ou junto ao detector. A latência do primeiro nível, medida do instante de uma colisão próton-próton até a decisão estar disponível para o nível seguinte, deve ser menor do que  $2,5\mu\text{s}$ . De forma a reduzir o volume de tráfego de dados para o segundo nível e, conseqüentemente, aumentar a banda passante entre esses dois níveis, o primeiro nível já marca as regiões de interesse (*Region of Interest - RoI*), que correspondem às regiões do detector em que houve efetivamente a incidência de algum evento, que foi aceito pelo primeiro nível. O nível seguinte de filtragem observará somente essas regiões de interesse e não toda a área do detector [78].

O LVL2 refina a decisão do primeiro nível, validando a etiquetagem (resultado da classificação do LVL1) de cada RoI, através da utilização da informação completa (resolução total dos detectores) de cada RoI. A etapa de extração de característica é realizada para cada sistema do detector. A combinação dessas características forma um objeto de segundo nível de filtragem que são candidatos ao objeto de interesse do canal. Finalmente uma decisão global é tomada baseada no *menu* de seleção. O LVL2 é implementado através de um software especializado, rodando em um conjunto de 500 processadores do

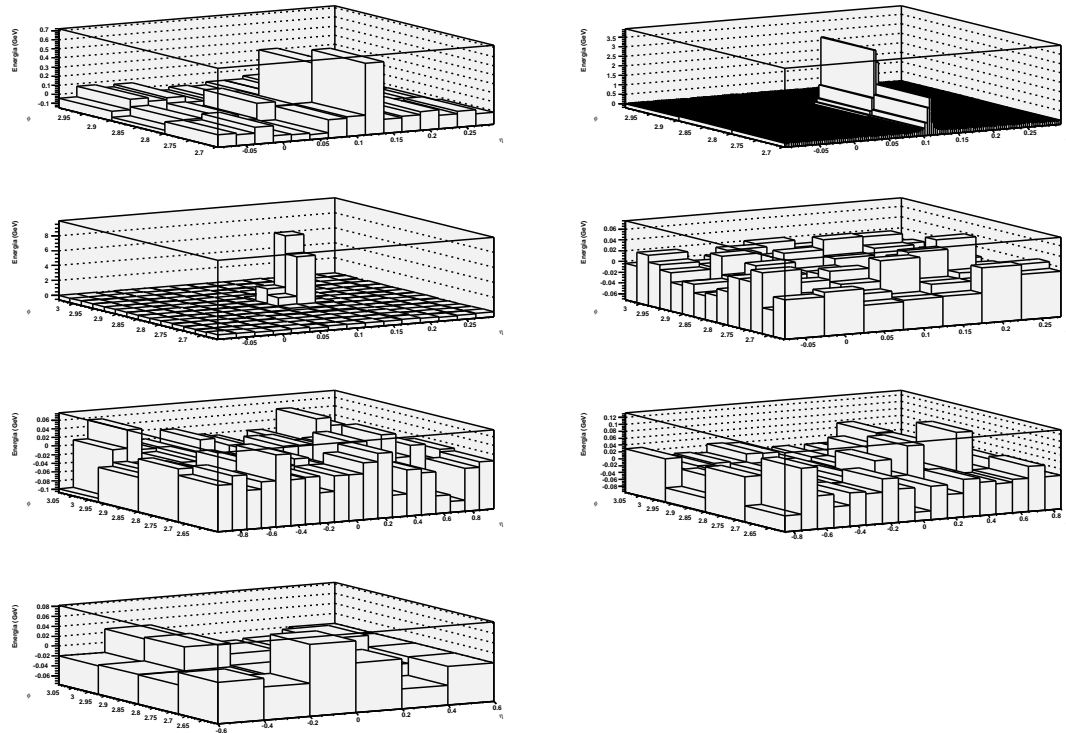


Figura 3.2: Exemplo de RoI de um elétron por camada: *Pre Sampler*, canto superior esquerdo; EM *front layer*, canto superior direito; EM *middle layer*, segunda linha no lado esquerdo; EM *back layer*, segunda linha no lado direito; HAD *layer 0*, terceira linha no lado esquerdo; HAD *layer 1*, terceira linha no lado direito; HAD *layer 2*, canto inferior esquerdo.

tipo PC (núcleo duplo), dedicados, operando em paralelo com praticamente todas as comunicações entre estes dispositivos feitas através de *switches Gigabit Ethernet*, devido à velocidade requerida [78].

Para o LVL2, foi definido um tempo de latência de, aproximadamente, 10ms, no qual a taxa dos eventos trafegando deste para o EF é da ordem de 1kHz [78]. Espera-se que a seleção de eventos do HLT [85] otimize os recursos computacionais disponíveis, a dimensão elevada dos dados de calorimetria motiva a realização de uma etapa de extração de características sobre as regiões de interesse. As Figuras 3.2, 3.3 e 3.4 mostram um exemplo de RoI para elétron, jato e *leptonjets*, respectivamente, aceitos pelo primeiro nível de filtragem.

O EF finalmente diminui a taxa de eventos de 1kHz para 100Hz. Os eventos que forem aceitos pelos 3 níveis de filtragem serão armazenados em mídia permanente para a análise *offline*. O tempo para a tomada de decisão no filtro de eventos está entre 1 e 2 segundos. Em muitos casos, não é desejável uma separação total entre o nível 2 de filtragem e o filtro de eventos, podendo haver uma estratégia conjunta para os dois níveis. Considerando os efeitos conjuntos do segundo nível e do filtro de eventos, a filtragem de alto nível deve reduzir em 1000 vezes a taxa de eventos [75]. As etapas de processamento do HLT são

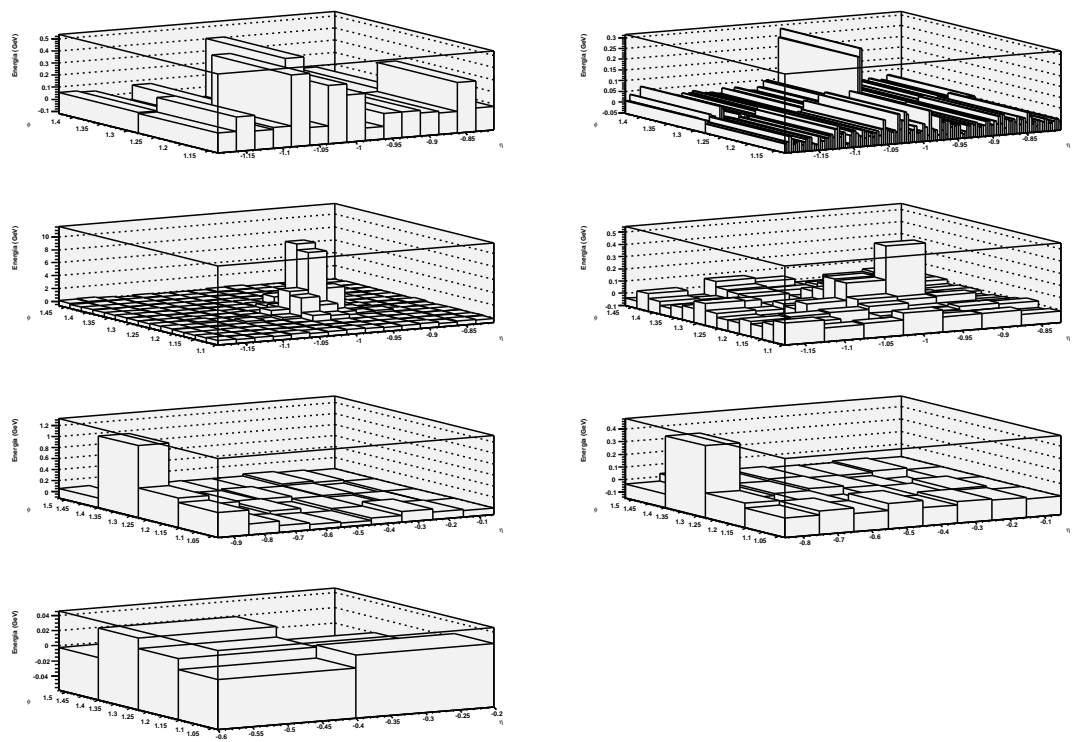


Figura 3.3: Exemplo de RoI de um jato por camada: *Pre Sampler*, canto superior esquerdo; *EM front layer*, canto superior direito; *EM middle layer*, segunda linha no lado esquerdo; *EM back layer*, segunda linha no lado direito; *HAD layer 0*, terceira linha no lado esquerdo; *HAD layer 1*, terceira linha no lado direito; *HAD layer 2*, canto inferior esquerdo.

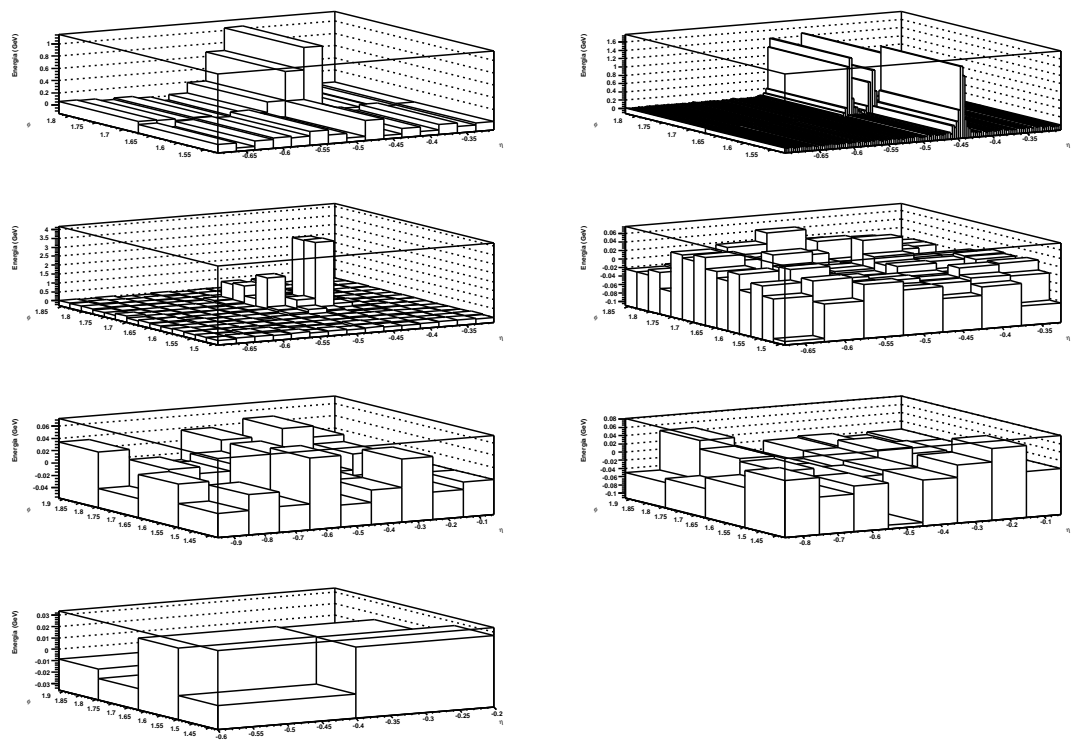


Figura 3.4: Exemplo de RoI de *leptonjets* por camada: *Pre Sampler*, canto superior esquerdo; *EM front layer*, canto superior direito; *EM middle layer*, segunda linha no lado esquerdo; *EM back layer*, segunda linha no lado direito; *HAD layer 0*, terceira linha no lado esquerdo; *HAD layer 1*, terceira linha no lado direito; *HAD layer 2*, canto inferior esquerdo.

iniciadas pelo acesso e preparação das informações das RoI do LVL1. A extração de características é efetuada em cada sistema do detector, iniciando pela confirmação da RoI no sistema em que foi originada (câmara de múons ou calorímetro). Os principais objetos de filtragem identificados no HLT são candidatos a múons, elétrons, fótons, taus, jatos e física dos hádrons B. O desempenho geral é medido a partir da eficiência na discriminação da física de interesse e do ruído de fundo.

A eficiência pode ser afetada por diferentes condições de operação. Um problema que pode afetar os algoritmos de extração de características é o efeito de empilhamento (*pile-up*), que ocorre quando há uma sobreposição de eventos em regiões do detector [75], i.e., um evento que ainda se desenvolve tem seu padrão de deposição de energia distorcido por um novo evento que chega e se sobrepõe. Outro problema a ser enfrentado vem dos eventos gerados a partir da alta escala de energia alcançada e da elevada taxa de colisões. Espera-se que algumas células de detecção sejam danificadas com o passar do tempo, seja deixando de registrar corretamente a informação de energia depositada, seja parando de funcionar por completo, o que se torna um problema real durante o tempo de operação do detector. É desejável que os algoritmos de filtragem de eventos tenham alguma imunidade à perda de informação dessas células.

## 3.2 Plataforma Computacional do Ambiente de Filtragem

A colaboração ATLAS desenvolveu um conjunto de ambientes computacionais e ferramentas de controle, conhecidos como Athena [86], que permitem aos membros da colaboração, independentemente de sua localização geográfica, acesso e análise dos dados gerados no detector. O Athena pode ser entendido como um estrutura de controle comum para os aplicativos necessários à colaboração. Para isso, ele oferece: simulador de colisões de feixes de prótons (Pythia); Interação entre partículas e partícula/detector (Geant); Filtragem (*trigger*); Reconstrução de evento e Análise da Física. Esse ambiente permite aos desenvolvedores o compartilhamento de uma estrutura homogênea onde seus códigos podem ser elaborados baseados em ferramental tecnológico e computacional atualizados, além de facilidade no reuso e atualização dos códigos.

## 3.3 Extração de Características e Teste de Hipótese

Os calorímetros são detectores altamente segmentados, de maneira que é possível estimar as classes das partículas a partir de seus perfis de deposição de energia. Esse perfil está contido em uma região de interesse selecionada pelo primeiro nível de filtragem. Essa região pode abranger mais 1000 células, considerando as diferentes camadas desses

elementos.

Os exemplos de perfis de deposição de energia de cada um dos eventos envolvidos no trabalho, mostrados nas Figuras 3.2,3.3 e 3.4, indicam algumas peculiaridades. Os elétrons são mais pontuais na deposição de energia como mostra a camada EM *middle layer* e quase não apresentam energia nas camadas hadrônicas. Os jatos hadrônicos, através de seus componentes eletromagnéticos falseiam os elétrons nas camadas eletromagnéticas do calorímetro, apresentando valor de energia maior nas camadas hadrônicas do calorímetro e, tipicamente, tendendo a espalhar a energia sobre múltiplas células em uma mesma camada do calorímetro. Entende-se, assim, que a informação relevante pode não necessariamente estar no ponto central de impacto, mas em uma vizinhança dele. Os *leptonjets* excitam uma região maior nas camadas eletromagnéticas do que o elétrons isolados, mas os acompanham no calorímetro hadrônico ao não apresentarem níveis de deposição de energia elevados para esses sensores.

A análise longitudinal demonstra também que a camada *Pre Sampler* amostra não somente mais energia para elétrons e *leptonjets*, como esses valores são mais concentrados, enquanto os jatos apresentam um maior espalhamento. Analogamente é o que se observa na camada EM *front layer*, a de maior resolução espacial.

A observação da interação dos objetos de interesse com os calorímetros permite assumir que esses interagem de forma (aproximadamente) isotrópica em relação ao eixo de penetração, i.e., dado um ponto de impacto inicial, a partícula tende a decair em objetos menos energéticos ao redor do eixo de penetração [49, 87]. O que os algoritmos de extração de característica fazem é capturar essas informações do ponto de vista de discriminantes.

O algoritmo de nível 1 trabalha com matrizes em  $\eta \times \phi$ , que armazenam a energia transversa ( $E_T$ ) por torre - são chamadas assim por transporem as diferentes camadas dos calorímetros. O procedimento ocorre separadamente para os calorímetros eletromagnético e hadrônico, com uma granularidade típica de  $\Delta\eta \times \phi = 0,1 \times 0,1$ . Os valores digitalizados de energia são truncados para oito bits, o que, efetivamente, aplica um limiar de 1GeV por torre. O algoritmo é baseado em uma janela de  $4 \times 4$  torres em ambos calorímetros na região  $|\eta| < 2,5$  e consiste de 4 elementos [88]:

- um agrupamento eletromagnético de  $2 \times 2$  torres usado para identificar a posição da RoI candidata.
- um agrupamento de  $2 \times 1$  ou  $1 \times 2$  torres eletromagnéticas usadas para medir a  $E_T$  dos chuveiros eletromagnéticos - há quatro desses agrupamentos dentro de uma RoI e o mais energético é usado.
- uma região de 12 torres eletromagnéticas ao redor dos agrupamentos, que é usada para teste de isolamento no calorímetro eletromagnético.

- as 16 torres hadrônicas atrás do agrupamento eletromagnético e da região de isolamento, que são usados para testes de isolamento no calorímetro hadrônico. A janela desliza em passos de uma torre em ambas direções  $\eta$  e  $\phi$ .

Para cada objeto aceito no nível um, a informação da RoI é transmitida para o nível dois com posição  $(\eta, \phi)$  e  $E_T$ . Devido a escolhas do algoritmo (tamanho do passo de 0, 1), a posição da RoI é fornecida com resolução  $\Delta\eta \times \Delta\phi = 0, 1 \times 0, 1$ . Os eventos selecionados são dominados por fótons de  $\pi^0$  e da troca de carga  $\pi^+/\pi^-$  [88].

Um dos algoritmos de referência de filtragem de segundo nível é o T2Ca1o [88–91]. Esse algoritmo realiza o teste de hipótese sobre a informação de calorimetria para eventos eletromagnéticos isolados e jatos hadrônicos. O T2Ca1o é um algoritmo de clusterização para chuveiros eletromagnéticos alimentado pelas RoI produzidas pelo primeiro nível de filtragem eletromagnética. Esse algoritmo pode selecionar objetos usando medidas de energia transversa e formato do chuveiro de deposição de energia. As RoI usadas nesse processo são compostas por células dos calorímetros eletromagnético e hadrônico e sua saída é uma classe específica contendo a energia e posição do *cluster*, e variáveis que descrevem o formato do chuveiro eletromagnético, todas usadas na seleção de eventos. O primeiro passo do algoritmo T2Ca1o é refinar a posição da RoI de nível 1 encontrando a célula de deposição de energia mais energética na segunda camada do calorímetro eletromagnético. Essa posição  $(\eta, \phi)$  é mais tarde recalculada ponderando a posição  $(\eta, \phi)$  em uma janela de  $3 \times 7$  células pela energia de cada uma delas. A partir daqui algumas variáveis são calculadas com o objetivo extrair informação de formato do chuveiro [88]:

- EM  $E_T$ : Energia total depositada nas camadas eletromagnéticas em uma região de  $3 \times 7$  em torno do centro  $\eta \times \phi$ . Devido à dependência de energia do ruído de fundo, um corte em energia fornece a melhor rejeição contra dados com alto momento transversa ( $P_T$ ) aceitos pelo primeiro nível.
- HAD  $E_T$ : Energia total depositada na camada hadrônica em uma janela  $\Delta\eta \times \Delta\phi = 0, 2 \times 0, 2$  ao redor do centro da RoI. Fótons e elétrons, tipicamente, depositam muito pouco de suas energias no calorímetro hadrônico. Contudo, para jatos com componentes eletromagnéticos aceitos pelo primeiro nível, aproximadamente 5% da energia é depositada no calorímetro hadrônico. Essa energia hadrônica pode ser usada para rotular jatos. A energia eletromagnética que vaza para o calorímetro hadrônico de genuínos *clusters*  $e/\gamma$ , aumenta com a energia incidente. Para baixa energia, ela não é significativa quando comparada com a contribuição do ruído e do *pile-up*. Para ser eficiente com *cluster*  $e/\gamma$  com alta  $E_T$ , os requerimentos do isolamento hadrônico precisam levar essa energia em consideração.
- Rshape: a razão energética  $E_{3 \times 7}/E_{7 \times 7}$  ( $E_{i \times j}$  é a energia depositada na região  $i \times j$ ) em torno do centro  $(\eta, \phi)$  da RoI. A maior parte da energia dos chuveiros eletro-



magnéticos (tipicamente, mais de 70%) é depositada na segunda camada do calorímetro eletromagnético. Essa camada é também menos afetada por ruído e pelo efeito *pile-up*. Dessa forma, o cálculo é realizado sobre ela. Chuveiros de fótons e elétrons têm (em média) um tamanho lateral de um raio de Molière, enquanto o tamanho lateral dos chuveiros iniciados por jatos é dominado pela fragmentação e o tamanho da interação hadrônica. Assim, a razão  $R_{\text{shape}}$  é calculada para distinguir entre *clusters*  $e/\gamma$  e jatos. Tipicamente esse valor é maior do que 0,9 para *clusters* eletromagnéticos.

- $R_{\text{strip}}$ : a razão energética  $(E_1 - E_2)/(E_1 + E_2)$  ( $E_i$  é o  $i$ -ésimo pico de energia) entre os dois maiores picos de energia em uma dada região  $\Delta\eta \times \Delta\phi = 0,125 \times 0,2$  em torno do centro da RoI. Após os cortes anteriores, somente jatos com pouca atividade hadrônica e de chuveiros estreitos permanecem. Esses jatos “eletromagnéticos” frequentemente consistem de um ou múltiplos  $\pi^0$  decaindo para dois fótons. Para rejeitar esses jatos, a granularidade fina da primeira camada eletromagnética pode ser observada para a análise do formato do chuveiro. No caso dos jatos pode-se encontrar dois máximos (ex. dois fótons do decaimento de um  $\pi^0$ ). De maneira a rejeitar esses eventos, uma variável que busca por máximo secundário é aplicada.

O teste de hipótese é feito através de cortes lineares sobre os descritores extraídos. Além do  $T2\text{Calo}$ , outro algoritmo de referência é o  $\text{RingerNeuralHypo}$ , também pertencente ao pacote de algoritmos de filtragem de nível dois do detector ATLAS [92]. No caso do  $\text{RingerNeuralHypo}$  outros discriminantes são usados para alimentar o classificador. Com o objetivo de ressaltar a informação topológica do decaimento, um pré-processamento baseado em somas de anéis [49, 87] é aplicado como algoritmo de extração de características. Nessa abordagem, a célula de máxima deposição de energia é considerada o centro da região de interesse em cada camada do calorímetro (sete no total). Então, um conjunto de anéis concêntricos são formados ao redor dele. É relevante ressaltar que, dependendo da granularidade da camada, os anéis podem ficar incompletos ou mesmo serem formados por apenas uma linha de células, como mostra a Figura 3.5. Por fim, as células pertencentes a um mesmo anel são somadas, reduzindo a dimensão final de cada evento, sem prejudicar suas interpretações físicas. Esse procedimento, realizado por camada do calorímetro, alcança um número final de 100 anéis distribuídos de acordo com a Tabela 3.1.

As características de deposição de energia de cada um dos eventos analisados são preservadas após a atuação do algoritmo  $\text{Ringer}$ . O sinal resultante, mostrado na Figura 3.6, explicita as semelhanças e diferenças entre os objetos de interesse.

O  $\text{RingerNeuralHypo}$  faz uso de uma rede neural artificial no teste de hipótese sobre os dados formatados em anéis (100, no total). A topologia usada, que pode sofrer variação ao longo do seu tempo de operação, conta com alguns neurônios na camada oculta e



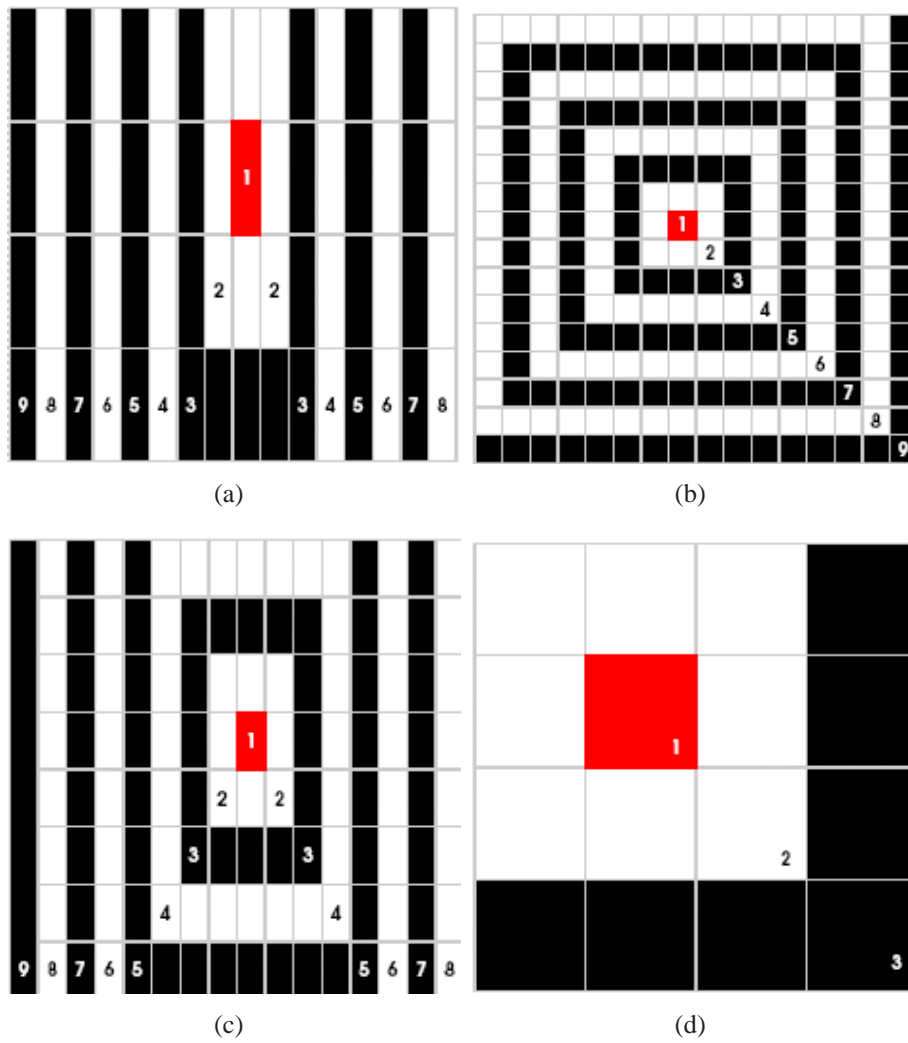


Figura 3.5: Formação de anéis por camada dos calorímetros: (a) *Pre Sampler*; (b) *EM middle layer*; (c) *EM back layer*; (d) *HAD layer 0*.

Tabela 3.1: Número de anéis por camada do calorímetro.

Camada	# de anéis
<i>Pre Sampler</i>	8
<i>EM front layer</i>	64
<i>EM middle layer</i>	8
<i>EM back layer</i>	8
<i>HAD layer 0</i>	4
<i>HAD layer 1</i>	4
<i>HAD layer 2</i>	4

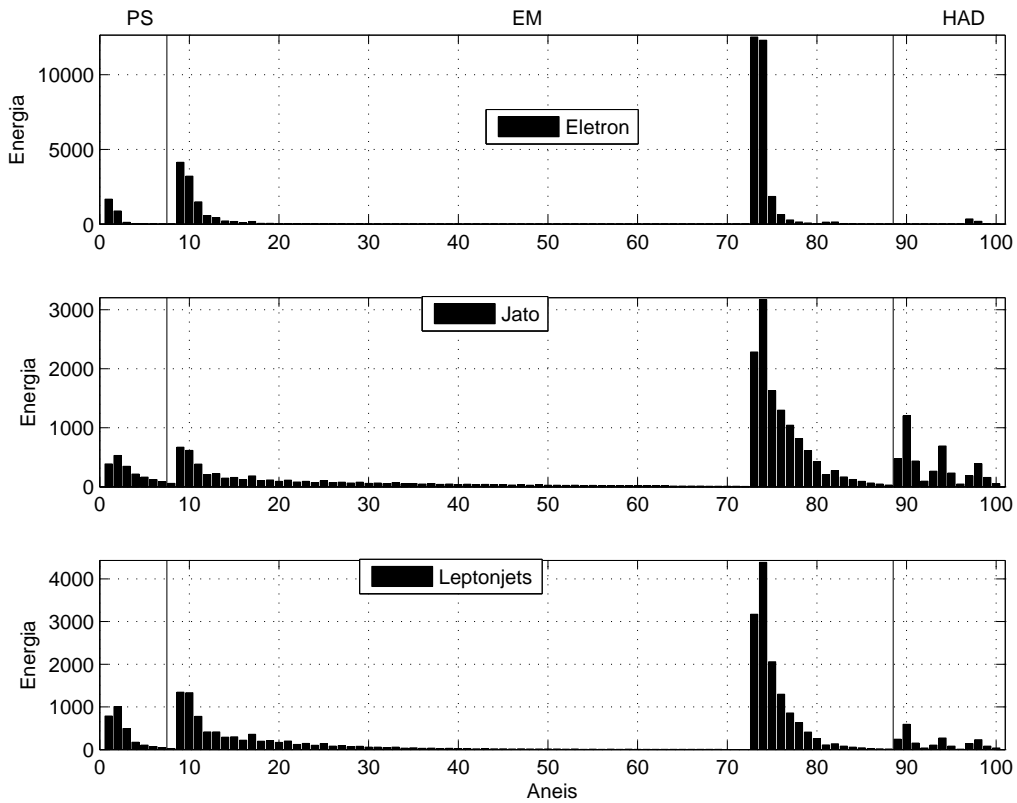


Figura 3.6: Eventos Formatados em Anéis.

um único neurônio na camada de saída. É definido, também, que o valor esperado para eventos da classe alvo é 1 enquanto  $-1$  para o ruído do canal. Todos os neurônios operam com função de ativação do tipo tangente hiperbólica.

Uma das mais recentes otimizações empregadas foi a incorporação do algoritmo de anelamento ao T2Calo. O T2Calo possui diversos sub-algoritmos, e, para as versões mais novas do Athena, o algoritmo de anelamento passa a ser um deles. A principal vantagem é a utilização de um objeto único que propaga tanto as variáveis do T2Calo, bem como os anéis produzidos, facilitando bastante a integração do algoritmo de anelamento com os outros módulos do Athena. Adicionalmente, sendo parte integrante do T2Calo, não mais se faz necessário, pelo algoritmo anelador, refazer os cálculos de refinamento

de  $E_T$ ,  $\eta$  e  $\phi$ , podendo esse delegar ao T2Calo tal tarefa. Em resumo, no que tange a extração de características, o algoritmo de anelamento proposto e o T2Calo tornam-se parceiros. A diferença entre os dois, entretanto, é que o T2Calo realiza a decisão sobre o tipo de partícula incidente observando as 4 variáveis calculadas por ele, enquanto o RingerNeuralHypo emprega redes neurais alimentadas com os anéis produzidos.

### 3.4 Tempo de Processamento

Um aspecto importante a se considerar no desenvolvimento dos algoritmos do sistema de filtragem *online* é o tempo de processamento. Isso inclui todos os passos desde o recebimento do sinal a ser tratado até a resposta do algoritmo de seleção. Para o segundo nível de filtragem, a janela de tempo permitida é de até 40 ms. Uma análise detalhada do tempo de processamento requerido pelo RingerNeuralHypo foi realizada no trabalho [52]. O RingerNeuralHypo, incorporado como uma subrotina do T2Calo, aproveita parte do processamento realizado por este último algoritmo.

As diferentes etapas necessárias ao RingerNeuralHypo e seus respectivos tempos de execução estimados no ambiente do LVL2 são listadas a seguir:

- 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 consome  $0,4927 \pm 0,0787$  ms, correspondendo a 40,7% do tempo total.
- Em seguida, o T2Calo realiza um pré-processamento nos sinais, em que são calculadas as suas 4 variáveis de decisão e os valores de  $E_T$ ,  $\eta$  e  $\phi$  com uma maior precisão do que no L1. Esse passo de processamento leva  $0,1408 \pm 0,0148$  ms, o que equivale a 11,6% do tempo total de execução. A partir daqui é iniciada a sequência de rotinas específicas do RingerNeuralHypo.
- Apenas parte da informação necessária para o RingerNeuralHypo é utilizada pelo T2Calo; então, o restante precisa ser solicitado nesse momento pelo *Region Selector* do RingerNeuralHypo. Um novo fluxo de informação é estabelecido entre o LVL2 e as memórias de armazenamento temporário, consumindo  $0,4375 \pm 0,0996$  ms, o que representa 36% do tempo total.
- Na etapa de pré-processamento do RingerNeuralHypo, as informações da RoI são utilizadas na construção dos anéis para cada camada do calorímetro. O tempo consumido nessa etapa é de  $0,0986 \pm 0,0165$  ms, correspondendo a 8% do total.
- Finalmente, a classificação dos sinais em anéis pela rede neural requer  $0,0387 \pm 0,0018$  ms, o que corresponde a 3% do tempo total.

Tabela 3.2: Tempo gasto nas diferentes etapas do processo filtragem implementado pelo RingerNeuralHypo.

Etapas	Tempo (ms)	% do total
<i>Region Selector</i> - T2Calo	$0,4927 \pm 0,0787$	40,7
Pré-processamento - T2Calo	$0,1408 \pm 0,0148$	11,6
<i>Region Selector</i> - Ringer	$0,4375 \pm 0,0996$	36,1
Pré-processamento (anelamento) - Ringer	$0,0986 \pm 0,0165$	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

Contabilizando todas as etapas de processamento chega-se a um tempo 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 LVL2 pode envolver a análise de sinais de dois subdetectores, 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 3.2, são mostradas as diversas etapas necessárias para a execução do RingerNeuralHypo, juntamente com o tempo requerido por cada uma delas (em valores absolutos e também percentuais).

Embora o RingerNeuralHypo 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, i.e., um aumento de 0,5883 ms, isso é compensado por uma eficiência de discriminação muito superior, conforme será mostrado no Capítulo 5. O excesso de tempo do RingerNeuralHypo em relação ao T2Calo pode ser compensado no Filtro de Eventos pois, como uma quantidade significativamente menor de ruído de fundo é enviado para o último nível, um menor tempo será despendido nesta etapa de filtragem.

# Capítulo 4

## Base de dados, Medidas de Desempenho, Metodologia Proposta e Algoritmos de Referência

Este capítulo apresenta e caracteriza a base de dados usada no trabalho. São informações de calorimetria e de distribuição espacial dos eventos dentro do detector ATLAS a partir da resposta disponibilizada pelo primeiro nível de filtragem. Sobre os dados de calorimetria são obtidas medidas de redundância da informação para cada um dos tipos de eventos considerados.

São apresentados, também, os procedimentos de normalização dos dados utilizados, bem como as implicações do ponto de vista da informação em cada caso.

Por fim, são apresentadas as medidas de desempenho usadas em cada uma das abordagens propostas nos capítulos de resultados e o desempenho alcançado pelos algoritmos que servem de referência. Esses valores, serão base de comparação.

### 4.1 Caracterização do Sinal

A base de dados disponível para o trabalho foi obtida através de simulação de Monte Carlo [93] para colisões próton-próton produzidas pela colaboração ATLAS utilizando Pythia (produção das colisões) e o Geant (interação com os calorímetros) da versão 14.2.10 do Athena. Esses dados representam a informação extraída do *Readout Driver* conectado a cada canal de leitura do ATLAS. Como eventos de sinal, elétrons isolados distribuídos uniformemente de 7 a 80 GeV de  $E_T$  foram produzidos para simular o decaimento conhecido como “canal de ouro” do Higgs, no qual um bóson de Higgs decai em dois bósons  $Z$  que sequentemente decaem em dois léptons cada um.

$$H \rightarrow ZZ^{(*)} \rightarrow l^+l^-l^+l^- \quad (4.1)$$

Como amostras de ruído de fundo, dijetos contendo pelo menos um componente eletromagnético (com  $E_T > 17$  GeV) foram utilizados, de maneira que não sejam eliminados no primeiro nível de filtragem. Os dijetos são, majoritariamente, amostras de jatos de processos QCD. O conjunto de *leptonjets* contém 4 elétrons no estágio final de decaimento cobrindo a mesma faixa de energia de elétrons isolados e do ruído de fundo. Nessa simulação, bósons de gauge escuros decaem no Modelo Padrão via um jato de léptons seguindo o modelo simples não abeliano [63]. No setor escuro, tem-se

$$Z' \rightarrow 2A' \rightarrow 4l \quad (4.2)$$

A produção desses bósons no LHC pode ocorrer através de um processo análogo a produção pontual de um fóton ou através de um raro canal de decaimento de  $Z$ .

A probabilidade de interação entre partículas diminui com o aumento do  $P_T$  dos *leptonjets*. O objetivo é ser sensível ao mais baixo  $P_T$  para maximizar essa probabilidade.

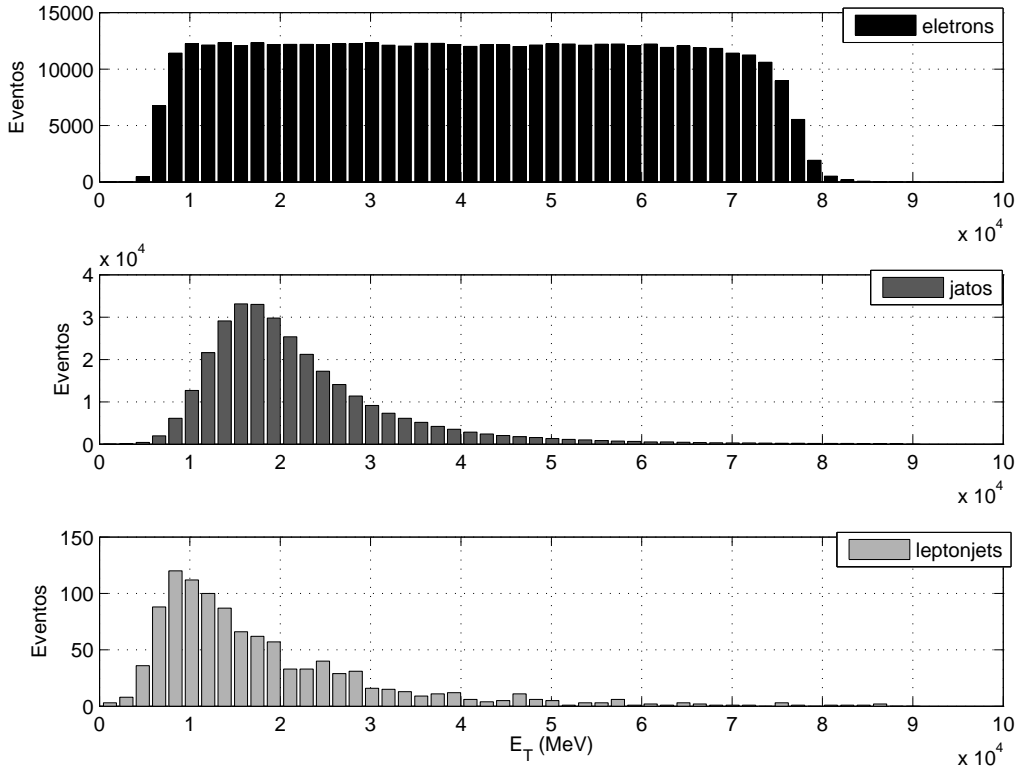


Figura 4.1: Faixa de energia coberta pelos eventos.

A Figura 4.1 caracteriza os tipos de eventos em função da  $E_T$ . A faixa de energia coberta pelos eventos simulados alcança aproximadamente 80 GeV e cada evento apresenta uma distribuição peculiar.

A distribuição dos eventos nas coordenadas  $\eta$  e  $\phi$  do detector é mostrada na Figura 4.2 e indica o espalhamento dos eventos dentro da área de estudo do detector. Nessa análise

foram desconsiderados a parte interna da tampa e o *forward calorimeter* ( $|\eta| > 2,5$ ), pois a probabilidade de eventos interessantes oriundos dessas regiões é bastante reduzida.

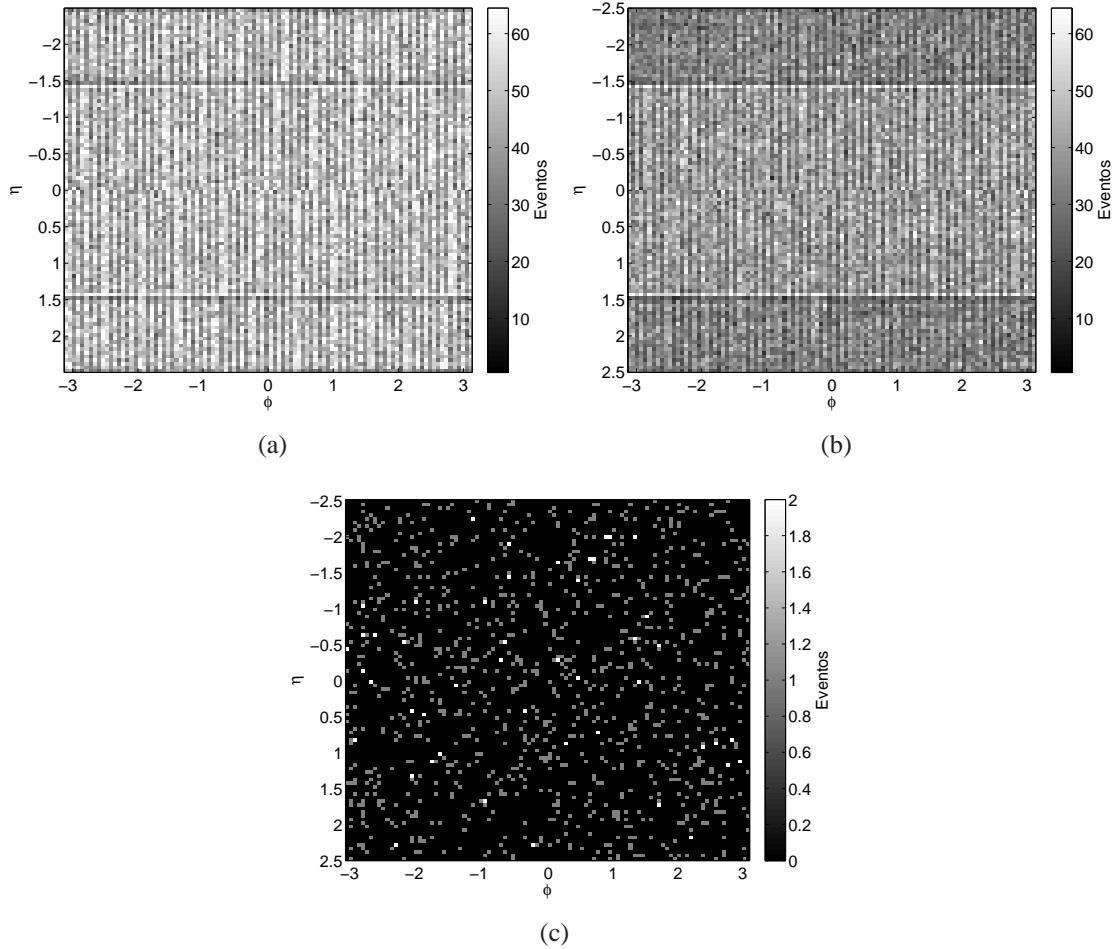


Figura 4.2: Distribuição em  $\eta$  e  $\phi$  dos eventos simulados: (a) elétrons; (b) jatos; (c) *leptonjets*.

O total de eventos usados compreende aproximadamente 470000 assinaturas de elétrons, 310000 assinaturas de jatos e 1000 assinaturas de *leptonjets* ( $4e$ ).

A simulação considera as características do detector e do primeiro nível de filtragem. Os dados foram (aproximadamente) igualmente divididos em treinamento, validação e teste, sendo os dados de *leptonjets* aplicados apenas na etapa de teste do menu  $e10$ . No caso do novo *menu* proposto,  $1j4e10$ , os dados de *leptonjets* foram divididos na seguinte proporção: 75% para treino e 25% para teste e validação. No treinamento dos algoritmos, a base de dados de treinamento foi replicada até alcançar uma quantidade similar à disponível de jatos. Além das probabilidades *a priori* serem desconhecidas, o objetivo foi defini-las com o mesmo valor, dado que o desbalanceamento de classes no treinamento de alguns dos algoritmos empregados é um problema relatado na literatura.

Tabela 4.1: Número de RoI obtidas para cada padrão após a filtragem de nível um com o uso do menu e10.

Eventos	Iniciais	Após LVL1	% de corte
elétron	479.402	470.282	2
jato	711.046	314.843	55,72
<i>leptonjets</i>	1.765	1.052	59,60

## 4.2 Filtragem de Nível 1

Mesmo que a pesquisa desenvolvida destine-se ao segundo nível de filtragem do ATLAS é necessário propagar os eventos simulados pelo primeiro nível de filtragem, conforme a cadeia de processamento definida no Capítulo 3. Como o sistema de filtragem *online* não pode desfazer uma rejeição de evento de um nível inferior, é de grande relevância que o processamento desses eventos esteja sendo feito como esperado, sem a introdução de tendências estatísticas ou outra anomalia que comprometa a operação do sistema. Dessa forma, para a validação do sistema de filtragem, antes do início de sua operação nominal, cortes mais brandos no primeiro nível são adotados, visando armazenar um conjunto maior de eventos que representem uma gama maior dos processos produzidos pelas colisões no LHC. Uma vez que os detectores e o sistema de filtragem estejam bem depurados pela colaboração, cortes mais rigorosos podem ser empregados para a operação nominal.

No âmbito deste trabalho, os dados simulados foram submetidos ao filtro de segundo nível denominado e10 no menu de seleção do canal elétron/jato, o que significa, para o nível 1 de filtragem, um corte sem isolamento que aprovará, basicamente, qualquer evento com energia transversa maior do que 7 GeV. Os cortes seguem a dinâmica detalhada na Seção 3.3.

A Tabela 4.1 mostra o resultado dessa filtragem. Os valores indicam que o algoritmo realiza a tarefa de maneira produtiva. Aceita quase a totalidade dos elétrons recebidos e rejeita mais da metade dos jatos hadrônicos que os falseiam. Pode-se afirmar ainda que os jatos hadrônicos aceitos compõem o conjunto de eventos que não poderiam ser diferenciados dentro das limitações impostas ao algoritmo de primeiro nível, tarefa a ser cumprida pelos níveis superiores de filtragem. O inconveniente encontra-se na taxa de rejeição dos *leptonjets*, que é superior a do ruído de fundo, i.e., a maior parte deles (apesar de serem jatos de elétrons) são mais destoantes de elétrons isolados do que os jatos hadrônicos, ao menos do ponto de vista do primeiro nível de filtragem. Esse resultado mostra que a tarefa de aceitação desse tipo de evento é considerada muito complexa, seja por meio do canal elétron/jato, seja por meio de um futuro canal *leptonjets/jato* criado. Além disso, o nível 1 diminui drasticamente a quantidade desses eventos para os testes dos algoritmos de segundo nível.



### 4.3 Normalização

Após o recebimento da RoI preparada pelo primeiro nível de filtragem, os algoritmos de extração de características do segundo nível são aplicados. Uma questão importante ao problema a ser tratada, no caso dos anéis, é a escala de energia dos eventos considerados. Essa escala ocupa um campo muito variado e como a maioria dos algoritmos empregados no tratamento dos dados é suscetível a amplitude (magnitude) do valor apresentado, um procedimento de normalização desses valores se torna necessário. A idéia é priorizar a análise do formato do chuveiro de partículas criado durante o decaimento de elétrons, jatos e *leptonjets* em detrimento da variação de energia desses eventos. Com isso, espera-se uma generalização da solução para diferentes valores de energia. Em cada uma das normalizações, uma forma de correlacionar os dados é inserida, seja localmente ou entre camadas. Essa relação, posteriormente, terá um papel importante, pois influenciara o resultado dos algoritmos de análise de componentes.

No trabalho foram aplicados três procedimentos:

**Normalização por energia** consiste em dividir o valor da energia de cada anel pelo valor da energia do evento. Nas Figuras 4.3, 4.4 e 4.5 temos a disposição dos anéis de elétrons, jatos e *leptonjets*, respectivamente, sob o efeito dessa normalização. São mostradas também medidas de redundância da informação entre os anéis: correlação, correlação não-linear e informação mútua.

**Normalização sequencial** para cada camada do calorímetro, a energia normalizada ( $E_N$ ) de cada anel da RoI é calculada a partir de

$$E_{N,i} = \frac{E_{l,i}}{E_{tot,l} - \sum_{j=1}^{i-1} E_{l,j}} \quad (4.3)$$

onde  $E_{l,i}$  é a energia original do  $i$ -ésimo anel pertencente a  $l$ -ésima camada e  $E_{tot,l}$  é a energia total da  $l$ -ésima camada. Como resultado, sucessivamente decrescentes fatores de atenuação são aplicados aos anéis exteriores. Como há um número grande de anéis dentro da região de análise e estes últimos, já na borda da RoI, apresentam um nível baixo de energia, um mesmo normalizador é aplicado. A normalização é resistente o suficiente para manter a razão sinal ruído, evitando a amplificação da informação irrelevante. Nas Figuras 4.6, 4.7 e 4.8 temos o resultado da energia normalizada sequencialmente em relação a cada anel de cada camada para os diferentes eventos analisados: elétrons, jatos e *leptonjets*, respectivamente. São mostradas também medidas de redundância da informação.

**Normalização por camada** consiste em dividir o valor da energia de cada anel pelo valor da energia depositada pelo evento apenas na camada a qual o ele pertence. O objetivo seria escalar o sinal em função das peculiaridades de cada camada e, de

uma maneira geral, garantir a mesma ordem de grandeza em todas as camadas. As Figuras 4.9, 4.10 e 4.11 apresentam o resultado da normalização por camada em relação a cada anel de cada camada de cada um dos diferentes eventos analisados: elétrons, jatos e *leptonjets*, respectivamente. Medidas de redundância da informação acompanham os resultados.

## 4.4 Avaliação de Desempenho

Dentre as diferentes medidas estatísticas de redundância da informação compartilhada por duas ou mais variáveis aleatórias, podemos destacar a correlação, a correlação não-linear e a informação mútua.

A correlação indica a força e a direção do relacionamento linear entre duas variáveis aleatórias. No uso estatístico geral, correlação se refere à medida da relação entre duas variáveis, embora correlação não implique causalidade [6]. A correlação é definida como

$$\rho_{x,y} = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y} \quad (4.4)$$

A correlação não-linear mede a força da relação entre duas variáveis em que as taxas de mudança não são constantes. O cálculo da correlação não-linear pode ser realizado ao transformarmos não linearmente pelo menos umas das variáveis

$$x_t = f(x) \quad (4.5)$$

$$y_t = g(y) \quad (4.6)$$

$$\nu_{x_t, y_t} = \frac{E[(x_t - \mu_{x_t})(y_t - \mu_{y_t})]}{\sigma_{x_t} \sigma_{y_t}} \quad (4.7)$$

onde  $f(\cdot)$  e  $g(\cdot)$  são funções não-lineares [7].

Em teoria da informação, a informação mútua de duas variáveis aleatórias é uma quantidade que mede a dependência mútua entre essas duas variáveis. Formalmente, a informação mútua de duas variáveis  $X$  e  $Y$  é definida por

$$I(X; Y) = \int_X \int_Y p(x, y) \log \left( \frac{p(x, y)}{p(x) p(y)} \right) dx dy \quad (4.8)$$

onde  $p(x, y)$  é a função densidade de probabilidade conjunta de  $X$  e  $Y$ , e  $p(x)$  e  $p(y)$  são as funções densidade de probabilidade marginais de  $X$  e  $Y$  [7]. Essas três medidas podem ser empregadas para avaliar a qualidade das componentes extraídas pelos blocos de pré-processamento nas diferentes abordagens propostas.

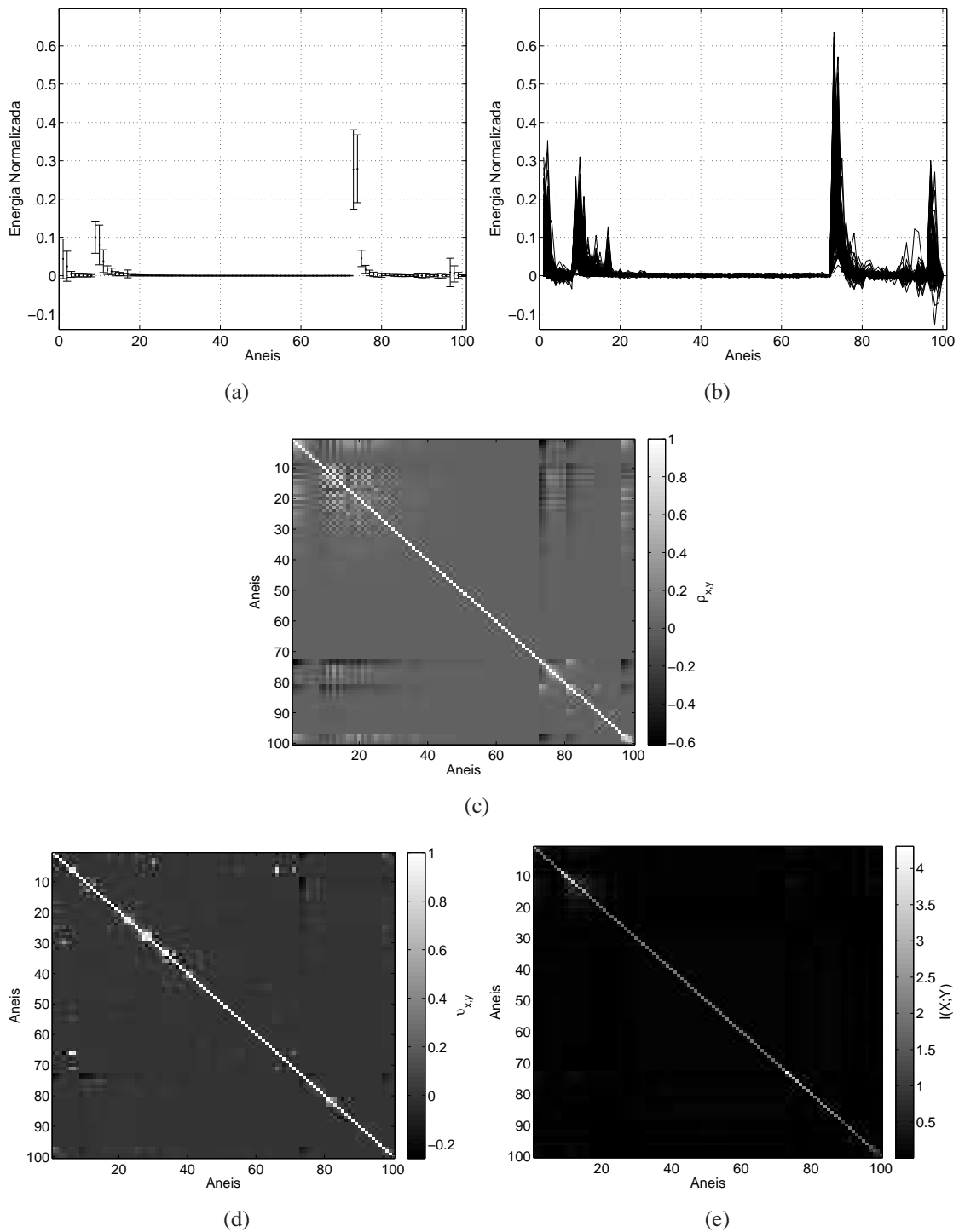
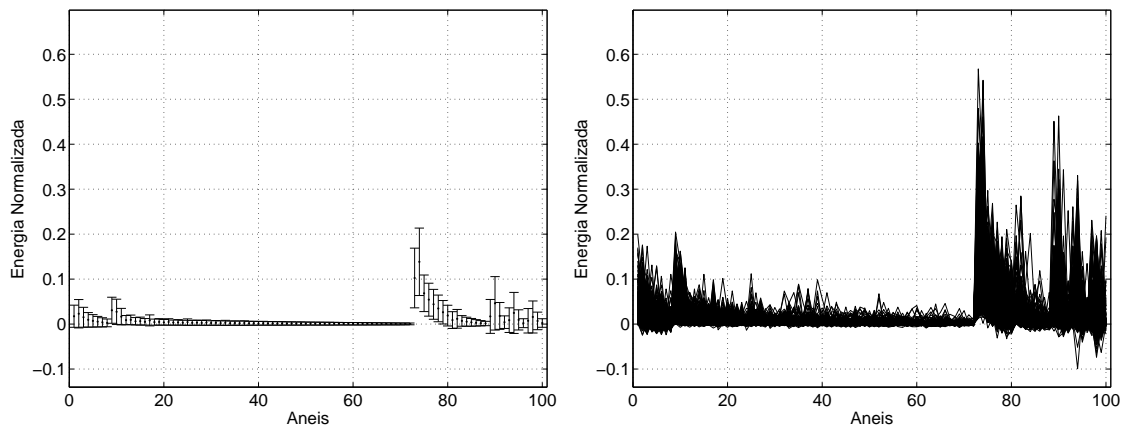
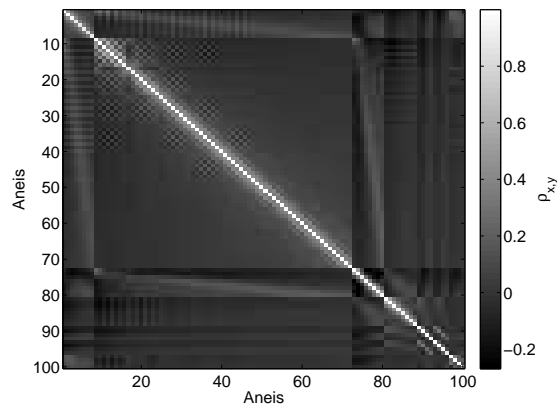


Figura 4.3: Elétron simulado normalizado por energia: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de elétrons: (c) Correlação linear; (d) Correlação não-linear; (e) Informação mútua.

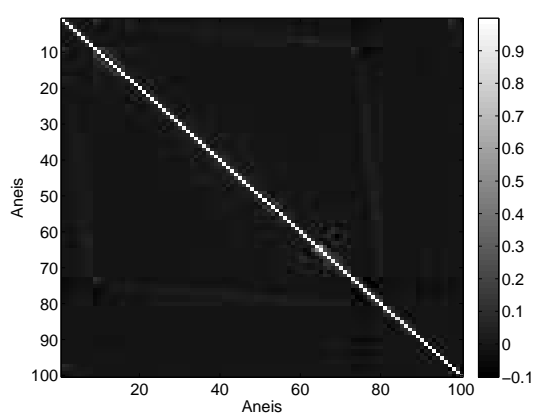


(a)

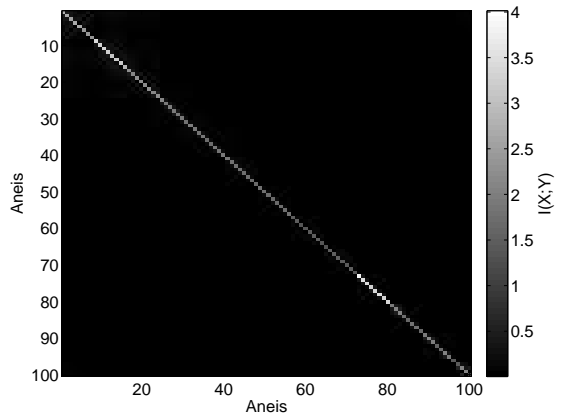
(b)



(c)



(d)



(e)

Figura 4.4: Jato simulado normalizado por energia: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de jatos: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua.

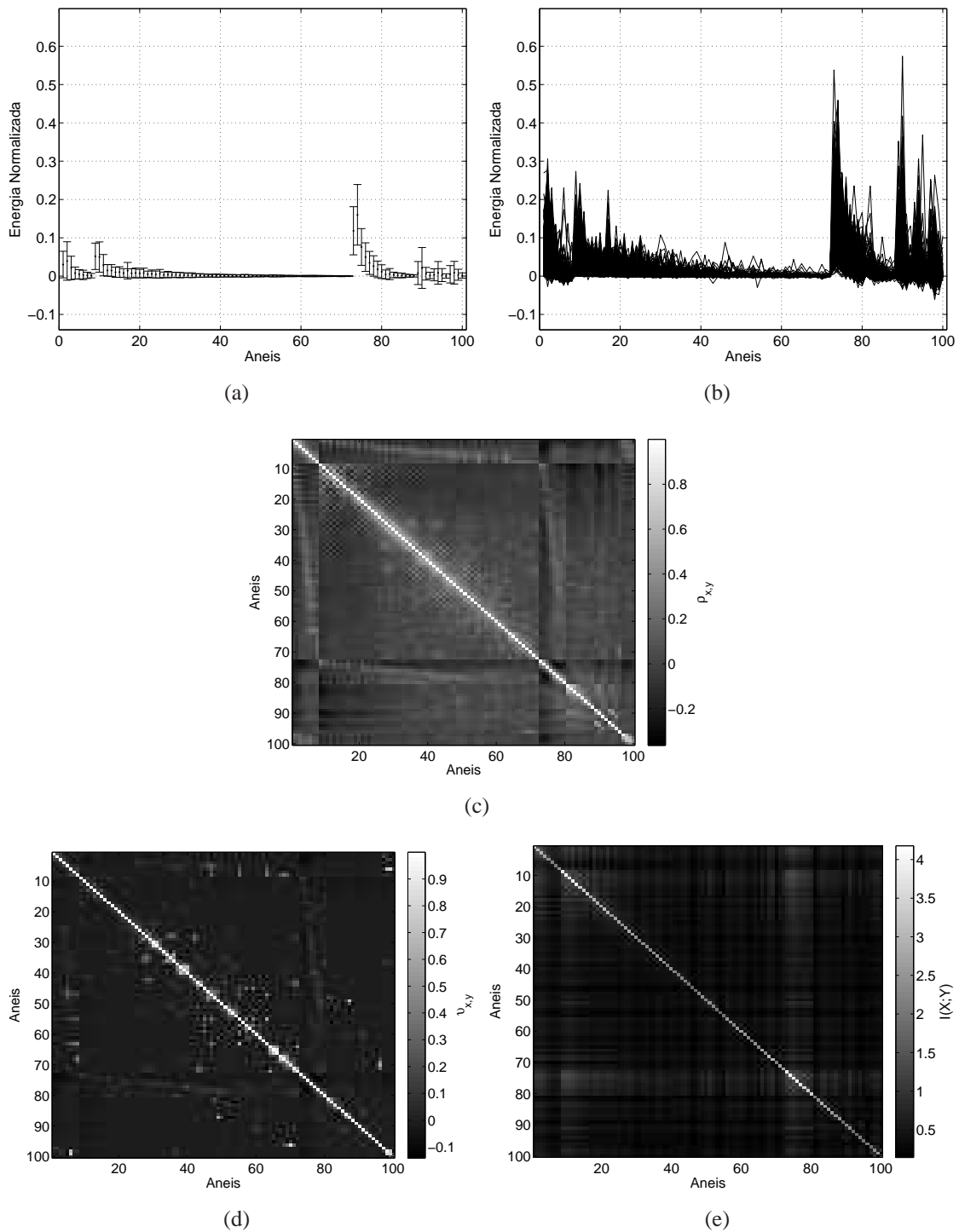
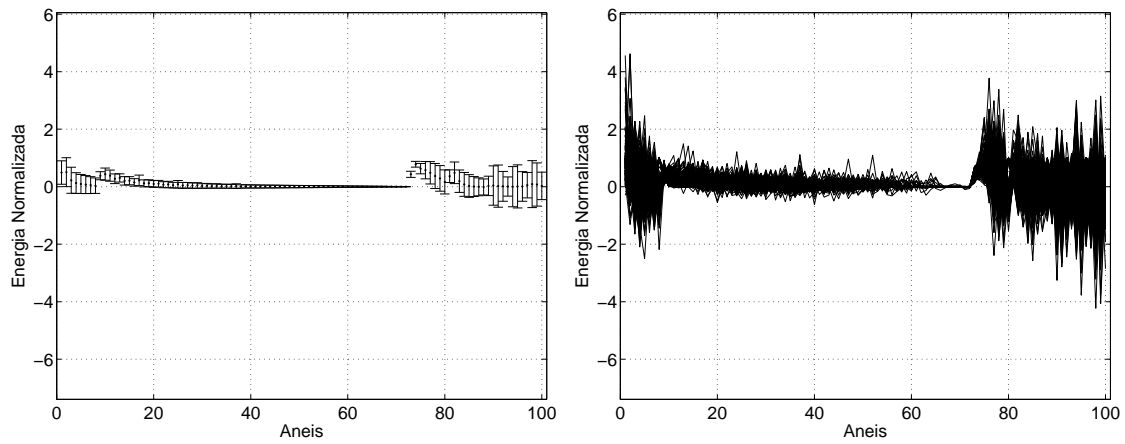
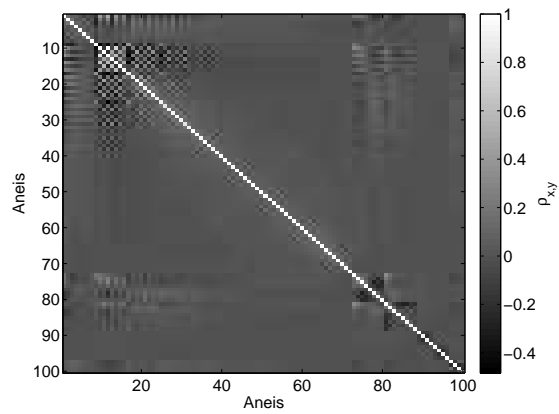


Figura 4.5: *Leptonjets* simulado normalizado por energia: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de *leptonjets*: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua.

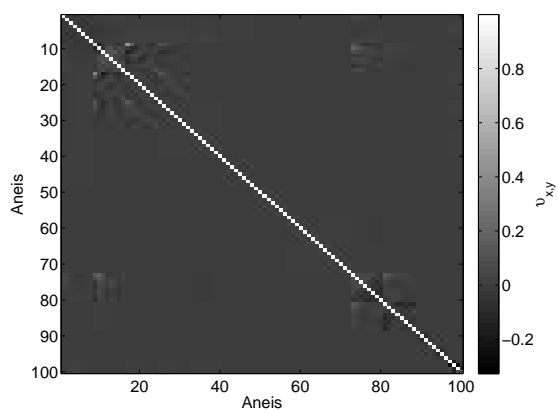


(a)

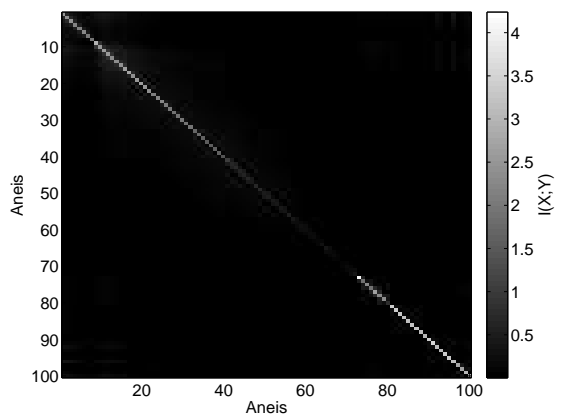
(b)



(c)

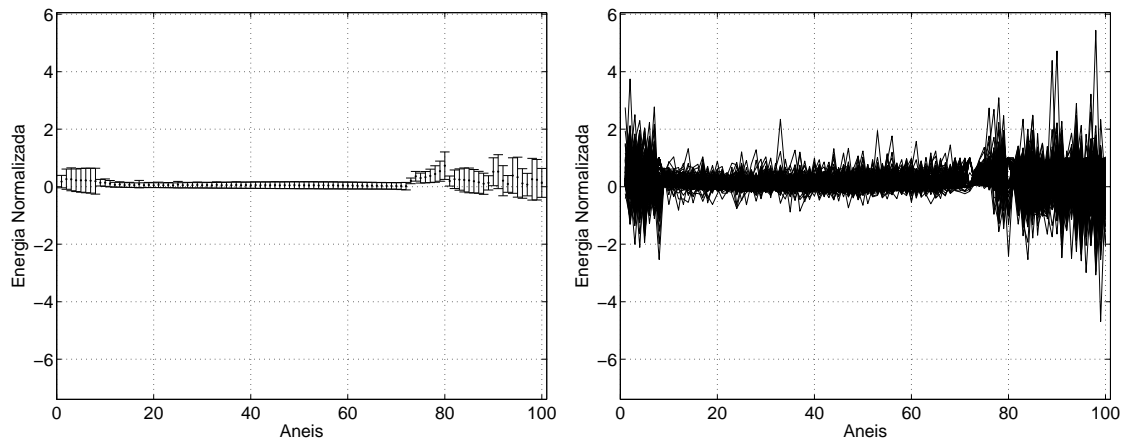


(d)



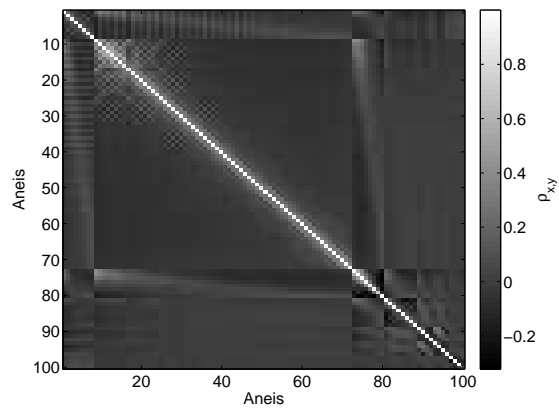
(e)

Figura 4.6: Elétron simulado normalizado sequencialmente: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de elétrons: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua.

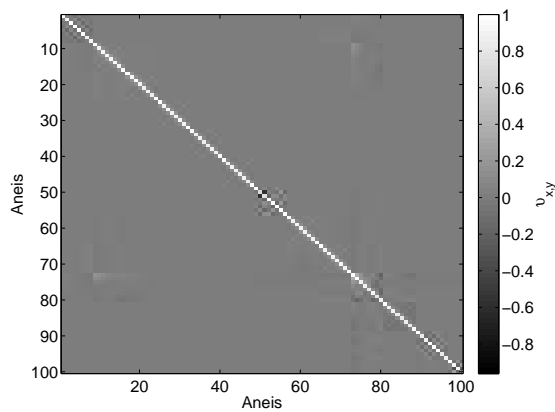


(a)

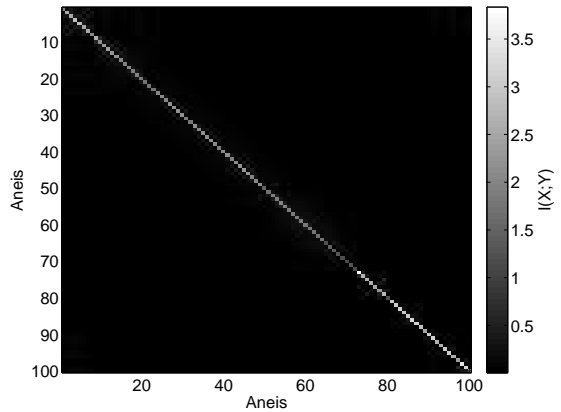
(b)



(c)

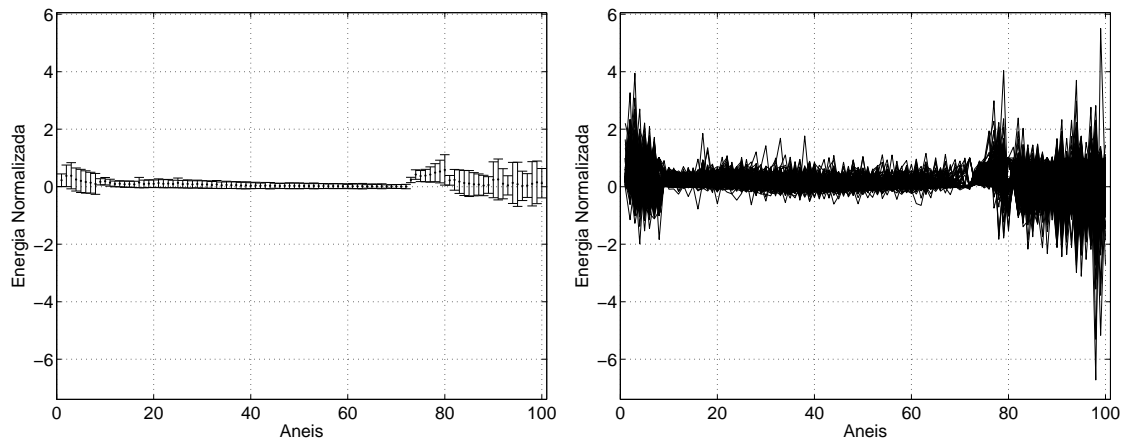


(d)



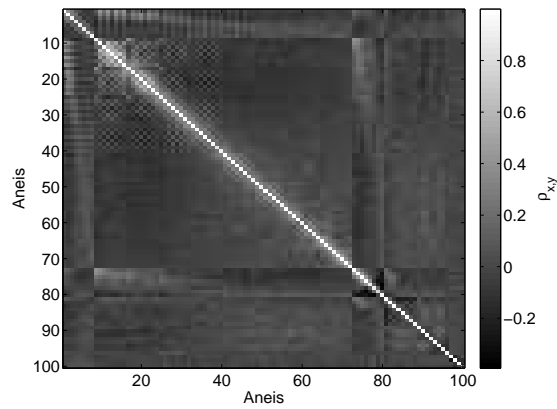
(e)

Figura 4.7: Jato simulado normalizado sequencialmente: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de jatos: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua.

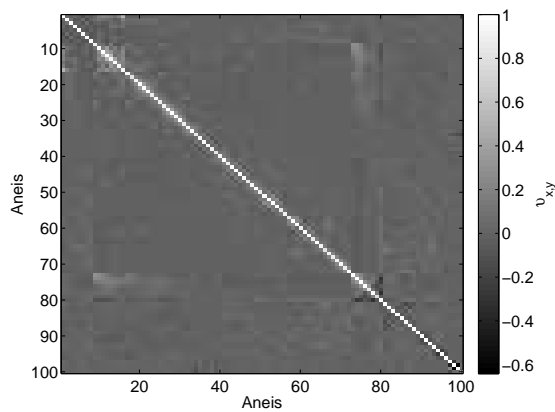


(a)

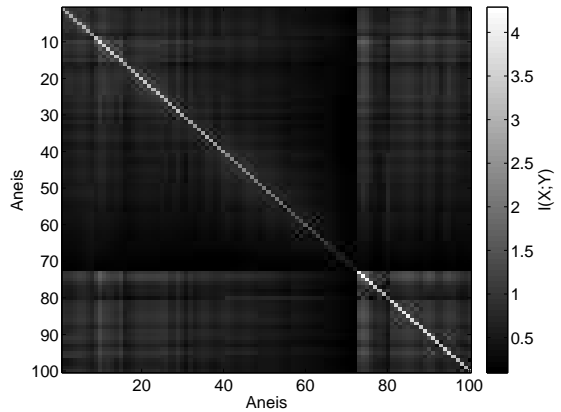
(b)



(c)



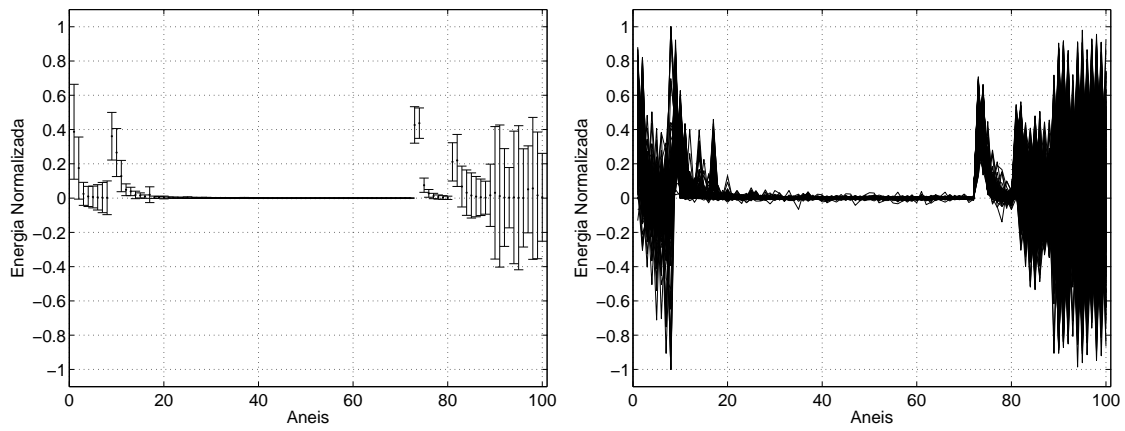
(d)



(e)

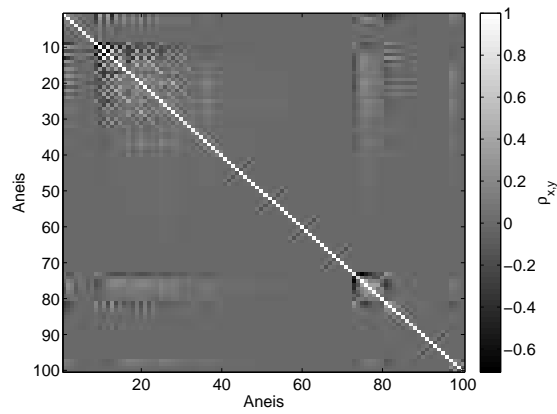
Figura 4.8: *Leptonjets* simulado normalizado sequencialmente: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de *leptonjets*: (a) Correlação linear. (b) Correlação não-linear. (c) Informação mútua.



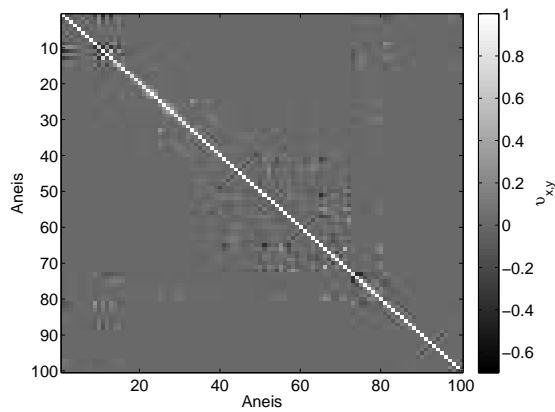


(a)

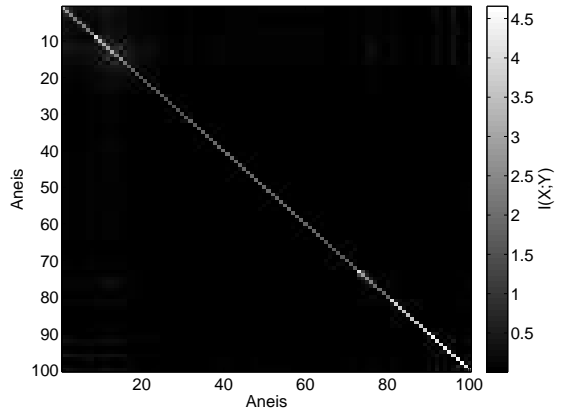
(b)



(c)

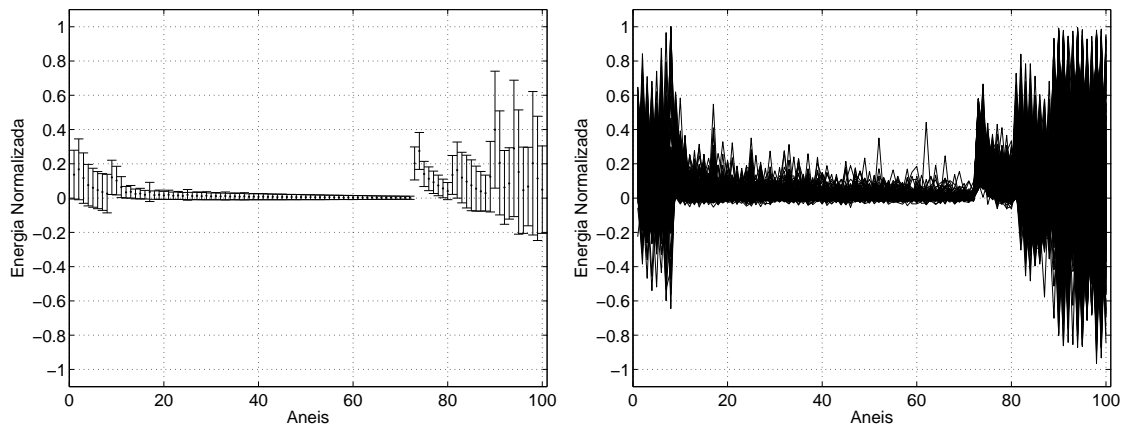


(d)



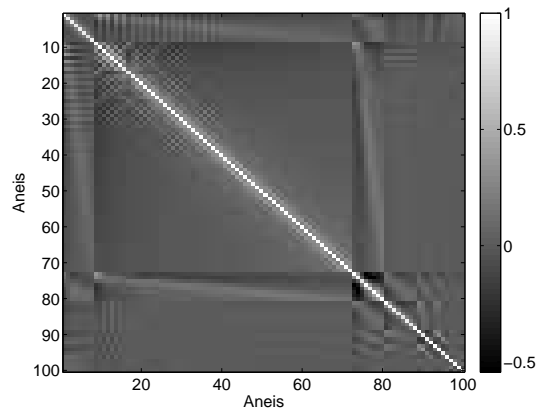
(e)

Figura 4.9: Elétron simulado normalizado por camada: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de elétrons: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua.

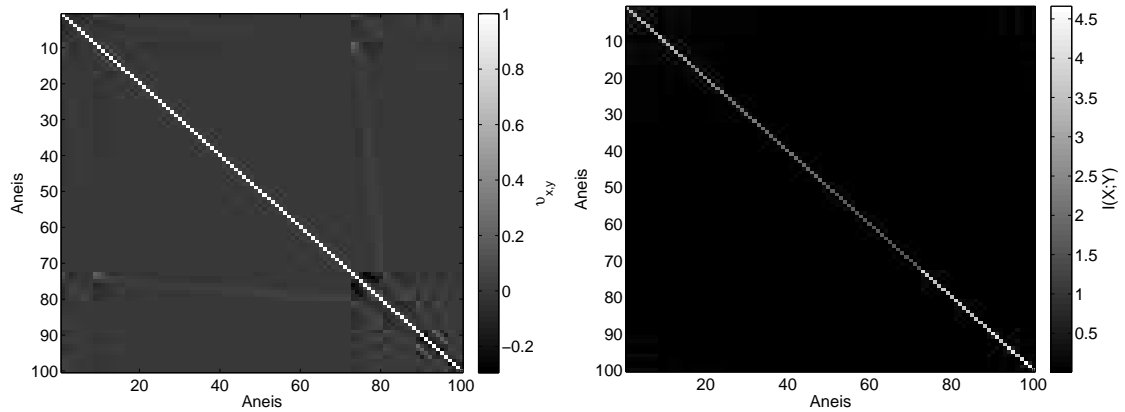


(a)

(b)



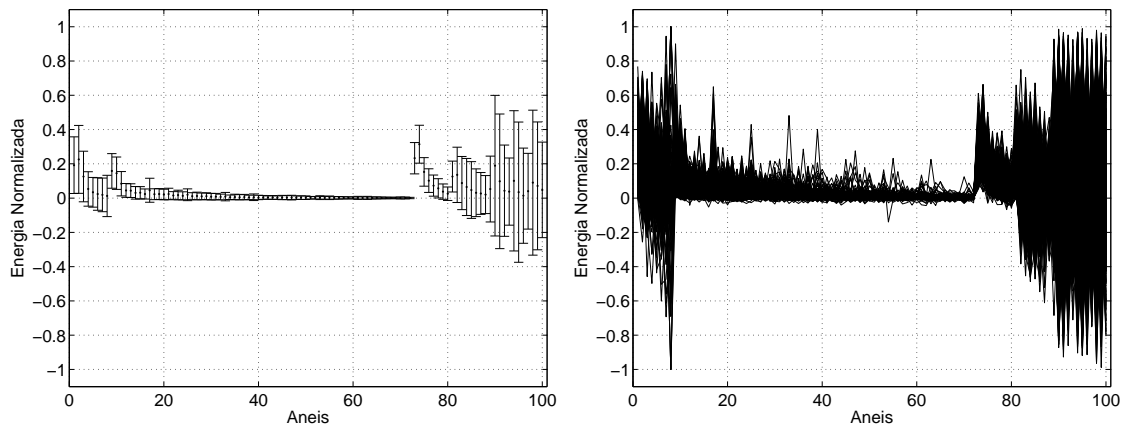
(c)



(d)

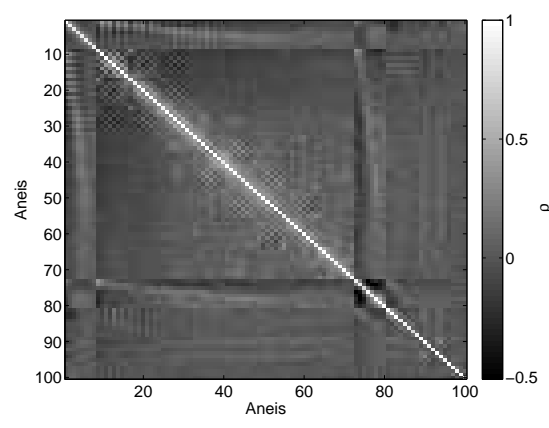
(e)

Figura 4.10: Jato simulado normalizado por camada: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de jatos: (c) Correlação linear. (d) Correlação não-linear. (e) Informação mútua.

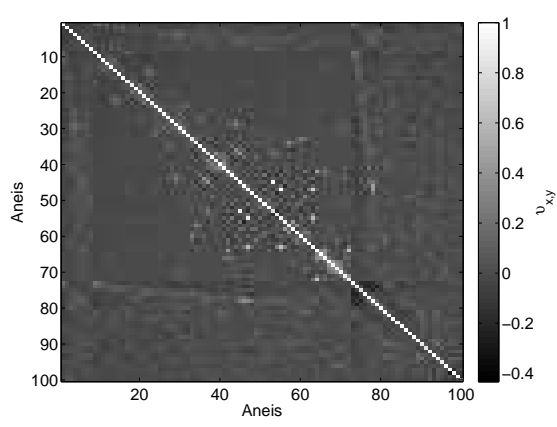


(a)

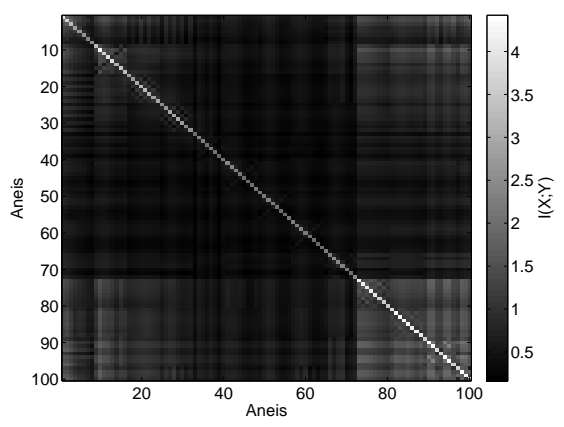
(b)



(c)



(d)



(e)

Figura 4.11: *Leptonjets* simulado normalizado por camada: (a) Evento médio; (b) Eventos sobrepostos. Medidas de redundância na informação contida no conjunto de dados de *leptonjets*: (a) Correlação linear. (b) Correlação não-linear. (c) Informação mútua.

No que diz respeito aos algoritmos de teste de hipótese, o desempenho de uma máquina de aprendizado pode ser mensurado durante e após o treinamento, segundo o tipo de aprendizado e método de treinamento utilizado. Na maioria dos casos, os algoritmos avançados de aprendizagem de máquina objetivam aprender representações de funções simbólicas que poderiam ser compreendidas e verificadas por peritos da aplicação tratada, os chamados especialistas. Os especialistas buscam anomalias nos resultados, tais como tendências, *outliers*, desconformidades com a teoria dentre outros. Quando essa tarefa não é trivial, a qualidade do treinamento de um algoritmo de aprendizado pode ser avaliada em termos do erro médio cometido durante o treinamento, além de medidas complementares na etapa de teste.

Mesmo a etapa de treinamento assumindo resultados dentro de um patamar dito aceitável, especificado pelo operador do sistema, sua completa validação só é obtida após a comprovação da generalização dos resultados para um conjunto de dados de entrada não presentes na fase de treinamento.

Para o caso específico de classificação de padrões, é comum o uso de medidas de probabilidade de detecção e de probabilidade de falso alarme. Como o cenário tratado nesse trabalho envolve mais de uma classe de detecção, teremos três medidas principais: probabilidade de detecção de elétrons ( $PD_e$ ), probabilidade de detecção de *leptonjets* ( $PD_lj$ ) e probabilidade de falso alarme de jatos hadrônicos ( $PF_j$ ), todas apresentadas em medidas percentuais (%). Esses valores também são apresentados em função de algumas importantes variáveis físicas de interesse como a energia transversa ( $E_T$ ) e as coordenadas  $\eta$  e  $\phi$  do evento dentro do detector ATLAS. Essa análise permite validar o sistema considerando o conhecimento especialista do problema e, nesse caso, são calculados histogramas em que 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.

Uma opção bastante difundida na literatura para medida de desempenho é a curva ROC (*Receiver Operating Characteristic*) [94, 95]. A curva ROC, para um problema de detecção binária, expõe no eixo  $x$  a taxa de falso alarme, enquanto, no eixo  $y$ , a taxa de detecção conforme se varia o patamar de decisão de exemplos corretamente classificados; com ela é possível mostrar uma relação entre falsos alarmes e o total de acertos para o classificador específico. O classificador perfeito corresponde ao ponto  $(0, 1)$ : nenhum falso positivo e 100% de precisão na classificação de exemplos da classe alvo. A Figura 4.12 mostra um exemplo de curva ROC.

Outra medida de comparação pode ser feita através do produto SP [48, 96], que é definido como o produto da soma das eficiências pelo produto destas, i.e.,

$$SP = \sqrt{\frac{\sum_{i=1}^j Ef_i}{j}} \times \sqrt{\prod_{i=1}^j Ef_i} \quad (4.9)$$

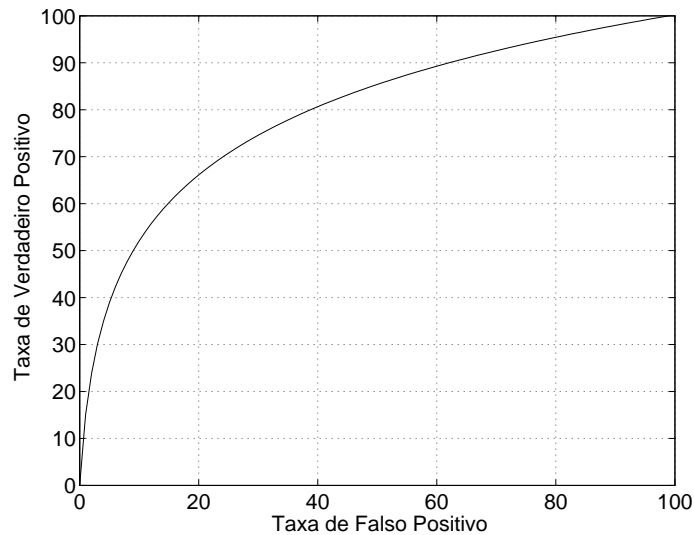


Figura 4.12: Curva ROC.

onde  $Ef_i$  é a eficiência de detecção na classe  $i$ . Ele assume seu valor máximo quando o algoritmo de classificação apresenta eficiência máxima para cada uma das  $j$  classes presentes.

O índice SP é usado como indicador do patamar ótimo de decisão para um dado discriminador. Ao excursionar o patamar de decisão dentro da régua limite, calcula-se correspondentes valores de desempenho, o SP máximo indica o ponto de operação ideal.

A Figura 4.13 exemplifica o comportamento de dois classificadores. O histograma de saída desses dois classificadores, dado um mesmo conjunto de dados particionado em duas classes, é mostrado nas Figuras 4.13(a) e 4.13(b). Para o classificador 1 há uma maior confusão entre as classes, enquanto o classificador 2 consegue separá-las melhor. Esse comportamento se reflete na curva ROC e no Produto SP, mostrados nas Figuras 4.13(c) e 4.13(d), respectivamente. Para o classificador 2, seu melhor desempenho se traduz em um valor maior de área sob a curva ROC. O valor SP máximo para cada classificador está associado a um respectivo patamar de decisão, que, por sua vez, está associado a um respectivo par de valores de detecção e falso alarme, marcados para ambos classificadores.

É importante ressaltar que o produto SP é sensível a degradação da eficiência de classificação de cada uma das classes e o valor obtido pode ser usado como métrica de comparação da eficiência de diferentes classificadores.

No que diz respeito à informação usada para a construção de algoritmos de hipótese, é importante inferir a relevância das variáveis consideradas no processo decisório, obtidas na fase inicial de extração de características. Para tal, o mapeamento de relevância [7] tem se mostrado eficiente, avaliando a resposta do classificador quando a variável em análise é fixada no seu valor médio. Assim,

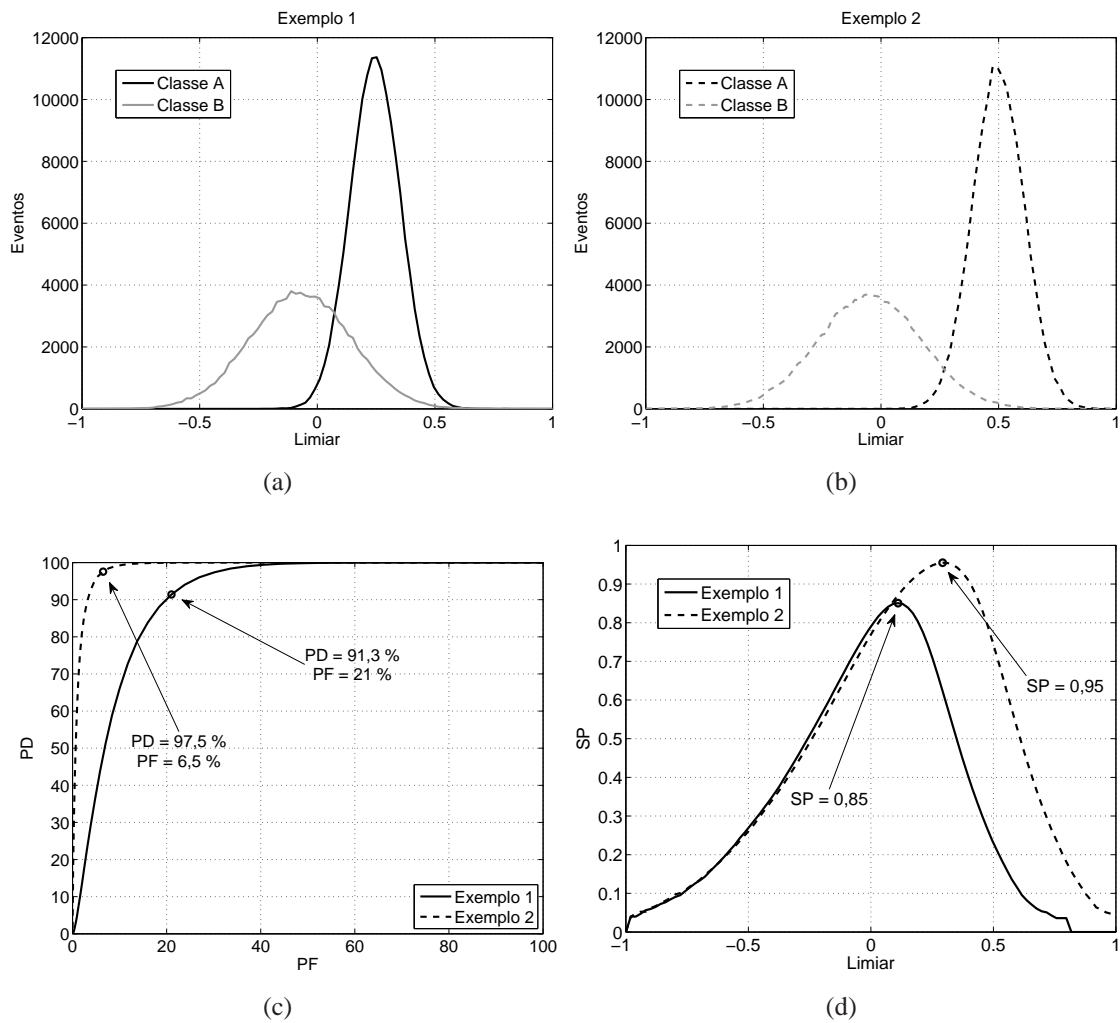


Figura 4.13: (a) Exemplo 1 de histograma de saída de classificador; (b) Exemplo 2 de histograma de saída de classificador; (c) Curvas ROC associadas aos histogramas mostrados em (a) e (b); (d) Produto SP associados aos histogramas mostrados nas Figuras (a) e (b).

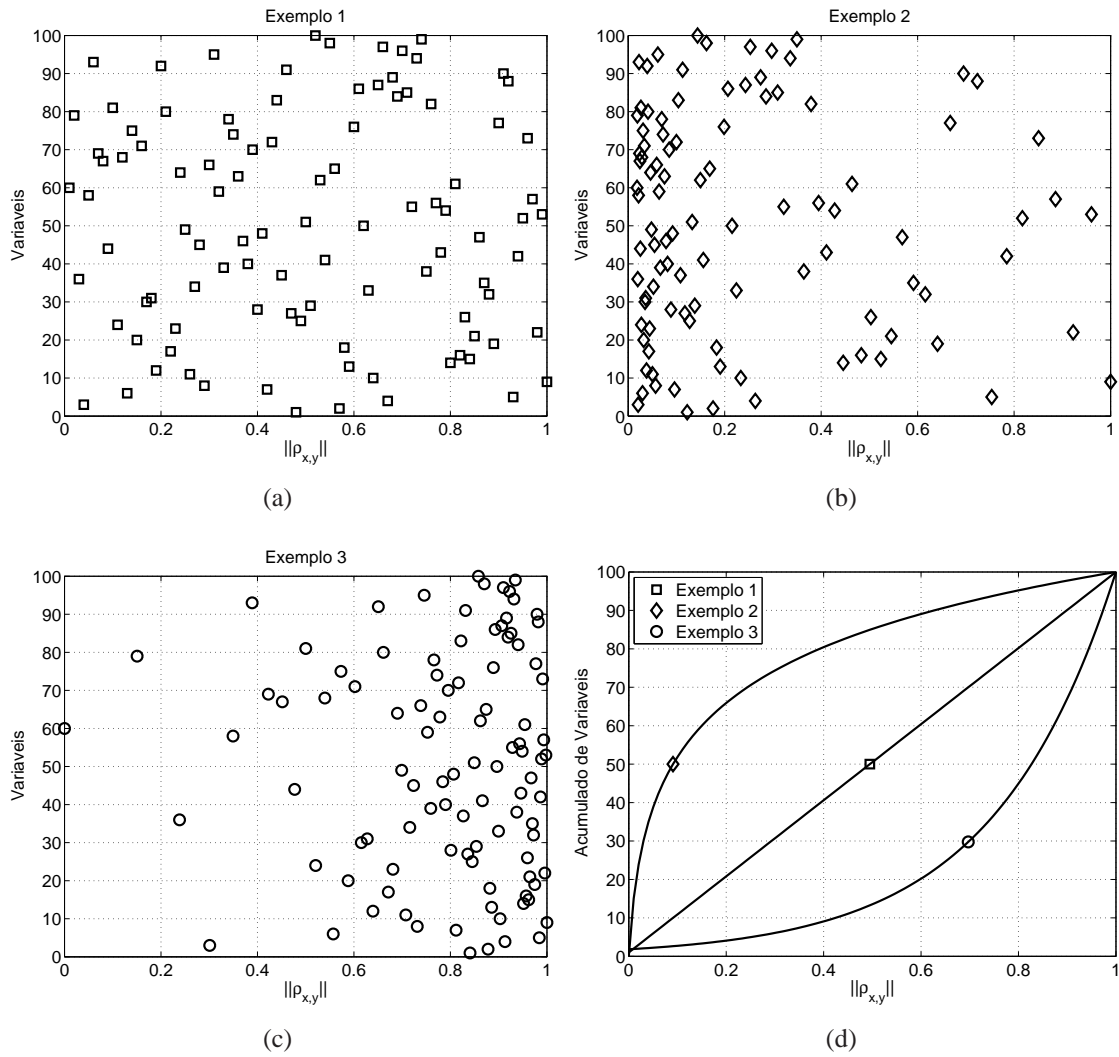


Figura 4.14: (a) (b) (c) três diferentes exemplos de correlação entre a componente de máxima relevância de um processo decisório e as variáveis usadas na sua geração. (d) Acumulado de variáveis por limiar de correlação.

$$R_i = \frac{1}{n} \sum_{j=1}^n (S(\mathbf{x}_j) - S(\mathbf{x}_j | x_{ji} = \bar{x}_i))^2 \quad (4.10)$$

onde  $S(\cdot)$  é a saída do classificador. Quanto maior o valor obtido para  $R_i$ , mais relevante é a variável, i.e., mais distante da média ela está e mais informação ela agrega ao processo decisório [7]. A indicação da relevância pode também ser usada para um processo de compactação *a posteriori*.

Uma medida que pode relacionar a relevância de uma componente que alimenta o processo decisório com as variáveis originais usadas na sua geração é o cálculo da correlação entre elas. Isso pode indicar quão frágil ou dependente é a componente de um sensor, anel ou camada do calorímetro. O cálculo é realizado como mostra a equação (4.4), onde  $x$  são os anéis originais e  $y$  a componente de maior relevância.

As Figuras 4.14(a), 4.14(c) e 4.14(b) ilustram a relação de três componentes diferentes

com respeito ao conteúdo de suas variáveis formadoras. Nesse exemplo, o número de variáveis formadoras é igual a 100 e os marcadores indicam o valor da correlação entre uma variável específica e a componente analisada. A Figura 4.14(d) ilustra o comportamento geral dessas relações ao apresentar em cada curva o número acumulado de variáveis formadoras que apresentam valor de correlação menor ou igual a um determinado patamar.

A linha contínua escura indica que a componente mostrada no exemplo da Figura 4.14(a) possui uma correlação crescente e aleatória com cada uma de suas variáveis formadoras, dado que todas, *a priori*, fornecem informação especializada do problema. Esse cenário pode ser interpretado como um sistema em que é possível medir o impacto da informação de cada variável na formação da componente além de explicitar que a informação na componente pode ser robusta, até certa medida, a uma falha de leitura em algumas das variáveis formadoras.

No caso da linha tracejada de coloração intermediária, a indicação é de que um número grande de variáveis formadoras possui uma pequena correlação com a componente em análise na Figura 4.14(b). Esse cenário indica que há um compartilhamento, em pequena escala, da contribuição dessas variáveis e uma falha de leitura em algumas delas não comprometeria o resultado observado na componente. Contudo, uma falha nas componentes de maior correlação - que não são muitas - poderia indicar uma operação com ressalvas para o sistema.

Para a linha de coloração mais clara, observa-se que a componente analisada na Figura 4.14(c) possui uma correlação forte com um número grande de variáveis formadoras. Nesse cenário, entende-se que a componente estudada tem alta sensibilidade a falhas, porém um alto poder de compactação, pois engloba boa parte da informação de muitas variáveis em uma só medida.

## 4.5 Metodologia Proposta

Considerando os sinais medidos nos calorímetros e formatados como descrito anteriormente, uma etapa de pré-processamento é aplicada antes do algoritmo de teste de hipótese fornecer uma resposta (rótulo) sobre o evento em análise. As principais variantes no trabalho residem no bloco de pré-processamento: com diferentes técnicas de análise de componentes usadas na extração de características; e na classificação, com o uso de classificadores lineares e não lineares. A Figura 4.15 ilustra tal procedimento.

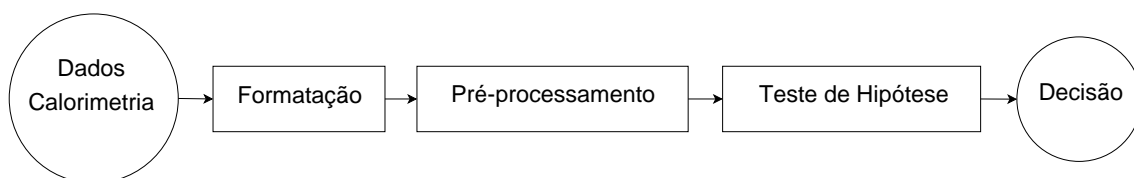


Figura 4.15: Procedimento de detecção/decisão de eventos.



Na extração de características foram usadas duas abordagens distintas:

- O modo não segmentado processa a informação presente nos 100 anéis das diferentes camadas dos calorímetros sem distinção entre elas. Para cada evento a informação é concatenada em um único vetor de alta dimensão que é, então, apresentado ao algoritmo de extração de características.
- O modo segmentado processa separadamente a informação presente em cada camada dos calorímetros. Os algoritmos de extração de características operam localmente sobre os anéis pertinentes.

Em todas as abordagens é esperada diminuição da dimensionalidade dos dados. Espera-se também manter ou mesmo aumentar os valores de eficiência alcançados pelos algoritmos de referência. Este trabalho está focado na utilização de técnicas de análise de componentes (principais, não-lineares e independentes) visando extrair características relevantes dos dados de calorimetria para dois canais de seleção de eventos: a discriminação elétron/jato com potencial de aceitação de eventos exóticos (*leptonjets*); e a discriminação, a partir de um canal independente, *leptonjets/jato*.

Considerando os diferentes algoritmos e modelos de extração de componentes em cada uma das vertentes, neste trabalho foram empregadas:

- A implementação da técnica PCA baseada no algoritmo “batch” que realiza a decomposição em autovalores/autovetores da matriz de correlação dos dados. Foi estipulado uma redução de dimensionalidade na transformação dos dados que preservasse, pelo menos, 80% da energia do sinal nas diferentes metodologias de pré-processamento aplicadas.
- A implementação da técnica ICA baseada no algoritmo FastICA. Assim como com a PCA, preservou-se, pelo menos, 80% da energia do sinal transformado.

Com respeito ao algoritmo de classificação, duas variantes são utilizadas, todas dentro da categoria de algoritmos supervisionados.

- O classificador linear de Fisher.
- O classificador neural do tipo percéptrons de múltiplas camadas. Nesse caso, seu treinamento foi realizado com o algoritmo *backpropagation*, onde o único parâmetro alterado durante os testes das diversas metodologias estudadas foi o número de neurônios na camada escondida da rede neural.

Detalhes de cada uma das técnicas de extração de características e dos algoritmos de classificação estão apresentadas no Apêndice A.

### 4.5.1 Metodologia Proposta para o Canal elétron/jato

A infraestrutura atual do detector, que apresenta adequada flexibilidade quanto aos recursos de implementação, é utilizada em diferentes abordagens que buscam uma solução experimental para problemas teóricos. Dentre esse ferramental está um dos canais de maior interesse do ATLAS na busca do Bóson de Higgs, o canal elétron/jato. Ele deve oferecer o melhor desempenho para uma massa estimada do Higgs entre 180 – 700 GeV. Esse canal é muitas vezes chamado de "canal de ouro" devido ao alto significado do sinal esperado, do qual espera-se a produção de um sinal de importância muito alta já nos primeiros anos de funcionamento do LHC.

Além da importância desse canal na busca do Higgs, ele oferece ainda a possibilidade de filtragem de outros eventos com algumas características nele potencializadas ou bem descritas. Um exemplo é a busca de um candidato a matéria escura, que seria indiretamente identificado por uma forte presença de objetos eletromagnéticos (*leptonjets*) em estado final de decaimento, como já descrito no Capítulo 2.

Diante desse cenário, criou-se três metodologias de trabalho. A primeira diz respeito a classificação dos dados formatados em anéis. A segunda e a terceira inserem um bloco de pré-processamento entre a formatação dos dados e a classificação deles. A diferença entre elas está na forma como o pré-processamento é realizado - segmentado ou não segmentado.

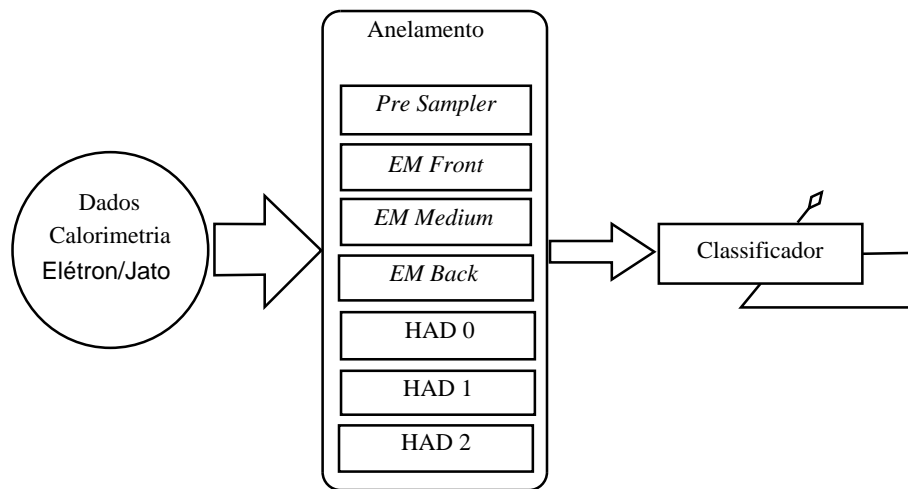
Na classificação baseada em anéis (Figura 4.16), dados de elétrons e jatos hadrônicos (pertinentes ao canal) são usados no treinamento do algoritmo de classificação (Figura 4.16(a)). Posteriormente, novos dados de elétrons e jatos (não usados no treinamento), além de *leptonjets*, são usados para avaliar o procedimento (Figura 4.16(b)).

Para o procedimento não segmentado, a Figura 4.17 mostra os passos seguidos. O algoritmo de extração de características é treinado com dados pertinentes ao canal (Figura 4.17(a)). Após o treinamento, o mesmo conjunto de dados, agora transformado pela técnica aplicada, é apresentado ao algoritmo de classificação que realiza então o seu treinamento (Figura 4.17(b)). Por fim, na fase de teste, dados pertinentes ao canal e dados exóticos são usados para avaliar o desempenho da abordagem (Figura 4.17(c)).

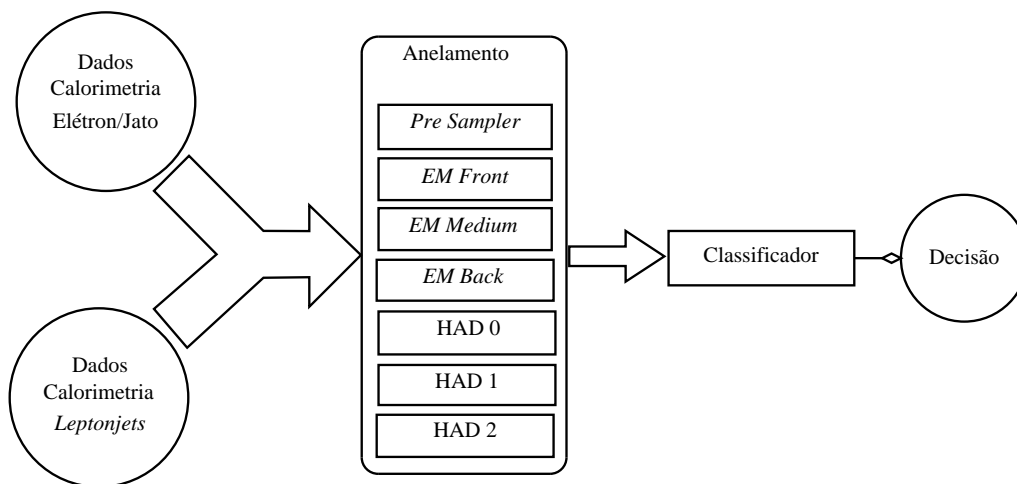
No caso do procedimento segmentado, o diferencial no treinamento dos algoritmos de pré-processamento está na sua aplicação sobre a informação de cada camada dos calorímetros, como ilustrado na Figura 4.18(a). A abordagem completa é apresentada na Figura 4.18.

### 4.5.2 Metodologia proposta para o canal *leptonjets*/jato

Nesse segunda proposta, um canal específico é criado para detecção e classificação de *leptonjets*. Foram considerados os mesmos requisitos demandados por um canal existente do detector (como o canal elétron/jato). Nesse cenário, foram realizados praticamente os



(a)



(b)

Figura 4.16: Procedimento baseado em anéis para o canal elétron/jato: (a) Treinamento do algoritmo de classificação; (c) Teste de hipótese.

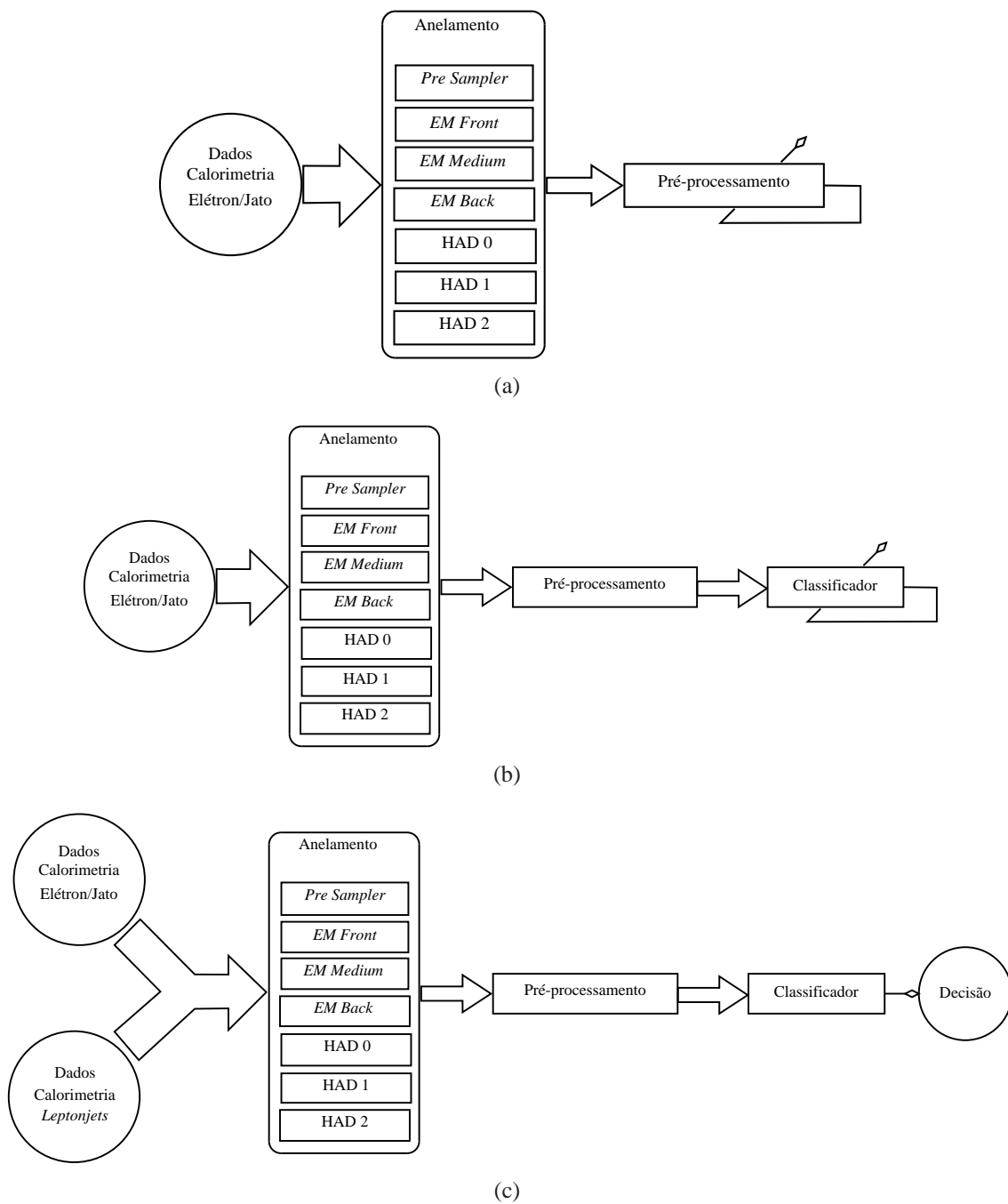


Figura 4.17: Procedimento não segmentado para o canal elétron/jato: (a) Treinamento do algoritmo de pré-processamento; (b) Treinamento do algoritmo de classificação; (c) Teste de hipótese.

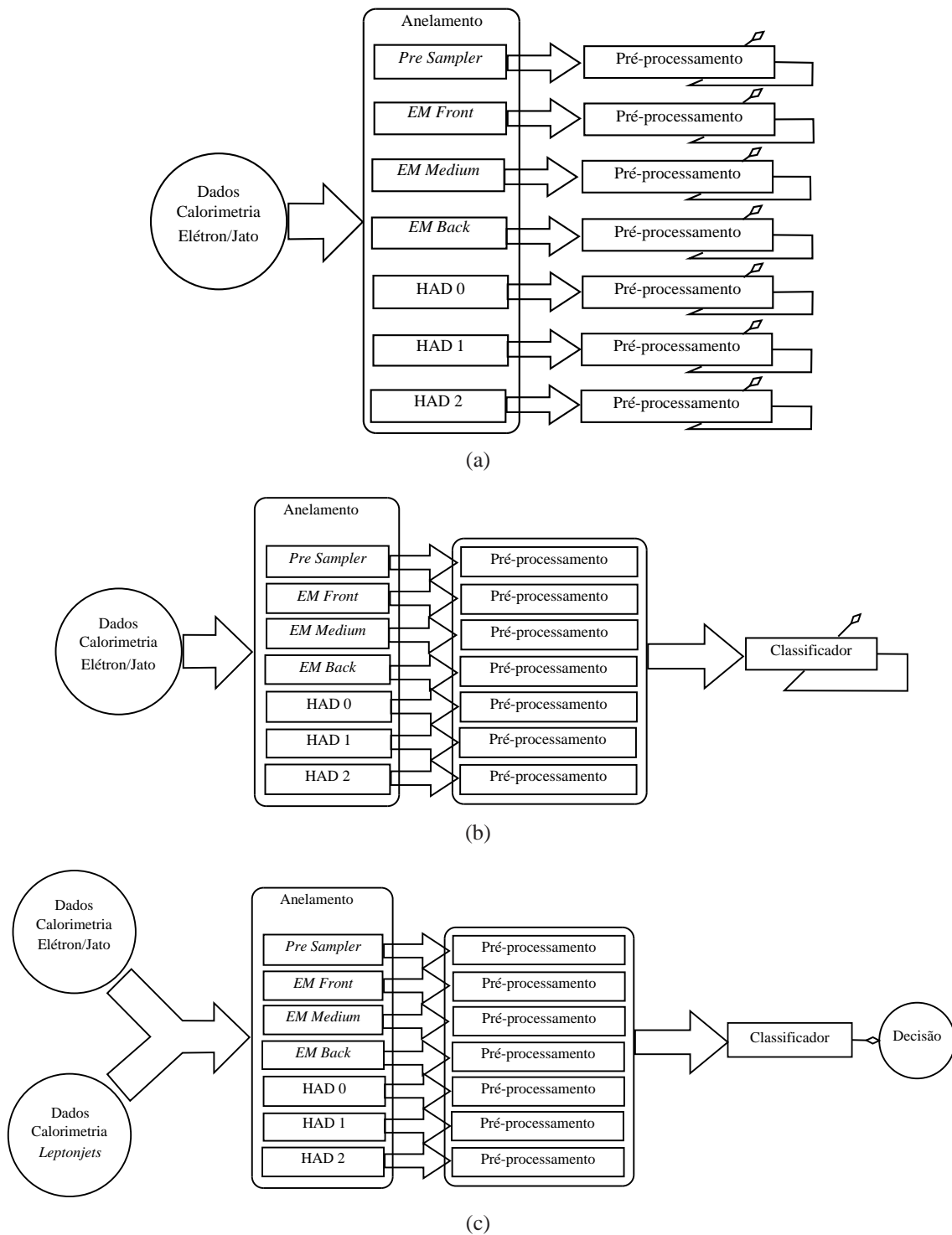


Figura 4.18: Procedimento segmentado para o canal elétron/jato: (a) Treinamento dos algoritmos de pré-processamento; (b) Treinamento do algoritmo de classificação; (c) Teste de hipótese.

mesmos testes discutidos anteriormente, com a mudança de elétrons por *leptonjets* em cada uma das etapas de treinamento.

## 4.6 Desempenho dos algoritmos de referência

O teste de hipótese implementado pelo algoritmo de referência T2Calo emprega uma série de cortes lineares sobre diferentes descritores extraídos dos eventos em análise, como descrito na Seção 3.3. Os cortes são subsequentes, i.e., os eventos aceitos são assim rotulados depois de passarem pelo corte HAD  $E_T$ , que é aplicado aos eventos que passaram pelo corte EM  $E_T$ , que por sua vez é aplicado aos eventos que passaram pelo corte Rstrip e que é aplicado sobre eventos que passaram pelo primeiro corte, denominado Rshape. O seu desempenho é resumido pela Tabela 4.2. As Figuras 4.19(a), 4.19(c) e 4.19(e) mostram a saída do algoritmo para cada um dos eventos estudados: elétron, jatos hadrônicos e *leptonjets*, respectivamente.

Tabela 4.2: Desempenho do algoritmo T2Calo.

$PD_e$	$PF_j$	$SP_{e/j}$	$PD_{lj}$	$SP_{e/lj}$
95,49	10,43	0,9251	29,75	0,6736

Observa-se que o T2Calo é muito eficiente na tarefa de aceitar o objeto de interesse do canal, o elétron. Uma minoria desses eventos é rejeitada pelos cortes Rshape e EM  $E_T$  (especificamente 4,51%). Ao rejeitar jatos hadrônicos, o T2Calo também mostra suas virtudes. Boa parte dos eventos são rejeitados já no primeiro corte (Rshape). A variável Rshape, ao calcular a razão energética de duas regiões de tamanho diferente em torno do centro da RoI, consegue rejeitar 69,21% dos jatos que alcançam o nível 2 de filtragem, enquanto um percentual menor de jatos é rejeitado pelos cortes subsequentes. Do total de jatos recebidos pelo nível 2, pouco mais de 10% é aceito pelo T2Calo e submetido ao nível superior de filtragem. De posse desses números reconhece-se no T2Calo um algoritmo com desempenho expressivo (resumido no produto  $SP_{e/j}$  de 0,9251) e apto a operar nesse canal do detector.

Já a investigação do potencial de filtragem de eventos exóticos, por parte do algoritmo de referência, revela que boa parte dos *leptonjets* (46,29%) é rejeitada também pelo corte Rshape. O corte Rshape mostra que, apesar de não ser tão punitivo quanto fora para os jatos hadrônicos, a razão energética calculada pela energia depositada na segunda camada eletromagnética é uma característica que aproxima jatos hadrônicos de *leptonjets*. O corte Rstrip, que capta a diferença entre máximos picos de deposição de energia a partir da informação disponível na primeira camada eletromagnética (a mais fina), também é bastante punitivo para a aceitação de *leptonjets*, respondendo por 15,68% de rejeição.

Esses valores acarretam uma aceitação de menos de 30% dos eventos exóticos (considerado baixo para os pesquisadores da área) e resumem o desempenho do T2Ca1o em um produto  $SP_{e/j/lj}$  de 0,6736.

As Figuras 4.19(b), 4.19(d) e 4.19(f) detalham o desempenho do algoritmo em função da energia transversa ( $E_T$ ), da coordenada  $\eta$  e da coordenada  $\phi$  do evento, respectivamente. Percebe-se que o desempenho do algoritmo para *leptonjets* apresenta uma tendência positiva em função do aumento da energia transversa do evento. O falso alarme também cresce com o aumento da energia transversa, contudo, de maneira menos acentuada, o que acaba não sendo um grande problema. Nessa condição, acredita-se que os dois eventos sejam mais colimados, o que aproxima suas características de deposição de energia da dos elétrons isolados. Pelo gráfico que mostra o desempenho em função de  $\eta$ , é visível que o T2Ca1o apresenta uma baixa na detecção de elétrons em torno da região de *crack*. Esse resultado já era esperado, tendo em vista que essa região apresenta dificuldades de amostragem do sinal, como descrito no Capítulo 3.

Como os cortes são aplicados de maneira serial, a única maneira de determinar quão eficientes eles são é através da submissão em separado de todos os eventos a cada um dos cortes (Figura 4.20). As Figuras 4.20(a), 4.20(b) e 4.20(c) apresentam o mecanismo de seleção dos cortes do T2Ca1o em relação a cada evento: elétron, jatos e *leptonjets*, respectivamente. Independentemente da ordem de aplicação, os cortes do T2Ca1o não diminuiriam o resultado alcançado na aceitação do elétron. No caso dos jatos, o que se verifica é que todos os cortes são capazes de eliminar, pelo menos, 30% dos eventos que falseiam o elétron. Outro destaque, que não é perceptível pelo ordenamento imposto pelo T2Ca1o, é o percentual de jatos que podem ser rejeitados através da observância da energia depositada nas camadas hadrônicas, como mostra o corte HAD  $E_T$ . De maneira geral, todas as variáveis criadas pelo T2Ca1o são relevantes para o bem sucedido processo decisório implementado. Contudo, essas mesmas variáveis indicam que além dos 45% de *leptonjets* rejeitados pelo corte Rshape, quase 35% não passariam do nível 2 de filtragem em virtude de sua semelhança com os jatos no que diz respeito aos múltiplos picos de deposição energética dentro da RoI (Rstrip). Os outros 22% não são aceitos devido ao nível de energia depositada nas camadas eletromagnéticas (EM  $E_T$ ), enquanto 17% são devido ao nível de energia depositada nas camadas hadrônicas (HAD  $E_T$ ). Esses resultados são importantes pois, além de detalhar o algoritmo de referência de filtragem de nível 2 do canal elétron/jato, servem para analisar as vantagens e desvantagens das abordagens propostas no trabalho.

Além do T2Ca1o, outro algoritmo de referência é o RingerNeuralHypo, também pertencente ao pacote de algoritmos de filtragem de nível dois do detector ATLAS [92]. Aqui, uma rede neural artificial realiza o teste de hipótese sobre os dados dos eventos formatados em anéis (100, no total). Os resultados desse algoritmo são detalhados na Seção 5.2.

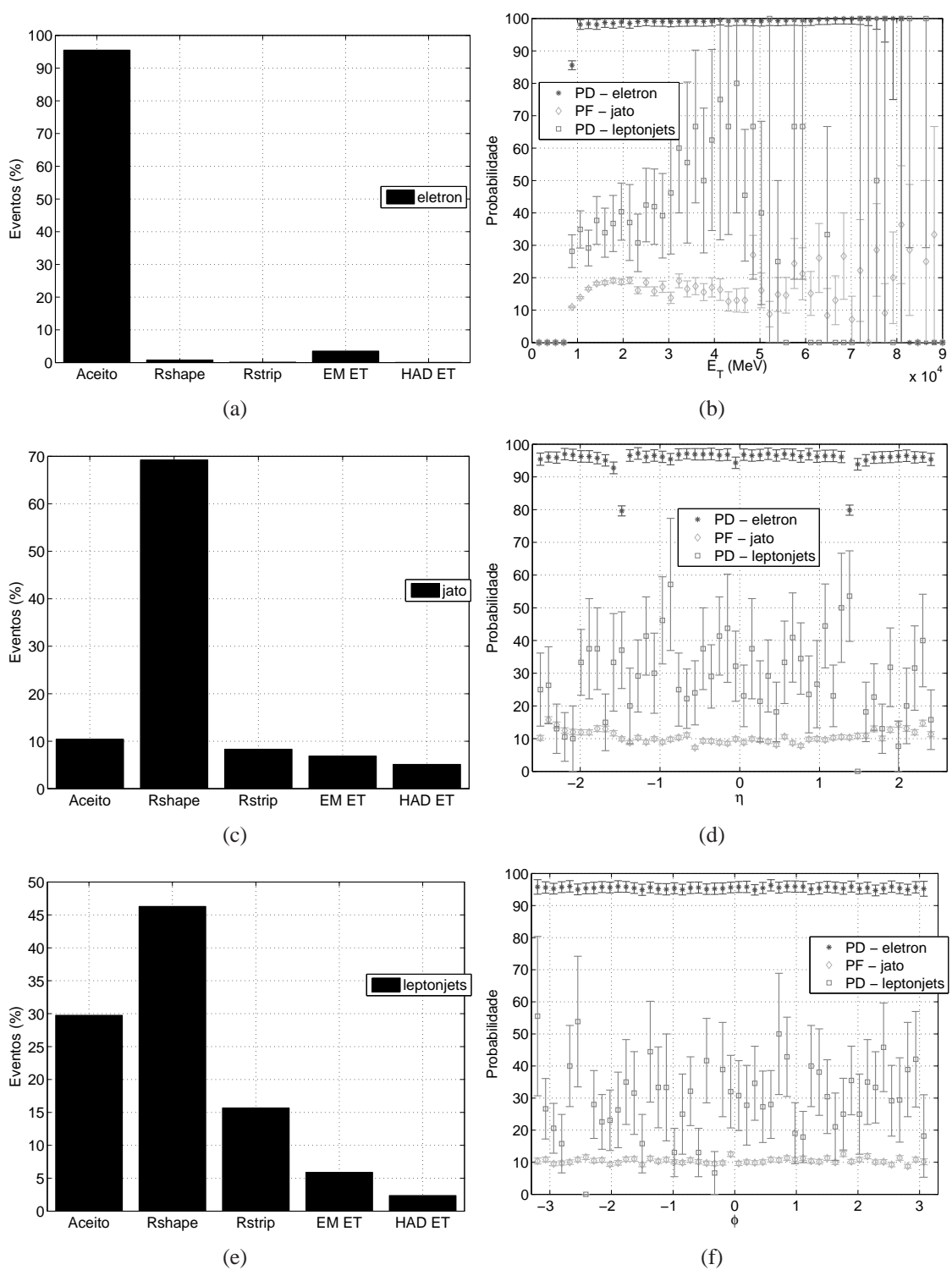
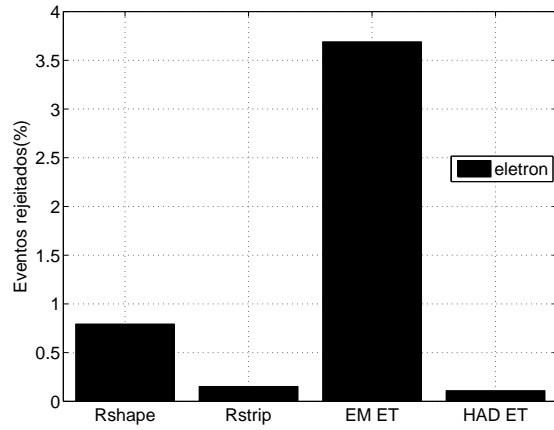


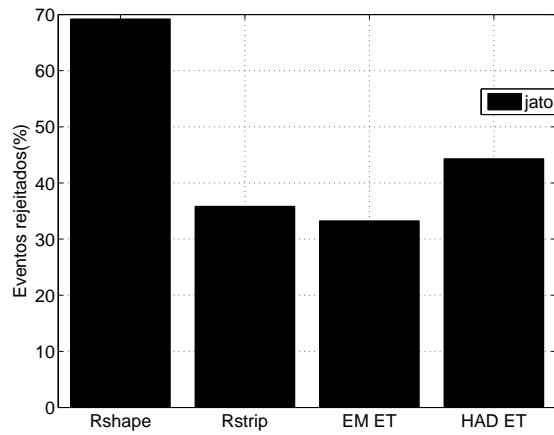
Figura 4.19: Desempenho do algoritmo T2Calo. (a) Histograma de saída para elétrons. (b) Probabilidades de detecção e falso alarme em função da  $E_T$ . (c) Histograma de saída para jatos. (d) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (e) Histograma de saída para *leptonjets*. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.



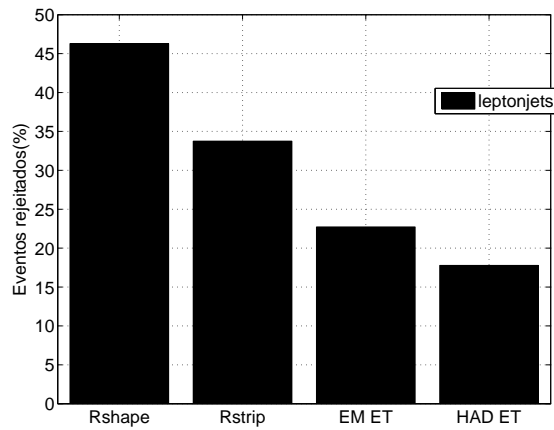
Algoritmos e abordagens com desempenho superior ao T2Ca1o vêm sendo propostos [48, 49, 51, 53, 97–101], contudo a implementação final ainda não está definida. Uma das contribuições do trabalho é exatamente nessa linha, com o diferencial da busca de aplicações com desempenho significativo também para o caso de objetos eletromagnéticos não isolados, como os *leptonjets*. A literatura mostra algumas publicações nesse sentido [68, 69, 102–104].



(a)



(b)



(c)

Figura 4.20: Desempenho independente dos cortes do algoritmo T2Ca1o. (a) Histograma de saída para elétrons. (b) Histograma de saída para jatos. (c) Histograma de saída para *leptonjets*.

## Capítulo 5

# Resultados da Filtragem *online* de Nível 2 - canal elétron/jato

Este capítulo apresenta os resultados dos procedimentos de filtragem de segundo nível propostos, todos a partir da informação de calorimetria disponibilizada pelo primeiro nível de filtragem. Comparações são feitas entre as diferentes implementações e as figuras de mérito, que são algoritmos de referência do detector ATLAS. A metodologia seguida na tarefa de classificação é dividida em operações sobre anéis e extração de características destes. A extração de característica, por sua vez, é implementada de maneira local e total, como comentado no Capítulo 4.

As técnicas de extração de características fazem uma prospecção sobre um conjunto de dados de elétrons e jatos pertinentes a esse canal de monitoramento do detector ATLAS. A partir da transformação obtida, que geralmente implica em uma redução de dimensionalidade, dados de elétrons e *leptonjets* são submetidos ao teste de hipótese, todos contra jatos hadrônicos.

### 5.1 *Menu* e10

A infraestrutura atual do detector é utilizada por diferentes estudos que buscam uma solução experimental para problemas teóricos bem ou mal compreendidos. O ferramental tecnológico (*Hardware* e *Software*) disponível, e em fase de desenvolvimento, apresenta boa flexibilidade quanto aos recursos de implementação. Considerando as características dos dados disponíveis (apresentadas no Capítulo 4) que, além de escassos no caso dos *leptonjets*, não são coincidentes em toda a faixa de energia coberta, optou-se pela seleção da assinatura e10 do *menu* de seleção do canal. Ela permite um maior número de eventos disponíveis para o nível 2 de filtragem, já que conta com um corte de primeiro nível menos rígido (corte baseado em energia com valor de corte baixo). A presença de um número maior de eventos para o nível seguinte de filtragem pode ser melhor explorada pelas

abordagens e algoritmos aqui implementados. Essa assinatura também indica valores de corte para cada uma das variáveis do T2Ca1o (algoritmo de filtragem *online* de nível 2) e determina seu desempenho, mostrado detalhadamente no Capítulo 4.

## 5.2 Anéis

A classificação dos dados nesta abordagem não emprega qualquer técnica de pré-processamento aos dados formatados e normalizados. Esses dados são apresentados ao algoritmo de classificação que infere o valor de saída, associando-o com uma das classes analisadas. O objetivo aqui é avaliar o poder de discernimento que os anéis fornecem aos algoritmos de classificação, mesmo em condições de dimensão elevada de dados. A variante neural dessa abordagem constitui o algoritmo de segundo nível de filtragem do ATLAS, RingerNeuralHypo, concorrente do T2Ca1o.

As similaridades e peculiaridades na deposição de energia por parte dos eventos de interesse para o canal determinam o potencial mínimo de separação obtido quando o algoritmo de classificação é um algoritmo linear. O algoritmo realiza a decisão de classificação com base na combinação linear das características dos eventos presentes na amostra analisada. Além disso, a partir desses resultados é possível determinar o ganho obtido com procedimentos de classificação mais complexos.

Tabela 5.1: Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	$PD_e$	$PF_j$	$SP_{e/j}$	limiar	$PD_{lj}$	$SP_{e/j/lj}$
Energia	98,19	2,64	0,9777	0,31	6,27	0,5132
Sequencial	98,03	3,36	0,9734	-0,01	14,26	0,5978
Camada	97,67	3,15	0,9726	0,35	11,60	0,5735

A Tabela 5.1 mostra o resultado do classificador linear operando sobre a base de dados normalizada de três diferentes maneiras. Os valores apresentados para dados do canal são melhores do que os alcançados pelo T2Ca1o, principalmente no que se refere ao falso alarme. Contudo, a detecção de *leptonjets* fica abaixo da alcançada pelo algoritmo de referência. Com dados formatados em anéis e sequencialmente normalizados, obteve-se o melhor resultado. A Figura 5.1 agrupa as diferentes medidas de desempenho obtidas para esse classificador. Os histogramas de saída (Figura 5.1(a)) mostram a separação alcançada pelos padrões. A partir dele obtém-se as curvas ROC (Figura 5.1(b)), com as respectivas probabilidades de detecção e falso alarme. O produto  $SP_{e/j}$  e a eficiência na detecção de *leptonjets*, ambos em função do limiar de saída do algoritmo, são mostrados na figura 5.1(c). O melhor valor de produto  $SP_{e/j}$  determina o limiar de decisão ideal que resulta em diferentes probabilidades de detecção e falso alarme em função da  $E_T$  (figura 5.1(d)), e das coordenadas  $\eta$  (figura 5.1(e)) e  $\phi$  (figura 5.1(f)) do detector.

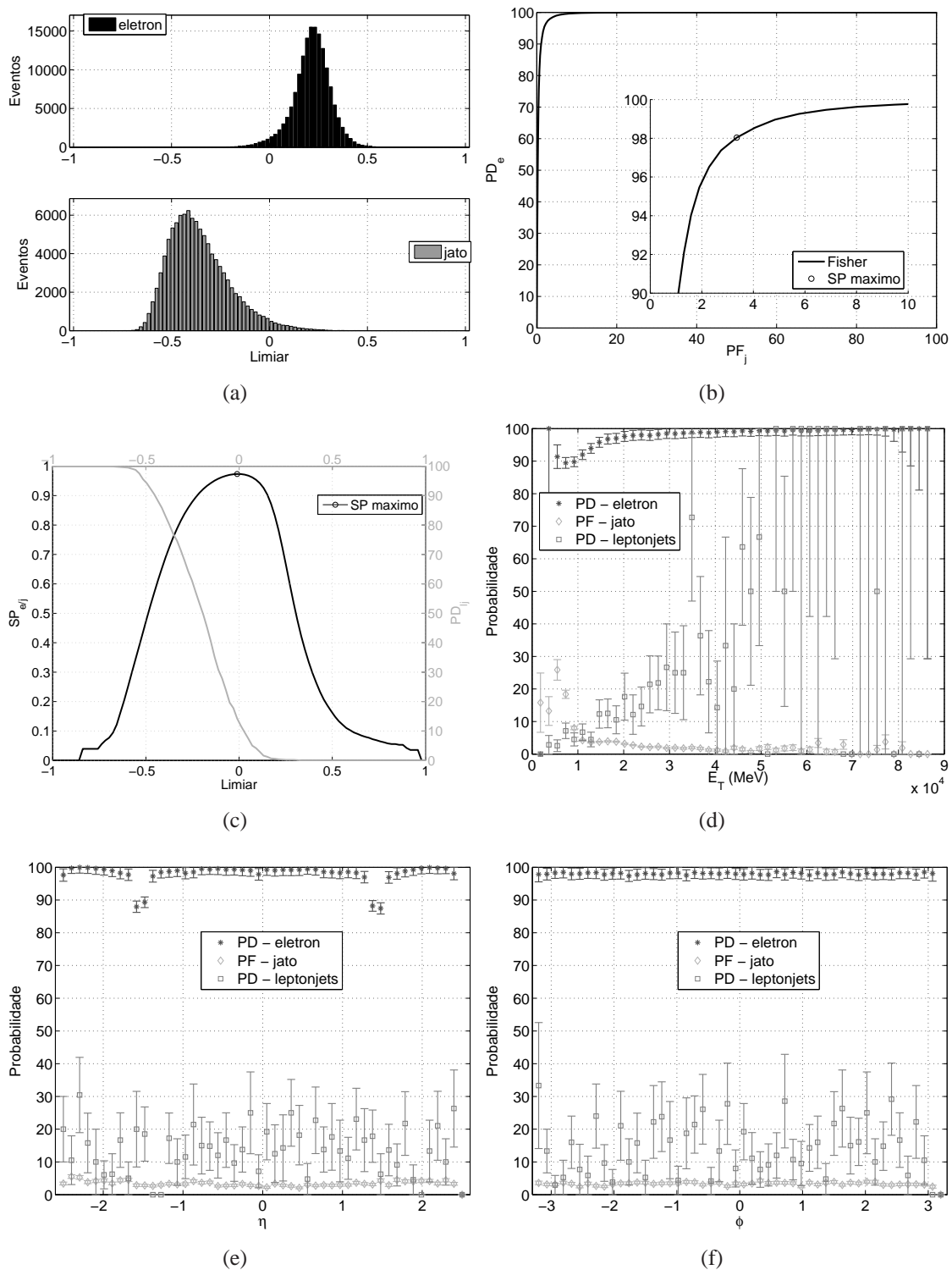


Figura 5.1: Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis e normalizados sequencialmente. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e/j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

Espera-se uma eficiência praticamente uniforme para as duas coordenadas, contudo é fato que a sensibilidade em  $\eta$  é afetada na região de *crack* ( $|\eta| \approx 1,5$ ), um espaço sem elementos de detecção, por onde passam cabos de leitura e manutenção dos detectores internos [75]. O classificador linear obtido mostra que é afetado por esse efeito, mas o resultado praticamente linear em  $\phi$  o afasta de alguma tendência.

A avaliação do potencial do classificador passa também pela análise da relevância da informação que lhe é fornecida. Nesse caso, a relevância dos 100 anéis é ilustrada na Figura 5.2. Os valores obtidos mostram que a informação presente nos primeiros anéis da primeira camada eletromagnética é determinante para a classificação de elétrons. Outra fonte de informação relevante para a separação de elétrons encontra-se nos anéis da primeira camada hadrônica. Essa característica também era prevista, dado que os elétrons depositam pouca ou nenhuma energia nessa seção, i.e., a baixa amplitude de energia depositada nessa região é determinante no processo de identificação de elétrons e/ou discernimento do ruído de fundo. Para os jatos, o segundo anel da segunda camada eletromagnética apresenta-se como o principal caracterizador do evento. Uma hipótese que explicaria esse resultado seria a de que sendo essa camada a mais profunda, as componentes eletromagnéticas presentes no jato depositariam aqui um percentual discriminante mínimo de energia. O mesmo acontece para os *leptonjets*, a informação que mais ajuda na construção da estrutura interna de decisão do classificador está presente nesse mesmo anel. Um outro destaque, são os anéis centrais da primeira camada eletromagnética. Acredita-se que, sendo essa a camada do calorímetro de maior resolução espacial, os objetos de interesse (principalmente os eletromagnéticos) tenham seus detalhes topológicos, aqui, melhor caracterizados. Além disso, observa-se que para os três eventos, vários anéis apresentam relevância com valores intermediários, seja na parte eletromagnética seja na parte hadrônica do calorímetro. Isso é importante pois o compartilhamento da informação relevante para o classificador dentre as várias variáveis de entrada torna o procedimento mais robusto do ponto de vista de resistência a falhas em alguns de seus sensores de leitura.

Os dados formatados em anéis também foram submetidos a um classificador não-linear; aqui, uma rede neural. Essa implementação representa o *RingerNeuralHypo*, algoritmo de referência de nível 2 mencionado no Capítulo 3. A topologia da rede neural usada no teste de hipótese não está definida, de maneira que testes com variação de apenas um dos seus parâmetros livres (números de neurônios na camada escondida da rede) foram implementados. A topologia da rede foi alterada significativamente no que se refere ao número de nós da camada intermediária, de um mínimo de 2 neurônios ao máximo de 15. Os valores alcançados a cada aumento do número de neurônios são indicados na Tabela 5.2.

Constata-se que para qualquer uma das normalizações escolhidas o resultado pouco se altera, mesmo nas condições mínima e máxima de topologia da rede. Pode-se justifi-

Tabela 5.2: Desempenho do classificador RingerNeuralHypo operando sobre dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Neurônios	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
Energia	2	99,00	2,13	0,9843	0,22	7,41	0,5320
	3	99,11	2,13	0,9849	0,26	7,98	0,5397
	4	99,10	2,14	0,9848	0,20	8,56	0,5466
	5	99,09	2,00	0,9855	0,22	8,08	0,5411
	6	99,08	1,98	0,9855	0,28	8,56	0,5469
	7	99,06	1,91	0,9857	0,30	7,98	0,5400
	8	99,04	1,96	0,9854	0,28	7,51	0,5338
	9	99,09	2,03	0,9853	0,22	7,70	0,5362
	10	99,19	2,07	0,9856	0,20	8,27	0,5435
	11	99,09	1,94	0,9857	0,26	8,08	0,5412
	12	99,11	2,03	0,9854	0,26	7,60	0,5350
	13	99,04	1,82	0,9861	0,34	7,03	0,5277
	14	99,18	2,03	0,9858	0,22	7,13	0,5288
	15	99,19	2,01	0,9859	0,18	8,17	0,5425
	Sequencial	2	98,33	3,07	0,9763	0,24	13,21
3		98,52	3,19	0,9766	0,16	12,07	0,5799
4		98,50	3,09	0,9770	0,24	12,55	0,5846
5		98,63	2,97	0,9783	0,24	12,93	0,5886
6		98,62	2,93	0,9785	0,18	12,93	0,5887
7		98,67	2,75	0,9796	0,24	11,31	0,5740
8		98,55	2,66	0,9794	0,26	9,98	0,5603
9		98,94	2,89	0,9802	0,08	11,69	0,5780
10		98,83	2,89	0,9797	0,16	12,17	0,5823
11		98,73	2,64	0,9804	0,20	9,79	0,5587
12		98,92	2,87	0,9803	0,06	10,36	0,5647
13		98,84	2,88	0,9798	0,14	11,03	0,5713
14		98,84	2,76	0,9804	0,18	10,08	0,5618
15		99,01	2,78	0,9811	0,06	10,84	0,5700
Camada		2	98,76	2,43	0,9816	0,18	12,36
	3	98,90	2,54	0,9818	0,04	12,17	0,5833
	4	98,77	2,39	0,9819	0,14	11,69	0,5788
	5	98,86	2,19	0,9833	0,18	11,60	0,5786
	6	98,91	2,15	0,9838	0,16	11,31	0,5760
	7	98,84	2,05	0,9840	0,24	11,03	0,5733
	8	98,85	2,11	0,9837	0,20	9,89	0,5612
	9	98,91	2,01	0,9845	0,22	10,93	0,5726
	10	98,98	2,26	0,9836	0,10	11,69	0,5796
	11	99,03	2,17	0,9843	0,10	11,12	0,5743
	12	99,02	2,29	0,9836	0,08	12,26	0,5851
	13	98,87	2,09	0,9839	0,18	10,27	0,5654
	14	98,91	2,21	0,9835	0,14	11,69	0,5796
	15	98,98	2,19	0,9839	0,12	11,12	0,5742

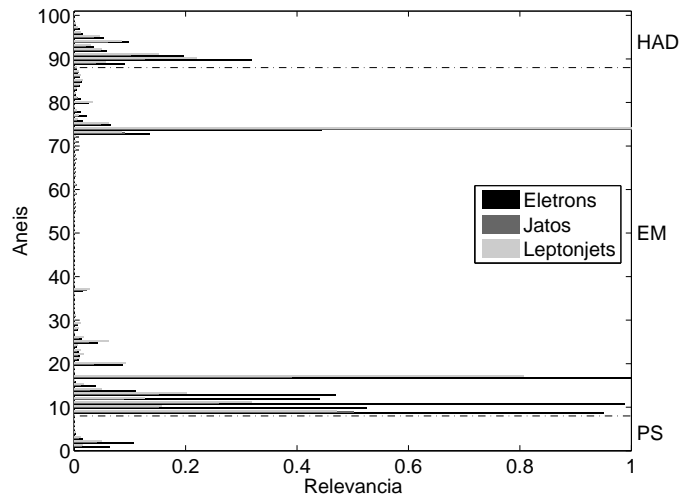


Figura 5.2: Relevância dos anéis normalizados sequencialmente para o processo decisório implementado pelo classificador linear de Fisher.

car esse resultado a partir dos parecidos perfis de deposição de cada classe de evento. A variação na estrutura de decisão interna alcançada pelo algoritmo não é sentida significativamente, i.e, a limitação imposta pela modelagem dos dados não pode ser contornada mesmo pelas modificações alcançadas entre as diferentes topologias dos classificadores.

Em todos os casos os valores para o canal elétron/jato são superiores ao obtido pelo T2Ca1o, sendo esse um dos motivos da incorporação do RingerNeuralHypo no conjunto de *softwares* do detector ATLAS. O ponto fraco da técnica está na detecção dos eventos exóticos. Os valores obtidos estão abaixo tanto do T2Ca1o quanto de patamares desejados pelos grupos de pesquisa interessados nesse tipo de evento.

A Figura 5.3 apresenta o comportamento das diferentes medidas de desempenho usadas neste trabalho em função do número de nós intermediários de cada classificador neural testado, todos alimentados com anéis sequencialmente normalizados (normalização que resultou no melhor classificador). Especificamente, a Figura 5.3(a) apresenta a variação do erro final de treinamento para 5 diferentes condições de inicialização. As marcações representam os menores, os maiores e os valores medianos de erro quadrático médio alcançados ao fim do treinamento. É notório que o aumento do número de nós intermediários proporcionou uma maior aproximação das saídas da rede às saídas desejadas, como mostra a trajetória da curva. O aumento do número de nós na camada intermediária da rede, aumenta o número de parâmetros livres ajustáveis ao problema tratado e, por conseguinte, ajustes mais finos na estrutura de decisão podem ser alcançados. A Figura 5.3(b) acompanha o comportamento do produto  $SP_{e/j}$  associado à melhor das redes finais obtidas. O comportamento da curva de treino sempre acima da curva de teste já era esperado tendo em vista que esses foram os dados com os quais a rede lidou diretamente no ajuste dos parâmetros livres, porém nota-se uma tendência de crescimento, o que evidencia a capacidade de generalização do algoritmo para dados não observados. As Figuras 5.3(c)



e 5.3(d) ilustram o comportamento das probabilidades de detecção e falso alarme, respectivamente. O comportamento dos diferentes classificadores para eventos exóticos é mostrado nas Figuras 5.3(e) e 5.3(f), com o produto  $SP_{e/j/lj}$  e a probabilidade de detecção  $PD_{lj}$ .

Considerando as três classes de eventos, o melhor valor de produto  $SP_{e/j/lj}$  alcançado entre os classificadores (0,5902) foi obtido com apenas 2 neurônios na camada intermediária da rede, o que implica 98,33% de detecção de elétrons, 3,07% de falso alarme e 13,21% de *leptonjets* aceitos. A Figura 5.4 apresenta o desempenho desse classificador. Os histogramas de saída da rede (Figura 5.4(a)) mostram uma excelente separação entre elétrons e jatos. A curva ROC é mostrada na Figura 5.4(b). O produto  $SP_{e/j}$  e a eficiência na detecção de *leptonjets* em função do limiar de saída da rede são mostrados na Figura 5.4(c). Já as probabilidades de detecção e falso alarme em função  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector ATLAS estão apresentadas nas Figuras 5.4(d), 5.4(e) e 5.4(f), respectivamente.

Diferentemente do que fora observado com o classificador linear, aqui, uma pequena variação do limiar de saída da rede pouco altera a eficiência do canal elétron/jato. Isso mostra robustez, pois está diretamente ligado à separação dos padrões. Infelizmente, isso implica um difícil ajuste, caso se deseje priorizar a seleção de *leptonjets*. Conclui-se que o RingerNeuralHypo é ótimo como algoritmo especialista do canal, superior ao T2Calo, mas com pouco destaque para a seleção de eventos exóticos.

Um ponto a ser ressaltado é quanto ao ganho de eficiência em  $\eta$  se comparado com o classificador linear anteriormente apresentado. O classificador neural consegue contornar, significativamente, o problema da perda de eficiência na região de *crack*, apresentando uma probabilidade de detecção praticamente uniforme em toda essa coordenada.

Observa-se ainda pela Figura 5.4(d) que o aumento da energia dos eventos implica aumento do falso alarme de jatos e da detecção de *leptonjets*. Uma explicação para esses dois movimentos é que o aumento da energia torna esses dois eventos mais colimados, ou seja, mais próximos dos elétrons isolados, mimetizando-o.

A relevância dos anéis (Figura 5.5(b)) para o classificador neural mostra principalmente a redução do compartilhamento da importância, no que se refere aos eventos da classe elétron, em relação ao apresentado pelo classificador linear. Enquanto com o classificador linear observava-se vários picos de relevância nas diferentes camadas (Figura 5.2), com o classificador neural destacam-se apenas dois, correspondentes aos anéis 1 e 9 da primeira camada eletromagnética. Anéis da camada hadrônica tiveram sua relevância minorada em relação ao classificador linear. Isso indica que o classificador neural concentrou relevância em poucos anéis; contudo, através de mapeamento não-lineares, foi capaz de extrair o máximo de rendimento dessas variáveis, como mostra o valor final de  $PD_e$ . Para os jatos, os anéis mais relevantes não sofreram alteração (EM *middle layer*), contudo ganham destaque uma maior contribuição dos primeiros anéis da primeira ca-

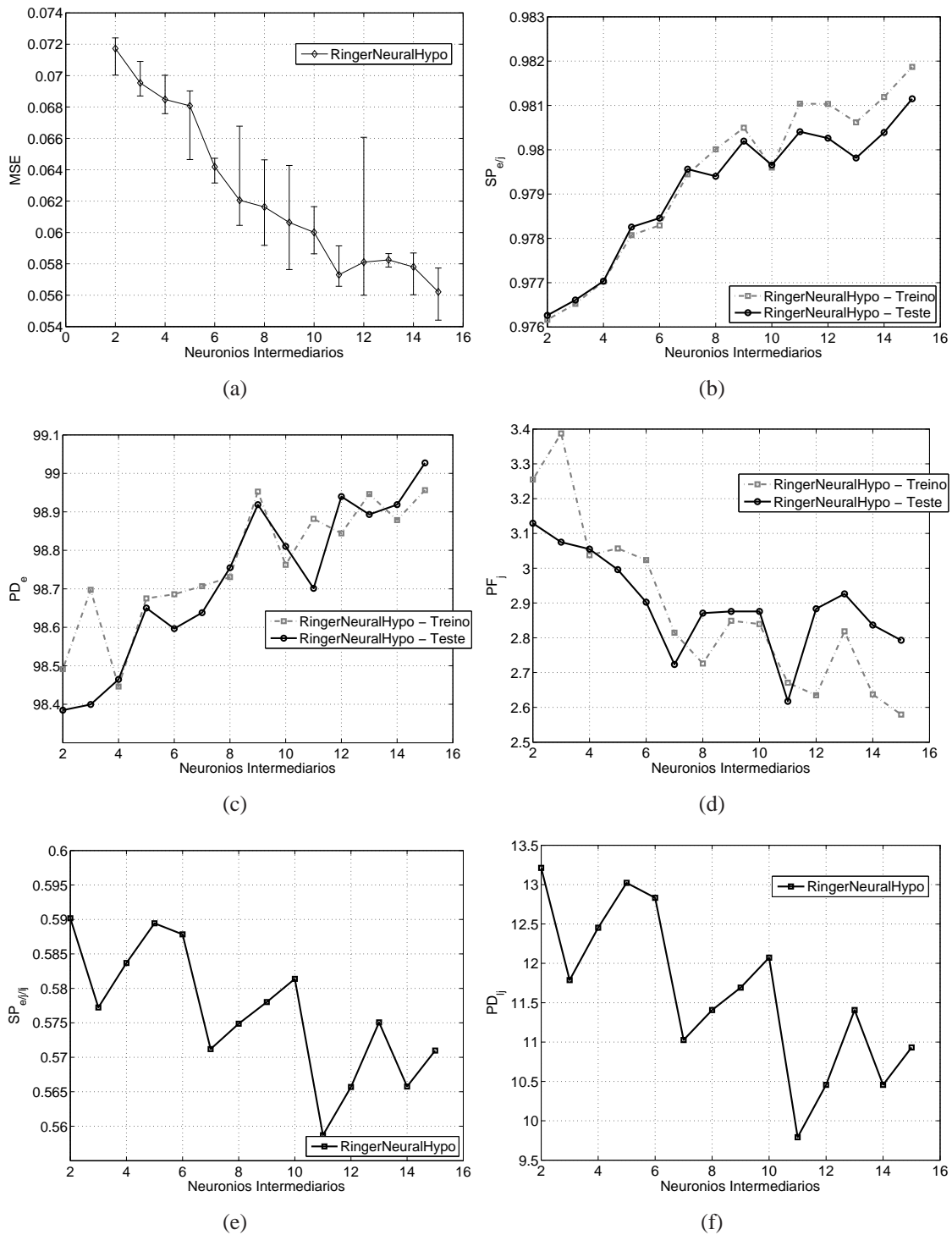


Figura 5.3: Variação de desempenho do RingerNeuralHypo operando sobre dados formatados em anéis e normalizados sequencialmente em função do número de neurônios na camada intermediária da rede: (a) Erro quadrático médio final de treinamento para diferentes condições de inicialização do algoritmo neural; (b) Produto  $SP_{e/j}$  associado à melhor rede; (c) Probabilidade de detecção  $PD_e$  associada à melhor rede; (d) Probabilidade de falso alarme  $PF_j$  associada à melhor rede; (e) Produto  $SP_{e/j/l_j}$  associado à melhor rede; (f) Probabilidade de detecção  $PD_{l_j}$  associada à melhor rede.

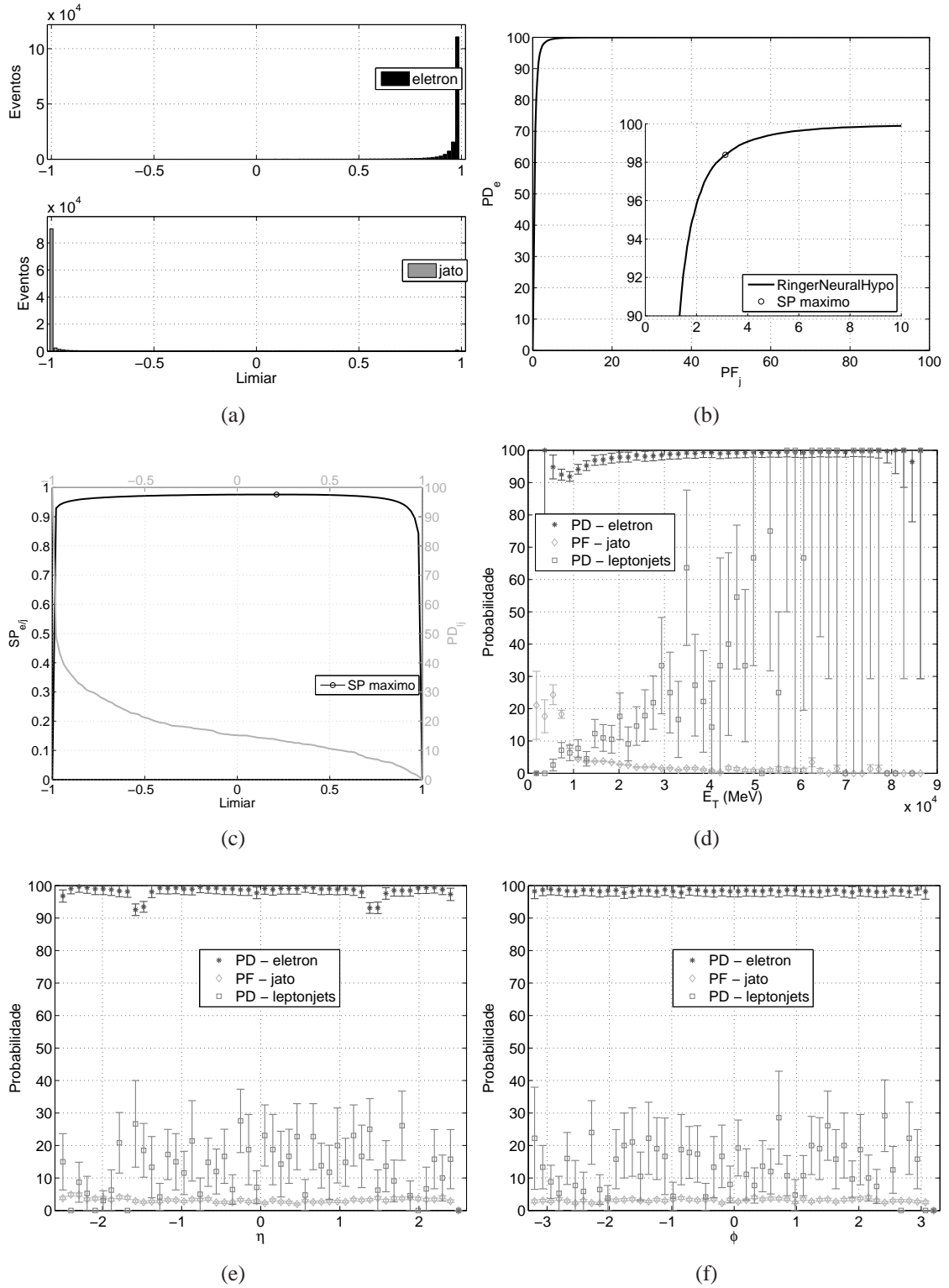


Figura 5.4: Desempenho do RingerNeuralHypo operando sobre dados formatados em anéis e normalizados sequencialmente. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e_j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

mada eletromagnética. Como já mencionado, essa primeira camada eletromagnética é a de maior resolução espacial e tal característica (importante na descrição do decaimento da partícula) pode ser potencializada pelo classificador neural na identificação desse evento. No caso dos *leptonjets* a informação que mais ajuda na construção da estrutura interna de decisão do classificador está presente no anel 9 da primeira camada eletromagnética (o mesmo é observado com o evento elétron). Essa é uma alteração significativa em relação ao classificador linear, cuja informação mais importante vinha da segunda camada eletromagnética. Dessa forma, o `RingerNeuralHypo` consegue colher de um mesmo anel a informação mais relevante para a caracterização tanto de elétrons quanto de *leptonjets*, um ponto bastante positivo. O restante da barra de relevância dos diferentes anéis para os *leptonjets* foi a que menos se modificou na comparação entre os dois tipos de classificadores estudados. Por fim a Figura 5.5(a) mostra o comportamento do algoritmo durante o seu treinamento.

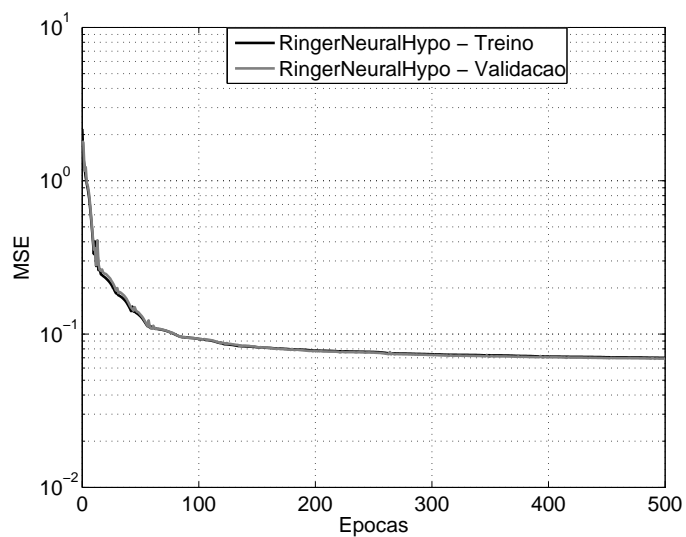
Os teste seguintes se referem a como manipular a informação dos anéis aqui analisados, ora de maneira não segmentada, ora de maneira segmentada, sempre do ponto de vista das camadas dos calorímetros. Para isso, são empregadas diferentes técnicas de análise de componentes.

## 5.3 Pré-processamento não segmentado

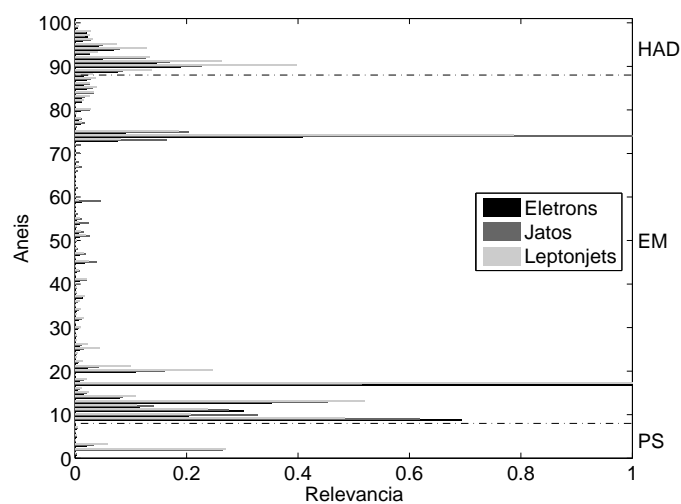
O pré-processamento não segmentado dos dados permite às técnicas de análise de componentes atuar sobre os dados de calorimetria sem levar em consideração as suas subdivisões e, de alguma maneira, realizar uma busca por correlação e redundância na informação entre camadas. Apesar dos dados de calorimetria estarem segmentados em camadas, a interpretação da informação como um todo é uma estratégia interessante na busca da melhor representação desses dados em um espaço de dimensão inferior. Espera-se nessa abordagem um número menor de componentes retidas em função da qualidade do sinal transformado.

### 5.3.1 PCA

Considerando que a PCA realiza compactação de sinais e a representação deles a partir de componentes de maior energia, a observação do comportamento da curva de carga (Figura 5.6) para cada um dos eventos de interesse mostra quão redundante é a informação presente no sinal, assim como o efeito de cada uma das normalizações sobre esses dados. Os elétrons são melhor representados com um número menor de componentes do que os jatos e os *leptonjets*, dado um mesmo percentual de energia retida do sinal quando normalizado por energia. Isso se explica pela característica de deposição de energia dos eventos dessa classe, tida como mais pontual. Já os jatos hadrônicos tendem a espalhar

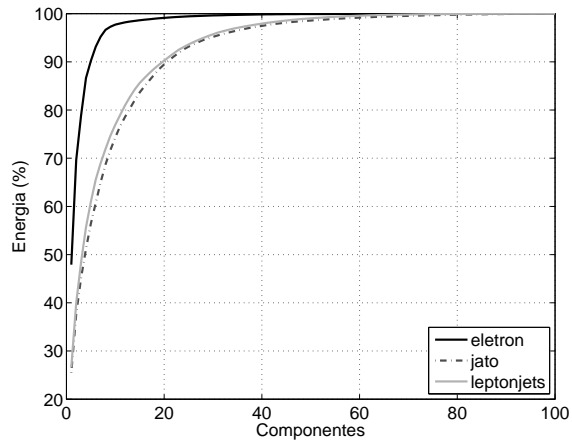


(a)

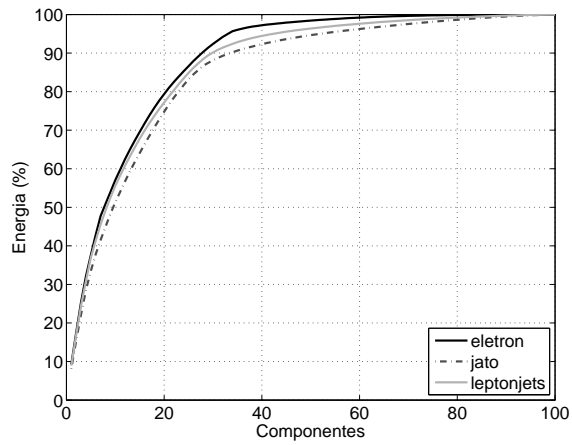


(b)

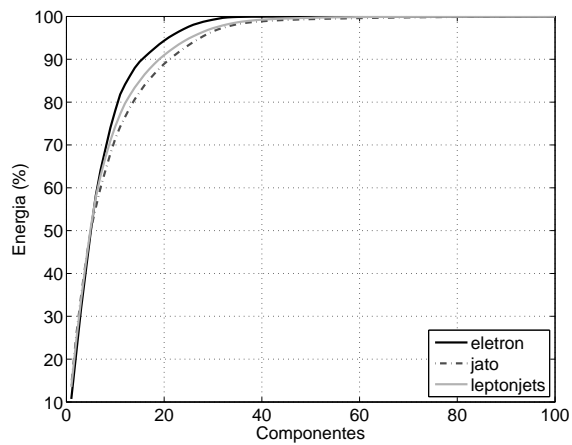
Figura 5.5: (a) Erro quadrático médio de treinamento do algoritmo neural. (b) Relevância dos anéis normalizados sequencialmente para o processo decisório implementado pelo classificador neural.



(a)



(b)



(c)

Figura 5.6: Curvas de carga: (a) eventos normalizados por energia; (b) eventos normalizados sequencialmente; (c) eventos normalizados por camada.

sua energia em uma vizinhança, o que requer um maior número de componentes para sua representação. Os *leptonjets* estão entre essas duas categorias, são elétrons, porém um número grande deles está em uma mesma RoI (mesmo que muito próximos) implicando um maior número de componentes retidas para a sua correta representação. Como a normalização por energia aplica apenas um normalizador sobre os dados, essas características são mantidas.

A normalização sequencial distribui a informação mais importante para a representação do evento entre mais anéis, i.e., a curva de carga cresce mais lentamente e mais componentes são necessária para uma adequada representação da partícula. Além disso, observa-se que essa característica se expressa para os três eventos, diferentemente do que ocorre com a normalização por energia. Ainda assim, a curva dos elétrons cresce mais rapidamente do que a dos jatos, como era esperado.

No caso da normalização por camada, nota-se que a curva de carga cresce tão rapidamente quanto o observado com a normalização por energia. Contudo, ela preserva uma representatividade aproximadamente igualitária (alcançada pela normalização sequencial) entre os eventos, dado um percentual de energia retida.

Nesse ponto é preciso ressaltar a diferenciação entre dois conceitos tratados ao longo dos resultados. A energia retida do sinal (medida em %) não tem relação direta com a energia (medida em MeV) do evento, trata-se do percentual de energia do sinal formatado em anéis e normalizado das diferentes maneiras propostas.

Desta forma, é possível dizer que com as normalizações sequencial e por camada, a retenção de 10 componentes principais representa, aproximadamente, 55% e 75% dos três eventos analisados, enquanto o mesmo número de componentes na normalização por energia significa uma representatividade quase total de elétrons e de 75% para jatos e *leptonjets*.

A estratégia seguida visa reter um número específico de componentes principais de maneira que nelas os *leptonjets* estejam o mais próximo (representativamente) possível dos elétrons pontuais e, ainda assim, haja uma diferenciação em relação aos jatos hadrônicos.

É importante ressaltar que a curva de carga aqui apresentada servirá de guia para as outras técnicas de análise de componentes, haja visto que o branqueamento dos dados e a redução de dimensionalidade, etapas comuns para ICA, estão diretamente associados com a PCA. Assim como na seção anterior, fez-se uso de um classificador linear e um não-linear nos testes. Em todos os teste foram retidos, aproximadamente, 80% da energia do sinal, o que representa uma grande compactação, seja qual for a normalização empregada.

A Tabela 5.3 quantifica o desempenho obtido com o classificador linear. É possível aferir que com um número pequeno de componentes principais a projeção dos *leptonjets* apresenta a similaridade desejada (mencionada anteriormente) em relação ao elétron. Os melhores resultados, independentemente da normalização, são alcançados com poucas

componentes retidas. Ganha destaque a classificação de elétrons e jatos com uso da normalização sequencial, enquanto a normalização por camada favorece a aceitação de *leptonjets*.

Tabela 5.3: Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Comp.	Energia(%) <sup>1</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
Energia	1	51,26	88,07	9,67	0,8919	-0,17	17,11	0,5790
	2	61,53	91,47	9,77	0,9085	-0,27	17,78	0,5922
	3	71,05	97,93	3,83	0,9705	0,27	11,41	0,5707
	4	75,28	98,20	3,88	0,9716	0,27	11,50	0,5721
	5	78,75	97,96	3,81	0,9707	0,11	9,79	0,5542
	6	82,14	97,46	3,30	0,9708	0,11	7,98	0,5334
Sequencial	1	11,34	92,58	12,58	0,8998	-0,11	44,39	0,7292
	2	19,38	92,70	12,50	0,9008	-0,11	44,49	0,7302
	3	26,53	92,37	11,70	0,9032	-0,13	43,16	0,7258
	4	32,58	92,39	11,74	0,9032	-0,13	43,06	0,7253
	5	38,60	92,32	11,37	0,9047	-0,05	42,78	0,7250
	6	43,27	94,84	9,86	0,9248	-0,11	38,40	0,7167
	7	47,50	95,75	10,15	0,9278	-0,19	37,36	0,7135
	8	50,92	95,38	9,76	0,9279	-0,17	36,03	0,7072
	9	53,84	94,93	9,29	0,9281	-0,17	34,41	0,6993
	10	56,57	94,94	9,58	0,9267	-0,21	34,41	0,6985
	11	59,23	94,05	8,85	0,9259	-0,21	32,89	0,6904
	12	61,68	95,08	9,46	0,9279	-0,21	34,03	0,6973
	13	63,89	95,34	9,72	0,9279	-0,21	34,60	0,7002
	14	65,98	95,06	9,29	0,9287	-0,19	34,32	0,6992
15	68,05	95,21	9,40	0,9289	-0,19	34,03	0,6979	
16	70,01	95,03	9,19	0,9291	-0,19	32,70	0,6912	
17	71,90	95,19	8,88	0,9315	-0,17	32,13	0,6897	
18	73,72	95,92	9,31	0,9329	-0,17	32,51	0,6924	
19	75,46	95,94	9,32	0,9329	-0,17	32,51	0,6924	
20	77,13	95,60	8,89	0,9334	-0,15	31,37	0,6868	
Camada	1	15,65	86,06	18,83	0,8360	-0,13	54,85	0,7332
	2	25,45	86,27	18,94	0,8365	-0,13	55,04	0,7343
	3	35,09	86,47	18,85	0,8379	-0,17	54,85	0,7344
	4	43,30	86,42	17,73	0,8433	-0,19	53,52	0,7326
	5	51,05	85,35	16,38	0,8448	-0,17	52,09	0,7279
	6	58,35	87,48	16,82	0,8532	-0,21	52,28	0,7338
	7	62,98	91,59	14,72	0,8841	-0,21	49,05	0,7396
	8	67,33	90,16	13,36	0,8839	-0,19	47,43	0,7327
	9	71,27	92,34	13,10	0,8960	-0,23	45,15	0,7302
	10	74,98	92,46	13,30	0,8956	-0,21	46,01	0,7336
	11	78,42	93,14	11,42	0,9085	-0,15	42,68	0,7268
	12	81,28	95,87	5,59	0,9514	0,01	21,39	0,6389

O valor de produto  $SP_{e/j/lj}$  obtido com a normalização por camada é o melhor dentre os valores observados, aproximando-se, inclusive, do obtido pelo algoritmo de referência T2Ca1o (Tabela 4.2) na detecção de elétrons e no falso alarme de jatos. O destaque da abordagem fica por conta do ganho de 20 pontos percentuais na detecção de *leptonjets*, quando observado os valores alcançados pelo T2Ca1o ou RingerNeuralHypo. Detalhes

<sup>1</sup>O percentual de energia se refere ao acumulado nas componentes principais.



da variação de desempenho (em função da energia retida do sinal) do algoritmo linear alimentado com componentes principais extraídas de anéis sob o efeito desse tipo de normalização são apresentados na Figura 5.7. O produto  $SP_{e/j}$  do canal é mostrado na Figura 5.7(a) e as respectivas taxas de eficiência na detecção de elétrons e falso alarme são mostradas nas Figuras 5.7(b) e 5.7(c). O produto  $SP_{e/j/lj}$  considerando os três tipos de eventos, Figura 5.7(d), resume o desempenho do classificador. Por fim, a aceitação de *leptonjets* é mostrada na Figura 5.7(e). Observa-se que o aumento da energia retida melhora a diferenciação entre elétrons e jatos e, por conseguinte, o desempenho do canal. Por outro lado, o desempenho para *leptonjets* piora, indicando que essas componentes acrescidas os aproximam mais dos jatos do que dos elétrons.

Considerando as três classes de eventos, o valor do produto  $SP_{e/j/lj}$  alcançado pelo algoritmo foi 0,7396. Isso implica 91,59% de elétrons detectados; 14,72% de falso alarme e 49,05% de *leptonjets* aceitos. A Figura 5.8 apresenta o resultado de desempenho do classificador. Os histogramas de saída (Figura 5.8(a)) mostram a separação entre elétrons e jatos alcançada. A curva ROC (Figura 5.8(b)) e o produto  $SP_{e/j}$  em conjunto com a aceitação de *leptonjets* (Figura 5.8(c)) mostram o comportamento do classificador em função do limiar de saída. Já as probabilidades de detecção e falso alarme em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector ATLAS são apresentadas nas Figuras 5.8(d), 5.8(e) e 5.8(f), respectivamente. Ao observarmos a eficiência em  $\eta$  (Figura 5.8(e)) notamos uma queda acentuada na detecção de elétrons e *leptonjets* para valores  $|\eta| \geq 1,5$ , além do aumento do falso alarme nessa mesma região. Entende-se que os eventos que atingiram o detector na região mencionada (que já compreende a tampa dos calorímetros) não foram fruto de um choque frontal entre as partículas pertencentes aos feixes de colisão. Mesmo sendo cobertas por sensores, essas regiões apresentam um número menor de canais de leitura, o que poderia explicar a queda de desempenho. Isso já era esperado em todos os testes, mas aqui, com uso das técnicas de análise de componentes e, por conseguinte, da eliminação da informação sob os critérios pertinentes à técnica, foi mais fortemente sentida pelo classificador linear. Esse é um resultado que sofrerá críticas do especialista do sistema para a sua consideração como alternativa viável de filtragem.

Analisando os resultados percebe-se que a melhora significativa na detecção de *leptonjets* implica uma pequena deterioração do desempenho do canal para elétrons e jatos; 91,59% de elétrons aceitos significa 4 pontos percentuais abaixo do obtido pelo algoritmo de referência T2Ca1o, enquanto o falso alarme de 14,72% representa um aumento de 4 pontos percentuais. Uma solução linear para o problema é apresentada com praticamente as mesmas cifras obtidas pelo T2Ca1o, porém com um ganho significativo na detecção de eventos exóticos. Vale ressaltar que o valor de detecção de *leptonjets* poderia ser ainda mais aumentado. Se houvesse a necessidade da priorização desses eventos, uma regulação do limiar de saída do classificador facilmente poderia ser implementada, sem sequer o retreinamento dos algoritmos, como mostra a Figura 5.8(c).

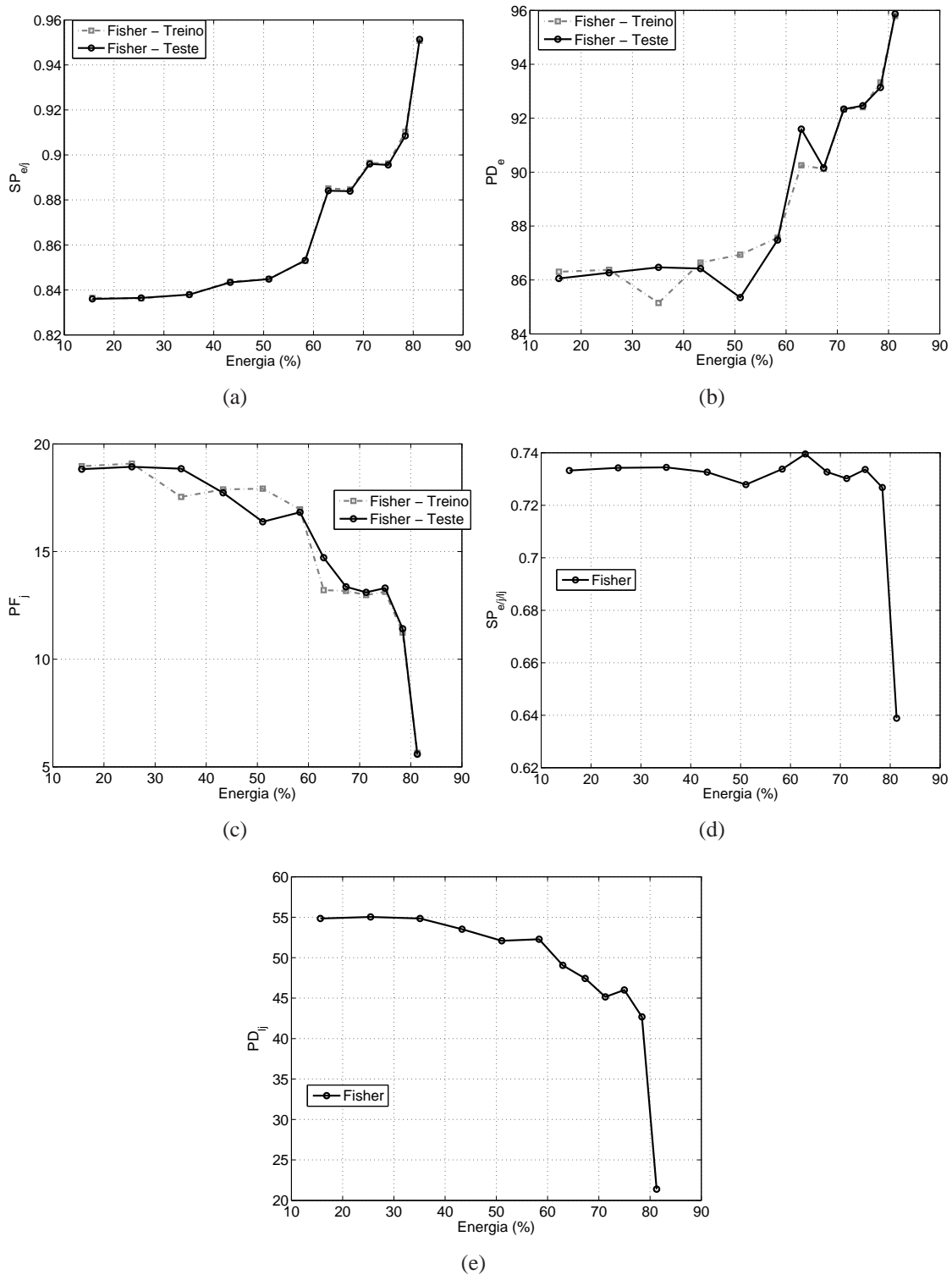


Figura 5.7: Variação de desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por camada em função da energia retida do sinal: (a) Produto  $SP_{e_j}$ ; (b) Probabilidade de detecção  $PD_e$ ; (c) Probabilidade de falso alarme  $PF_j$ ; (d) Produto  $SP_{e_j/l_j}$ ; (e) Probabilidade de detecção  $PD_{l_j}$ .

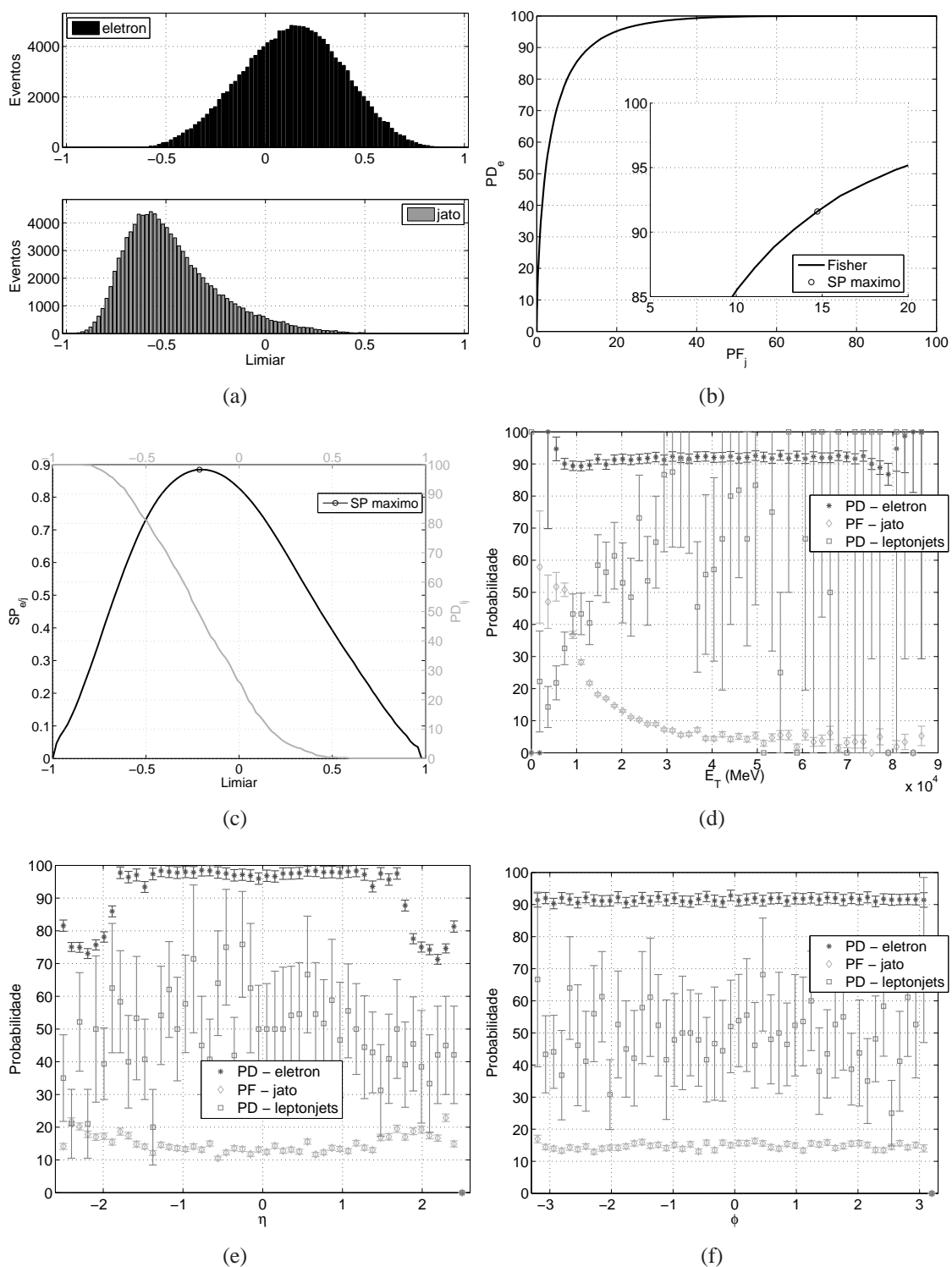


Figura 5.8: Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por camada. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e/j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

Ainda avaliando o comportamento da abordagem proposta, a relevância das componentes extraídas (Figura 5.9) mostra, principalmente, a concentração da informação decisiva para o processo decisório na primeira componente, seja para elétrons, jatos ou *leptonjets*. Já a Figura 5.9(b) ilustra o quão correlacionadas estão as componentes extraídas. Dessa maneira, conclui-se que as componentes são uma boa aproximação da informação que realmente rege o processo e que uma delas é crucial para o sistema decisório implementado.

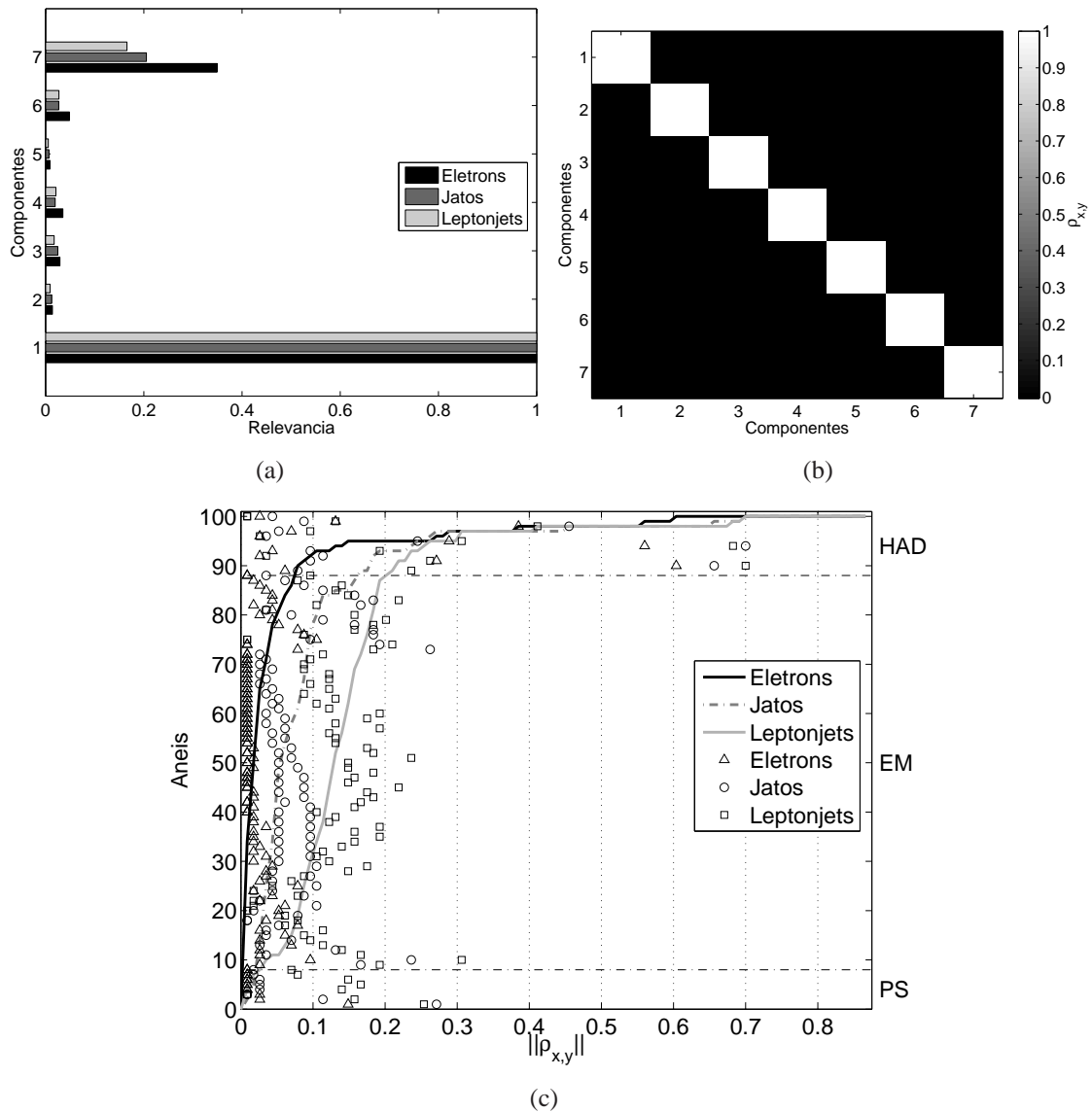


Figura 5.9: Medidas qualitativas das variáveis de entrada do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por camada: (a) Relevância das componentes principais no processo decisório implementado; (b) Correlação entre as componentes principais extraídas; (c) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes principais.

A partir da relevância calculada (Figura 5.9) é possível fazer um levantamento da

relação entre a componente mais significativa e os anéis originais, como mostra a Figura 5.9(c). As curvas indicam o número acumulado de anéis com valor igual ou inferior a um determinado patamar de correlação. Os marcadores indicam o valor absoluto específico de correlação entre o anel e a componente de máxima relevância. O crescimento abrupto dessa curva indica um número grande de anéis compartilhando o mesmo nível de correlação com a componente extraída, i.e., um compartilhamento da importância de cada anel na formação da informação presente na componente de máxima relevância. Essa propriedade é bastante interessante quando se analisa a robustez a falhas do sistema.

Analisando os elétrons, verifica-se que um grande número de anéis estão fracamente correlacionados com a componente extraída (especificamente os anéis externos da camada *EM front layer*), i.e., poucos contribuem de maneira contundente para o descritor extraído. O fato da informação proveniente de alguns anéis das camadas hadrônicas serem fortemente correlacionados com a informação da componente relevante indica que boa parte da decisão tomada pelo classificador para aceitar os elétrons encontra-se nessas variáveis. Não se trata de dizer que os elétrons depositam mais energia nessa camada, e sim que a informação aqui depositada é importante na construção da componente extraída e, por conseguinte, da formação da estrutura interna de decisão criada pelo classificador. No caso dos *leptonjets*, observa-se o oposto, um número menor de anéis se destacam na formação da componente principal extraída, como mostra o crescimento mais lento da curva. Os anéis das camadas eletromagnéticas são mais importantes para a aceitação desse tipo de evento do que foram para o elétron isolado. Os jatos hadrônicos apresentam um comportamento intermediário.

O valor da correlação entre a componente de máxima relevância e os anéis indica, além da contribuição de cada um deles na formação do descritor, a dependência dessa variável em relação a cada região dos calorímetros. Ganha destaque os primeiros anéis das camadas *EM front layer*, *EM middle layer*, *EM back layer* e *HAD layer 1*.

O uso da PCA na etapa de pré-processamento dos dados também foi submetido ao teste de hipótese implementado por um classificador neural. Algumas melhoras foram notadas. A partir da curva de carga (Figura 5.6) e com um número específico de componentes principais associado a um percentual de energia retido, diferentes topologias de classificadores neurais foram testadas, todas com variação apenas no número de neurônios na camada intermediária da rede. A variação do número de componentes retidas e do número de nós na camada escondida da rede resultou em diferentes valores finais de desempenho. O número de neurônios variou de 2 a 15, enquanto o percentual de energia retida do sinal alcançou valores próximos a 80%.

A Tabela 5.4 apresenta o resultado das melhores configurações para cada uma das normalizações de dados empregada. Assim como o observado com o classificador linear, os melhores resultados foram obtidos com um número pequeno de componentes principais. A melhora observada nos valores finais deve-se ao uso da não-linearidade aplicada pelo

classificador. Vale destacar que o mérito maior pelos valores finais alcançados fica por conta do uso das componentes principais, pois o ganho observado pelo algoritmo neural não é tão superior ao obtido com o algoritmo linear. Como um dos requisitos do sistema é a alta eficiência, o mínimo ganho obtido é bem visto, principalmente quando se trata de detecção de eventos raros e/ou exóticos.

Tendo em vista os valores de referência até aqui apresentados, o uso da normalização sequencial favorece a informação do canal (elétron/jato) e a normalização por camada favorece a aceitação de *leptonjets*. A Figura 5.10 ilustra o comportamento da abordagem usada em função da energia retida do sinal e do número de neurônios na camada escondida da rede, enquanto a Figura 5.11 mostra dados de desempenho do melhor classificador.

Tabela 5.4: Desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Neurônios	Comp.	Energia(%) <sup>2</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
Energia	7	2	61,53	93,87	9,66	0,9210	-0,24	30,32	0,6744
Sequencial	5	5	38,60	93,50	12,06	0,9070	0,15	45,63	0,7389
Camada	10	5	51,05	89,68	18,19	0,8570	0,13	55,42	0,7486

O erro quadrático médio (Figura 5.10(b)) ao término do treinamento da rede neural mostra a tendência de aproximação dos valores desejados de saída com aumento tanto do número de neurônios na camada intermediária da rede quanto da energia retida do sinal. A Figura 5.10(b) mostra o gráfico de comportamento do produto SP<sub>e/j</sub> do canal. A variação do número de neurônios intermediários da rede, assim como do percentual de energia retido do sinal indicam que quanto maiores forem esses valores melhor o resultado final do classificador. O destaque fica para o percentual de energia do sinal retido, que impacta mais fortemente o resultado final do que a complexidade do classificador não-linear usado. É visível que para um mesmo percentual de energia a curva em função do número de neurônios intermediário sofre uma leve variação, principalmente com baixa energia, haja visto a pouca informação disponível para modelar a estrutura interna de decisão. Quando o percentual de energia retido é maior, o número elevado de parâmetros livres do classificador neural ganha certo protagonismo no desempenho final, como pode-se constatar com, por exemplo,  $\approx 80\%$  de energia da retida.

Os diferentes valores de produto SP<sub>e/j</sub> obtidos estão associados, respectivamente, às probabilidades de detecção e falso alarme mostradas nas Figuras 5.10(c) e 5.10(d).

No que se refere a dados exóticos, é identificado um movimento inverso da curva, quanto menos energia retida na transformação melhor é a sua detecção. A Figura 5.10(f) apresenta a eficiência do sistema, PD<sub>lj</sub>, para esse tipo de evento. Acredita-se que com a compactação realizada (e sua transformação de base), *leptonjets* tenham suas características aproximadas em relação aos elétrons pontuais. Ao eliminar as componentes menos energéticas as informações que causariam confusão no discernimento entre elétrons

<sup>2</sup>O percentual de energia se refere ao acumulado nas componentes principais.

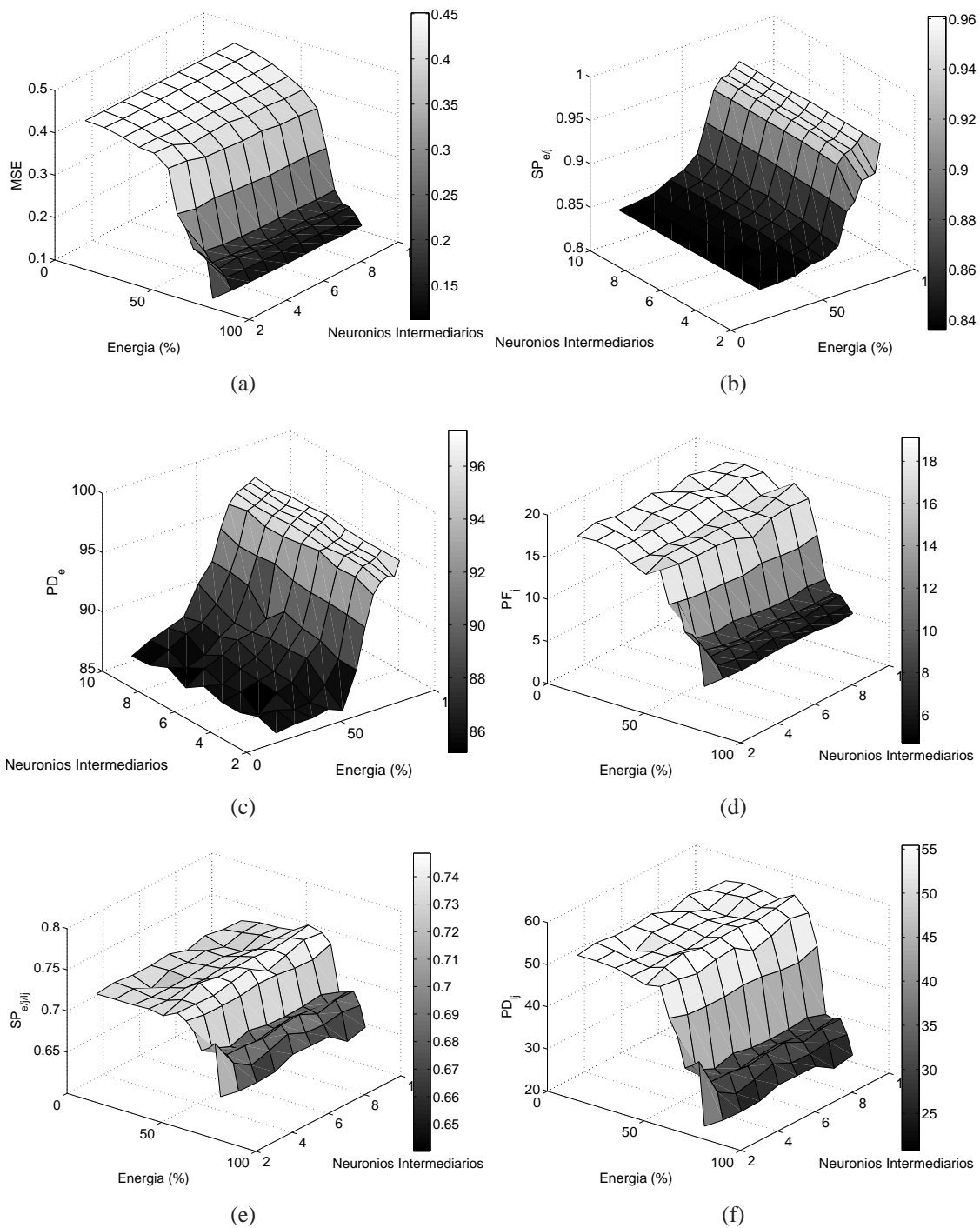


Figura 5.10: Variação de desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis normalizados por camada em função do número de neurônios na camada intermediária da rede e da energia retida do sinal: (a) Erro quadrático médio (MSE) final de treinamento; (b) Produto  $SP_{e/j}$  associado à melhor rede; (c) Probabilidade de detecção  $PD_e$  associada à melhor rede; (d) Probabilidade de falso alarme  $PF_j$  associada à melhor rede; (e) Produto  $SP_{e/j/l_j}$  associado à melhor rede; (f) Probabilidade de detecção  $PD_{l_j}$  associada à melhor rede.



e *leptonjets* são reduzidas, sem deixar a taxa de falsos positivos (jatos rotulados como elétrons) crescer de maneira descontrolada.

Na avaliação do produto  $SP_{e/j/l_j}$ , que é o medidor de desempenho final, Figura 5.10(e), a configuração que atende o melhor classificador é obtida com 10 neurônios intermediários e 5 componentes principais extraídas (ou 51,05% da energia do sinal). Com esse classificador é possível discriminar 55,42% dos *leptonjets*, enquanto 89,68% dos elétrons são aceitos para uma taxa de falso alarme de 18,19%, resumidos em um produto  $SP_{e/j/l_j}$  de 0,7486.

No que se refere a detecção de elétrons ( $PD_e$ ), a comparação com o classificador linear e o algoritmo de referência T2Ca1o indica um desempenho próximo, levemente inferior a este último. Contudo, acompanhando o observado com o classificador linear, o percentual de detecção para *leptonjets* é bastante significativo, i.e., mais de 25 pontos percentuais em relação ao apresentado pelo T2Ca1o, mostrando uma grande evolução. Pode-se considerar esse resultado promissor, dado que a configuração final do classificador (topologia 5-10-1, mesmo com um acréscimo na dimensão interna da rede) e a compactação implementada (95% da dimensão original) implicam operações simples do ponto de vista computacional. Mesmo que o resultado oferecido, especificamente para a taxa de  $PD_e$ , esteja levemente abaixo do T2Ca1o e um pouco mais distante do RingerNeuralHypo, a metodologia aplicada pode sim ter suas potencialidades exploradas com o intuito de alcançar números finais ainda mais satisfatórios.

Os detalhes de desempenho do classificador são apresentados na Figura 5.11. Os histogramas de saída da rede são exibidos pela Figura 5.11(a). A curva ROC e o produto  $SP_{e/j}$  em conjunto com aceitação de *leptonjets* são apresentados nas Figuras 5.11(b) e 5.11(c), respectivamente. A eficiência em função da  $E_T$ , de  $\eta$  e de  $\phi$  e mostrada nas Figuras 5.11(d), 5.11(e) e 5.11(f). Fica evidente o poder do classificador neural na melhora da aceitação de elétrons e *leptonjets* na região além do *crack* ( $|\eta| \geq 1,5$ ), em relação ao observado com o classificador linear.

Na Tabela 5.4 encontramos outra peculiaridade do resultado, o limiar de decisão correspondente ao melhor classificador está deslocado para o lado positivo do eixo (0,13). Isso indica que os jatos (cuja saída deseja é  $-1$ ), por representarem aparentemente um evento de maior complexidade do ponto de vista da cascata de decaimento, com o uso das componentes principais retidas precisaram ser mapeados dentro de um intervalo de decisão maior do que o demandado pelos elétrons isolados. Isso levou a reduzir a eficiência na detecção dos *leptonjets*, que mesmo assim foi a mais alta até aqui atingida. É fácil observar também que para uma maior aceitação de *leptonjets* um simples ajuste no limiar de decisão de saída da rede pode ser realizado, sem a necessidade de retreinamento dos algoritmos. Essa característica é importante para períodos em que o acelerador e o detector estão disponíveis, por um pequeno período de tempo, para testes livres dos pesquisadores.

A análise de relevância das variáveis que alimentam o classificador neural, mostrada



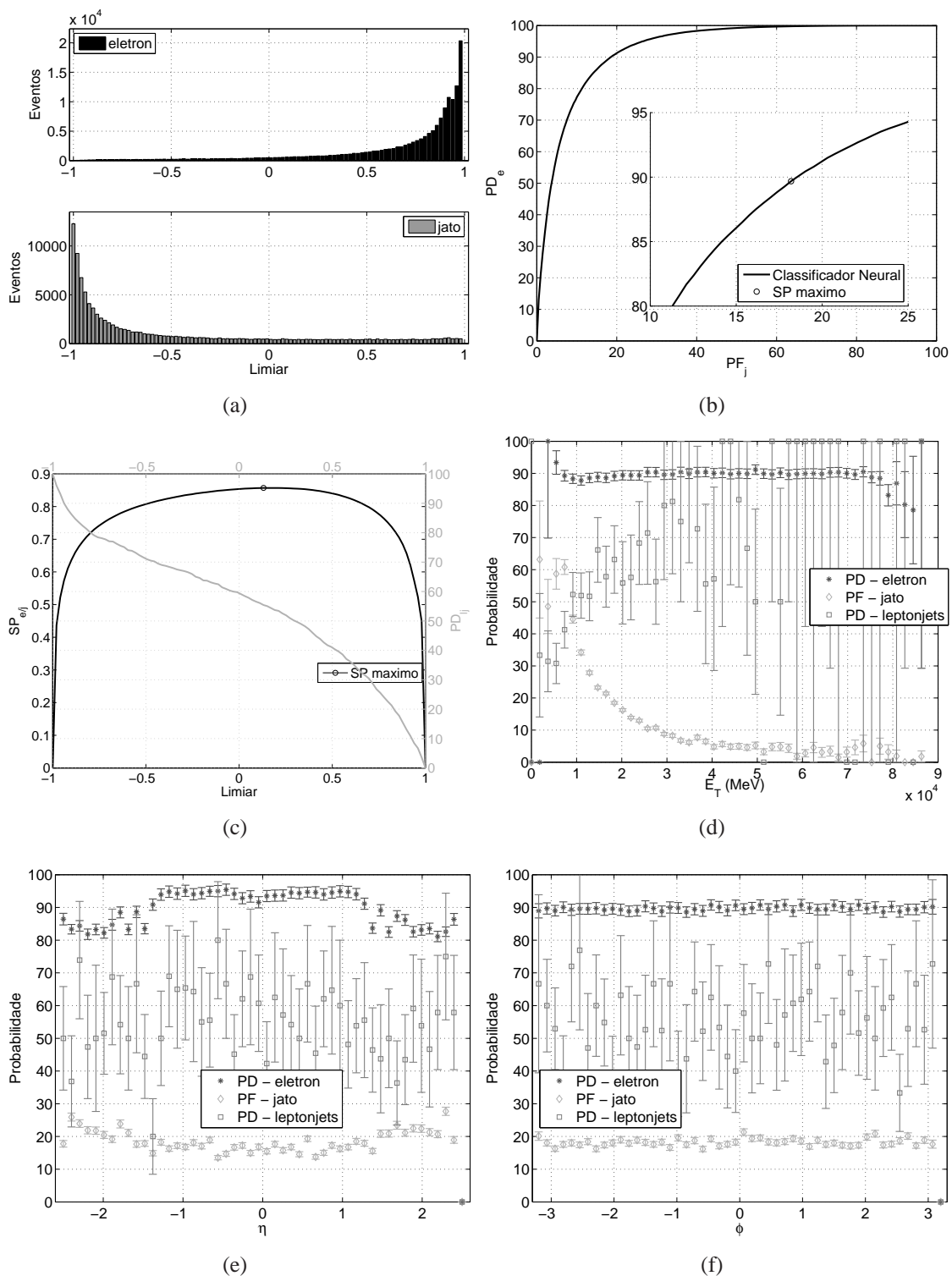


Figura 5.11: Desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por camada. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e/j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

na Figura 5.12(b), indica que das componentes retidas, uma detém papel preponderante. Porém, mesmo com pouca relevância relativa no processo decisório, as componentes restantes apresentaram um valor maior do que o observado com o classificador linear, o que indica que o ganho alcançado (em relação àquele classificador) pode ser creditado ao melhor uso, por parte do algoritmo neural, da informação nelas contida.

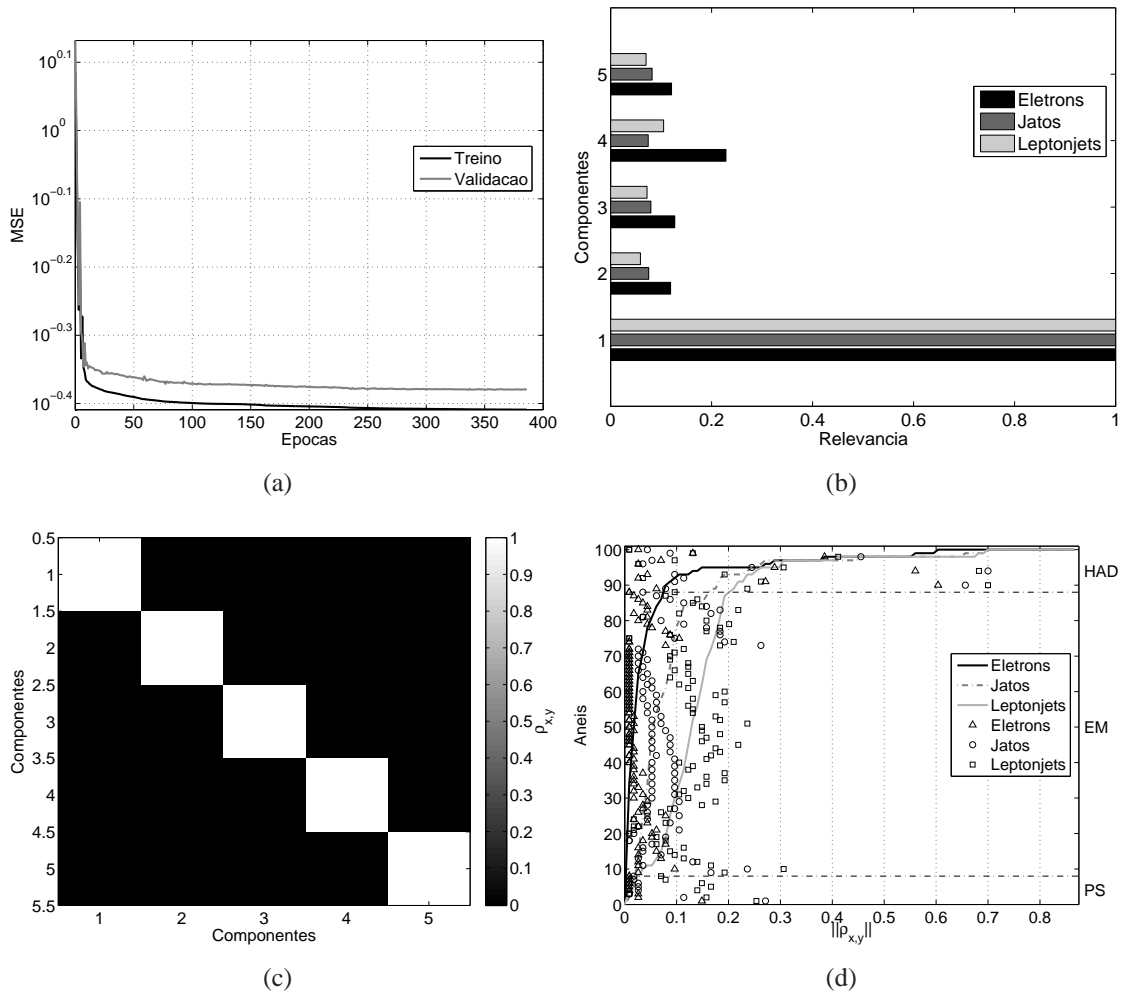


Figura 5.12: (a) Erro quadrático médio de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por camada: (b) Relevância das componentes principais no processo decisório implementado; (c) Correção entre as componentes principais extraídas; (d) Correção entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes principais.

A análise da correção, entre a componente de maior relevância e os anéis originais (Figura 5.12(d)), mostra que essa componente apresenta o mesmo comprometimento que já fora observado no resultado do classificador linear (Figura 5.9(c)). Mesmo com um número diferente de componentes retidas alimentando cada uma das abordagens (7 com classificador linear contra 5 com classificador neural), a componente de máxima relevância é a mesma, indicando que os classificadores lidam com a mesma informação principal

e de maneira diferente com a informação secundária. O classificador neural, como era esperado, tira maior proveito dessa situação.

### 5.3.2 ICA

É sabido que as técnicas de ICA apresentam uma série de dificuldades ao operar em ambientes de alta dimensionalidade, ora por instabilidade numérica ora por dificuldade de convergência, dentre outros [7]. Para contornar esse problema, a maioria das técnicas empregam uma etapa de branqueamento dos dados, como mostra o Apêndice A. Esse passo, além de garantir a desconexão linear dos dados transformados permite a redução da dimensionalidade com algum controle sobre a qualidade dos mesmos. Os algoritmos ICA operam sobre essa base de dados de dimensão reduzida. Dependendo da propriedade estatística dos dados transformados, o algoritmo consegue implementar alguma rotação nessa base de maneira a tornar os dados o mais independente possível. Em muitos casos a base ortonormal obtida no branqueamento já exibe um valor de independência estatística das variáveis transformadas que não permite mais manobras por parte do algoritmo ICA.

Como a etapa de branqueamento dos dados está diretamente ligada à técnica PCA, isso nos habilita a usar a mesma curva de carga (Figura 5.6) empregada anteriormente para a análise dos resultados dessa seção. Além disso, esse fato nos leva a conclusões previamente esperadas, como uma aproximação muito grande dos resultados ou mesmo resultados similares em vários testes de dimensão reduzida.

A Tabela 5.5 quantifica o desempenho obtido com o primeiro classificador testado, o classificador linear de Fisher. É possível aferir que com um número pequeno de componentes independentes extraídas, a projeção dos *leptonjets* apresenta a similaridade desejada em relação aos elétrons, como observado com o uso da técnica PCA (Tabela 5.3). O pequeno número de componentes independentes não representam um mesmo percentual de energia retida do sinal original, na verdade ele varia bastante entre as normalizações empregadas. O melhor resultado é alcançado com 7 componentes ou 62,98% da energia retida do sinal quando submetido à normalização por camada. A projeção dos jatos hadrônicos nessas componentes é suficiente para a diferenciação dos eventos, o que resulta em um falso alarme próximo do apresentado pelo algoritmo de referência T2Ca1o. Os valores aqui obtidos são: 91,59% de  $PD_e$ ; 14,72% de  $PF_j$ ; 0,8841 de  $SP_{e/j}$ ; 49,05% de  $PD_{lj}$  e 0,7396 de  $SP_{e/lj}$ . Novamente o destaque fica para alta taxa de detecção de *leptonjets*. Um outro ponto a destacar nesse resultado é o valor do limiar de decisão associado ao ponto de operação do classificador,  $-0,21$ . Isso indica que após a técnica de pré-processamento ICA o classificador precisou deslocar o seu limiar ótimo de decisão em função de uma maior aproximação dos perfis de cada eventos (elétrons e jatos), i.e., os elétrons ficaram mais espalhados ao longo do eixo de decisão do classificador, enquanto os jatos se comportaram de maneira mais homogênea. Essa conclusão pode ser estendida

para todas as outras normalizações e valores de componentes retidas, como mostra a Tabela 5.5. Examinando os valores finais de desempenho obtidos, podemos considerar esse deslocamento do limiar de decisão como uma característica positiva, pois permitiu que os *leptonjets* tivessem mais liberdade de projeção e ainda assim fossem detectados como elétrons.

Tabela 5.5: Desempenho do classificador linear operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Comp.	Energia(%) <sup>3</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
Energia	1	51,26	88,07	9,67	0,8919	-0,17	17,11	0,5790
	2	61,53	91,47	9,77	0,9085	-0,27	17,78	0,5922
	3	71,05	90,43	9,26	0,9058	-0,11	18,35	0,5948
	4	75,28	98,20	3,88	0,9716	0,27	11,50	0,5721
	5	78,75	97,96	3,81	0,9707	0,11	9,79	0,5542
	6	82,14	97,46	3,30	0,9708	0,11	7,98	0,5334
Sequencial	1	11,34	92,58	12,58	0,8998	-0,11	44,39	0,7292
	2	19,38	35,77	35,12	0,4924	-0,01	38,21	0,4544
	3	26,53	92,37	11,70	0,9032	-0,13	43,16	0,7258
	4	32,58	92,39	11,74	0,9032	-0,13	43,06	0,7253
	5	38,60	92,32	11,37	0,9047	-0,05	42,78	0,7250
	6	43,27	94,84	9,86	0,9248	-0,11	38,40	0,7167
	7	47,50	95,75	10,15	0,9278	-0,19	37,36	0,7135
	8	50,92	95,38	9,76	0,9279	-0,17	36,03	0,7072
	9	53,84	94,93	9,29	0,9281	-0,17	34,41	0,6993
	10	56,57	94,94	9,58	0,9267	-0,21	34,41	0,6985
	11	59,23	94,05	8,85	0,9259	-0,21	32,89	0,6904
	12	61,68	95,08	9,46	0,9279	-0,21	34,03	0,6973
	13	63,89	95,34	9,72	0,9279	-0,21	34,60	0,7002
	14	65,98	95,06	9,29	0,9287	-0,19	34,32	0,6992
	15	68,05	95,21	9,40	0,9289	-0,19	34,03	0,6979
	16	70,01	95,03	9,19	0,9291	-0,19	32,70	0,6912
	17	71,90	95,19	8,88	0,9315	-0,17	32,13	0,6897
	18	73,72	95,92	9,31	0,9329	-0,17	32,51	0,6924
	19	75,46	95,94	9,32	0,9329	-0,17	32,51	0,6924
	20	77,13	95,60	8,89	0,9334	-0,15	31,37	0,6868
Camada	1	15,65	86,06	18,83	0,8360	-0,13	54,85	0,7332
	2	25,45	41,63	42,54	0,4922	0,15	38,97	0,4568
	3	35,09	86,47	18,85	0,8379	-0,17	54,85	0,7344
	4	43,30	86,42	17,73	0,8433	-0,19	53,52	0,7326
	5	51,05	85,35	16,38	0,8448	-0,17	52,09	0,7279
	6	58,35	87,48	16,82	0,8532	-0,21	52,28	0,7338
	7	62,98	91,59	14,72	0,8841	-0,21	49,05	0,7396
	8	67,33	90,16	13,36	0,8839	-0,19	47,43	0,7327
	9	71,27	92,34	13,10	0,8960	-0,23	45,15	0,7302
	10	74,98	92,46	13,30	0,8956	-0,21	46,01	0,7336
	11	78,42	93,14	11,42	0,9085	-0,15	42,68	0,7268
	12	81,28	95,87	5,59	0,9514	0,01	21,39	0,6389

A Figura 5.13 mostra a evolução do desempenho obtido com o classificador linear operando sobre componentes independentes extraídas na etapa de pré-processamento. Os

<sup>3</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.

gráficos são exibidos em função da energia retida do sinal para dados normalizados por camada. O produto  $SP_{e/j}$  do canal é mostrado na Figura 5.13(a) e as respectivas taxas de eficiência na detecção de elétrons e falso alarme de jatos são mostradas nas Figuras 5.13(b) e 5.13(c). O produto  $SP_{e/j/lj}$  (Figura 5.13(d)) resume o desempenho final da abordagem. Por fim, evolução da aceitação de *leptonjets* é mostrada na Figura 5.13(e). Fica evidente que com um número pequeno de componentes extraídas, a aproximação do perfil de *leptonjets* e elétron é alcançada, resultando em um valor de produto  $SP_{e/j/lj}$  alto. Contudo, confrontando os valores obtidos pelos algoritmos de referência T2Ca1o e RingerNeuralHypo, no que diz respeito ao canal elétron/jato, os valores finais da abordagem em questão representam uma perda significativa de rendimento, de maneira que a utilização de poucas componentes só se justificaria para a priorização da detecção dos eventos exóticos (i.e., num período de testes livres exclusivos para esse tipo de decaimento). Pode-se contornar esse problema com uma perda compartilhada, i.e., escolhe-se um número maior de componentes extraídas, ganha-se o suficiente para igualar os valores do canal elétron/jato com o apresentado pelos algoritmos de referência e perde-se um pouco na detecção de *leptonjets*, mas ainda assim com cifras bem superiores ao destes últimos.

A Figura 5.14 mostra o comportamento do melhor classificador obtido. Os histogramas de saída do classificador (Figura 5.14(a)) exibem a separação alcançada pelo algoritmo de Fisher na separação entre elétrons e jatos. A partir da variação do limiar de decisão, é possível construir a curva ROC (Figura 5.14(b)), além do produto  $SP_{e/j}$  e da aceitação de *leptonjets*  $PD_{lj}$  (Figura 5.14(c)). As probabilidades de detecção e falso alarme em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector ATLAS estão dispostas nas Figuras 5.14(d), 5.14(e) e 5.14(f), respectivamente.

Um outro aspecto avaliado na metodologia executada foi a relevância das componentes extraídas e a relação delas com os anéis originais (Figura 5.15). A Figura 5.15(a) que detalha a importância das variáveis de entrada do classificador indica, principalmente, o valor de uma componente específica - independentemente do evento em questão - na construção da estrutura de decisão interna do classificador. Já a Figura 5.15(b) mostra o quão independentes são as componentes extraídas. Dessa maneira, ao relacionar-se as duas figuras, conclui-se que as componentes são uma boa aproximação da informação que realmente rege o processo estudado e que uma delas, a de número 4, é crucial para o algoritmo decisório implementado, enquanto a de número 3 é praticamente desprezível. Por fim, a Figura 5.15(c) mostra a correlação dessa componente de maior relevância com os anéis originais. Em uma comparação com o classificador linear alimentada por componentes principais (Figura 5.9), que também apresentou o melhor resultado com 7 componentes retidas, verificamos que a relação da componente de maior relevância com os anéis usados na sua construção dá-se de forma diferente. Apesar de a ICA não manter qualquer critério de ordem na extração de suas componentes (diferentemente do que ocorre com

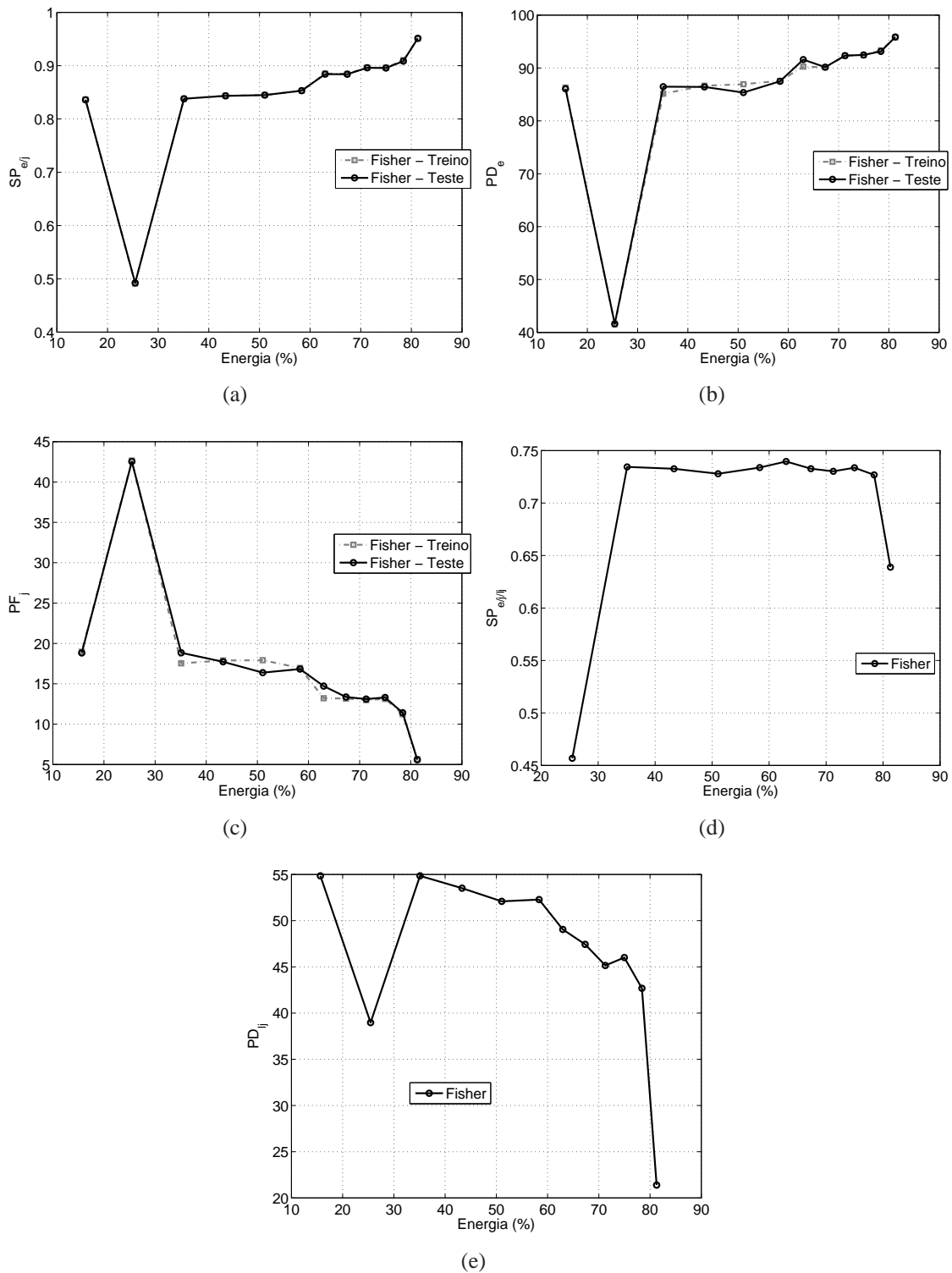


Figura 5.13: Variação de desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por camada. Resultados em função da energia retida do sinal: (a) Produto  $SP_{e/j}$ ; (b) Probabilidade de detecção  $PD_e$ ; (c) Probabilidade de falso alarme  $PF_j$ ; (d) Produto  $SP_{e/j/l_j}$ ; (e) Probabilidade de detecção  $PD_{l_j}$ .

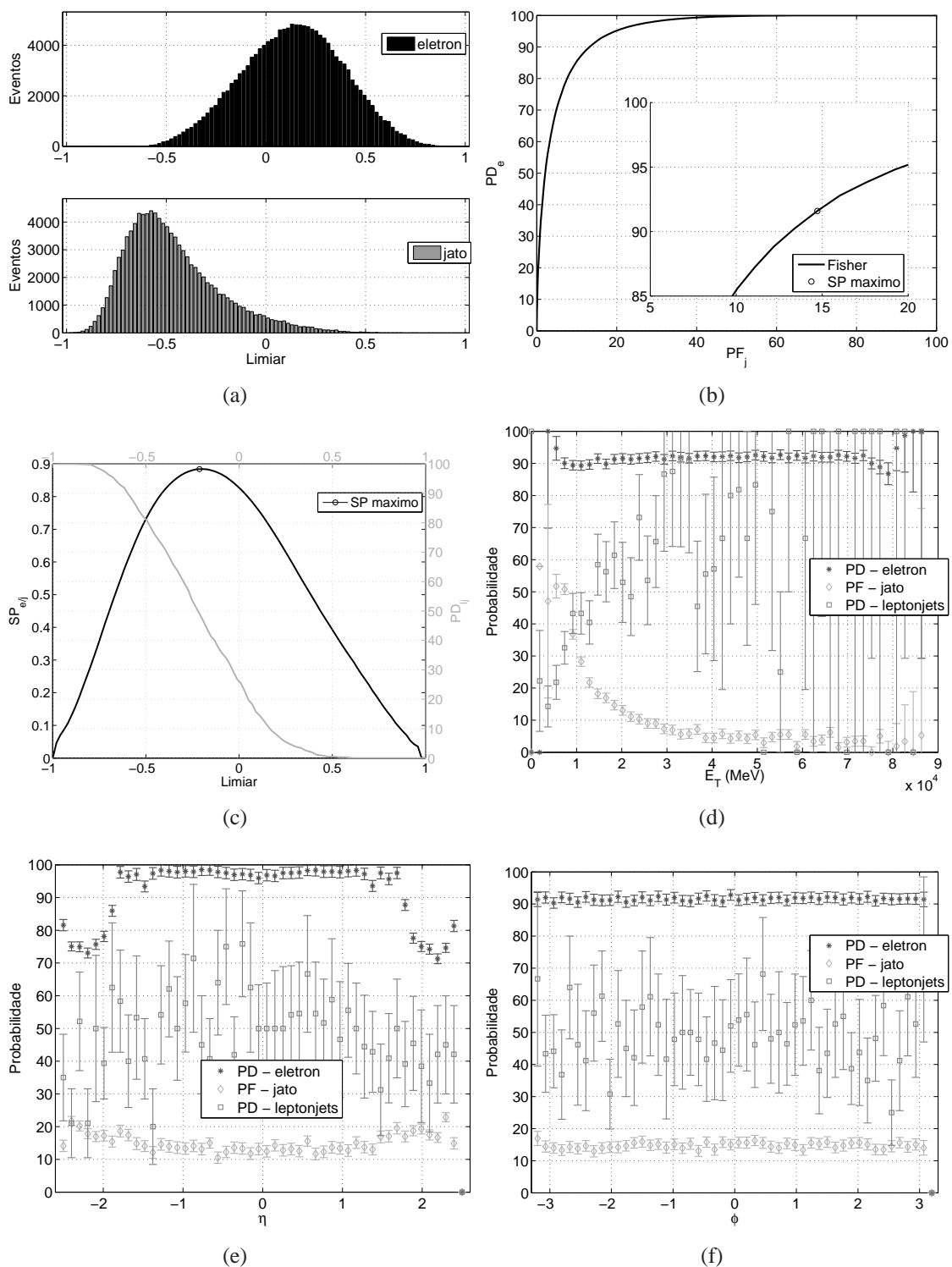


Figura 5.14: Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por camada. (a) Histograma de saída. (b) Curva ROC. (c) Produto  $SP_{e/j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da energia transversa ( $E_T$ ). (e) Probabilidades de detecção e falso alarme em função da corrdenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.



a PCA), podemos fazer uma comparação entre as componentes de maior relevância em cada abordagem para tirar algumas conclusões. A componente ICA de maior relevância baseia sua construção na informação proveniente das camadas hadrônicas. Essa propriedade já tinha sido observada na formação da componente PCA de máxima relevância, porém, com uma participação maior da informação das outras camadas. Um outro aspecto pertinente ao caso ICA, é a possibilidade de inferir que a detecção de cada um dos eventos é afetada da mesma maneira em caso de falha nos sensores de qualquer camada do detector, haja visto a semelhança no crescimento da curva que indica o acumulado de anéis correlacionados por limiar.

No aspecto geral, essa abordagem mostrou características positivas para um algoritmo de filtragem de nível 2. O classificador linear conseguiu lidar satisfatoriamente com a informação pré-processada fornecida, além de prover explicação para a sua estrutura interna de decisão. Aproveitando tais virtudes, o uso de um classificador neural foi testado.

A partir dos valores de energia retida do sinal (e por conseguinte de diferentes componentes extraídas), variações topológicas de classificadores neurais foram testadas, todas com alteração no número de neurônios na camada intermediária da rede. A variação desses dois parâmetros resultou em classificadores com diferentes valores finais de desempenho. O número de neurônios da camada escondida da rede variou de 2 a 14 neurônios, enquanto o percentual de energia retida alcançou valores em torno de 80%.

Seguindo o raciocínio até aqui exposto, selecionamos os melhores resultados em função do produto  $SP_{e/j/l_j}$  com cada uma das normalizações testadas, como indica a Tabela 5.6.

Tabela 5.6: Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Neurônios	Comp.	Energia(%) <sup>4</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>l<sub>j</sub></sub>	SP <sub>e/j/l<sub>j</sub></sub>
Energia	2	2	61,53	94,86	11,21	0,9180	0,12	21,10	0,6195
Sequencial	5	5	38,60	93,80	12,45	0,9064	0,11	46,20	0,7410
Camada	3	4	32,58	92,23	12,56	0,8982	0,31	44,58	0,7291

Nestes resultados nota-se a melhora oferecida pelo algoritmo de classificação não-linear (para algumas normalizações), resultando em um aumento da eficiência de detecção dos eventos de interesse. O aumento, mesmo não tão expressivo, só evidencia que o pré-processamento inserido realiza a tarefa mais difícil, que é a disposição dos padrões de maneira menos complexa do ponto de vista da classificação. O pré-processamento é tão bem explorado, que o algoritmo não-linear de classificação poderia ser preterido em detrimento ao algoritmo linear de Fisher (mostrado anteriormente) se cada percentual ganho na detecção não fosse tão valioso como no caso de eventos raros e/ou exóticos.

<sup>4</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.



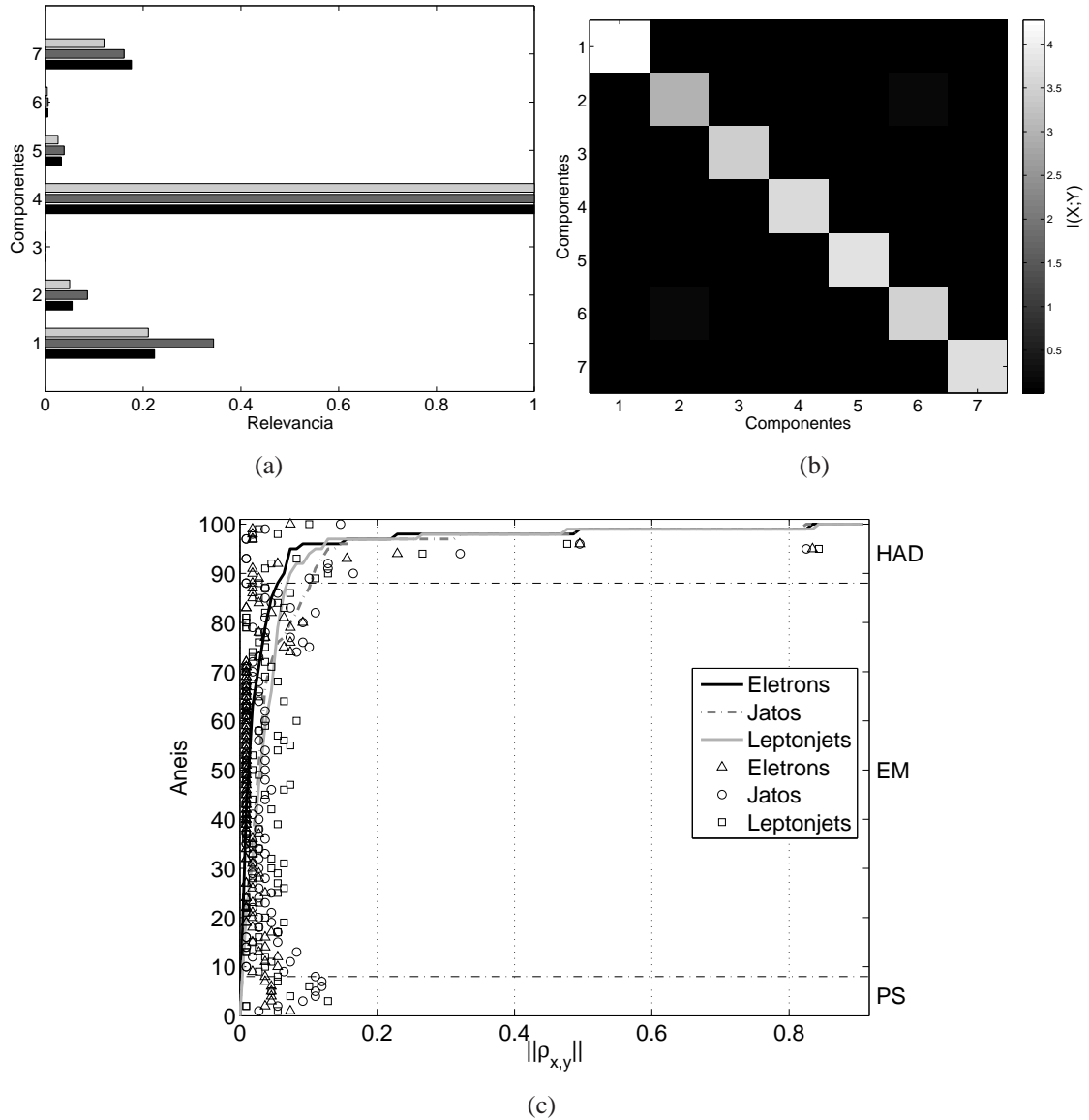


Figura 5.15: Medidas qualitativas das variáveis de entrada do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por camada: (a) Relevância das componentes independentes no processo decisório implementado; (b) Informação mútua entre as componentes independentes extraídas; (c) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes.

Assumindo a normalização sequencial como aquela que oferece o melhor resultado final, detalhes do comportamento da variação dos parâmetros livres (energia retida do sinal e número de nós intermediários da rede) são ilustrados pelas superfícies de desempenho para cada uma das figuras de méritos, como mostra a Figura 5.16.

Da Figura 5.16(a) é possível concluir que o aumento do valor dos parâmetros variáveis no teste aproxima a saída da rede das respostas desejadas, diminuindo o erro quadrático médio (MSE). A Figura 5.16(b) mostra o gráfico de comportamento do produto  $SP_{e/j}$  do canal. A variação do número de neurônios intermediários da rede, assim como do percentual de energia retido do sinal indicam que quanto maior esses valores melhor o resultado final do classificador. O destaque fica para o percentual de energia do sinal retido, que impacta mais fortemente o valor alcançado do que a complexidade do classificador não-linear usado. É observado que para um mesmo percentual de energia, a curva em função do número de neurônios intermediários sofre uma leve variação, principalmente com baixa energia retida do sinal, haja visto a pouca informação disponível para modelar a estrutura interna de decisão. Quando o percentual de energia retido é maior, o número elevado de parâmetros livres do classificador neural ganha certo protagonismo no desempenho final, como pode-se constatar, por exemplo, com  $\approx 80\%$  de energia. Os valores finais variaram de um mínimo de 0,9010, com 2 neurônios intermediários e  $\approx 20\%$  da energia retida até 0,9614, com 10 neurônios intermediários e  $\approx 80\%$  da energia retida.

Os diferentes valores de produto  $SP_{e/j}$  obtidos estão associados, respectivamente, às probabilidades de detecção e falso alarme mostradas nas Figuras 5.16(c) e 5.16(d).

No que se refere a dados exóticos, é identificado um movimento inverso, quanto menos energia retida na transformação melhor é a sua detecção. A Figura 5.16(f) apresenta a eficiência do sistema para esse tipo de evento. Acredita-se que com a compactação realizada (e sua transformação de base), *leptonjets* tenham suas características aproximadas em relação aos elétrons pontuais. Ao eliminar as componentes menos energéticas as informações que causariam confusão no discernimento entre elétrons e *leptonjets* são suprimidas, sem deixar a taxa de falsos positivos (jatos rotulados como elétrons) crescer de maneira descontrolada.

Na avaliação do produto  $SP_{e/j/lj}$  final (Figura 5.16(e)), a configuração que contempla o melhor classificador é alcançada com 5 neurônios intermediários e 5 componentes principais extraídas (ou 38,60% da energia do sinal). Com esse classificador é possível discriminar 46,20% dos *leptonjets*, enquanto 93,80% dos elétrons são aceitos para uma taxa de falso alarme de 12,45%.

A abordagem ICA apresentou dois pontos interessantes: manteve os excelentes índices de aceitação de *leptonjets* já alcançados com a PCA e não degradou tanto os valores para o canal elétron/jato. O percentual de detecção de elétron (93,80%) rivaliza com o apresentado pelo classificador linear operando sobre anéis e pelo algoritmo de referência T2Ca1o. O falso alarme de 12,45% está na mesma faixa de valor do obtido com a PCA e

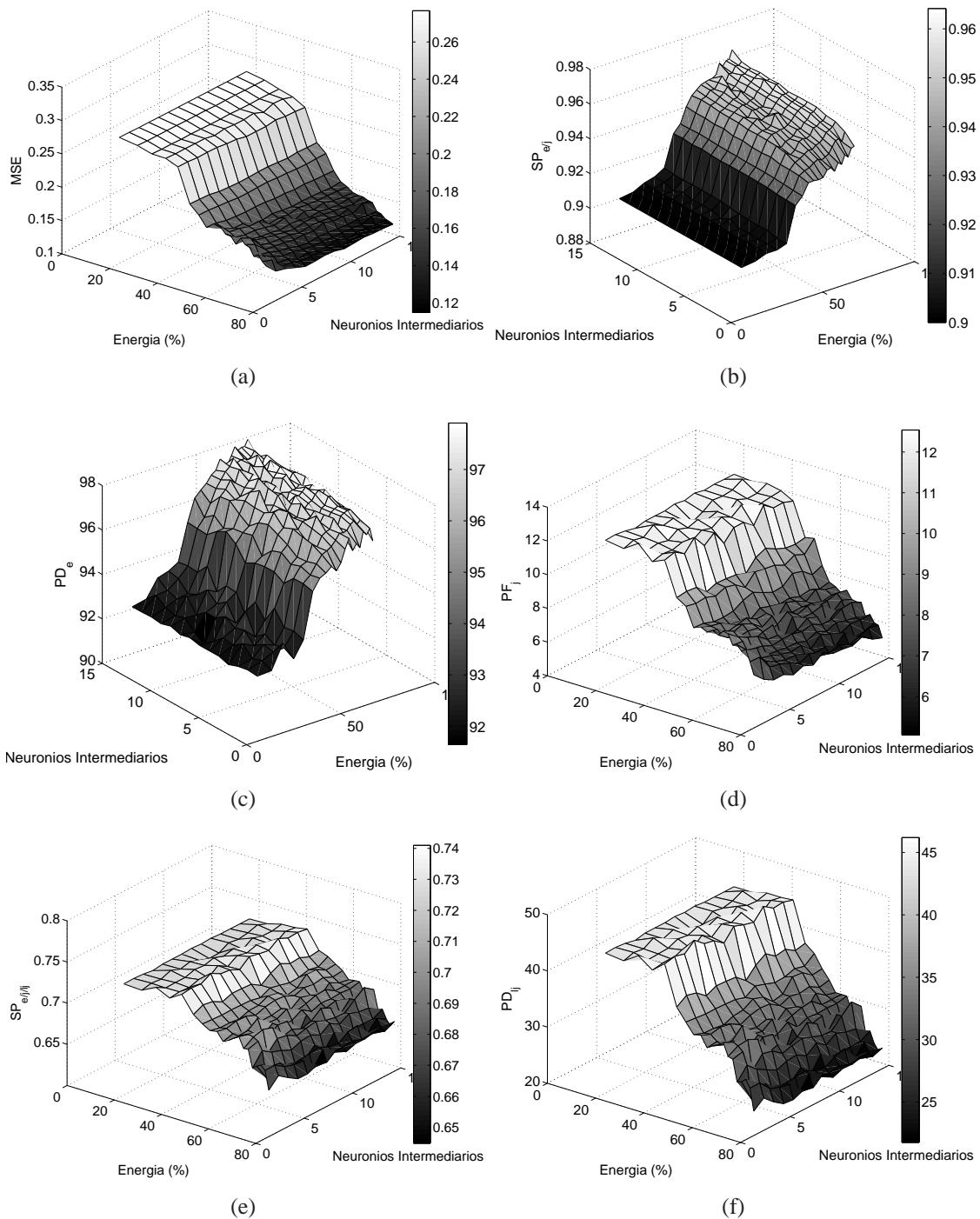


Figura 5.16: Variação de desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados sequencialmente em função do número de neurônios na camada intermediária da rede e da energia retida do sinal: (a) Erro quadrático médio (MSE) final de treinamento; (b) Produto  $SP_{e|j}$  associado à melhor rede; (c) Probabilidade de detecção  $PD_e$  associada à melhor rede; (d) Probabilidade de falso alarme  $PF_j$  associada à melhor rede; (e) Produto  $SP_{e|j/l_j}$  associado à melhor rede; (f) Probabilidade de detecção  $PD_{l_j}$  associada à melhor rede.

com o T2Ca1o. O produto  $SP_{e/j/lj}$  de 0,7410 coloca o resultado dessa abordagem entre os melhores alcançados.

A configuração final obtida pelo classificador (topologia 5-5-1) e a compactação implementada (ICA) implicam operações bastante simples do ponto de vista computacional. Essas duas operações podem, inclusive, ser combinadas dentro do ambiente do detector ATLAS, de maneira que a compactação ICA represente, na estrutura de dados criadas, uma camada (frontal) adicional para o algoritmo neural. Os detalhes do resultado do classificador são apresentados na Figura 5.17. Assim como na PCA, acredita-se que com a compactação realizada *leptonjets* tenham suas características aproximadas em relação aos elétrons pontuais.

A análise de relevância das componentes independentes extraídas, mostrada na Figura 5.18(b), indica que das quatro componentes independentes retidas apenas uma delas detém relevância significativa, em linha com o apresentado pela PCA. Esse resultado permite supor que há a possibilidade de uma compactação ainda maior da informação fornecida ao classificador. Porém, essa componente independente (mais relevante) retida, *a priori*, não é a mesma obtida com o algoritmo ICA configurado para a extração de uma única componente. Aqui, cinco variáveis transformadas através do branqueamento dos dados originais e da retenção de aproximadamente 38% da energia do sinal foram necessárias para a estimação alcançada, o que difere do algoritmo ICA operando sobre uma componente branqueada com a retenção de 11% da energia. Isso pode ser constatado através de uma comparação direta entre valores finais alcançados para as figuras de mérito em cada um dos casos (Tabela 5.5).

## 5.4 Pré-processamento Segmentado

A redução de dimensionalidade dos dados de entrada na abordagem segmentada considera a divisão entre as diferentes camadas dos calorímetros como diferentes processos aleatórios (sete no total). A classificação é realizada sobre os dados transformados em cada uma das camadas.

A extração de características de modo segmentado tem a vantagem de explorar as diferentes potencialidades de cada camada do calorímetro de maneira independente. Espera-se uma maior robustez a falhas nesse procedimento, uma vez que é de conhecimento que os sensores podem parar de funcionar ao longo do tempo e a abordagem segmentada isola essa perda de informação na camada em questão. Contudo, espera-se uma taxa de compactação menor em relação a alcançada pela abordagem não segmentada, dado que as componentes finais obtidas fazem uso apenas de informação local e não consideram a relação dessa com a informação de outras camadas.

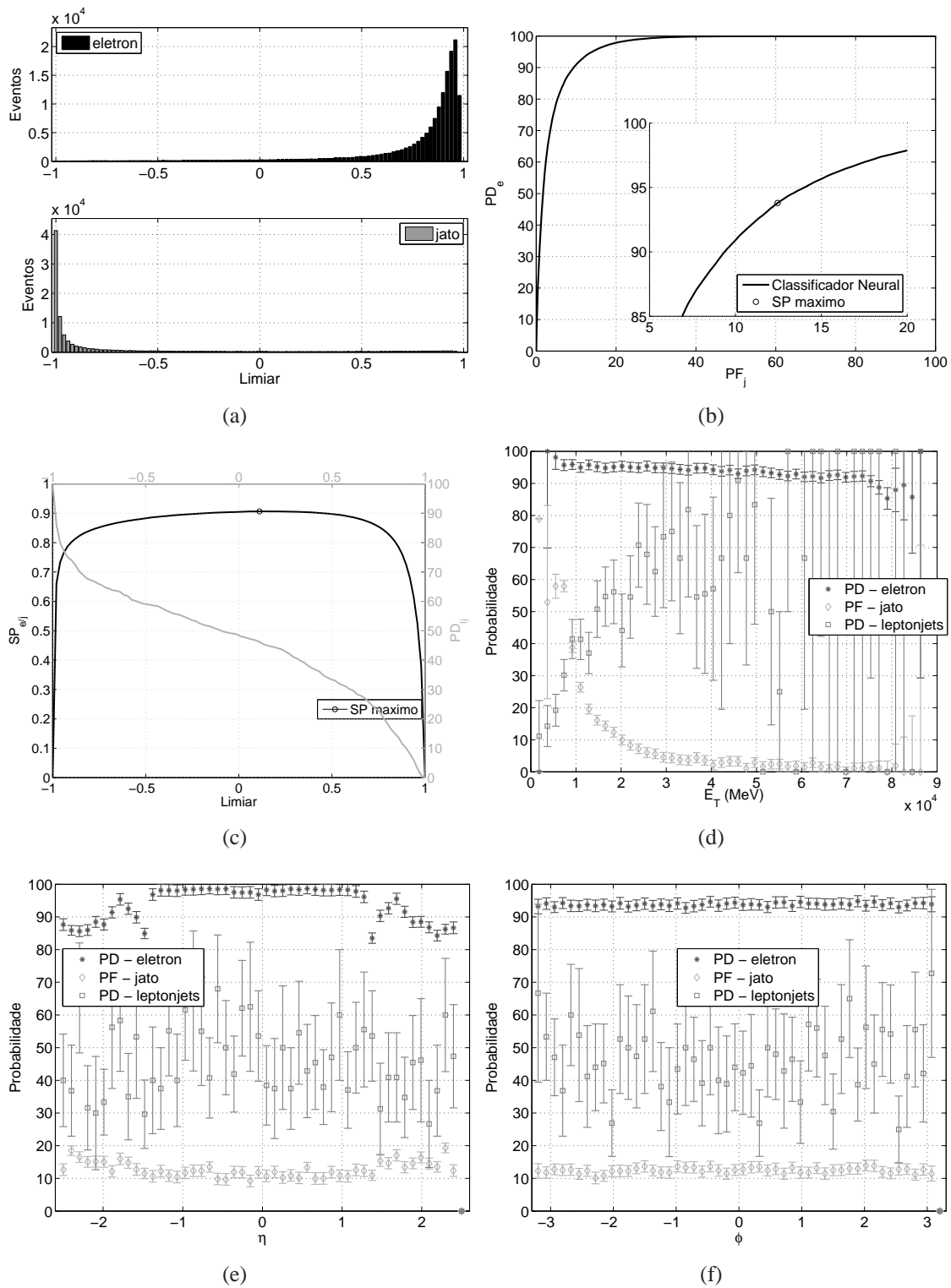


Figura 5.17: Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados sequencialmente. (a) Histograma de saída. (b) Curva ROC. (c) Produto  $SP_{e/j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

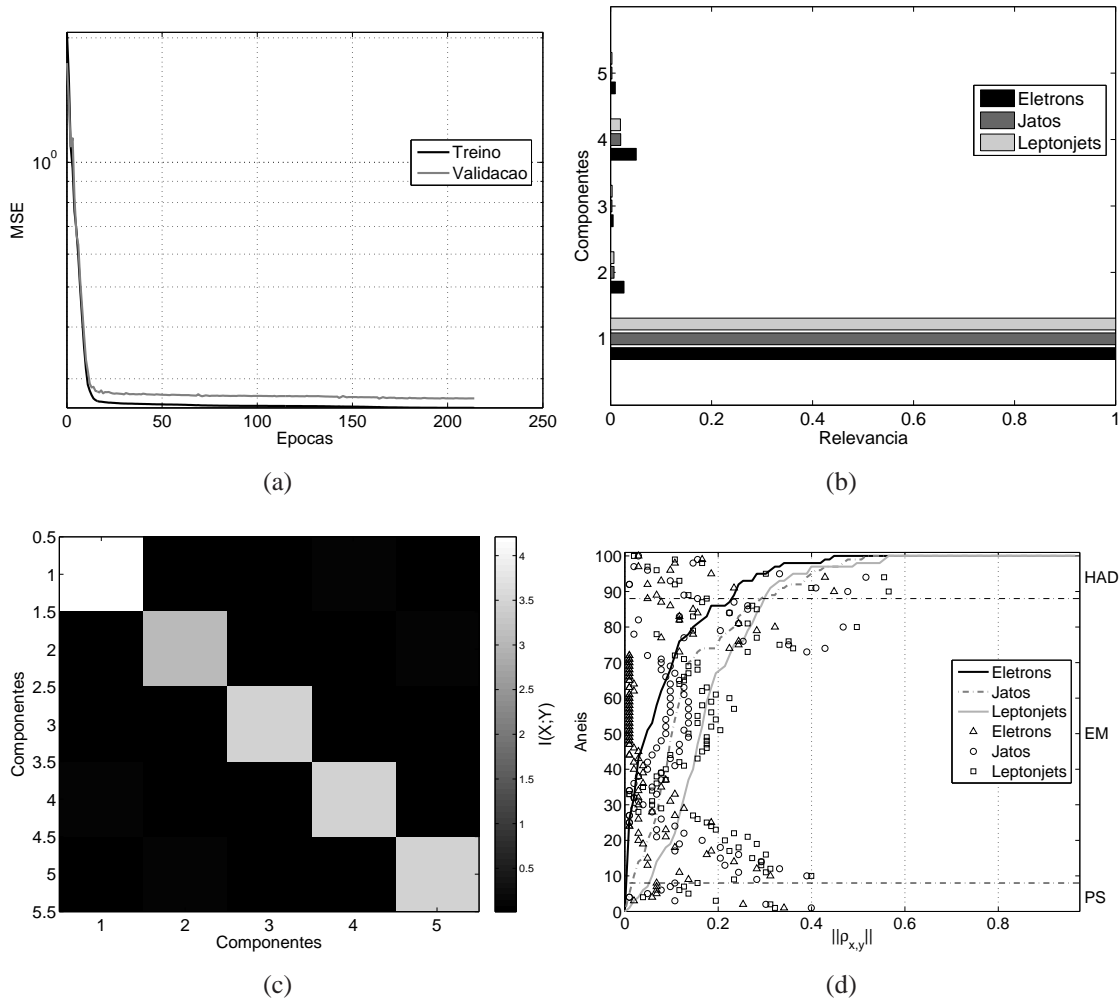


Figura 5.18: (a) Erro quadrático médio de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados sequencialmente: (b) Relevância das componentes independentes no processo decisório implementado; (c) Informação Mútua entre as componentes independentes extraídas; (d) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes.

### 5.4.1 PCA

Em analogia à curva de carga apresentada na Figura 5.6 para o pré-processamento não segmentado, apresentamos, aqui, as correspondentes curvas em relação a cada camada do calorímetro. As Figuras 5.19, 5.20, 5.21 e 5.22 mostram a energia acumulada para as diferentes camadas dos calorímetros em função do número de componentes retidas. O que se observa com essas curvas de cargas é que o comportamento de cada evento não é homogêneo entre as camadas. Apesar de, em uma maneira geral, o elétron ser mais pontual na deposição de sua energia, o jato hadrônico ser mais espalhado e os *leptonjets* serem um evento intermediário, especificamente, as camadas dos calorímetros fazem leituras peculiares de cada um deles.

Ao analisarmos os dados sob o ponto de vista da normalização por energia, verificamos que a camada *Pre Sampler*, apesar de ser composta de sensores extremamente finos (como faixas) a frente do calorímetro eletromagnético, consegue diferenciar bem as três classes de eventos (Figura 5.19(a)). A técnica PCA, operando sobre os dados amostrados nessa camada, fornece o resultado esperado, i.e., dado um mesmo percentual de energia retida do sinal, um número menor de componentes é requerido para a representação do elétron enquanto o jato hadrônico exige um número maior de componentes. O mesmo é observado com a camada EM *front layer* (Figura 5.19(b)). Para a segunda camada eletromagnética, EM *middle layer* (Figura 5.20(a)), o que se observa é que a diferenciação entre jatos hadrônicos e *leptonjets* não é tão acentuada. No caso da última camada eletromagnética, EM *back layer* (Figura 5.20(b)), a diferenciação entre os eventos vem com algumas peculiaridades. O elétron necessita de um número maior de componentes retidas para uma correta representação nessa camada. Acredita-se que boa parte da energia do evento já tenha sido absorvida nas primeiras camadas eletromagnéticas, restando apenas a energia do ruído para ser absorvida, o que o torna um evento muito complexo do ponto de vista de representatividade. O jato hadrônico ainda deposita uma dose de energia nessa camada - resultado de diferentes decaimentos intra jato - enquanto os *leptonjets*, por serem mais de uma partícula em um jato colimado, ainda não tiveram toda a sua componente eletromagnética absorvida pelo calorímetro. Na seção hadrônica do calorímetro, especificamente na primeira camada, HAD *layer 0* (Figura 5.21(a)), o comportamento das curvas de carga dos eventos analisados se assemelha ao da última camada eletromagnética. O que se esperava, nesse ponto, era que os *leptonjets* apresentassem uma curva de carga mais próxima da apresentada pelo elétron isolado, dado que para uma partícula eletromagnética toda sua energia é absorvida nas camadas anteriores. Além disso, essa camada do calorímetro apresenta uma resolução mais baixa, i.e., com menos detalhes. Nesse ponto, o jato hadrônico é um evento mais rico do ponto de vista da topologia e da energia, mas sob as condições de resolução espacial dessa camada a informação topológica parece ser sub-utilizada. De toda maneira, para um determinado valor de energia retida, elétron

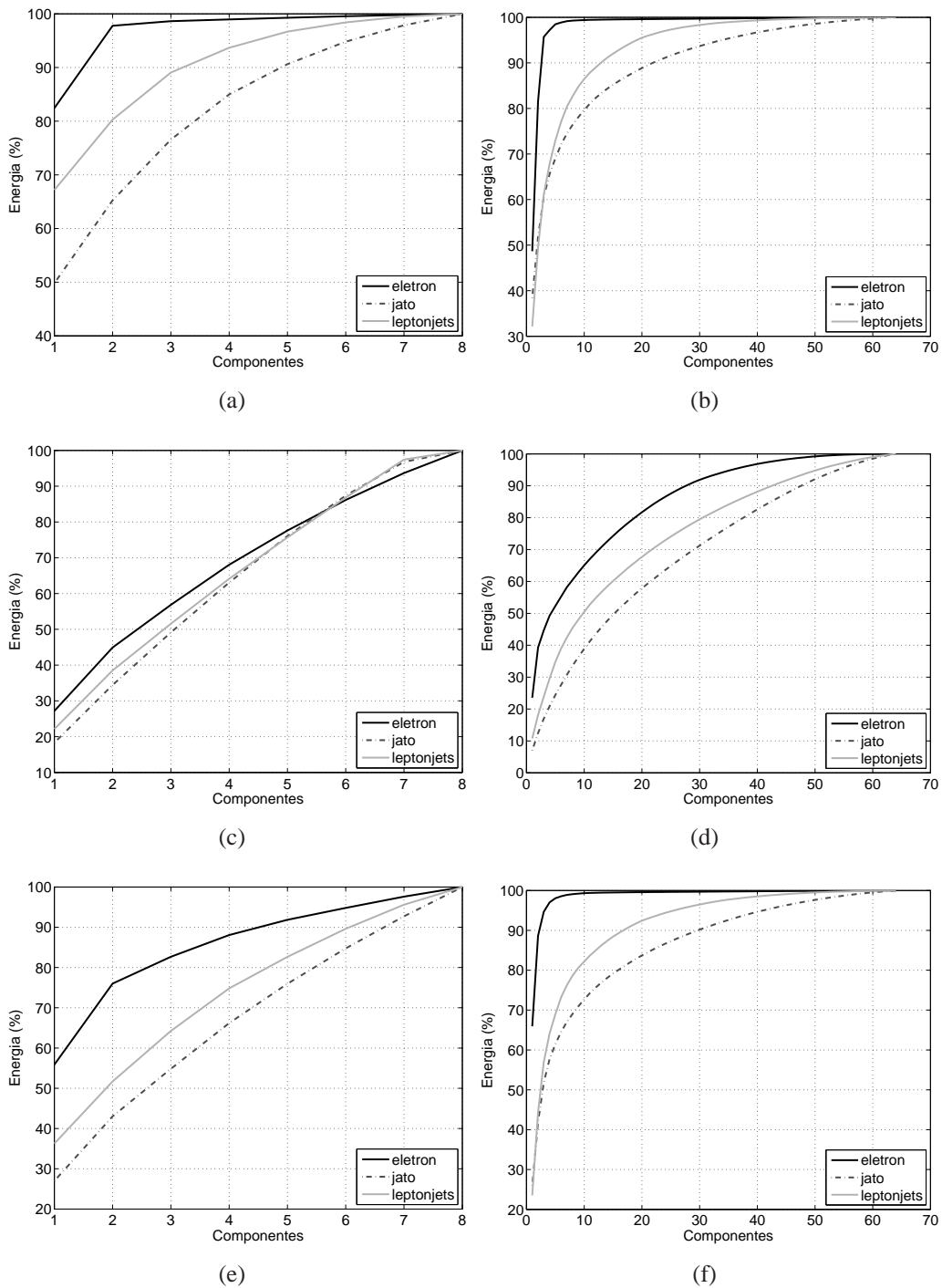


Figura 5.19: Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) *Pre Sampler*; (b) *EM front layer*. Normalização sequencial: (c) *Pre Sampler*; (d) *EM front layer*. Normalização por camada: (e) *Pre Sampler*; (f) *EM front layer*.



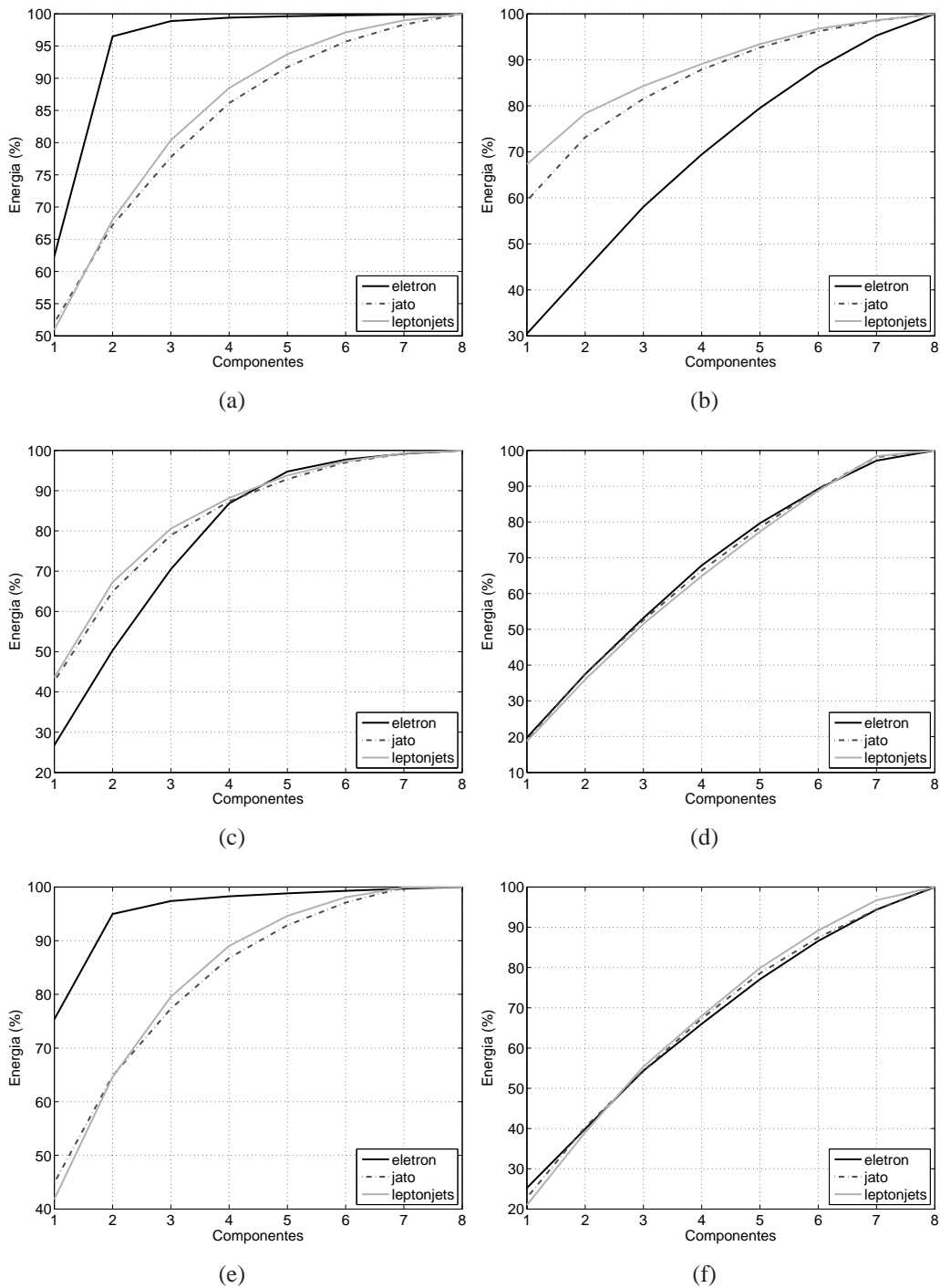


Figura 5.20: Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) EM *middle layer*; (b) EM *back layer*. Normalização sequencial: (c) EM *middle layer*; (d) EM *back layer*. Normalização por camada: (e) EM *middle layer*; (f) EM *back layer*.

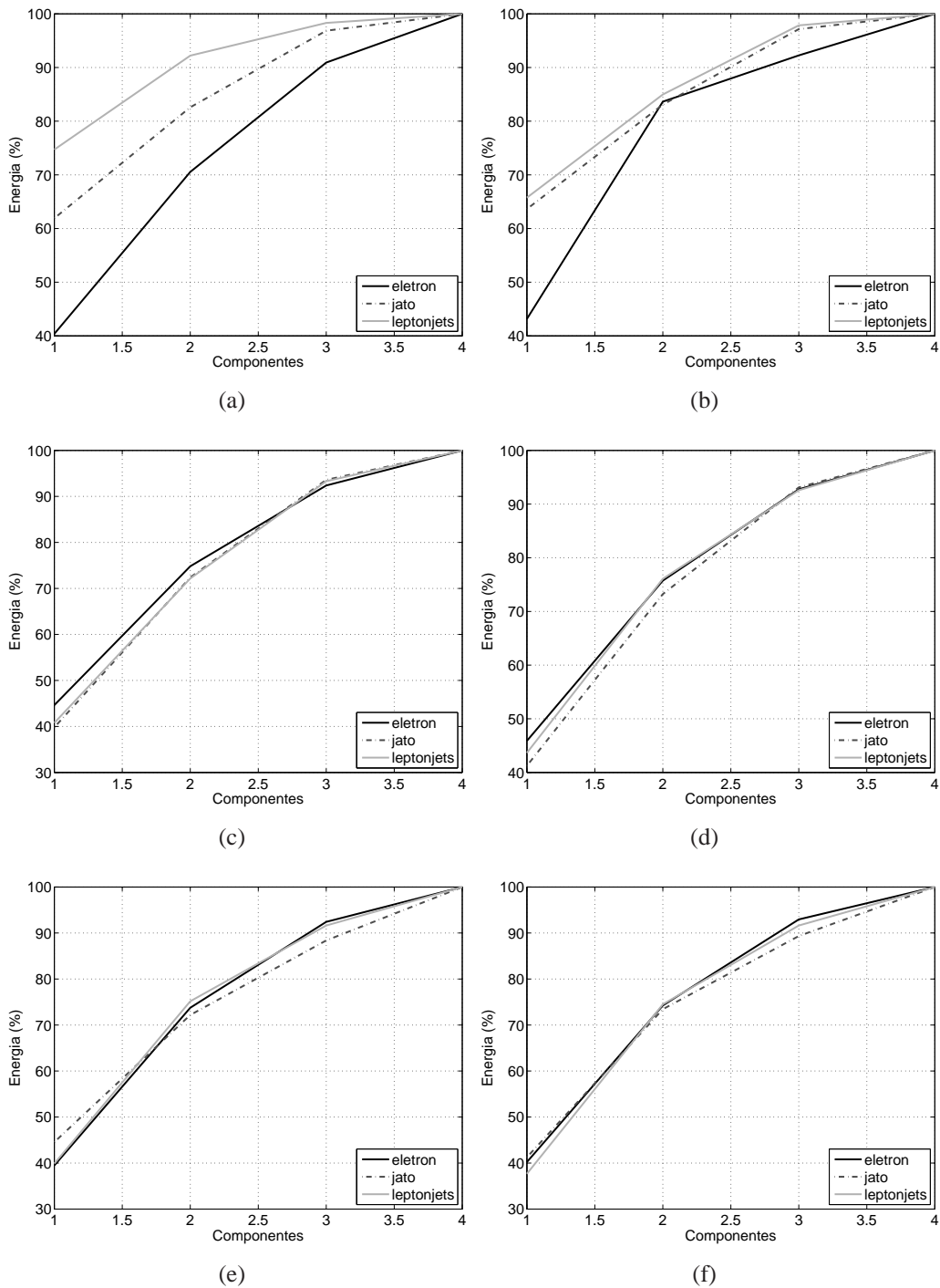
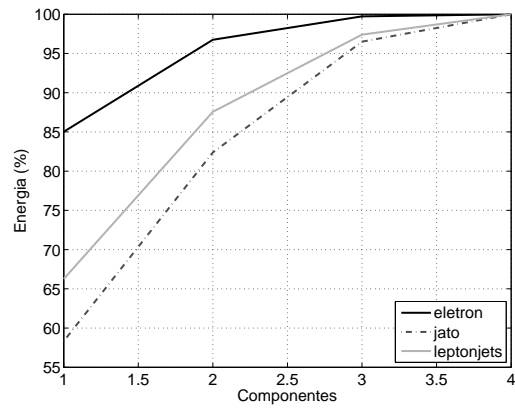
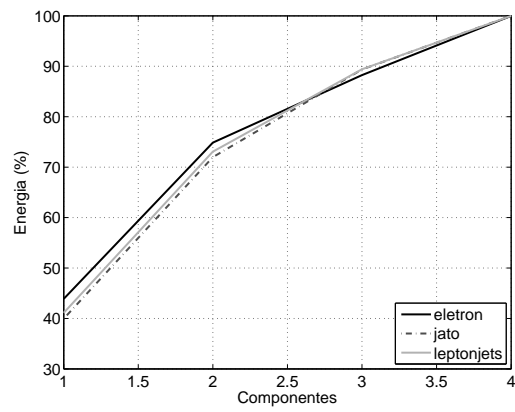


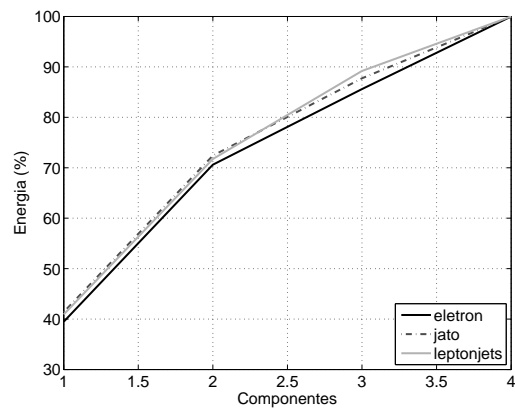
Figura 5.21: Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) *HAD layer 0*; (b) *HAD layer 1*. Normalização sequencial: (a) *HAD layer 0*; (b) *HAD layer 1*. Normalização por camada: (a) *HAD layer 0*; (b) *HAD layer 1*.



(a)



(b)



(c)

Figura 5.22: Curva de carga por camada do calorímetro em função da normalização usada. Normalização por energia: (a) *HAD layer 2*. Normalização sequencial: (b) *HAD layer 2*. Normalização por camada: (c) *HAD layer 2*.

e *leptonjets* apresentam uma sobreposição de curvas, o que é extremamente desejável, pois garante o mesmo nível de reconstrução desses eventos em função das componentes principais extraídas.

Sob a ótica da normalização sequencial, observamos uma aproximação do perfil das curvas de carga entre os eventos estudados. O efeito não é particular a uma camada ou seção do calorímetro, ele é observado nas diferentes curvas de carga mostradas. Na camada *Pre Sampler* (Figura 5.19(c)), a diferença no número de componentes necessárias para a representação dos eventos, dado um mesmo percentual de energia retida, é praticamente a mesma ao longo da curva. A exceção encontra-se no intervalo final, em que os normalizadores já não agem, implicando uma sobreposição. Na região ótima, o resultado é uma aproximação maior dos perfis desejados, elétron e *leptonjets*, mas com algum grau de separabilidade dos jatos hadrônicos. De maneira geral, enxerga-se o efeito da normalização sobre os dados dessa camada como algo que auxilia o classificador na tarefa final. Para a camada EM *front layer* (Figura 5.19(d)) se observa algo semelhante. Com exceção do início e do fim da curva, o efeito da normalização aproxima elétron e *leptonjets*, de maneira que a informação produzida nessa camada contribui de maneira significativa para o desempenho final obtido pelos classificadores. Nas duas últimas camadas do calorímetro eletromagnético (Figuras 5.20(c) e 5.20(d)) não é esperado uma eficiência tão significativa. Das camadas hadrônicas (Figuras 5.21(c), 5.21(d) e 5.22(b)) a que apresenta uma maior aproximação entre os perfis de elétron e *leptonjets* é a camada HAD *layer 0* (Figura 5.21(d)). Nesse caso observa-se a mesma e desejada sobreposição das curvas de carga já mencionada com a normalização por energia.

No caso da normalização por camada, o que se observa é uma aproximação do que foi observado com a normalização por energia. O normalizador aplicado em cada camada, principalmente nas camadas eletromagnéticas, atenua o sinal localmente, mas gera eventos com o mesmo perfil do observado com a normalização por energia. A diferença começa a se destacar a partir da última camada eletromagnética, EM *back layer* (Figura 5.20(f)), quando os níveis de energia local são bem pequenos, alterando o peso do normalizador sobre os dados.

Diante destas informações, a estratégia implementada dentro do conceito da abordagem segmentada foi denominada aglomerativa. Nessa estratégia, o uso da PCA como extrator local de característica foi implementado em cada camada dos calorímetros e essa informação apresentada para um algoritmo de classificação. O objetivo foi fornecer uma estimativa do potencial de separação da informação de cada camada isoladamente. Posteriormente, as componentes que forneceram os melhores resultados foram combinados em um só classificador.

A estratégia da abordagem segmentada foi realizada com o uso de classificadores lineares e neurais. A Tabela 5.7 apresenta os melhores resultados do classificador linear de Fisher ao lidar isoladamente com as componentes extraídas em cada camada.

Tabela 5.7: Melhor desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Norm.	Camada	Energia(%) <sup>5</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
Energia	<i>Pre Sampler</i>	94,92	97,40	38,86	0,7821	0,37	49,43	0,6791
	<i>EM front layer</i>	80,08	89,90	15,04	0,8741	-0,56	46,58	0,7232
	<i>EM middle layer</i>	98,09	93,86	8,74	0,9256	0,23	13,69	0,5695
	<i>EM back layer</i>	98,34	95,99	36,29	0,7902	0,66	73,00	0,7700
	<i>HAD layer 0</i>	84,72	96,15	26,92	0,8422	0,58	68,06	0,7865
	<i>HAD layer 1</i>	97,21	92,63	39,95	0,7545	0,39	72,05	0,7432
	<i>HAD layer 2</i>	98,11	84,56	54,27	0,6365	0,25	69,68	0,6561
Sequencial	<i>Pre Sampler</i>	38,69	92,74	41,87	0,7442	-0,18	56,08	0,6804
	<i>EM front layer</i>	37,96	96,70	5,92	0,9539	0,00	17,02	0,6099
	<i>EM middle layer</i>	59,71	95,99	29,23	0,8290	-0,45	45,91	0,6934
	<i>EM back layer</i>	97,60	80,00	30,54	0,7464	-0,13	49,81	0,6579
	<i>HAD layer 0</i>	92,87	82,55	28,31	0,7703	-0,31	60,27	0,7121
	<i>HAD layer 1</i>	92,76	72,88	36,22	0,6825	-0,30	62,07	0,6616
	<i>HAD layer 2</i>	88,76	62,01	41,64	0,6017	-0,23	51,62	0,5724
Camada	<i>Pre Sampler</i>	91,08	92,25	33,08	0,7908	0,03	43,25	0,6592
	<i>EM front layer</i>	43,42	92,77	10,66	0,9104	-0,35	16,35	0,5829
	<i>EM middle layer</i>	91,70	92,88	7,11	0,9289	0,49	10,27	0,5397
	<i>EM back layer</i>	94,37	76,62	33,65	0,7139	0,05	42,59	0,6094
	<i>HAD layer 0</i>	74,89	72,01	25,03	0,7348	-0,37	52,38	0,6604
	<i>HAD layer 1</i>	39,14	69,63	35,63	0,6697	-0,29	61,03	0,6496
	<i>HAD layer 2</i>	86,67	66,99	44,28	0,6123	-0,43	55,13	0,5916

Observa-se que isoladamente a primeira camada hadrônica é a que fornece um melhor resultado final na diferenciação dos eventos, independentemente da normalização. Já a camada com menor expressão no resultado final variou de normalização para normalização. De maneira geral, a normalização por energia foi a que apresentou números mais expressivos.

Aproveitando os resultados mostrados na Tabela 5.7, implementou-se um classificador linear cujas entradas são as componentes nela descritas. O intuito foi verificar se essas componentes, que pontualmente apresentaram desempenhos expressivos, conjuntamente poderiam produzir resultados ainda melhores. Parte-se das componentes pertencentes a camada com melhor desempenho e, gradativamente, se agrega as componentes das camadas sub-sequentes. O procedimento é interrompido quando o acréscimo de componentes piora o resultado anterior. A Tabela 5.8 mostra os resultados do melhor desempenho para cada uma das normalizações. Constata-se que a agregação da informação de algumas camadas, no processo decisório, implica uma melhora do resultado final para os dados normalizados por energia, enquanto os valores obtidos com a normalização sequencial e a normalização por camada indicam que, pontualmente, algumas camadas complicam a tarefa de classificação mais do que ajudam. Esse resultado pode sinalizar a necessidade da supressão da informação de alguma camada - sob supervisão de um especialista - em

<sup>5</sup>O percentual de energia se refere ao acumulado nas componentes principais.

detrimento de um ganho de desempenho. O melhor classificador foi obtido com as informações das componentes principais extraídas da última camada eletromagnética e das duas primeiras camadas hadrônicas. Além disso, a configuração final obtida indica um classificador muito simples do ponto de vista computacional. As doze variáveis de entrada representam um espaço de dimensão bem inferior ao original. A Figura 5.23 detalha o comportamento do classificador final obtido. Os histogramas de saída (Figura 5.23(a)) sugerem que a tarefa de classificação nesse espaço de dimensão reduzida é bastante complexa. Os padrões de elétrons e jatos hadrônicos ficam muito próximos na régua de decisão, o que resulta em um classificador com margem de decisão (limiar de funcionamento) muito estreita, o que é considerado um ponto fraco dessa abordagem. Mesmo com essa peculiaridade, o resultado final obtido é o melhor em comparação com as outras abordagens realizadas. A Figura 5.23(b) mostra a curva ROC do classificador. A Figura 5.23(c) mostra o comportamento do produto  $SP_{e/j}$  e da detecção de *leptonjets*,  $PD_{lj}$ , em função do limiar de decisão. As Figuras 5.23(d), 5.23(e) e 5.23(f) exibem, respectivamente, a eficiência do classificador em função da  $E_T$ , da coordenada  $\eta$  e da coordenada  $\phi$  do evento. Nota-se que para o limiar de trabalho escolhido não há qualquer tipo de tendência nos resultados do classificador. A maior complicação na detecção dos eventos de interesse é na faixa de energia de  $E_T < 10$  GeV e quando os eventos se depositam próximos da região de *crack*. Espera-se que o uso de um classificador neural possa minimizar essa perda de eficiência caracterizada.

Para o classificador neural, a Tabela 5.9 resume os melhores resultados. No estudo do classificador neural foi variado o número de neurônios na camada escondida da rede. Um mínimo de 2 e um máximo de 5 neurônios foi estipulado para todas as camadas, com exceção das camadas hadrônicas, onde a dimensão é menor. Nessas camadas a rede neural foi fixada com 2 neurônios escondidos.

Assim como feito com o classificador linear, as componentes que produziram o melhor resultado, foram apresentadas a um único classificador neural. A variação na topologia do classificador neural foi feita em relação ao número de neurônios na camada escondida da rede, de 2 a 5 neurônios. O resultado é mostrado na Tabela 5.10.

Os números obtidos são superiores aos apresentados pelo classificador linear, independentemente da normalização. O melhor classificador opera com dados da última camada eletromagnética e da primeira camada hadrônica normalizados por energia. Verifica-se que 3 componentes principais da camada EM *back layer* e 3 componentes principais da camada HAD *layer 0* alimentando um classificador neural com topologia 6 – 5 – 1 (5 neurônios na camada escondida) são suficientes para realizar a tarefa desejada, i.e., encontrou-se uma abordagem com alternativa ao T2Ca1o na seleção do evento de interesse do canal, mas que permite um paralelo no que diz respeito a informação usada nessa seleção. Os dados provenientes da primeira camada hadrônica podem representar um

---

<sup>6</sup>O percentual de energia se refere ao acumulado nas componentes principais.

Tabela 5.8: Desempenho do classificador linear de Fisher operando sobre as componentes principais extraídas de forma segmentada mostradas na Tabela 5.7.

Normalização	Componentes							PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
	Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2						
Energia	0	0	0	7	2	3	0	94,57	15,11	0,8966	0,66	60,74	0,7939
Sequencial	5	0	2	7	3	3	0	92,76	12,25	0,9024	-0,25	37,45	0,6993
Camada	6	0	0	0	2	1	0	88,39	14,60	0,8689	-0,09	42,78	0,7037

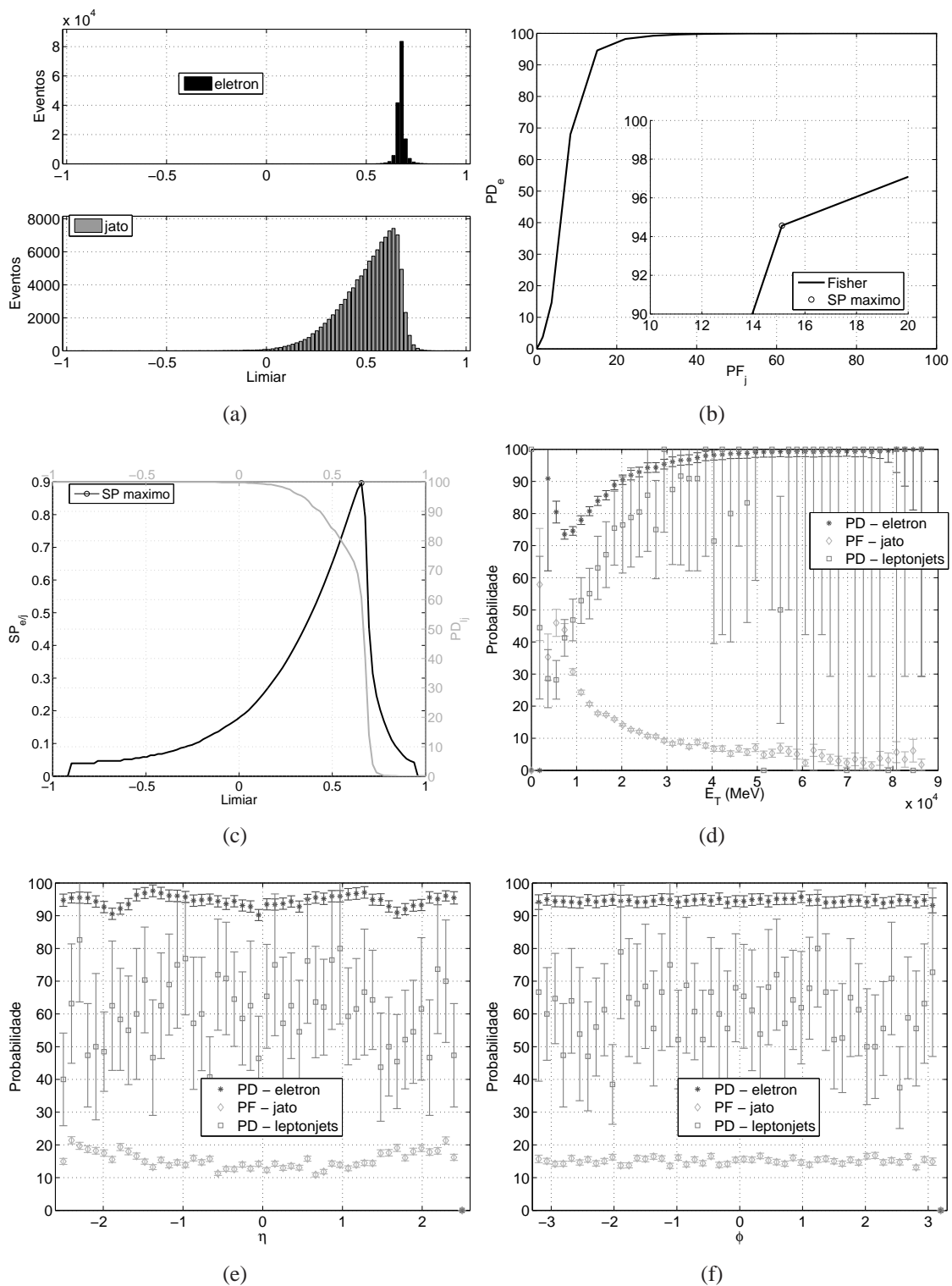


Figura 5.23: Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e|j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.



Tabela 5.9: Desempenho do classificador neural operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Norm.	Camada	Neur.	Comp.	Energia(%) <sup>6</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>ij</sub>	SP <sub>e/j/lj</sub>
Energia	<i>Pre Sampler</i>	3	3	92,08	94,75	36,31	0,7844	0,13	48,48	0,6766
	<i>EM front layer</i>	4	2	80,08	89,97	14,78	0,8758	0,07	45,72	0,7206
	<i>EM middle layer</i>	2	1	74,08	87,09	11,61	0,8774	0,17	17,11	0,5715
	<i>EM back layer</i>	3	3	81,67	90,32	27,09	0,8139	0,27	62,74	0,7490
	<i>HAD layer 0</i>	2	3	97,29	93,42	19,31	0,8694	0,31	60,93	0,7775
	<i>HAD layer 1</i>	2	3	97,21	87,30	33,32	0,7664	0,27	65,11	0,7270
Sequencial	<i>HAD layer 2</i>	2	2	89,81	70,54	42,92	0,6363	0,19	59,89	0,6237
	<i>Pre Sampler</i>	3	2	38,69	88,29	35,79	0,7577	0,25	49,71	0,6648
	<i>EM front layer</i>	2	2	31,49	94,73	7,02	0,9385	0,13	18,44	0,6123
	<i>EM middle layer</i>	3	1	39,53	96,59	28,86	0,8338	-0,37	46,48	0,6986
	<i>EM back layer</i>	3	7	97,60	82,53	31,87	0,7516	0,11	54,56	0,6792
	<i>HAD layer 0</i>	2	3	92,87	82,10	27,87	0,7704	0,07	60,17	0,7118
Camada	<i>HAD layer 1</i>	2	3	92,76	72,56	35,70	0,6836	0,01	61,98	0,6620
	<i>HAD layer 2</i>	2	3	88,76	58,11	38,38	0,5985	0,15	48,57	0,5596
	<i>Pre Sampler</i>	5	2	64,10	88,41	28,55	0,7970	0,21	44,30	0,6672
	<i>EM front layer</i>	4	1	43,42	93,45	11,29	0,9107	0,19	17,87	0,5940
	<i>EM middle layer</i>	4	5	97,41	94,24	5,91	0,9416	0,27	13,02	0,5717
	<i>EM back layer</i>	4	7	94,37	79,66	33,28	0,7305	0,17	43,16	0,6219
	<i>HAD layer 0</i>	2	3	90,88	74,02	21,93	0,7603	0,13	50,95	0,6710
	<i>HAD layer 1</i>	2	3	91,74	68,76	32,25	0,6826	0,17	59,03	0,6511
	<i>HAD layer 2</i>	2	1	40,75	71,47	58,42	0,5551	0,17	68,54	0,5967

corde similar ao HADE<sub>T</sub> do T2Ca10, enquanto a informação da última camada eletromagnética pode auxiliar a estrutura interna de decisão do classificador a construir elementos de diferenciação, como os cortes EME<sub>T</sub> ou Rshape do T2Ca10.

A Figura 5.24 mostra o comportamento do classificador para as diferentes figuras de mérito avaliadas. Os histogramas de saída (Figura 5.24(a)) mostram a separação alcançada entre elétron e jatos hadrônicos (contornando o problema identificado com o uso do classificador linear, Figura 5.23(a)). A rede neural foi capaz de isolar consideravelmente esses padrões. Isso se reflete em uma curva ROC suave (Figura 5.24(b)) e em um regime de operação (intervalo do limiar de decisão) mais folgado, como mostram o produto SP<sub>e/j</sub> e a eficiência de detecção de *leptonjets* (Figura 5.24(c)).

Considerando a eficiência em energia (Figura 5.24(d)), observa-se que, a abordagem apresenta alguns pontos positivos: a curva de detecção cresce, tanto para elétron quanto para *leptonjets*, com o aumento da energia do evento, enquanto o falso alarme apresenta uma lógica inversa. Essa característica permite prever e calibrar o comportamento do classificador em função da faixa de energia dos eventos analisados. Para energias muito baixas ( $E_T < 10$  GeV), a eficiência na detecção do elétron é relativamente baixa e a de *leptonjets* é baixa. Além disso, o falso alarme é considerado alto. Na faixa de energia seguinte ( $10 \text{ GeV} < E_T < 20$  GeV), as curvas de eficiência mostram, de maneira acentuada, valores mais próximos dos desejados para a operação de filtragem de nível 2. A partir desse limiar ( $E_T > 20$  GeV), o comportamento do classificador é melhor e está em linha com o observado para a maioria das abordagens bem sucedidas testadas. Esses resultados são respaldados também pela eficiência obtida em função das coordenadas  $\eta$  e  $\phi$  dos eventos. O comportamento praticamente uniforme elimina qualquer tendência na

Tabela 5.10: Desempenho do classificador neural operando sobre as componentes principais extraída de forma segmentada mostradas na Tabela 5.9.

Normalização	Neurônios	Componentes							PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	PD <sub>e/j/lj</sub>
		Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2						
Energia	5	0	0	0	3	3	0	0	94,85	15,21	0,8975	0,27	61,31	0,7966
Sequencial	2	2	0	1	7	3	3	0	94,23	11,89	0,9115	0,13	40,87	0,7204
Camada	4	2	0	0	0	3	3	0	90,78	11,87	0,8945	0,13	44,39	0,7260

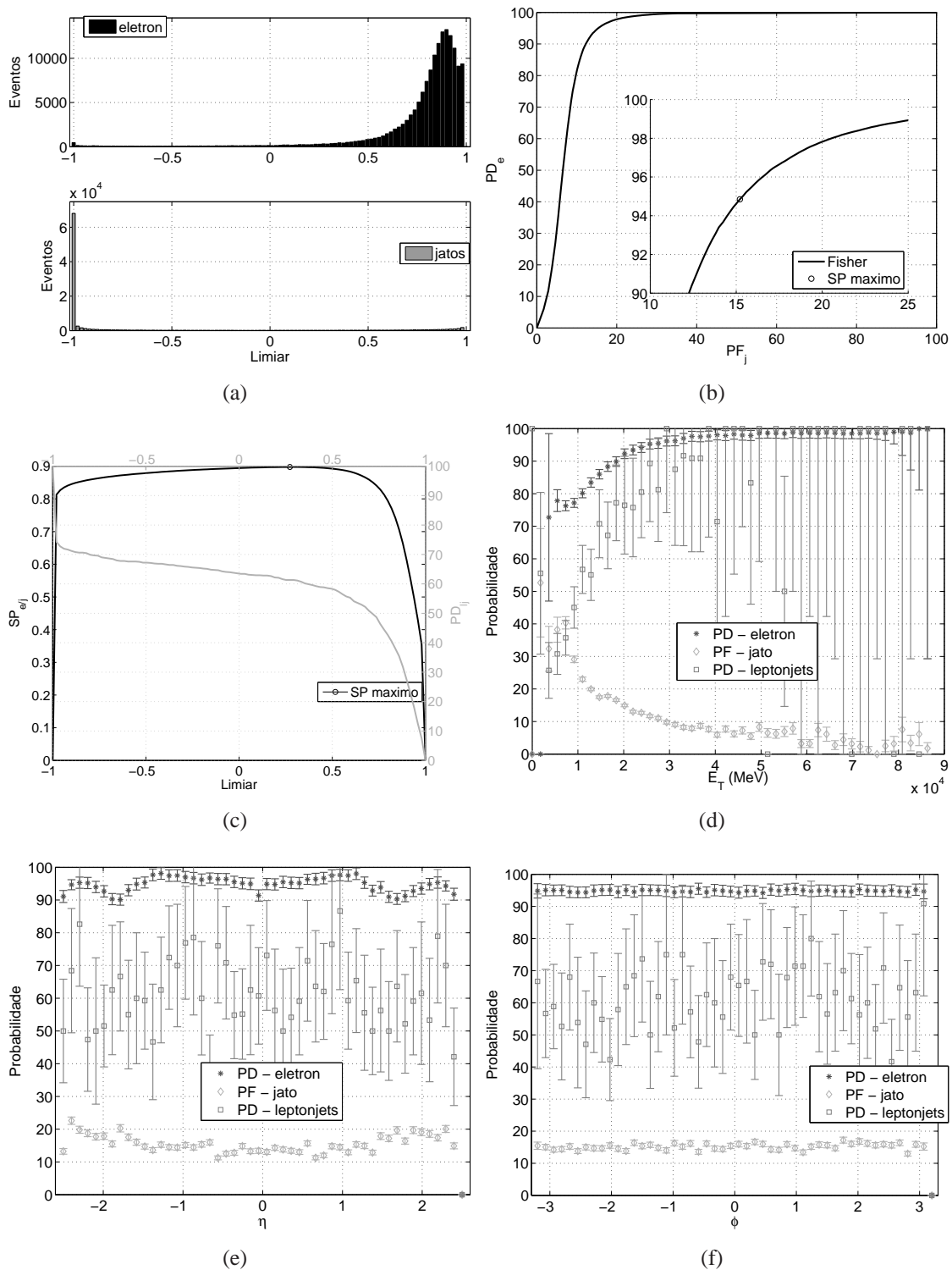


Figura 5.24: Desempenho do classificador neural operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e|j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

Tabela 5.11: Melhor desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Norm.	Camada	Energia(%) <sup>7</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
Energia	<i>Pre Sampler</i>	99,21	97,07	37,09	0,7906	0,38	48,00	0,6786
	<i>EM front layer</i>	88,91	90,31	15,05	0,8761	-0,58	45,82	0,7212
	<i>EM middle layer</i>	89,95	87,38	11,93	0,8772	-0,28	17,87	0,5767
	<i>EM back layer</i>	98,34	94,75	33,48	0,8001	0,66	69,87	0,7656
	<i>HAD layer 0</i>	97,29	95,58	22,36	0,8638	0,52	63,31	0,7829
	<i>HAD layer 1</i>	83,41	94,84	47,73	0,7197	0,56	77,66	0,7383
	<i>HAD layer 2</i>	98,11	76,66	44,57	0,6562	0,26	60,08	0,6375
Sequencial	<i>Pre Sampler</i>	38,69	93,01	42,30	0,7430	-0,18	57,41	0,6845
	<i>EM front layer</i>	37,96	96,80	6,01	0,9539	-0,02	17,21	0,6113
	<i>EM middle layer</i>	39,53	96,32	28,61	0,8338	-0,78	45,82	0,6959
	<i>EM back layer</i>	97,60	80,05	30,56	0,7465	-0,42	49,81	0,6580
	<i>HAD layer 0</i>	92,87	76,80	22,74	0,7703	-0,22	53,80	0,6881
	<i>HAD layer 1</i>	92,76	71,15	34,61	0,6824	-0,26	61,03	0,6579
	<i>HAD layer 2</i>	88,76	63,01	42,60	0,6017	-0,38	52,00	0,5738
Camada	<i>Pre Sampler</i>	91,08	92,25	33,08	0,7908	0,03	43,25	0,6592
	<i>EM front layer</i>	43,42	92,77	10,66	0,9104	-0,35	16,35	0,5829
	<i>EM middle layer</i>	95,23	91,74	44,17	0,7267	0,27	51,71	0,6531
	<i>EM back layer</i>	94,37	77,58	34,50	0,7141	0,04	43,54	0,6134
	<i>HAD layer 0</i>	90,88	69,73	20,38	0,7459	-0,36	47,15	0,6473
	<i>HAD layer 1</i>	91,74	68,76	32,66	0,6805	-0,30	59,03	0,6497
	<i>HAD layer 2</i>	86,67	67,30	44,60	0,6121	-0,44	55,23	0,5918

detecção desses eventos.

Assim, é obtida uma satisfatória redução de dimensionalidade dos dados a serem classificados, uma adequada separação entre os padrões do canal elétron/jato, com o maior valor de aceitação de *leptonjets* entre os testes realizados sem qualquer presença de tendências por parte do classificador e uma topologia final de decisão considerada computacionalmente simples.

## 5.4.2 ICA

Seguindo a metodologia usada com uso da PCA, esta seção apresenta os resultados obtidos com o uso da ICA como extrator de características. É pertinente lembrar que os percentuais de energia retida a que se referem os resultados relacionam-se com a energia retida na etapa de branqueamento dos dados de cada camada. Como já dito, essa etapa é pertinente à técnica de análise de componentes independentes escolhida. Novamente, foram realizados testes com classificadores lineares e não-lineares.

Usando o classificador linear de Fisher em cada uma das 7 camadas dos calorímetros, obteve-se o número ideal de componentes retidas que maximizavam o produto  $SP_{e/j/lj}$ . A Tabela 5.11 mostra o desempenho obtido em cada camada dos calorímetros.

<sup>7</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.

Observa-se que com o uso da normalização por energia, a primeira camada do calorímetro hadrônico apresenta o melhor desempenho. Esse resultado já tinha sido observado com o uso da PCA como extrator de características. Contudo, o valor final indica que ao realizar a busca pela independência das componentes dessa camada a técnica ICA produz um resultado inferior ao obtido com a PCA. Um outro ponto observado, foi a divergência das camadas com melhor poder de diferenciação de eventos em cada uma das normalizações aplicadas. Com o uso da normalização sequencial a informação mais importante está na camada EM *middle layer*, enquanto com a normalização por camada a informação que se destaca provém da camada *Pre Sampler*. De maneira geral, os resultados em cada uma das camadas estão próximos ao obtido com a PCA. A normalização por energia foi a que apresentou números finais mais expressivos.

Aproveitando os resultados mostrados na Tabela 5.11, implementou-se um classificador linear cujas entradas são as componentes nela descritas (mesma metodologia aplicada com a PCA). O intuito foi verificar se essas componentes conjuntamente poderiam produzir resultados ainda melhores do que isoladamente. Parte-se das componentes pertencentes a camada com melhor desempenho e gradativamente se agrega as componentes das camadas sub-sequentes. A Tabela 5.12 mostra o melhor desempenho para cada uma das normalizações. Constatou-se que a agregação da informação de algumas camadas no processo decisório implica uma melhora significativa do resultado final.

O melhor classificador foi obtido com as informações das componentes independentes extraídas da última camada eletromagnética e das duas primeiras camadas hadrônicas. Além disso, a configuração final obtida indica um classificador muito simples do ponto de vista computacional. As onze variáveis de entrada representam um espaço de dimensão, aproximadamente, dez vezes inferior ao original. A Figura 5.25 ilustra o comportamento do classificador final obtido. Os histogramas de saída (Figura 5.25(a)) sugerem que a tarefa de classificação nesse espaço de dimensão reduzida é bastante complexa. Os padrões de elétrons e jatos hadrônicos ficam muito próximos na régua de decisão, o que resulta em um classificador com margem de decisão (limiar de funcionamento) muito estreita, o que pode representar uma complicação futura, além de sofrer fortes críticas por parte do especialista no sistema. Vale ressaltar, ainda, que como a estatística disponível para os *leptonjets* não é tão rica, uma base de dados mais ampla poderia reduzir o desempenho final do sistema. Mesmo com essa peculiaridade, o resultado obtido é um dos melhores já alcançados. A Figura 5.25(b) mostra a curva ROC do classificador. A Figura 5.25(c) mostra o comportamento do produto  $SP_{e_j}$  e da detecção de *leptonjets*,  $PD_{lj}$ , em função do limiar de decisão. As Figuras 5.25(d), 5.25(e) e 5.25(f) exibem, respectivamente, a eficiência do classificador em função da  $E_T$ , da coordenada  $\eta$  e da coordenada  $\phi$  do evento. Nota-se que para o limiar de trabalho escolhido não há qualquer tipo de tendência nos resultados do classificador. A maior complicação na detecção dos eventos de interesse está na faixa de energia de  $E_T < 10$  GeV, e quando os eventos se depositam próximos da

Tabela 5.12: Desempenho do classificador linear de Fisher operando sobre as componentes independentes extraídas de forma segmentada mostradas na Tabela 5.11.

Normalização	Componentes							PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>
	Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2						
Energia	0	0	0	7	1	3	0	94,48	15,02	0,8967	0,66	60,74	0,7939
Sequencial	2	0	1	7	3	3	0	92,71	12,81	0,8993	-0,31	38,12	0,7006
Camada	6	0	4	0	3	3	0	89,68	14,61	0,8752	0,05	41,83	0,7033

região de *crack*. Esses resultado sugere que o classificador não-linear pode explorar ainda mais as vantagens trazidas pela técnica ICA e tentar contornar o problema da separação dos padrões de elétrons e jatos na régua de decisão do classificador.

Para o classificador neural, a Tabela 5.13 resume os melhores resultados. Nesse estudo variou-se o número de neurônios na camada escondida da rede. Um mínimo de 2 e um máximo de 5 neurônios foi estipulado para todas as camadas, com exceção das camadas hadrônicas, onde a dimensão é menor. Nessas camadas a rede neural foi fixada com 2 neurônios escondidos.

Tabela 5.13: Desempenho do classificador neural operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Norm.	Camada	Neur.	Comp.	Energia(%) <sup>8</sup>	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>ij</sub>	SP <sub>e/j/ij</sub>
Energia	<i>Pre Sampler</i>	2	4	94,92	93,63	36,13	0,7803	0,07	52,57	0,6900
	<i>EM front layer</i>	3	1	54,96	88,79	15,20	0,8678	0,01	46,77	0,7202
	<i>EM middle layer</i>	3	2	89,95	87,61	11,93	0,8784	-0,01	17,30	0,5734
	<i>EM back layer</i>	5	3	81,67	91,22	30,13	0,8019	0,21	64,92	0,7493
	<i>HAD layer 0</i>	2	3	97,29	93,04	19,90	0,8645	0,33	61,50	0,7766
	<i>HAD layer 1</i>	2	3	97,21	85,93	35,17	0,7501	0,31	65,97	0,7194
	<i>HAD layer 2</i>	2	2	89,81	72,49	44,13	0,6391	0,27	61,79	0,6320
Sequencial	<i>Pre Sampler</i>	5	2	38,69	87,98	35,32	0,7588	0,19	50,19	0,6673
	<i>EM front layer</i>	2	2	31,49	94,88	7,11	0,9389	0,17	18,82	0,6151
	<i>EM middle layer</i>	5	5	92,37	89,25	28,94	0,7989	0,98	47,24	0,6804
	<i>EM back layer</i>	3	7	97,60	81,59	30,85	0,7524	0,17	53,52	0,6759
	<i>HAD layer 0</i>	2	3	92,87	80,73	26,57	0,7703	0,09	59,41	0,7090
	<i>HAD layer 1</i>	2	3	92,76	69,78	33,38	0,6819	0,19	59,79	0,6533
Camada	<i>HAD layer 2</i>	2	3	88,76	64,67	44,02	0,6025	0,15	55,32	0,5858
	<i>Pre Sampler</i>	3	2	64,10	88,94	30,94	0,7868	0,13	45,63	0,6665
	<i>EM front layer</i>	3	27	96,90	95,15	15,90	0,8954	0,17	32,51	0,6713
	<i>EM middle layer</i>	3	4	95,23	86,81	22,61	0,8204	0,21	25,67	0,5935
	<i>EM back layer</i>	5	7	94,37	78,59	32,23	0,7308	0,23	50,29	0,6500
	<i>HAD layer 0</i>	2	2	74,89	68,71	18,38	0,7502	0,27	46,96	0,6492
	<i>HAD layer 1</i>	2	3	91,74	68,73	32,46	0,6813	0,23	59,22	0,6509
<i>HAD layer 2</i>	2	1	40,75	71,47	58,42	0,5551	0,17	68,54	0,5967	

Assim como feito com o classificador linear, as componentes que produziram o melhor resultado, foram apresentadas a um único classificador neural. A variação na topologia do classificador neural foi feita em relação ao número de neurônios na camada escondida da rede, de 2 a 5 neurônios. O resultado é mostrado na Tabela 5.10.

Com exceção da normalização por camada, os números obtidos não são superiores aos apresentados pelo classificador linear. Isso pode indicar que a disposição dos dados favorece a queda em mínimos locais por parte do algoritmo de classificação em sua etapa de treinamento, apesar das múltiplas inicializações a que é submetida a rede neural. O melhor classificador opera com dados da última camada eletromagnética e da primeira camada hadrônica normalizados por energia. Especificamente, 3 componentes independentes da camada *EM back layer* e 3 componentes da camada *HAD layer 0*. A informação que alimenta um classificador neural com topologia 6 – 2 – 1 (2 neurônios na camada escondida) é suficiente para realizar a tarefa desejada.

<sup>8</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.

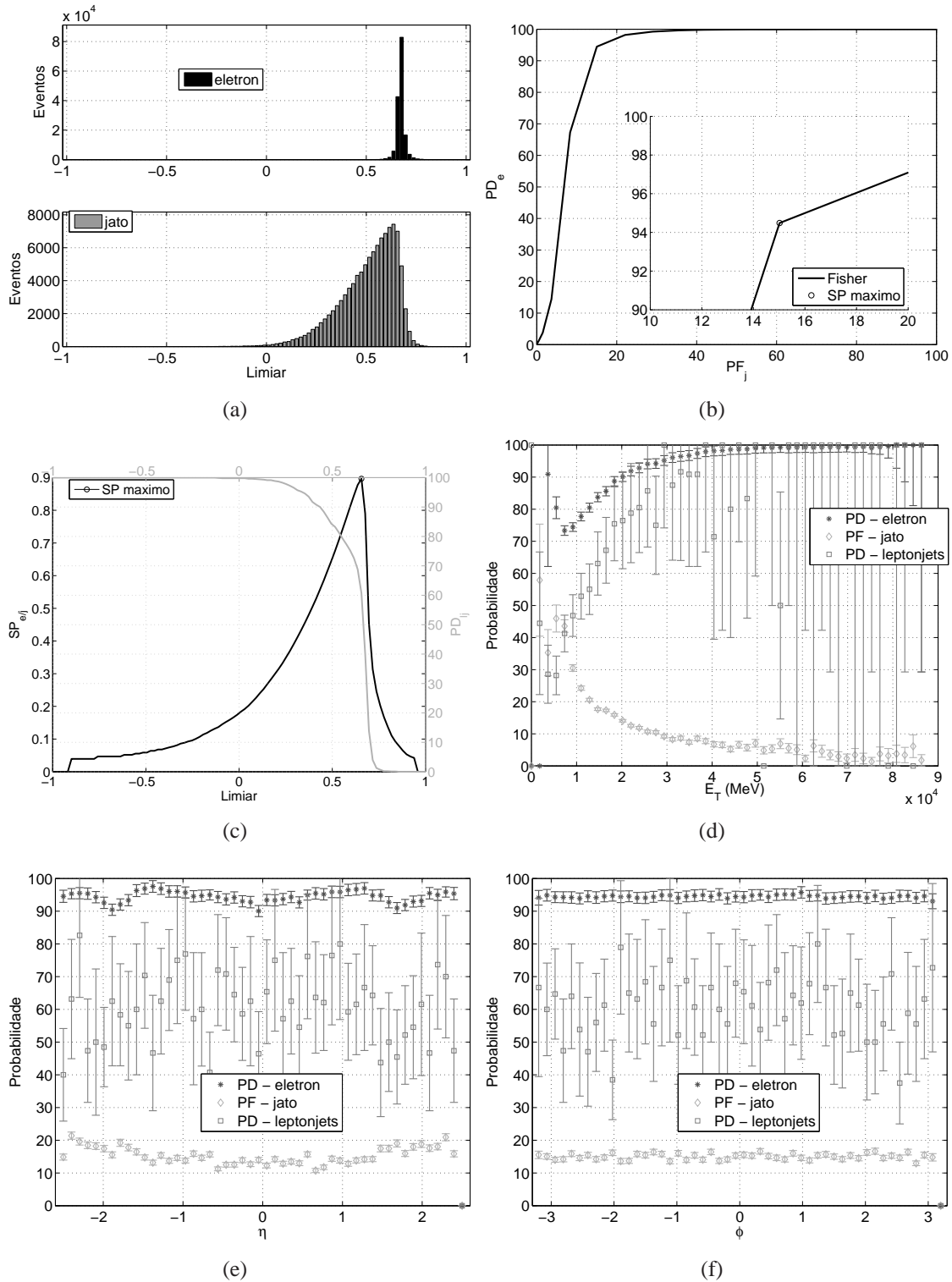


Figura 5.25: Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e/j}$  e probabilidade de detecção de  $leptonjets$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.



Tabela 5.14: Desempenho do classificador neural operando sobre as componentes independentes extraída de forma segmentada mostradas na Tabela 5.13.

Normalização	Neurônios	Componentes							PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	PD <sub>e/j/lj</sub>
		Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2						
Energia	2	0	0	0	3	3	0	0	94,72	14,54	0,9003	0,31	59,51	0,7914
Sequencial	4	2	0	5	7	3	3	0	95,69	9,12	0,9327	0,17	33,46	0,6971
Camada	2	2	27	0	7	2	3	1	96,37	6,89	0,9474	0,11	32,41	0,7001

A Figura 5.26 mostra o comportamento do classificador para as diferentes figuras de mérito avaliadas. Os histogramas de saída (Figura 5.26(a)) mostram a boa separação alcançada entre elétron e jatos hadrônicos. A rede neural foi capaz de isolar consideravelmente esses padrões. Isso se reflete em uma curva ROC suave (Figura 5.26(b)) e em um regime de operação (intervalo do limiar de decisão) mais folgado, como mostram o produto  $SP_{e/j}$  e a eficiência de detecção de *leptonjets* (Figura 5.26(c)).

Considerando a eficiência em energia (Figura 5.26(d)), observa-se que, a abordagem apresenta similaridades com o observado no uso da PCA: a curva de detecção cresce, tanto para elétron quanto para *leptonjets*, com o aumento da energia do evento, enquanto o falso alarme apresenta uma lógica inversa. Essa característica permite prever e calibrar o comportamento do classificador em função da faixa de energia dos eventos analisados. Para energias muito baixas ( $E_T < 10$  GeV), a eficiência na detecção do elétron é relativamente baixa e a de *leptonjets* é baixa. Além disso, o falso alarme é considerado alto. Na faixa de energia seguinte ( $10 \text{ GeV} < E_T < 20$  GeV), as curvas de eficiência mostram, de maneira acentuada, valores mais próximos dos desejados para a operação de filtragem de nível 2. A partir desse limiar ( $E_T > 20$  GeV), o comportamento do classificador é melhor e está em linha com o observado para a maioria das abordagens bem sucedidas testadas. Uma diferenciação encontrada em relação ao processamento com a PCA é o fato de o classificador ser mais sensível a região de *crack* do detector (Figura 5.26(e)). A queda na eficiência de detecção nessa região (minorada nas abordagens anteriores) é observada tanto para o elétron isolado quanto para os *leptonjets*. A informação resiliente nessa região dos calorímetros não é suficiente para a extração de bons descritores pela técnica ICA. Fica *a priori* afastada qualquer tendência na detecção desses eventos, dado o comportamento praticamente uniforme dos resultados para a coordenada  $\phi$  (Figura 5.26(e)).

Desta forma, obtém-se uma grande redução de dimensionalidade dos dados a serem classificados, uma boa separação entre os padrões do canal elétron/jato, com um excelente valor de aceitação dos eventos desejados, além de uma topologia final de decisão considerada computacionalmente simples.

## 5.5 Comparação

Dos testes implementados com diferentes abordagens podemos identificar potencialidades em cada um deles, pontos positivos e negativos são, ora encontrados na detecção de eventos de interesse do canal (elétron), ora encontrados na detecção dos eventos exóticos (*leptonjets*) buscados.

A Tabela 5.15 resume os valores finais alcançados, enquanto as Figuras 5.27 e 5.28 apresentam a curva ROC e a  $PD_{lj}$  dos correspondentes classificadores. A partir da tabela e das figuras é possível extrair algumas conclusões. Observa-se que os classificadores que operam alimentados apenas com a informação dos anéis de entrada, tanto o linear

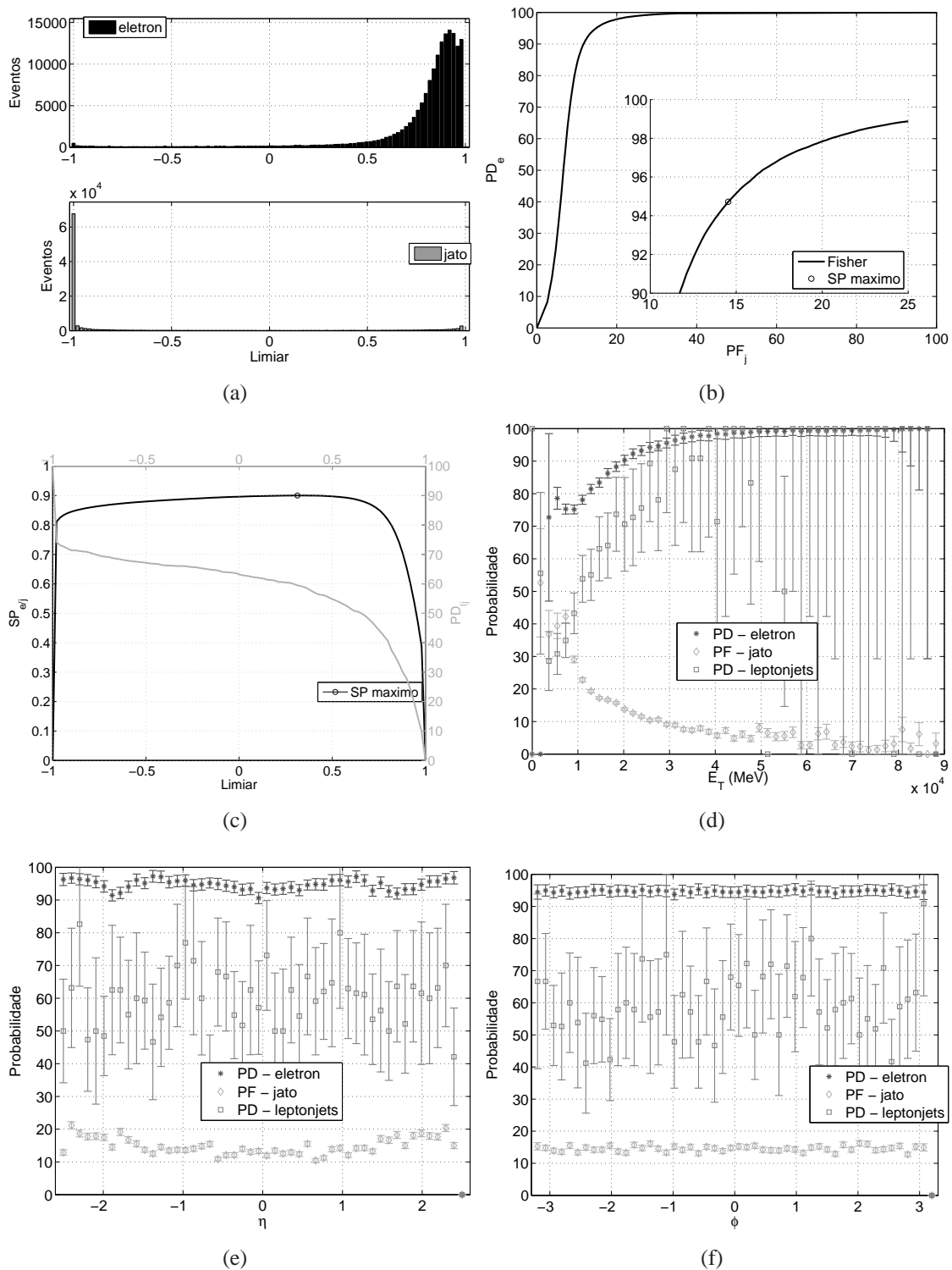


Figura 5.26: Desempenho do classificador neural operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{e/j}$  e probabilidade de detecção de *leptonjets*. (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

Tabela 5.15: Comparação dos resultados entre as diferentes abordagens propostas.

Técnica	Anéis(Comp.)	Energia(%)	Neurônios	PD <sub>e</sub>	PF <sub>j</sub>	SP <sub>e/j</sub>	limiar	PD <sub>lj</sub>	SP <sub>e/j/lj</sub>	PD <sub>e</sub> = T2Ca1o	
										PF <sub>e</sub>	PD <sub>lj</sub>
Fisher - Anéis	100			98,03	3,36	0,9734	-0,01	14,26	0,5978	1,58	5,04
RingerNeuralHypo	100		2	98,33	3,07	0,9763	0,25	13,21	0,5902	1,89	6,94
Fisher - PCA (Não Seg.)	(7)	62,98		91,59	14,72	0,8841	-0,21	49,05	0,7396	20,96	59,32
Neural - PCA (Não Seg.)	(5)	51,05	10	89,68	18,19	0,8570	0,13	55,42	0,7486	27,81	66,92
Fisher - ICA (Não Seg.)	(7)	62,98		91,59	14,72	0,8841	-0,21	49,05	0,7396	20,96	59,32
Neural - ICA (Não Seg.)	(5)	38,60	5	93,80	12,45	0,9064	0,11	46,20	0,7410	14,85	49,90
Fisher - PCA (Seg.)	(12)			94,57	15,11	0,8966	0,66	60,74	0,7939	22,17	68,82
Neural - PCA (Seg.)	(6)		5	94,85	15,21	0,8975	0,27	61,31	0,7966	16,31	62,83
Fisher - ICA (Seg.)	(11)			94,48	15,02	0,8967	0,66	60,74	0,7939	15,02	60,74
Neural - ICA (Seg.)	(6)		2	94,72	14,54	0,9003	0,31	59,51	0,7914	17,25	63,50
T2Ca1o				95,49	10,43	0,9251		29,75	0,6736		

quanto o neural (RingerNeuralHypo), são excelentes na tarefa de detectar elétron e rejeitar jatos hadrônicos (Figura 5.27). Seus números finais não são superados por qualquer outra abordagem, nem mesmo pelo algoritmo de referência T2Ca1o. Esses valores credenciam qualquer um dos classificadores a operarem nesse canal do detector ATLAS. A complexidade computacional dos dois algoritmos também é considerada simples. O algoritmo linear de Fisher, que tem suas características descritas no Apêndice A, é uma das estruturas de decisão mais simples de serem implementadas, enquanto um classificador neural com 2 neurônios na camada escondida da rede acrescenta, sutilmente, a não linearidade necessária na hipersuperfície de decisão para superar o desempenho do seu concorrente linear. Contudo, as duas abordagens apresentam valores considerados baixos para a detecção de *leptonjets*. A partir das Figura 5.28 verifica-se que é possível alterar o limiar de decisão e, por conseguinte, melhorar a aceitação de *leptonjets*, porém com efeitos negativos na rejeição de falso alarme.

Dos testes implementados com diferentes técnicas de análise de componentes, verificou-se que a extração de característica realizada pelos algoritmos de PCA e ICA são eficientes no fornecimento de informação relevante para o processo decisório.

A transformação dos dados de alta dimensionalidade dos sinais dos calorímetros do ATLAS em variáveis com poder de descrição do processo analisado foi identificada tanto na abordagem não segmentada quanto na segmentada.

Na abordagem não segmentada, foi possível atingir um nível de compactação bem elevado. Um outro benefício produzido, foi a aproximação das características de elétron e *leptonjets* nas variáveis resultantes. Isso fica explícito pelos significativos valores finais obtidos para a detecção de *leptonjets*. O que se nota, é que, independentemente da técnica de análise de componentes empregada, o custo final nos resultados, em relação ao algoritmo de referência T2Ca1o, é de uma leve redução na detecção de elétrons e um pequeno aumento de falso alarme. Se olharmos as colunas mais a direita da Tabela 5.15, verificamos que para uma taxa de  $PD_e$  igual a do T2Ca1o e assumindo a  $PF_j$  como referência, a melhor abordagem aumenta o falso alarme em menos de 5 pontos percentuais, enquanto no pior dos casos o aumento é de 17 pontos percentuais. Mesmo com ganhos entre 20 e 37 pontos percentuais (respectivamente) na  $PD_{lj}$ , esse segundo classificador se torna inviável para operar no segundo nível de filtragem do detector, enquanto o primeiro fornece um ganho substancial na detecção de *leptonjets* com a preservação dos níveis de detecção dos objetos de interesse para o canal  $e/j$ , tornando-se assim um candidato a entrar em operação.

No caso da abordagem segmentada, os números são ainda mais expressivos. A aplicação das técnicas de análise de componentes localmente (extração de características em cada camada dos calorímetros) permitiu potencializar a transformação da informação disponível. A alimentação do classificador pela informação realmente discriminante de cada camada ou de apenas algumas camadas, eliminou outras variáveis (consideradas nos tes-

tes anteriores) que complicavam a tarefa de identificação do evento de interesse mais do que auxiliavam. Aqui também as taxas de compactação obtidas foram expressivas. Os valores de detecção de *leptonjets* são ainda mais elevados do que os obtidos com o processamento não segmentado. Com valores de  $PD_e$  muito próximos do alcançado com o T2Calo, obteve-se ganhos na  $PD_{lj}$  de mais de 30 pontos percentuais para taxas de falso alarme levemente superiores. Tudo isso com classificadores lineares ou neurais de complexidade computacional baixa.

Todas as abordagens testadas foram bem superiores aos algoritmos de referência na detecção de *leptonjets*. Em algumas abordagens esse ganho alcançado não significou uma deterioração tão expressiva nas taxas de  $PD_e$  e  $PF_j$  como resume o  $SP_{e/lj}$ . O melhor resultado foi o obtido na abordagem segmentada com o uso da PCA como algoritmo de extração de características e do classificador neural como algoritmo de decisão. Com praticamente o mesmo valor de  $PD_e$  do T2Calo e um aumento de menos de 5 pontos percentuais no valor de  $PF_j$ , a  $PD_{lj}$  teve um ganho de mais de 30 pontos percentuais. Considerando esses valores finais de desempenho, restou avaliar o impacto desta abordagem no tempo de processamento total da filtragem de nível 2. Os algoritmos implementados foram então inseridos no ambiente computacional do detector, Athena. Na Tabela 5.16, são mostrados os tempos gastos nos processos de seleção da informação da região de interesse (RoI), 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, seguida das camadas hadrônicas. A explicação para a camada eletromagnética está relacionada com a maior granularidade dessa, i.e., mais sensores, conseqüentemente, mais anéis e mais tempo para a formação deles. No caso das camadas hadrônicas, a explicação vem do fato calorímetro hadrônico ser segmentado em um número maior de módulos e dependendo da posição de interação (onde a partícula o intercepta) a RoI pode conter informação de vários desses módulos, o que impacta diretamente no tempo de formação dos anéis. 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, i.e., se apenas uma camada for utilizada, é necessário solicitar os dados de todas as camadas; porém, na utilização de mais de uma camada esse tempo é contado apenas uma vez.

Assim, com eficiência de discriminação equivalente ou superior à dos algoritmos de referência, mesmo sem utilizar a informação de algumas das camadas, a abordagem escolhida implica uma redução da dimensionalidade, quando são mantidas apenas as informações das camadas EM b. e HAD 0, e um algoritmo de decisão menos complexo, com um tempo final de processamento de  $0,8101 \pm 0,0317$ ms. Esse valor representa apenas 66,9% do tempo despendido pelo RingerNeuralHypo e é 25,22% superior ao tempo requerido pelo T2Calo, como mostrado na Tabela 3.2.

Além da redução da dimensionalidade dos dados, da configuração simples do classifi-

Tabela 5.16: Tempo requerido, em ms, para selecionar região de interesse e construir topologicamente os sinais em anéis para cada camada dos calorímetros.

Camada	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
EM f.	$0,1264 \pm 0,0309$	$0,0311 \pm 0,0077$	$0,1575 \pm 0,0318$	21,8
EM m.	$0,0794 \pm 0,0091$	$0,0203 \pm 0,0021$	$0,0997 \pm 0,0093$	13,8
EM b.	$0,0425 \pm 0,0056$	$0,0111 \pm 0,0018$	$0,0536 \pm 0,0059$	7,4
HAD 0	$0,1230 \pm 0,0775$	$0,0080 \pm 0,0029$	$0,1309 \pm 0,0776$	18,1
HAD 1	$0,1214 \pm 0,0775$	$0,0071 \pm 0,0027$	$0,1285 \pm 0,0776$	17,8
HAD 2	$0,1292 \pm 0,0778$	$0,0071 \pm 0,0018$	$0,1364 \pm 0,0778$	18,9

cador e do aceitável tempo de execução da abordagem, a saída do algoritmo está amparada pela interpretação detalhada realizada sobre os eventos usados na etapa de teste.

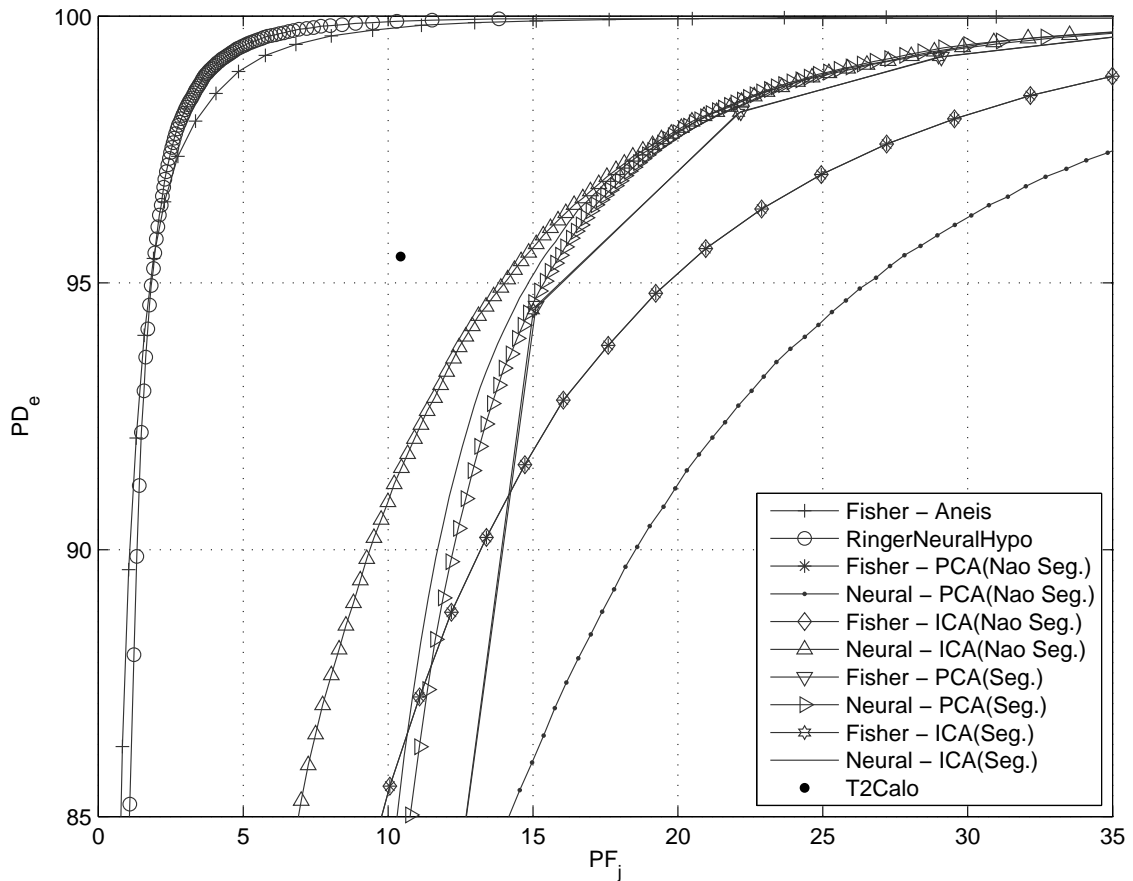


Figura 5.27: Curvas ROC para o canal e/j obtidas com cada uma das abordagens propostas.

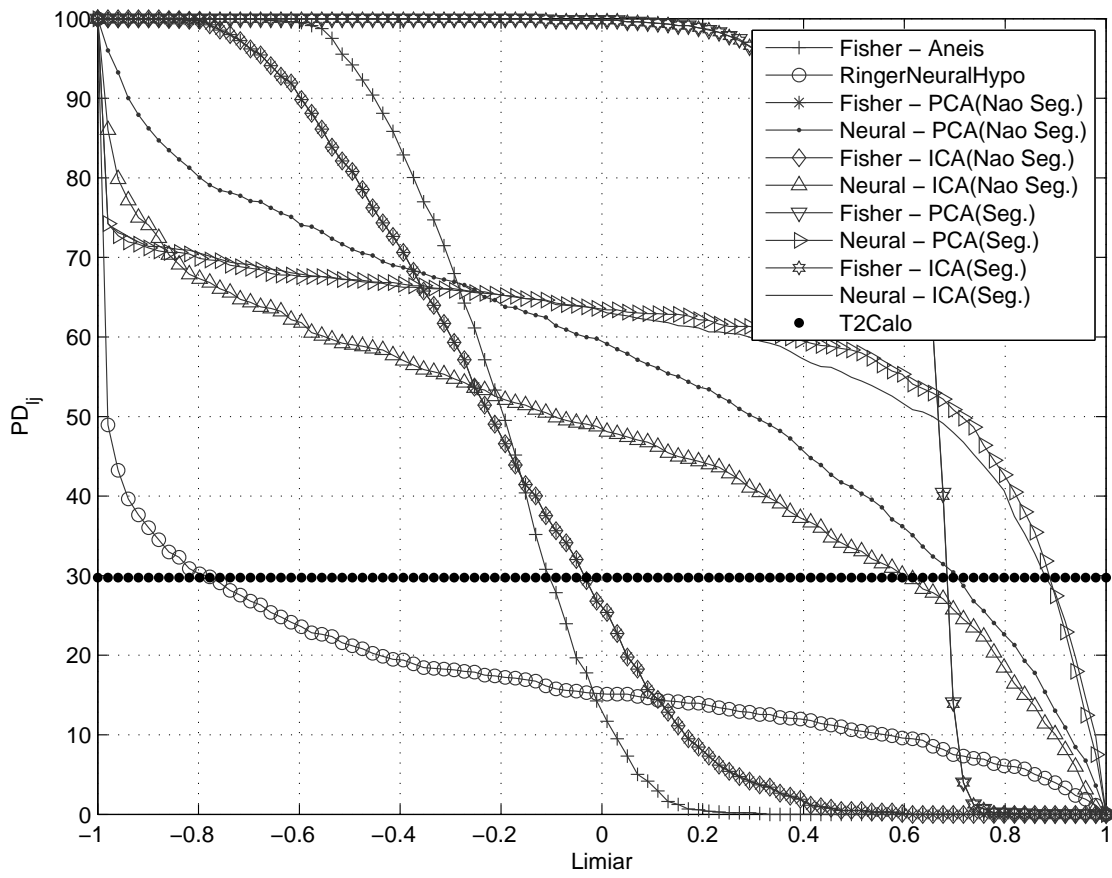


Figura 5.28: Curvas de eficiência na detecção de *leptonjets* obtidas com cada uma das abordagens propostas.



# Capítulo 6

## Resultados da Filtragem *online* de Nível 2 - canal *leptonjets*/jato

As abordagens propostas no Capítulo 5 fizeram uso da infraestrutura já existente no canal elétron/jato do detector ATLAS para alcançar um aumento na eficiência de detecção de *leptonjets*, sem, para isso, comprometer a detecção de eventos de interesse do canal. As intervenções propostas se resumiram a investigar o potencial de algoritmos lineares e não lineares de classificação, uso de diferentes técnicas de normalização de dados, além, da inserção de blocos de pré-processamento, como detalhado no Capítulo 4.

Neste capítulo, apresentamos algumas propostas e resultados referentes a hipótese de colocação de um novo canal de seleção de eventos, uma tarefa que demanda tempo e resultados significativos, principalmente com projetos em regime de colaboração como os realizados no CERN. Nesse canal proposto, as assinaturas de interesse são de *leptonjets*, enquanto o ruído de fundo é povoado por jatos hadrônicos.

Os testes foram conduzidos com o uso de *leptonjets* e jatos com propriedades descritas na Seção 4.1. O ruído de fundo hadrônico é um dos possíveis estados resultantes concorrentes, haja visto que os Higgs escuros (partícula pertencente a cadeia de decaimento do candidato a matéria escura) podem aparecer com estados hadrônicos de decaimentos. Além disso, acredita-se que *leptonjets* podem ter alguns estados hadrônicos em si, por exemplo, se o Higgs decair em píons carregados.

Os resultados apresentados referem-se ao procedimento de filtragem de segundo nível operando a partir da informação de calorimetria disponibilizada pelo primeiro nível de filtragem. Comparações são feitas entre as diferentes implementações propostas.

### 6.1 *Menu* 1j4e10

Como mencionado no estudo do canal elétron/jato, a infraestrutura atual do detector apresenta certa flexibilidade quanto aos recursos de implementação ofertados. Aprovei-

tando essa característica, as proposições de estratégias de filtragem de nível 2, para um futuro canal *leptonjets/jato*, foram precedidas das considerações a respeito do nível 1, de maneira a preservar a realidade de todos os níveis de filtragem do detector.

Como estamos filtrando objetos eletromagnéticos, resolvemos preservar os conceitos e considerações do primeiro nível de filtragem do canal elétron/jato. Assim, para o *menu* e10, objetos com energia transversa maior ou igual a 7GeV no primeiro nível eram selecionados como candidatos a objetos de interesse para os níveis seguintes de filtragem. Para o canal *leptonjets/jato* foi feito o mesmo procedimento e o correspondente nome dado para esse *menu* foi 1j4e10.

## 6.2 Anéis

A classificação dos dados nesta abordagem não emprega qualquer técnica de pré-processamento aos dados formatados e normalizados. Esses dados são apresentados ao algoritmo de classificação que infere o valor de saída, associando-o com uma das classes analisadas. O objetivo é avaliar o poder de discernimento que os anéis fornecem aos algoritmos de classificação, mesmo em condições de dimensão elevada de dados.

Para isso, investigou-se o uso de diferentes tipos de normalização de anéis além dos algoritmos linear e não-linear como classificadores de padrão. A versão neural dessa abordagem pode ser interpretada como uma variante do algoritmo RingerNeuralHypo, apresentado para o canal elétron/jato.

A Tabela 6.1 mostra o desempenho obtido com o uso do classificador linear de Fisher. Os resultados obtidos contribuem, primeiramente, para criação de algum parâmetro de comparação de resultados, dado o ineditismo dos testes realizados e os resultados obtidos. Pode-se afirmar que a agregação da informação contribuiu positivamente para quase todos os cenários de normalização. O melhor resultado foi obtido com o uso da normalização por camada. O valor de 0,7324 do produto  $SP_{lj/j}$  reflete uma taxa de detecção de *leptonjets* de 63,88% contra um falso alarme de 16,76%. Além de ser um referencial para as demais abordagens utilizadas no decorrer do capítulo, o resultado mostra que a tarefa de separar os padrões de *leptonjets* e jatos hadrônicos, do ponto de vista dos anéis normalizados, é bem complexa. Os valores finais alcançados são promissores e, ao mesmo tempo, mostram o desafio empreendido. Promissores por conseguirem a partir, apenas, da formatação topológica dos dados e uma estrutura linear de decisão alcançar valores significativos de desempenho. E desafiantes por exprimirem a complexidade tratada na classificação dos eventos. Ao se tratar de classificação de eventos raros, em que o custo da perda de um desses eventos é bastante significativo para a tarefa como um todo (a validação de uma teoria experimentalmente), o mínimo ganho alcançado é suficiente para promover uma metodologia de trabalho.

A Figura 6.1 apresenta o desempenho do classificador linear com melhor resultado. A

Tabela 6.1: Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	$PD_{lj}$	$PF_j$	$SP_{lj/j}$	limiar
Energia	46,39	46,42	0,4992	-0,09
Sequencial	59,32	12,19	0,7286	0,33
Camada	63,88	16,76	0,7324	0,52

Figura 6.1(a) mostra os histogramas de saída do classificador e a separação dos padrões alcançada. Nas Figuras 6.1(b) e 6.1(c) são mostradas a curva ROC e o produto  $SP_{lj/j}$  em função do limiar de decisão do classificados. As Figuras 6.1(d), 6.1(e) e 6.1(f) detalham o desempenho do classificador em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector ATLAS, respectivamente.

Considerando os valores esperados de saída para cada um dos eventos ( $-1$  para jatos hadrônicos e  $1$  para *leptonjets*), pode-se concluir que o perfil dos eventos em análise causa grande confusão para o classificador, haja visto a grande sobreposição de dados na régua de decisão do classificador. Isso reflete tanto na curva ROC quanto no produto  $SP_{lj/j}$ , em que fica evidente a escolha do limiar ótimo de decisão. Ao operar nesse patamar ótimo de decisão, o classificador mostra-se pouco exitoso na distinção de eventos com baixa energia ( $E_T < 20\text{GeV}$ ), menor eficiência na  $PD_{lj}$  e um aumento na  $PF_j$ . Esse resultado evidencia (como já mencionado em vários dos testes realizados no Capítulo ??) que os *leptonjets* e jatos compartilham muitas características nessas condições, i.e., os *leptonjets* são menos colimados e topologicamente mais parecidos com os jatos hadrônicos. Essa peculiaridade nos resultados é esperada para os de mais testes executados ao longo do capítulo, ora mais evidente ora não.

Para analisar a qualidade da informação fornecida ao classificador, é necessário estudar a relevância de cada um dos anéis usados como informação de entrada. Na Figura 6.2, ganha destaque a informação contida na segunda camada eletromagnética, que contribui significativamente para o resultado alcançado. A segunda camada eletromagnética é a mais profunda e conseqüentemente a responsável por absorver a maior parte da energia dos eventos eletromagnéticos que a atravessam. Aparentemente não é tanto a topologia do evento e sim o nível (magnitude) da energia absorvida a responsável pela diferenciação dos eventos, haja visto que a resolução da primeira camada eletromagnética (maior detalhamento espacial) é bem superior à da segunda. Contudo não se pode ignorar a informação espacial, pois os primeiros anéis da primeira camada eletromagnética também contribuem para o resultado final, porém sem grande destaque.

Os dados formatados em anéis também foram submetidos a um classificador não-linear; aqui, uma rede neural. Na expectativa de avaliar a contribuição do algoritmo de classificação para uma decisão final sobre o problema exposto, foi realizado um teste de desempenho com a variação de apenas um dos seus parâmetros livres. A topologia da rede

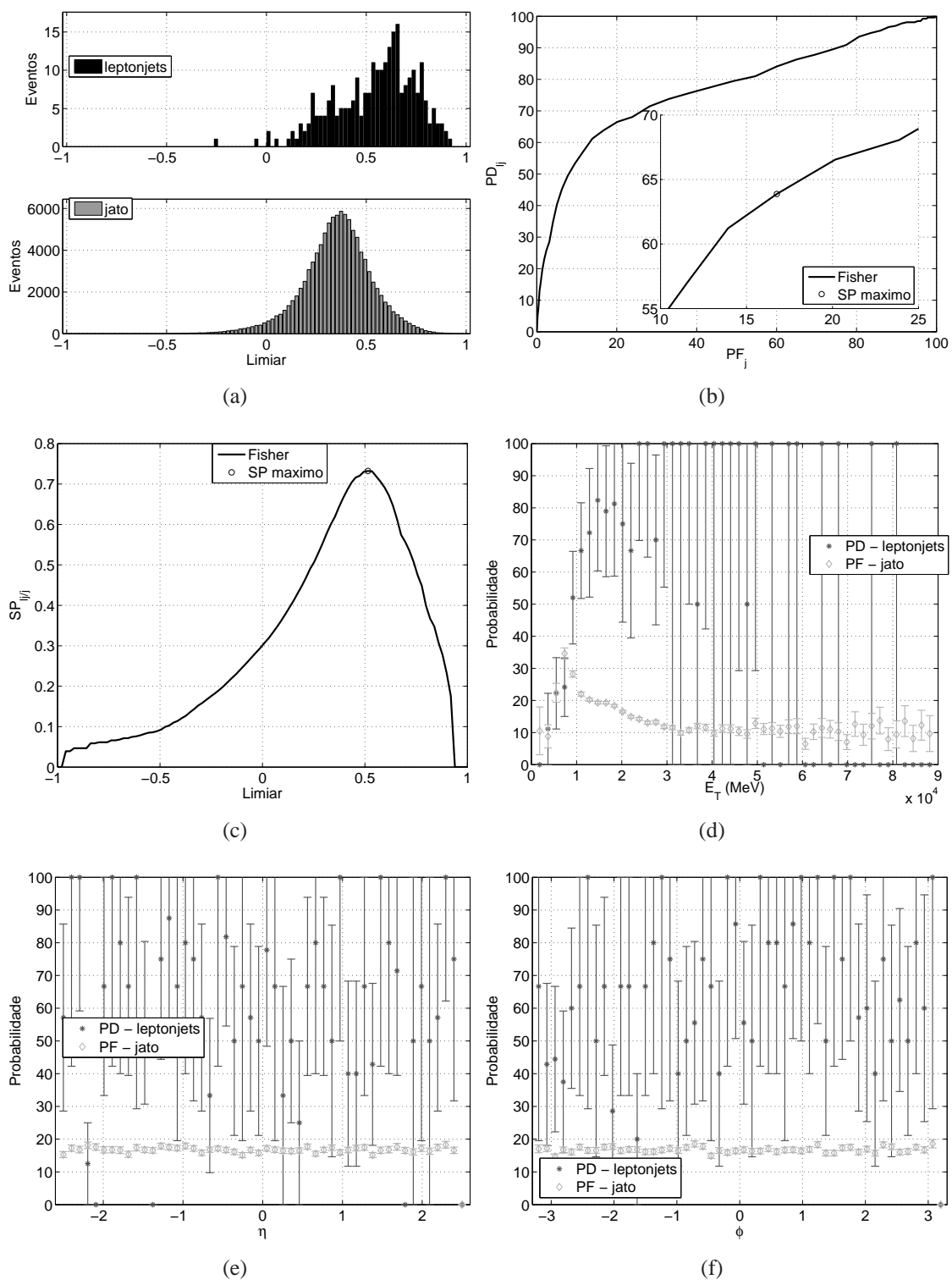


Figura 6.1: Desempenho do classificador linear de Fisher operando sobre dados formados em anéis e normalizados por camada. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

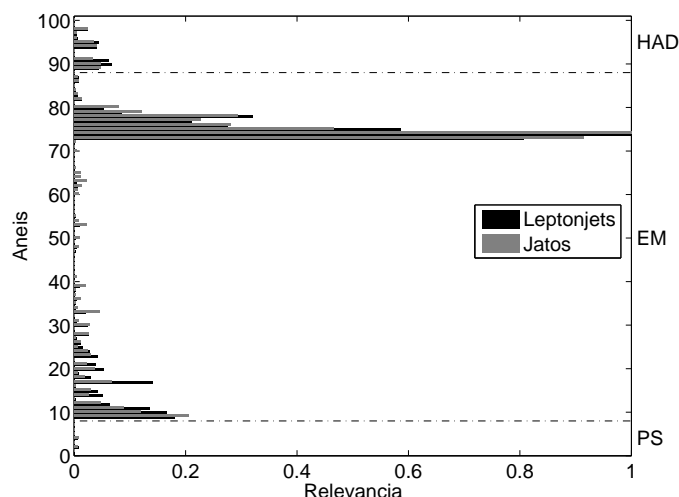


Figura 6.2: Relevância dos anéis normalizados por camada para o processo decisório implementado pelo classificador linear de Fisher.

foi alterada significativamente no que se refere ao número de nós da camada intermediária, de um mínimo de 2 neurônios ao máximo de 15. Os valores alcançados a cada aumento do número de neurônios da camada intermediária da rede estão quantificados na Tabela 6.2.

Ao analisarmos os valores de produto  $SP_{ij}$ , verificamos que para qualquer uma das normalizações escolhidas o resultado pouco se altera e dentro de uma mesma regra de normalização as condições antagônicas de topologia da rede apresentam valores finais próximos. Pode-se justificar essa característica a partir dos parecidos perfis de deposição de cada classe de evento, i.e., a variação na estrutura de decisão interna alcançada pelo algoritmo, mesmo com um acréscimo significativo nos parâmetros livres da rede neural, não influi de maneira contundente na separação de padrões tão parecidos. Entende-se o resultado como uma limitação imposta pela estrutura dos dados que pode parcialmente ser contornada pelas características das modificações alcançadas entre os classificadores. Por outro lado, é visível, mesmo que em menor escala, a evolução em relação aos números obtidos com o classificador linear de Fisher. Nesse ponto destaca-se o ganho que a estrutura de decisão não-linear trouxe principalmente para os dados normalizados por energia. Com o classificador linear essa normalização apresentou o pior resultado, enquanto aqui foi a melhor avaliada.

A Figura 6.3 apresenta o comportamento de diferentes medidas de desempenho usadas no trabalho, todas para o caso da normalização por energia. Especificamente, a Figura 6.3(a) apresenta a variação do erro final de treinamento para 5 diferentes inicializações do algoritmo neural em função do número de neurônios na camada intermediária da rede. As marcações representam os menores, os maiores e os valores medianos de erro quadrático médio alcançados ao fim do treinamento. Novamente fica explícito que o aumento do número de parâmetros ajustáveis por parte do algoritmo de classificação não é uma garantia da diminuição do erro quadrático médio final de treinamento, i.e., uma

Tabela 6.2: Desempenho dos classificadores neural sobre dados formatados em anéis e normalizados de diferentes maneiras.

Técnica	Neurônios	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ij/j</sub>	limiar
Energia	2	63,50	14,09	0,7428	0,23
	3	65,78	14,90	0,7513	0,23
	4	69,96	20,01	0,7489	0,01
	5	64,26	13,57	0,7493	0,23
	6	68,82	17,87	0,7533	0,15
	7	68,44	17,28	0,7541	0,07
	8	64,64	17,07	0,7350	0,23
	9	65,78	15,66	0,7477	0,19
	10	63,50	13,46	0,7457	0,27
	11	66,16	16,72	0,7447	0,17
	12	69,20	20,75	0,7414	0,09
	13	66,92	18,15	0,7419	0,13
	14	73,00	24,38	0,7430	-0,01
	15	62,36	12,85	0,7423	0,23
	Sequencial	2	63,50	15,66	0,7355
3		65,40	16,81	0,7403	0,11
4		65,78	17,75	0,7378	0,13
5		65,78	18,44	0,7346	0,13
6		63,50	18,11	0,7240	0,17
7		70,34	24,55	0,7287	0,05
8		63,12	16,83	0,7280	0,21
9		62,36	17,04	0,7229	0,19
10		65,40	19,75	0,7263	0,13
11		63,88	16,92	0,7316	0,19
12		69,96	21,57	0,7413	0,11
13		65,40	18,75	0,7311	0,15
14		64,26	19,38	0,7221	0,13
15		60,84	14,62	0,7259	0,27
Camada		2	63,12	20,18	0,7123
	3	68,82	25,11	0,7182	0,05
	4	62,36	19,43	0,7117	0,17
	5	63,50	20,14	0,7145	0,19
	6	60,08	16,37	0,7137	0,31
	7	59,32	16,01	0,7112	0,27
	8	63,50	17,15	0,7285	0,23
	9	66,92	23,53	0,7161	0,11
	10	65,02	22,94	0,7091	0,13
	11	71,10	26,21	0,7244	0,05
	12	69,58	25,47	0,7203	0,05
	13	61,98	17,73	0,7176	0,23
	14	68,82	23,13	0,7279	0,09
	15	66,54	21,72	0,7229	0,13

maior aproximação das saídas da rede para as saídas desejadas, como mostra a trajetória da curva. A Figura 6.3(b) mostra o comportamento do produto  $SP_{1j/j}$  associado à melhor das redes finais obtidas. O comportamento da curva de treino sempre acima da curva de teste já era esperado, tendo em vista que esses foram os dados com os quais a rede lidou diretamente no ajuste dos parâmetros livres. Porém, nota-se aqui um reflexo da evolução do erro treinamento, pois mesmo com valores inferiores, a curva de teste nem sempre acompanha a tendência da curva de treino. Pode-se atribuir esses resultados à metodologia usada no treinamento e testes dos dados, limitada pelo número reduzido de eventos disponíveis. Isso se reflete nas Figuras 6.3(c) e 6.3(d) que mostram o comportamento das probabilidades de detecção e falso alarme, respectivamente.

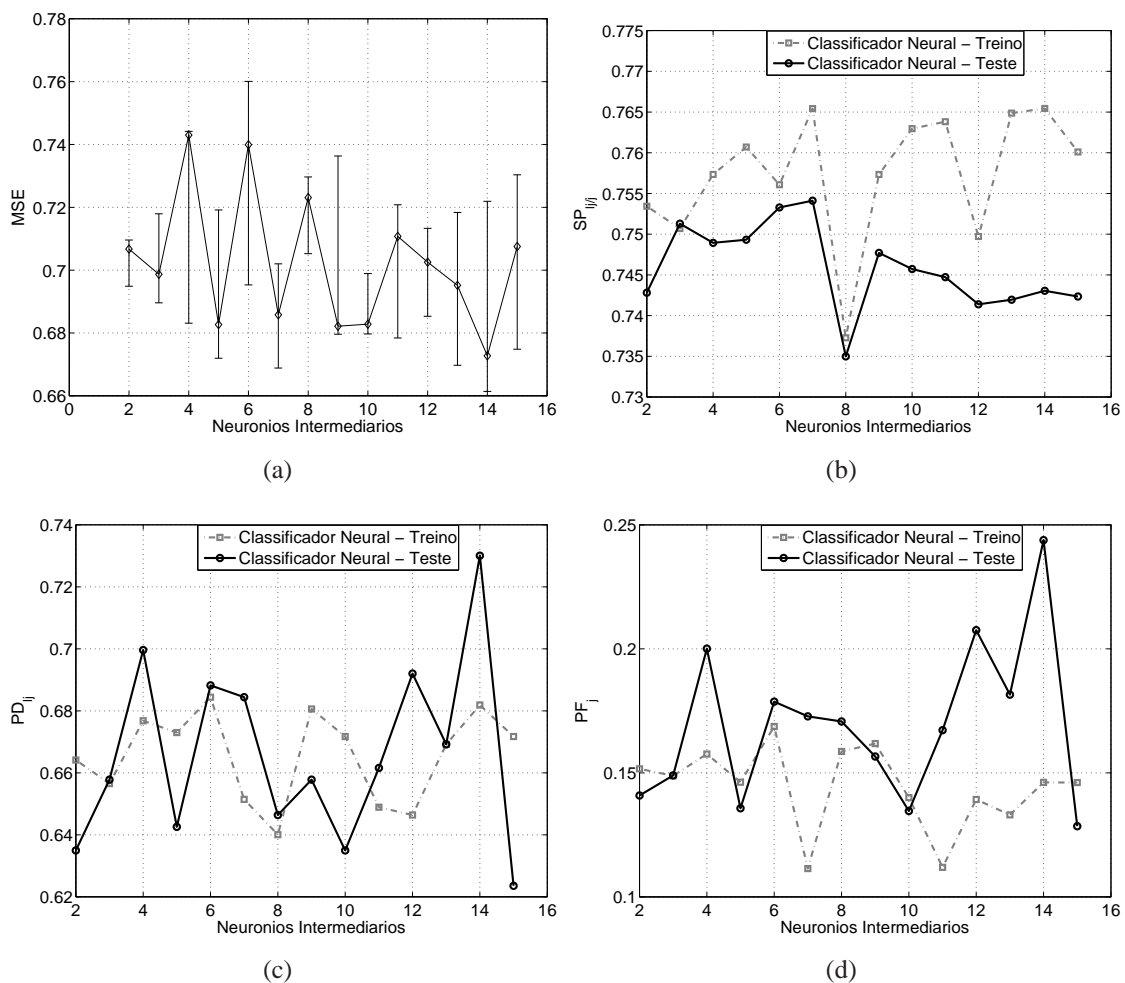


Figura 6.3: Variação de desempenho do classificador neural operando sobre dados formatados em anéis e normalizados por energia: (a) Erro quadrático médio final de treinamento para diferentes inicializações do algoritmo neural; (b) Produto  $SP_{1j/j}$  associado à melhor rede; (c) Probabilidade de detecção  $PD_{1j}$  associada à melhor rede; (d) Probabilidade de falso alarme  $PF_{1j}$  associada à melhor rede.

A topologia do detector (100 – 7 – 1) com melhor resultado é considerada computacionalmente simples, o que motiva a sua implementação. Para esse classificador foi alcançado um valor de produto  $SP_{1j/j}$  de 0,7541, o que implica 68,44% de detecção de



*leptonjets* e 17,28% de falso alarme. A Figura 6.4 apresenta o desempenho desse classificador. Os histogramas de saída da rede (Figura 6.4(a)) mostram a separação alcançada entre *leptonjets* e jatos. É notória a evolução obtida quando comparamos com a saída do classificador linear. Os padrões estão bem mais separados. A curva ROC é mostrada na Figura 6.4(b). O produto  $SP_{l;j}$  em função do limiar de saída da rede é mostrado na Figura 6.4(c). Já as probabilidades de detecção e falso alarme em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector ATLAS são apresentadas nas Figuras 6.4(d), 6.4(e) e 6.4(f), respectivamente.

As Figuras 6.5(a) e 6.5(b) mostram a evolução do erro quadrático médio de treinamento e a relevância da informação de cada anel para o processo decisório. O valor de relevância indica que os primeiros anéis das camadas *Pre Sampler*, *EM front layer* e *HAD layer 0* desempenham um papel primordial para a identificação dos eventos e que a importância da informação de entrada do classificador está mais compartilhada entre diferentes camadas do que a observada com o uso do algoritmo linear.

Os resultados da análise de relevância dos anéis pertencentes as RoIs, indicam que: a região central (o ponto de penetração e de deposição máxima de energia da partícula) e imediatamente periférica de algumas camadas dos calorímetros são potencialmente discriminantes para a tarefa de separação de padrões buscada. Uma investigação, então, dessa região é feita na sub-seção a seguir.

### 6.2.1 Regiões dentro da RoI

Ao trabalhar a classificação de eventos exóticos buscou-se na literatura variantes da informação de interesse a ser processado. Tendo em vista o relato de alguns pesquisadores [102] de que uma pequena região em torno do eixo de penetração dos *leptonjets* poderia fornecer informação suficiente para discriminar esse objeto do ruído de fundo, foi realizada uma análise do potencial de separação que um conjunto de anéis delimitados pela equação 6.1 em torno do centro de uma RoI produziria.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (6.1)$$

Foram consideradas regiões com  $\Delta R < 0,1$ ,  $\Delta R < 0,2$  e  $\Delta R < 0,3$ . Como há uma diferença de resolução espacial de camada para camada do calorímetro, o número final de células e anéis em cada uma das abordagens determina a dimensão dos dados estudados. Além disso, devido ao número reduzido de anéis para alguns valores de  $\Delta R$ , as técnicas de normalização também foram adaptadas para esses cenários. A Figura 6.6 ilustra a cobertura dessas regiões que, para efeito prático, tiveram arredondado sempre para mais o número de células contidas quando o cálculo indicava uma fração.

Analisar a região  $\Delta R$  mencionada, implica observar os anéis especificados na Tabela 6.3 por camada do calorímetro (considerando as configurações da região do barril



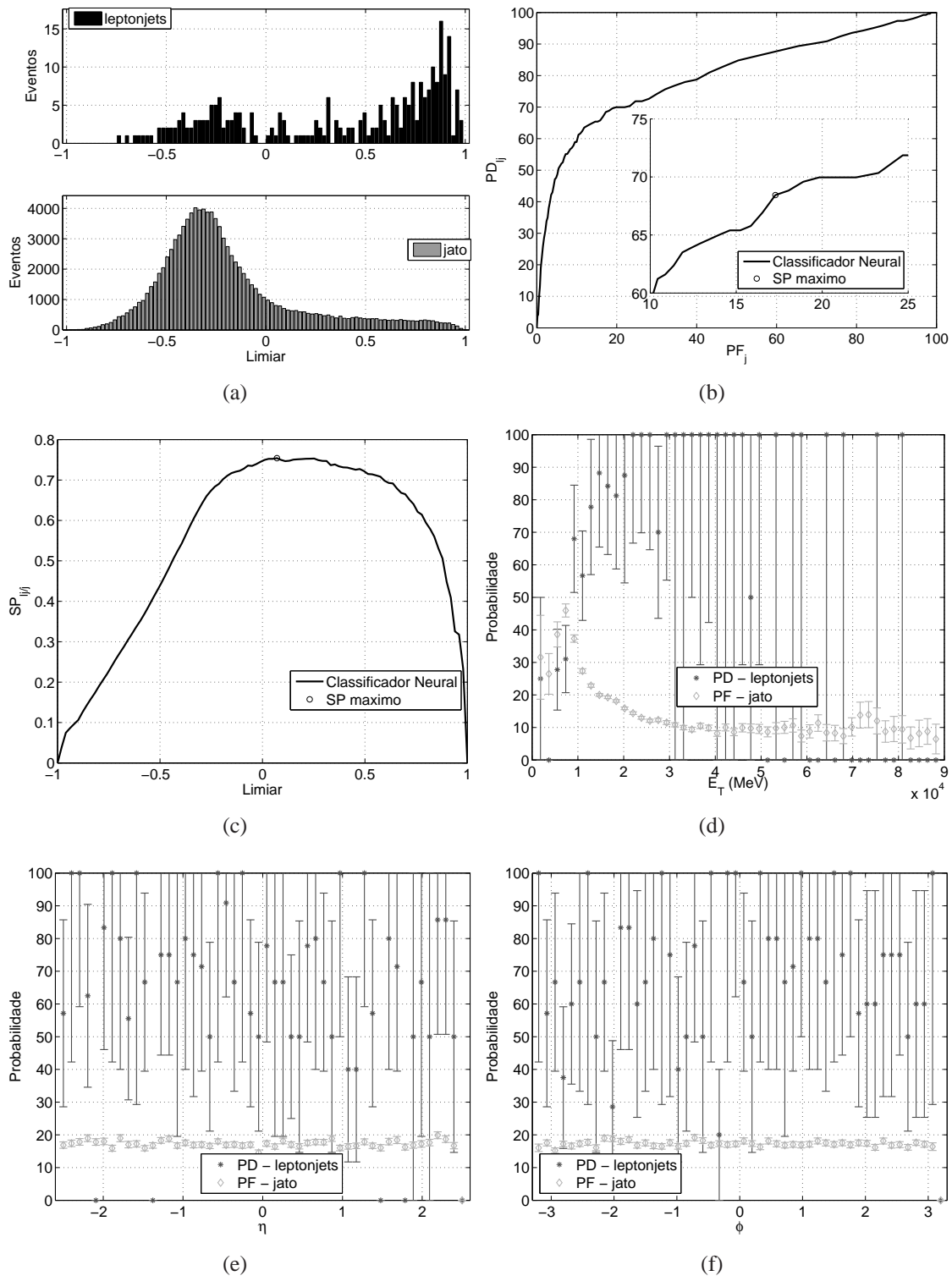
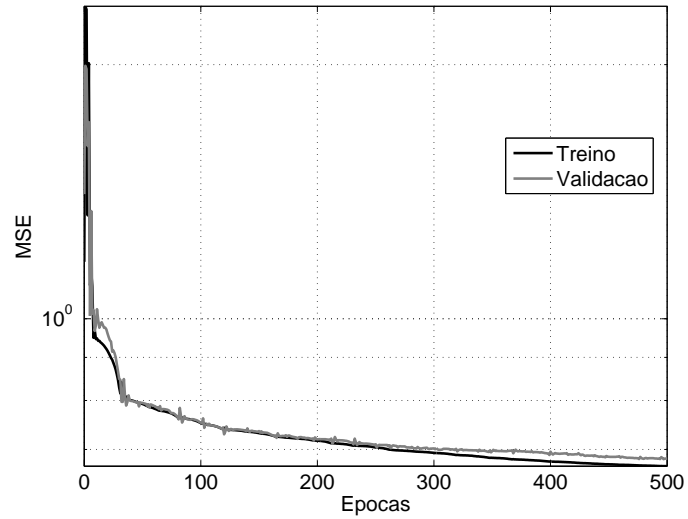
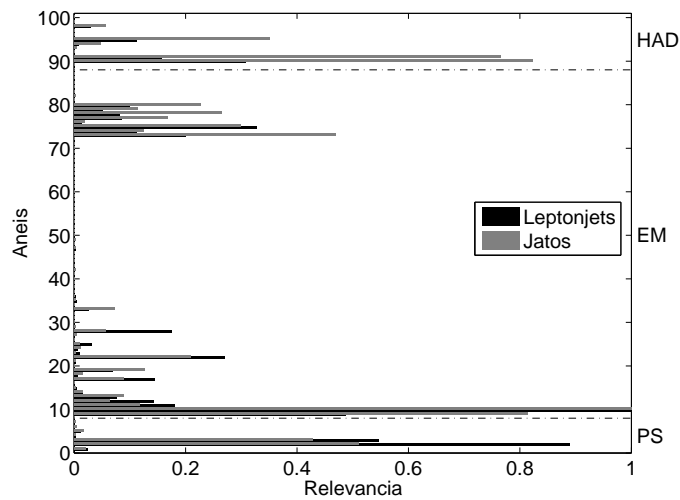


Figura 6.4: Desempenho do classificador neural operando sobre dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.



(a)



(b)

Figura 6.5: (a) Erro quadrático médio de treinamento do algoritmo neural. (b) Relevância dos anéis normalizados por energia para o processo decisório implementado pelo classificador neural.

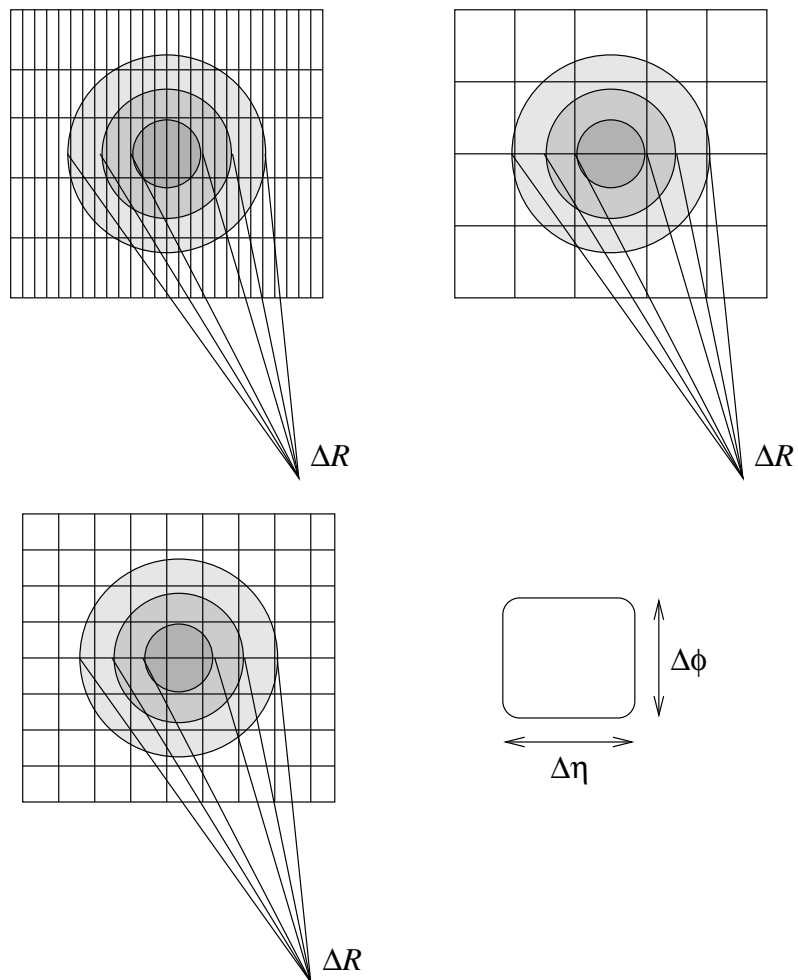


Figura 6.6: Formação de regiões de análise dentro da RoI dado as diferentes resoluções espaciais por camada dos calorímetros.

do detector, como indicado na Tabela 2.2).

Tabela 6.3: Anéis dentro da região  $\Delta R$  por camada do calorímetro.

Camada	Anéis		
	$\Delta R < 0,1$	$\Delta R < 0,2$	$\Delta R < 0,3$
<i>Pre Sampler</i>	1	2	3
<i>EM front layer</i>	1	2	3
<i>EM middle layer</i>	3	6	8
<i>EM back layer</i>	1	2	3
<i>HAD layer 0</i>	1	2	3
<i>HAD layer 1</i>	1	2	3
<i>HAD layer 2</i>	1	1	2

Um ponto a ser destacado no conjunto de testes realizados é a normalização aplicada aos dados. No caso dos anéis que estão inseridos no limite espacial  $\Delta R < 0,1$ , em função do número reduzido de anéis, foram testadas algumas variantes das normalizações até aqui aplicadas. A normalização por energia, já mencionada, teve alterado o cálculo do seu normalizador (ora considerando a energia contida nos sensores pertencentes à região  $\Delta R$ , ora considerando a energia contida na RoI). A normalização híbrida, que consiste em aplicar a normalização por energia para as camadas onde há apenas um anel e a normalização sequencial onde há mais do que um anel (ora considerando a energia contida em  $\Delta R$ , ora considerando a energia contida na RoI), foi uma das soluções encontradas. Devido à formulação da equação de normalização (4.3), o último anel de cada camada é normalizado pelo seu próprio valor de energia, o que destrói a informação nele contida. No caso dos canais elétron/jato e *leptonjets*/jato em que a região de análise é a própria RoI, esse efeito não é sentido, pois o normalizador, a partir de certo anel, é o mesmo para os anéis restantes. Aqui, na análise das regiões  $\Delta R$ , o último anel de cada camada ainda é muito energético e não se encaixa na exceção da regra da normalização. Para isso, foram propostas as normalizações que consideram em seus cálculos a energia da RoI. Estudos futuros podem se debruçar na busca de novos patamares de corte em função do tamanho das regiões  $\Delta R$ , com já é feito com a normalização sequencial por parte do algoritmo Ringer do canal  $e/j$ . Na abordagem com anéis inseridos no limite espacial  $\Delta R < 0,2$  e  $\Delta R < 0,3$ , a normalização reproduziu os testes realizados para a região  $\Delta R < 0,1$ .

Em todos os procedimentos foram usados o classificador linear de Fisher e um classificador não-linear (uma rede neural) como algoritmo de classificação. Por se tratar de um espaço de baixa dimensão, os classificadores neurais testados variaram sua topologia apenas no número de neurônios da camada escondida, de 2 a 5.

A Tabela 6.4 mostra o desempenho do classificador linear de Fisher ao lidar com os dados das regiões  $\Delta R$  submetidos às diferentes normalizações. Como já dito, há uma grande dificuldade em aplicar uma normalização que preserve as características de deposição de

energia de cada um dos eventos analisados. Os resultados mostram que os eventos normalizados considerando a energia da RoI são melhor avaliados pelo classificador. Além disso, percebe-se que o aumento da região  $\Delta R$  considerada, aumenta a eficiência do classificador. Assim, os anéis centrais, os primeiros anéis, fornecem discriminantes relevantes (como se esperava) e os anéis periféricos auxiliam nessa tarefa mais complicada com a captação dos detalhes que diferenciam os jatos hadrônicos dos *leptonjets* ora mais ora menos colimados. Os testes apontam uma interessante linha de pesquisa pois mostram que o melhor resultado se aproxima dos números obtidos com o classificador de Fisher operando sobre todos os anéis da RoI (Tabela 6.1). Fica evidente também que algumas normalizações não têm aplicação viável, pelo menos do ponto de vista da classificação linear. É o caso da normalização híbrida empregada para dados pertencentes a região  $\Delta R = 0,2$  em que os normalizadores são calculados em função da energia pertencente a essa mesma região. Como a região engloba muitas camadas com apenas dois anéis, o cálculo sequencial, nesses casos, requer um ajuste dos valores de corte (valor de energia a partir do qual os anéis restantes são submetidos a um mesmo normalizador), assim como é feito com a RoI.

Tabela 6.4: Desempenho do classificador linear de Fisher operando sobre dados formatados em anéis, delimitados pela região  $\Delta R$ , e normalizados de diferentes maneiras.

$\Delta R$	Normalização	PD <sub>l<sub>j</sub></sub>	PF <sub>j</sub>	SP <sub>l<sub>j</sub>/j</sub>	limiar
0,1	Energia ( $\Delta R$ )	60,08	49,93	0,5496	-0,09
	Energia (RoI)	66,92	27,17	0,6984	-0,19
	Híbrido ( $\Delta R$ )	75,67	42,88	0,6507	0,17
	Híbrido (RoI)	71,86	34,17	0,6881	-0,27
0,2	Energia ( $\Delta R$ )	63,88	36,46	0,6371	0,47
	Energia (RoI)	63,50	21,42	0,7084	0,01
	Híbrido ( $\Delta R$ )	1,14	0,04	0,2323	0,14
	Híbrido (RoI)	57,79	14,68	0,7089	-0,07
0,3	Energia ( $\Delta R$ )	59,70	44,75	0,5745	0,05
	Energia (RoI)	58,56	12,22	0,7243	0,29
	Sequencial ( $\Delta R$ )	68,06	26,94	0,7054	-0,15
	Sequencial (RoI)	68,44	24,39	0,7198	-0,11

Detalhes do resultado obtido com o uso do classificador linear alimentado por anéis contidos na região  $\Delta R = 0,3$  normalizados pela energia da RoI são apresentados na Figura 6.7. A relevância para o classificador desses anéis de entrada é mostrada na Figura 6.8. O classificador identifica a informação proveniente de um anel da camada hadrônica como muito importante para a rejeição dos jatos hadrônicos. Para a aceitação dos *leptonjets*, o classificador baseia, principalmente, sua decisão na informação dos anéis das camadas *Pre Sampler* e *EM front layer*. A afirmação de que os anéis periféricos são relevantes para diferenciar eventos muito parecidos (a minoria), fica visível pelo valor que

esses anéis recebem em relação aos mais relevantes.

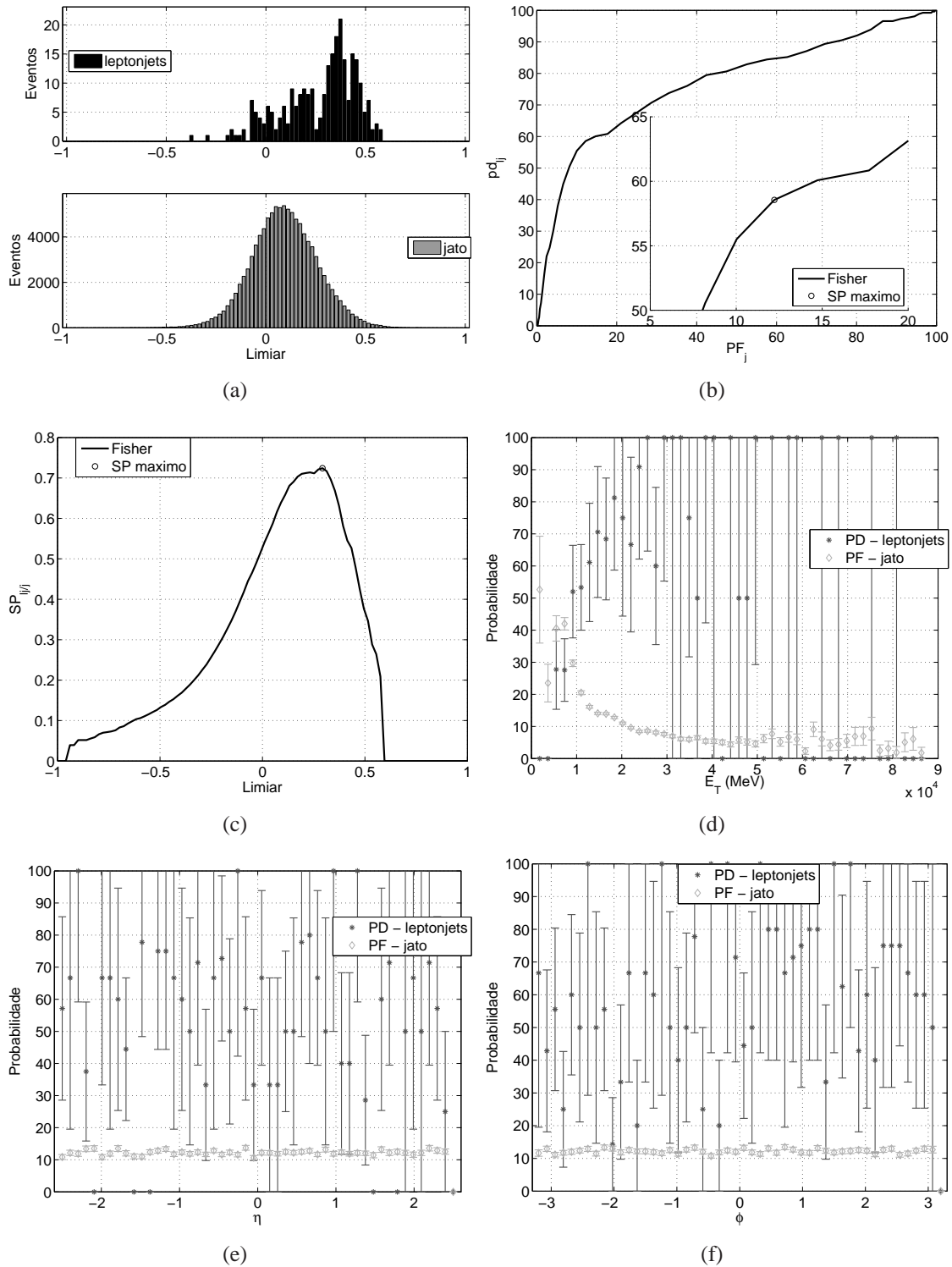


Figura 6.7: Desempenho do classificador linear de Fisher operando sobre dados formados em anéis pertencentes a região  $\Delta R = 0,3$  e normalizados pela energia de RoI. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{l/j}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

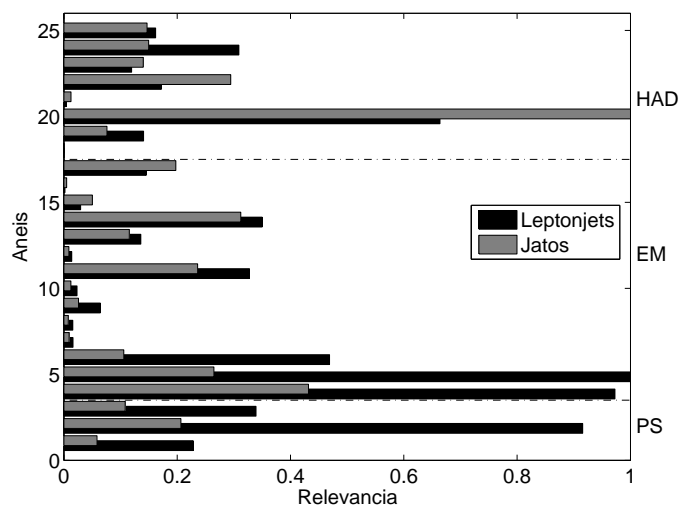


Figura 6.8: Relevância dos anéis normalizados por energia para o processo decisório implementado pelo classificador linear de Fisher.

Com o piso da eficiência definido pelo classificador linear, o uso de um classificador neural foi necessário para verificar o potencial de separação da informação fornecida pelos anéis estudados. A Tabela 6.5 indica que o melhor resultado se aproximou significativamente do melhor resultado obtido com o classificador neural operando sobre todos os anéis da RoI (Tabela 6.2). O mérito desse melhor resultado (com valor de produto  $SP_{l_j/j}$  de 0,7503) pode ser creditado ao classificador neural. Com anéis normalizados pela energia da região  $\Delta R = 0,3$  (configuração essa que com o classificador linear apresentou resultados questionáveis para um classificador de nível 2 de filtragem) foi possível detectar 63,5% dos *leptonjets* para uma taxa de falso alarme de 12,47%. A melhoria trazida pelo uso específico do classificador neural pôde ser observada em outras formatações dos anéis de entrada (outros valores de  $\Delta R$ ). De maneira geral, os resultados seguiram o observado anteriormente: o aumento da região  $\Delta R$  (e por conseguinte dos anéis), resultou em aumento da eficiência final do sistema.

O valor de 0,7503 para o produto  $SP_{l_j/j}$  alcançado com o classificador neural mostra a força de discriminação trazida pelo processamento da informação central da RoI. Levando-se em conta a dimensão dos dados analisada, 25, o ganho no tempo de processamento é considerável, haja visto que a etapa de maior consumo de tempo no segundo nível de filtragem é exatamente a etapa de construção dos anéis. A topologia final do classificador,  $25 - 5 - 1$ , é um outro ponto a ser destacado (bastante simples computacionalmente de ser implementada). Os detalhes do desempenho do classificador podem ser observados na Figura 6.9. Pelos histogramas de saída (Figura 6.9(a)), fica evidente que a construção da estrutura interna de decisão se baseou nos mesmos dados do classificador neural operando sobre os anéis da RoI (Figura 6.4(a)). Há uma semelhança na moda estatística da saída do classificador para cada uma das classes de eventos, além do número de *leptonjets* associados à classe *jatos* hadrônicos. A curva ROC (Figura 6.9(b))

Tabela 6.5: Desempenho do classificador neural operando sobre dados formatados em anéis, delimitados pela região  $\Delta R$  e normalizados de diferentes maneiras.

$\Delta R$	Normalização	Neurônios	$PD_{lj}$	$PF_j$	$SP_{lj/j}$	limiar
0,1	Energia ( $\Delta R$ )	5	68,44	31,68	0,6838	0,19
	Energia RoI	4	72,62	31,95	0,7032	0,11
	Híbrido ( $\Delta R$ )	4	69,20	34,06	0,6756	0,17
	Híbrido (RoI)	2	70,72	31,18	0,6977	0,09
0,2	Energia ( $\Delta R$ )	3	63,50	14,57	0,7406	0,33
	Energia (RoI)	5	62,36	16,20	0,7268	0,25
	Híbrido ( $\Delta R$ )	2	56,65	38,41	0,5910	0,21
	Híbrido (RoI)	5	63,50	18,03	0,7244	0,15
0,3	Energia ( $\Delta R$ )	5	63,50	12,47	0,7503	0,31
	Energia (RoI)	5	68,06	20,46	0,7369	0,13
	Sequencial ( $\Delta R$ )	4	65,02	32,85	0,6608	-0,09
	Sequencial (RoI)	4	63,12	27,23	0,6786	-0,43

e o produto  $SP_{lj/j}$  (Figura 6.9(c)) resumem a eficiência do classificador para toda a faixa do limiar de decisão, assim como o indicativo do ponto de operação ideal da abordagem. Para o limiar de decisão selecionado, verificamos que o classificador apresenta uma uniformidade na rejeição do ruído de fundo em função das coordenadas  $\eta$  (Figura 6.9(e)) e  $\phi$  (Figura 6.9(f)). A maior riqueza estatística dessa classe de evento, em função do maior número de exemplos, aponta uma ausência de tendência nos resultados, como desejado. Verifica-se ainda que a taxa de falso positivo é maior do que o valor final do classificador (12,47%) para eventos com energia transversa  $< 20\text{GeV}$  (Figura 6.9(d)). Além do comportamento já ter sido observado com o classificador neural operando com todos os anéis da RoI (Figura 6.4(d)) evidencia-se que a informação considerada para alimentar o classificador, seja aqui seja na abordagem mencionada, não consegue eliminar a confusão causada na diferenciação de *leptonjets* e jatos hadrônicos com baixa energia, o pior cenário de decisão.

Ao olhar com mais cuidado a informação de entrada do classificador (Figura 6.10(b)), constatamos o que foi dito anteriormente: os primeiros anéis são bastante relevantes para o processo decisório, enquanto os anéis secundários apresentam valores intermediários de importância, porém não menos significativos para que o classificador seja bem sucedido. A Figura 6.10(a) mostra a curva do erro quadrático médio de treinamento da rede neural.

Apesar de não estabelecer os melhores valores finais para as figuras de mérito, a redução da dimensionalidade, a simplicidade da configuração obtida, a interpretação da informação relevante de entrada do classificador, indicam potencialidades dessa abordagem para a solução do problema estudado.



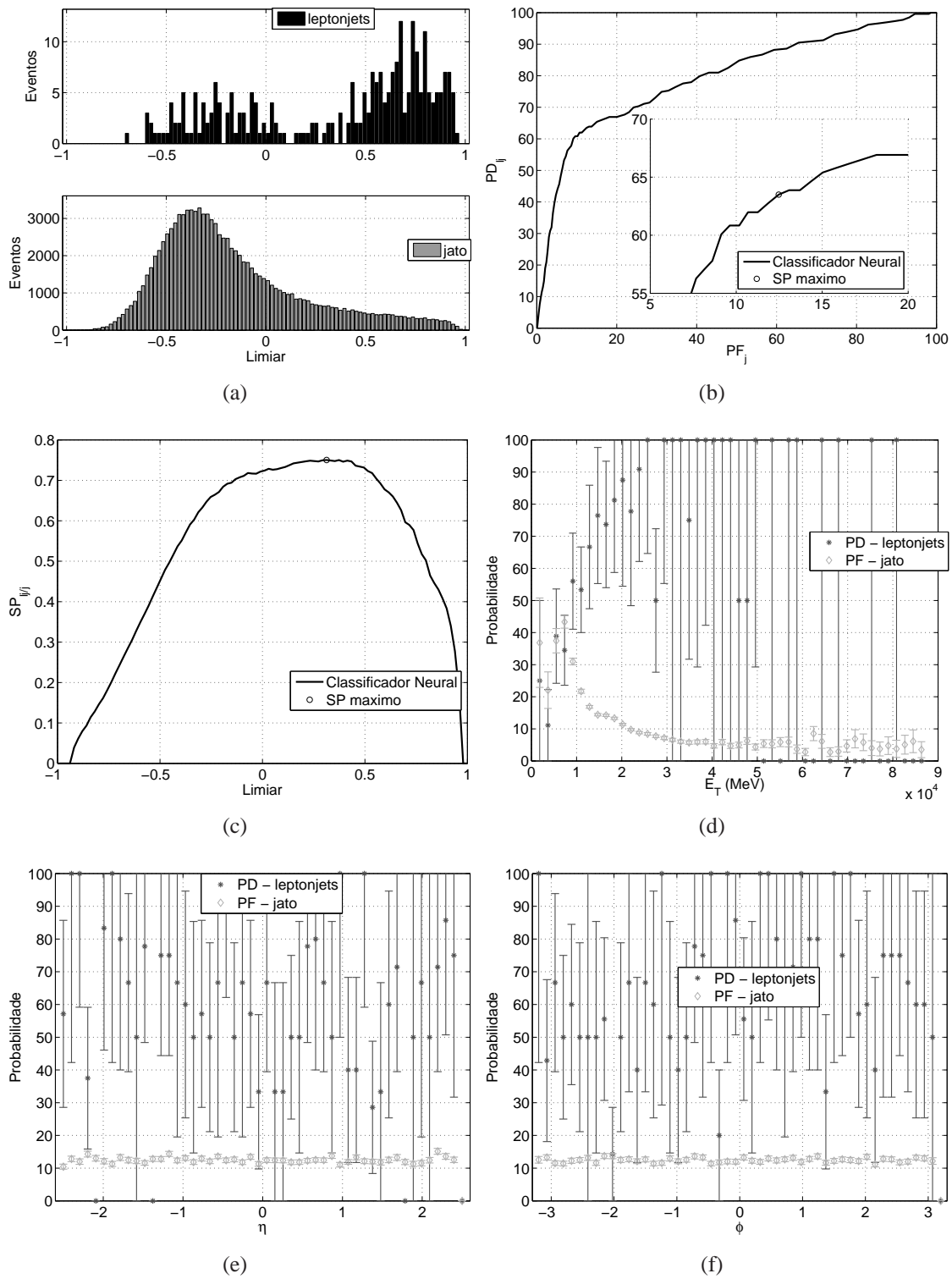


Figura 6.9: Desempenho do classificador neural operando sobre dados formatados em anéis pertencentes a região  $\Delta R = 0,3$  e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{ij}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

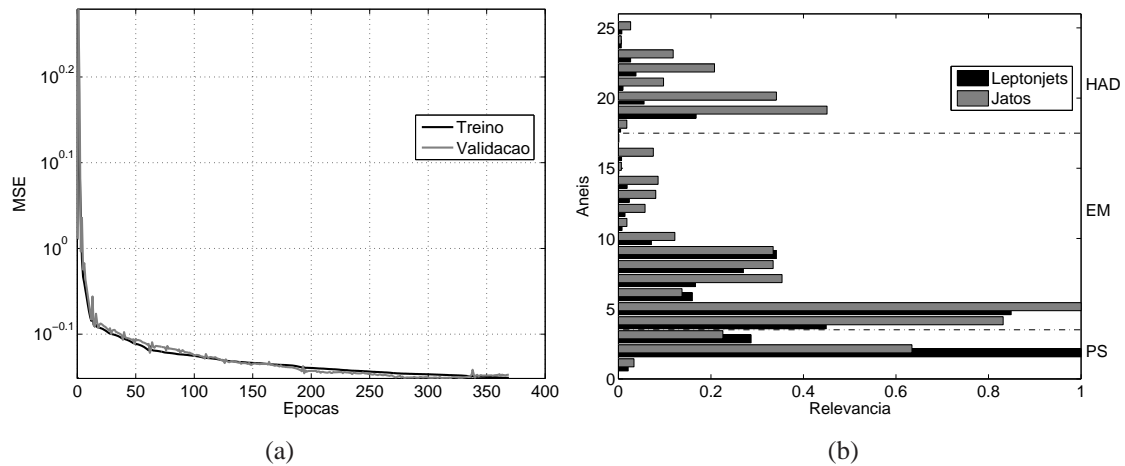


Figura 6.10: (a) Erro quadrático médio de treinamento do algoritmo neural. (b) Relevância dos anéis normalizados por energia para o processo decisório implementado pelo classificador neural.

### 6.3 Pré-processamento não Segmentado

O pré-processamento não segmentado dos dados permite às técnicas de análise de componentes atuarem sobre os dados de calorimetria sem levar em consideração as suas subdivisões e, de alguma maneira, uma busca por correlação e redundância na informação entre camadas. Apesar de os dados de calorimetria estarem segmentados em camadas, a interpretação da informação como um todo traz alguns benefícios na busca da melhor representação desses dados em um espaço de dimensão inferior. Espera-se nessa abordagem um número final menor de componentes retidas.

As diferentes análises de componentes empregadas fizeram uso do conjunto de dados de treinamento de *leptonjets* e de jatos na construção da matriz de mapeamento do espaço original de dimensão elevada para o espaço de dimensão reduzida. Essas projeções são usadas no treinamento dos algoritmos da classificação. Na etapa de teste, novos dados foram submetidos ao procedimento sendo este, finalmente, avaliado.

#### 6.3.1 PCA

Considerando que a PCA realiza compactação de sinais e sua representação a partir de componentes de maior energia, a observação do comportamento das curvas de carga (Figuras 5.6) para o evento de interesse mostra uma pequena diferenciação desse em relação ao ruído de fundo, independentemente da normalização aplicada. Esperam-se resultados muito próximos com vistas aos tipos de normalização empregados. Os *leptonjets* são elétrons, porém um número grande desses em uma mesma RoI, mesmo que muito próximos, implica um maior número de componentes retidas para a sua correta representação. Já os jatos hadrônicos tendem a espalhar sua energia em uma vizinhança, o que requer um maior número de componentes para sua representação. O que era considerado, no ca-

pítulo anterior, um tarefa difícil, a diferenciação do elétron isolado do jato hadrônico a partir de componentes representativas, se mostra menos complexa do que a que se deseja implementar nesse capítulo, a separação de *leptonjets* de jatos hadrônicos.

A estratégia seguida visa reter um número específico de componentes principais de maneira que a projeção de *leptonjets* nessas componentes seja o mais diferenciável possível do ruído de fundo do canal.

Novamente, é importante ressaltar que a curva de carga servirá de guia para as outras técnicas de análise de componentes. Haja visto que o branqueamento dos dados e a redução de dimensionalidade, etapa comum na ICA, estão diretamente associados com a PCA. Assim como nos testes anteriores desse capítulo, empregamos um classificador linear e um não-linear na etapa de classificação dos dados.

Na Tabela 6.6, são apresentados os resultados obtidos com o classificador linear de Fisher construído a partir de diferentes percentuais de energia retida do sinal formatado em anéis. Infere-se que os valores finais obtidos são muito próximos, independentemente da normalização aplicada. A evolução do produto  $SP_{l/j}$  mostra ainda que esse é proporcional ao número de componentes retidas. Assim, o acréscimo de informação (aumento de detalhamento) é importante para a diferenciação do evento de interesse do ruído de fundo. Dessa maneira a redução de dimensionalidade, almejada com o uso das técnicas de análise de componentes usadas como extratores de características, deve ser usada com parcimônia no processo. Com um número não pequeno de componentes principais (o equivalente a pelo menos 50% da energia retida) a projeção dos *leptonjets* apresenta a máxima diferenciação dos jatos hadrônicos, para cada uma das normalizações.

Esses resultados, ainda que bem próximos, apontam (a partir da figura de mérito) que o melhor classificador é obtido com o uso da normalização por energia sobre os anéis. Detalhes da variação de desempenho dos classificadores submetidos a essa normalização são apresentados na Figura 6.11. A evolução do algoritmo linear de decisão em função da energia retida do sinal é apresentada para cada uma das figuras de mérito. O produto  $SP_{l/j}$  do canal é mostrado na Figura 6.11(a), e as respectivas taxas de eficiência na detecção de *leptonjets* e falso alarme são apresentadas nas Figuras 6.11(b) e 6.11(c). Através das curvas é possível visualizar que o aumento da energia retida melhora a diferenciação entre *leptonjets* e jatos e, por conseguinte, o desempenho do canal. Além disso, nota-se que os valores alcançados com os dados de teste seguem de perto os valores alcançados com os dados usados no treinamento do algoritmo de decisão, demonstrando a generalização final da técnica.

O maior valor do produto  $SP_{l/j}$  alcançado pelo algoritmo foi 0,7287. Isso implica 68,82% de *leptonjets* detectados e 22,96% de falso alarme. Apesar da significativa redução de dimensionalidade dos dados alcançada, o resultado final é inferior aos números obtidos com ambos classificador operando sobre dados em anéis normalizados. Apesar

---

<sup>1</sup>O percentual de energia se refere ao acumulado nas componentes principais.

Tabela 6.6: Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Comp.	Energia(%) <sup>1</sup>	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ijj</sub>	limiar
Energia	1	25,91	54,75	31,41	0,6148	-0,27
	2	38,96	51,91	46,53	0,5268	-0,21
	3	46,81	66,16	22,44	0,7174	0,31
	4	53,54	62,36	18,80	0,7147	0,21
	5	58,34	68,82	23,38	0,7267	0,07
	6	62,91	66,92	22,64	0,7204	-0,11
	7	66,63	68,06	23,21	0,7236	-0,07
	8	69,71	69,58	24,79	0,7237	-0,13
	9	72,58	68,06	22,97	0,7247	-0,11
	10	75,25	66,92	21,67	0,7251	-0,01
	11	77,47	66,16	21,34	0,7227	-0,05
	12	79,33	68,82	22,96	0,7287	-0,11
	13	80,95	61,98	16,44	0,7237	0,19
Sequencial	1	9,38	70,34	35,43	0,6743	-0,09
	2	16,91	62,74	27,61	0,6748	-0,05
	3	24,07	65,40	27,87	0,6873	-0,03
	4	30,39	73,38	35,71	0,6876	-0,09
	5	35,80	73,38	35,75	0,6874	-0,09
	6	40,29	72,62	31,72	0,7043	-0,13
	7	44,31	68,82	26,30	0,7124	0,03
	8	47,75	70,34	27,38	0,7148	0,01
	9	51,04	68,44	22,85	0,7273	-0,01
	10	54,07	67,68	23,12	0,7221	-0,01
	11	56,73	69,96	26,06	0,7194	-0,09
	12	59,30	69,58	25,92	0,7181	-0,09
	13	61,68	68,44	24,70	0,7183	-0,09
	14	64,01	67,68	24,25	0,7166	-0,07
	15	66,24	65,02	21,34	0,7168	-0,05
	16	68,40	66,92	22,73	0,7200	-0,03
	17	70,41	66,92	22,84	0,7195	-0,03
	18	72,42	65,40	21,44	0,7183	-0,01
	19	74,31	66,16	21,50	0,7220	-0,01
	20	76,11	64,64	20,53	0,7186	-0,01
Camada	1	14,89	66,92	36,33	0,6528	-0,17
	2	25,67	68,06	38,72	0,6463	-0,19
	3	35,07	62,36	29,27	0,6648	-0,23
	4	43,32	58,94	23,47	0,6745	-0,15
	5	50,88	70,34	34,02	0,6815	-0,31
	6	57,45	63,88	27,55	0,6810	-0,25
	7	62,05	64,26	27,63	0,6825	-0,25
	8	66,31	71,86	35,82	0,6797	-0,35
	9	69,91	72,24	35,13	0,6851	-0,35
	10	73,26	67,68	26,00	0,7081	-0,25
	11	76,21	68,06	26,36	0,7082	-0,27
	12	78,63	67,30	25,46	0,7088	-0,21
	13	80,71	62,74	20,85	0,7071	-0,15

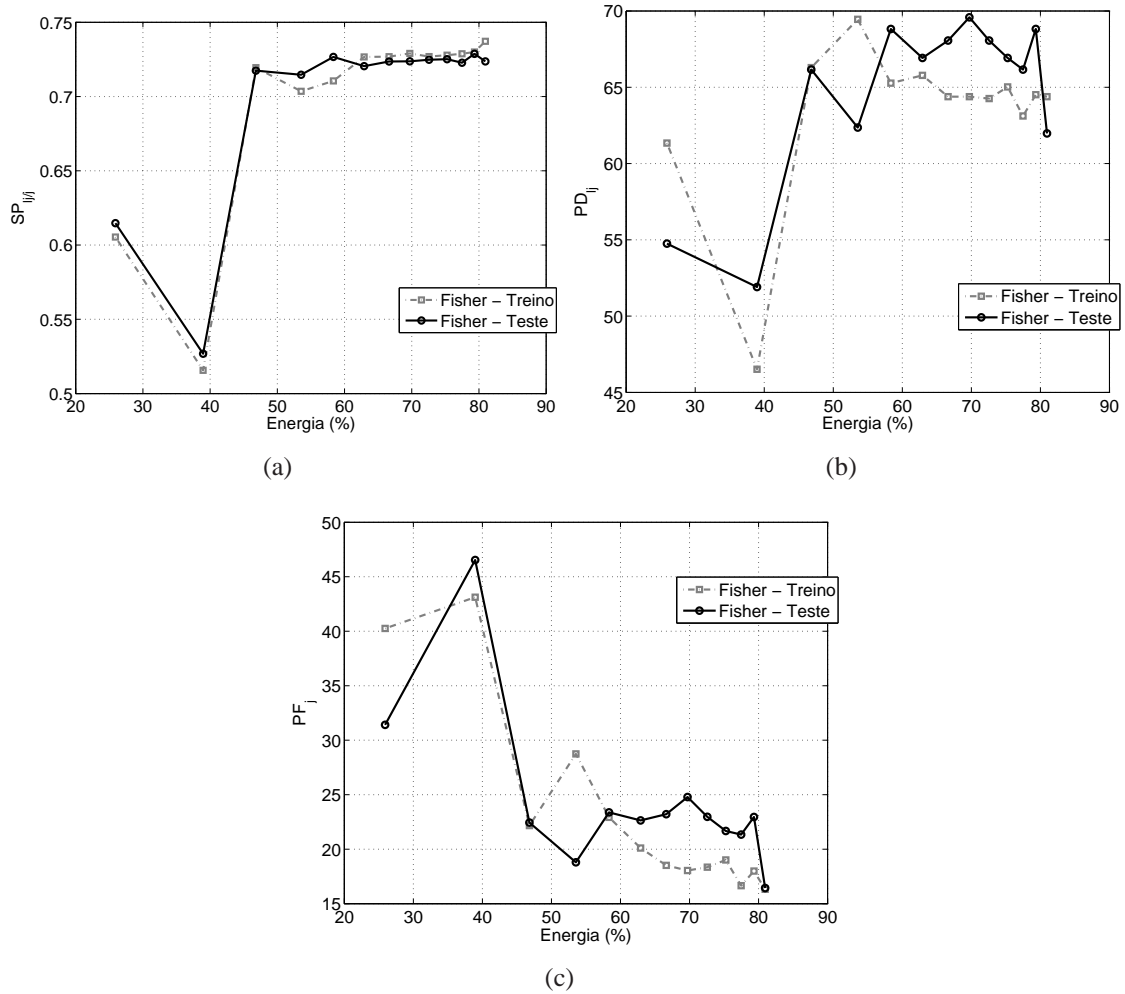


Figura 6.11: Variação de desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia: (a) Produto  $SP_{ij}$ ; (b) Probabilidade de detecção  $PD_{ij}$ ; (c) Probabilidade de falso alarme  $PF_j$ .

disso, é importante detalhar as características do classificador final obtido, conforme mostra a Figura 6.12. Os histogramas de saída (Figura 6.12(a)) mostram certa similaridade ao observado com o uso do classificador linear operando sobre anéis normalizados. Há uma grande sobreposição das saídas dos eventos em análise na régua de decisão e nenhum dos valores desejados são alcançados. A curva ROC (Figura 6.12(b)) e o produto  $SP_{l;j}$  (Figura 6.12(c)) refletem o comportamento do classificador em função desse limiar de saída. Já as probabilidades de detecção e falso alarme em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector ATLAS são mostradas nas Figuras 6.12(d), 6.12(e) e 6.12(f), respectivamente.

Analisando os resultados e comparando-os com os valores obtidos na abordagem anterior em que a dimensão dos dados observados também era reduzida (regiões  $\Delta R$ ), os valores aqui apresentados são superiores aos obtidos com  $\Delta R < 0,1$  e  $0,2$ . Apesar de ser mais custoso construir os anéis de uma RoI inteira nas diferentes camadas para só posteriormente reduzir a dimensão dos dados, os números finais justificam esse trabalho, ainda mais quando se trata de eventos raros e exóticos como é o caso dos *leptonjets*. No geral, entende-se o resultado como positivo, pois apresenta-se uma solução linear para o problema com uma significativa redução da dimensionalidade dos dados de entrada.

Ainda avaliando o comportamento da abordagem proposta, o valor da relevância das variáveis de entrada (Figura 6.13) mostra, principalmente, que três componentes extraídas concentram a informação importante para a tomada de decisão, indicando que o problema poderia ser resolvido com um número ainda menor de variáveis de entrada. A qualidade das componentes principais, i.e., o quanto ortogonais são entre si ou o quão correlacionadas elas estão é ilustrado na Figura 6.13(b). Os baixos valores de correlação das componentes extraídas indicam que elas são uma boa aproximação da informação que realmente rege o processo. O baixo grau de redundância da informação presente entre elas também reforça a tese de que o problema poderia ser resolvido com ainda menos variáveis de entrada.

A partir da relevância calculada é possível fazer um levantamento da relação entre a componente de maior valor e os anéis originais, como mostrado na Figura 6.13(c). As curvas indicam o número acumulado de anéis com valor igual ou inferior a um determinado patamar de correlação. Os marcadores indicam o valor absoluto específico de correlação entre o anel e a componente de máxima relevância. A análise das curvas mostra que a variável de maior relevância no processo decisório está relacionada com os anéis originais das diferentes camadas dos calorímetros praticamente da mesma maneira, seja para *leptonjets* seja para jatos hadrônicos.

Analisando os *leptonjets*, verifica-se que um grande número de anéis estão fracamente correlacionados com a componente extraída (especificamente os anéis externos da camada EM *front layer*, com exceção dos últimos), i.e., poucos contribuem de maneira contundente para o descritor extraído, porém o restante dos anéis divide o protagonismo, principalmente os primeiros anéis das camadas *Pre Sampler*, EM *front layer* e HAD *layer 1*,

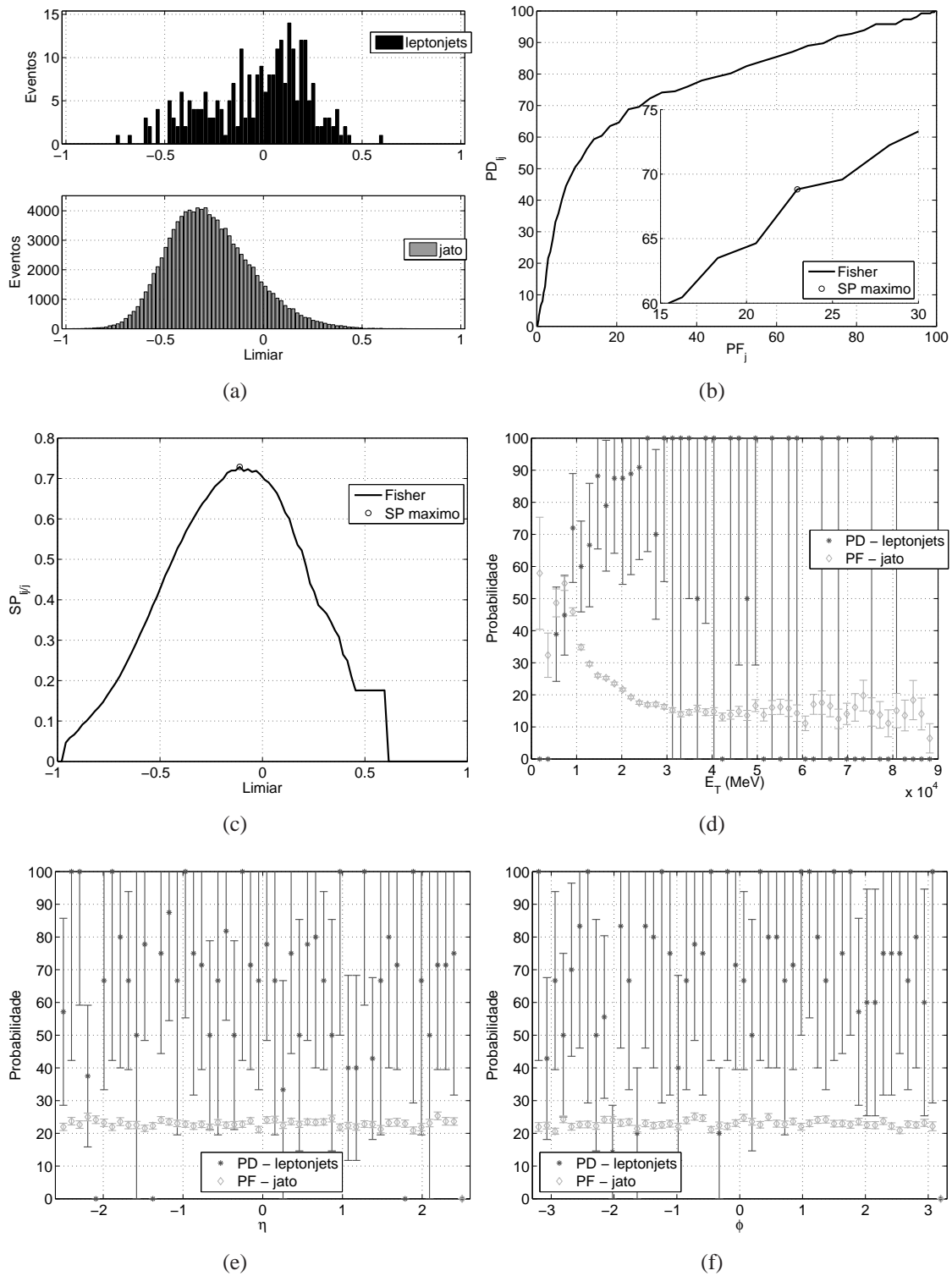


Figura 6.12: Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{l/j}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

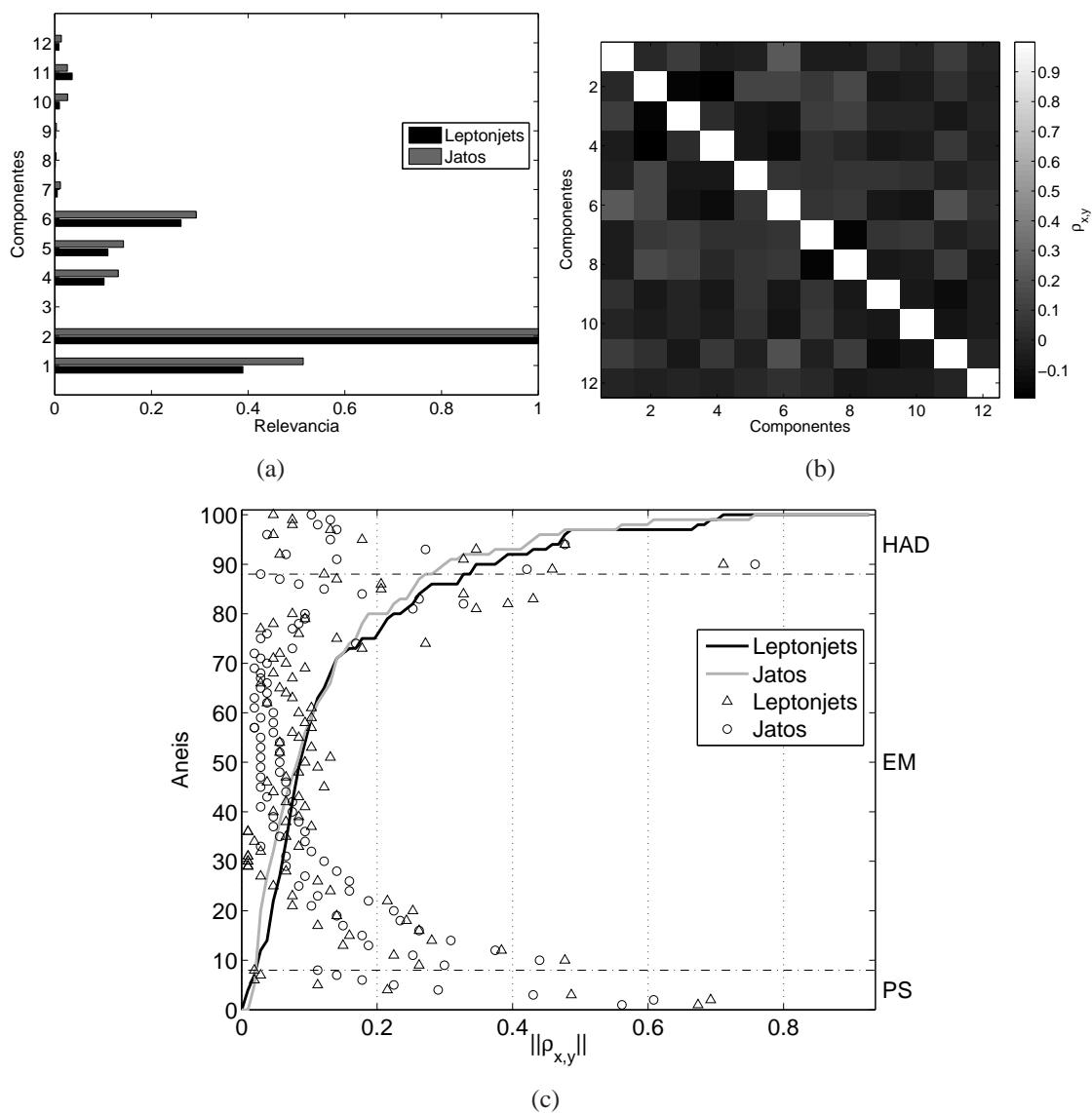


Figura 6.13: Medidas qualitativas das variáveis de entrada do classificador linear de Fisher operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia: (a) Relevância das componentes principais no processo decisório implementado; (b) Correlação entre as componentes principais extraídas; (c) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes principais.



similar ao que se observara com o classificador neural operando sobre os anéis originais normalizados.

A implementação da mesma abordagem com um classificador neural apresentou melhores resultados. A partir das curvas de carga (Figura 5.6) e com um número específico de componentes principais associado a um percentual de energia retido, diferentes topologias de classificadores neurais foram testadas, todas com variação apenas no número de neurônios na camada intermediária da rede. A variação desses dois parâmetros resultou em diferentes valores finais de desempenho. O número de neurônios da camada escondida da rede variou de 2 a 10 neurônios, enquanto o percentual de energia retida alcançou valores próximos a 80%.

A Tabela 6.7 apresenta os resultados das melhores configurações para cada uma das normalizações de dados empregada. Assim como o observado com o classificador linear, os melhores resultados foram obtidos com um número não tão pequeno de componentes principais. A melhora nos valores finais - em relação ao classificador Fisher - deve-se ao uso da não-linearidade aplicada pelo classificador neural. Os valores finais alcançados apresentam um ganho significativo ao obtido com o algoritmo linear.

Tendo em vista os valores de referência até aqui apresentados, o uso da normalização por energia favorece a informação do canal (*leptonjets/jato*), sendo a que apresentou melhores resultados na maioria dos testes. A Figura 6.14 apresenta o comportamento do classificador neural em função das componentes retidas e do número de neurônios na camada escondida da rede para eventos normalizados por energia.

Tabela 6.7: Desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Neurônios	Comp.	Energia(%) <sup>2</sup>	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ij/j</sub>	limiar
Energia	9	9	72,58	70,72	18,09	0,7621	0,11
Sequencial	9	13	61,68	73,00	26,89	0,7306	0,07
Camada	4	12	78,63	66,16	22,15	0,7188	0,15

O gráfico com a evolução do erro quadrático médio de treinamento (Figura 6.14(a)) indica que o valor final alcançado foi muito mais determinado pela energia retida do sinal de entrada do que pelo número de neurônios na cada escondida da rede. Essa conclusão só reforça o que já tinha sido observado em testes anteriores, os resultados são muito próximos, independentemente do classificador utilizado. A Figura 6.14(b) mostra o comportamento do produto SP<sub>ij/j</sub> do canal. A variação do número de neurônios intermediários da rede, assim como do percentual de energia retido do sinal indicam que quanto maior esses valores melhor o resultado final do classificador. O destaque fica para o percentual de energia do sinal retido, que impacta mais fortemente o resultado final do que a complexidade do classificador não-linear usado. É visível que para um mesmo percentual de

<sup>2</sup>O percentual de energia se refere ao acumulado nas componentes principais.

energia a curva em função do número de neurônios intermediários sofre uma leve variação, principalmente com baixa energia, haja visto a pouca informação disponível para modelar a estrutura interna de decisão. Quando o percentual de energia retido cresce, o número elevado de parâmetros livres do classificador neural ganha certo protagonismo no desempenho final, prova disso é que todos resultados dessa abordagem foram superiores aos obtidos com o classificador linear de Fisher.

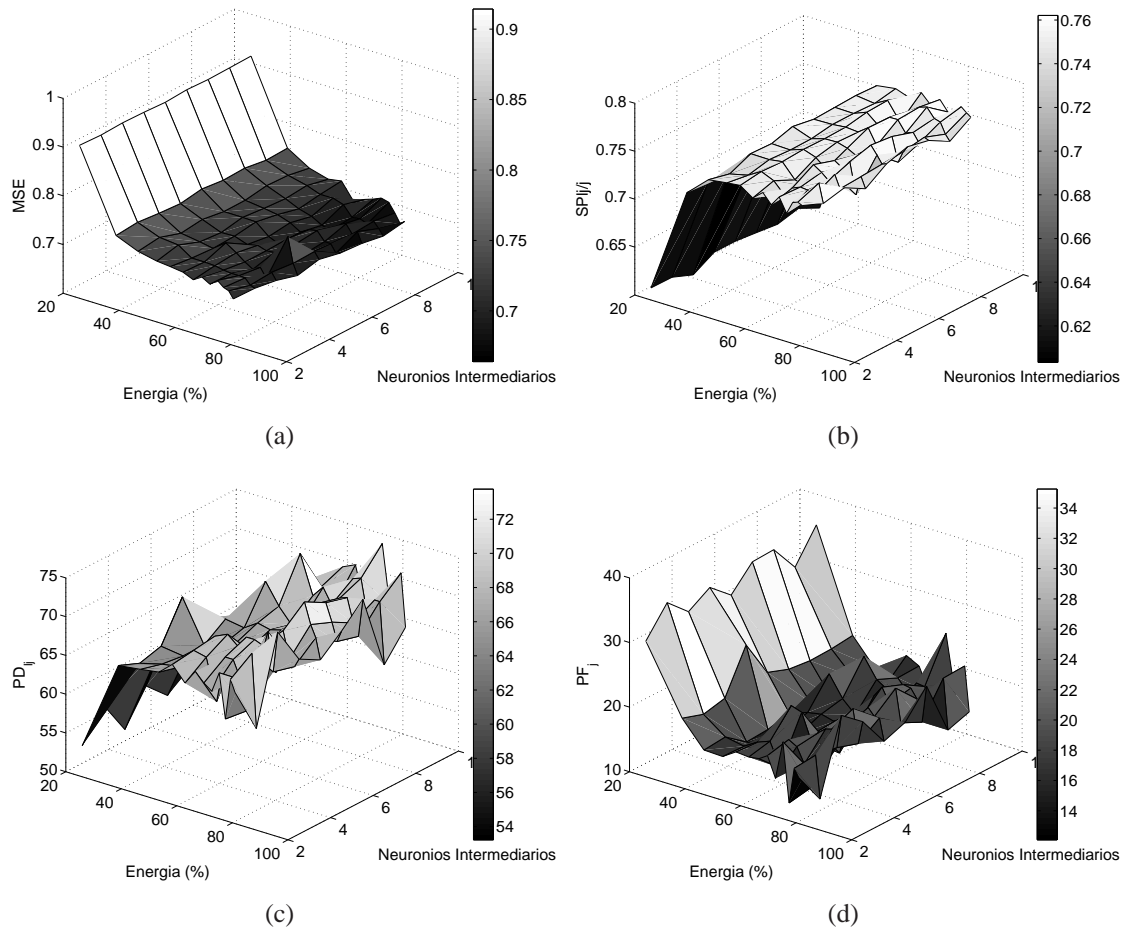


Figura 6.14: Variação de desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia, em função do número de neurônios na camada intermediária da rede e da energia retida do sinal: (a) Erro quadrático médio (MSE) final de treinamento; (b) Produto  $SP_{1j/j}$  associado à melhor rede; (c) Probabilidade de detecção  $PD_{1j}$  associada à melhor rede; (d) Probabilidade de falso alarme  $PF_j$  associada à melhor rede.

Os diferentes valores de produto  $SP_{1j/j}$  obtidos estão associados, respectivamente, às probabilidades de detecção e falso alarme mostradas nas Figuras 6.14(c) e 6.14(d).

Na avaliação do produto  $SP_{1j/j}$ , que é o medidor de desempenho final, a configuração que contempla o melhor classificador é alcançada com apenas 9 neurônios intermediários e 9 componentes principais extraídas (ou 72,58% da energia retida do sinal). Com esse classificador é possível discriminar 70,72% dos *leptonjets*, para uma taxa de falso alarme de 18,09%, resumidos em um produto  $SP_{1j/j}$  de 0,7621. A Figura 6.15 ilustra o de-

sempenho do classificador. Os histogramas de saída são apresentados na Figura 6.15(a). Novamente fica evidente o ganho oferecido pelo algoritmo neural ao aproximar os valores de saída de seus respectivos valores desejados e, dessa maneira, maximizar a separação dos padrões na régua de decisão do classificador. A curva ROC e o produto SP são apresentados nas Figuras 6.15(b) e 6.15(c), respectivamente. A eficiência em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector são mostrada nas Figuras 6.15(d), 6.15(e) e 6.15(f). Confirma-se o já se observara nos testes anteriores sobre a uniformidade apresentada pelo falso alarme em função das coordenadas do detector. Esse resultado elimina tendências do classificador em função da região de penetração das partículas dos eventos analisados. Considera-se o resultado excelente, dado, inclusive, que a configuração final obtida do classificador (topologia 9-9-1) e a compactação implementada (PCA) implicam operações bastante simples do ponto de vista computacional, além dos resultados apresentados serem os melhores obtidos até aqui. Tudo isso com uma taxa de compactação de dados de 91%.

A análise de relevância feita para o classificador neural (Figura 6.16(b)) indica que das componentes retidas, duas detêm valor significativo e que a maior parte das outras componentes são mais relevantes para o classificador identificar jatos hadrônicos do que *leptonjets*. A correlação entre todas as componentes extraídas é mostrada na Figura 6.16(c). Já a correlação entre a componente de maior relevância e cada um dos anéis originais é apresentada na Figura 6.16(d).

### 6.3.2 ICA

Como mencionado no capítulo anterior, as técnicas de ICA apresentam uma série de dificuldades ao operar em ambientes de alta dimensionalidade, ora por instabilidade numérica ora por dificuldade de convergência. Para contornar esse problema, a maioria das técnicas empregam uma etapa de branqueamento de dados, que além de garantir a decorrelação linear dos dados transformados permite a redução da dimensionalidade com algum controle sobre a qualidade deles. Os algoritmos ICA operam então sobre essa base de dados dimensão reduzida. Dependendo da propriedade estatística dos dados transformados, o algoritmo consegue implementar alguma rotação nessa base de maneira a tornar os dados o mais independente possível. Como a etapa de branqueamento dos dados está diretamente ligada à técnica PCA, isso nos habilita a usar a mesma curva de carga empregada no capítulo anterior para a análise dos resultados desta seção. Novamente, os testes compreenderam a análise de classificadores lineares e não-lineares.

A Tabela 6.8 quantifica o desempenho obtido com o classificador linear. Fica evidente que o algoritmo ICA teve pouca margem de manobra sobre os dados branqueados e compactados. Assim como na PCA, nota-se que o desempenho do algoritmo de classificação melhora com o aumento do número de componentes independentes fornecidas e, por con-

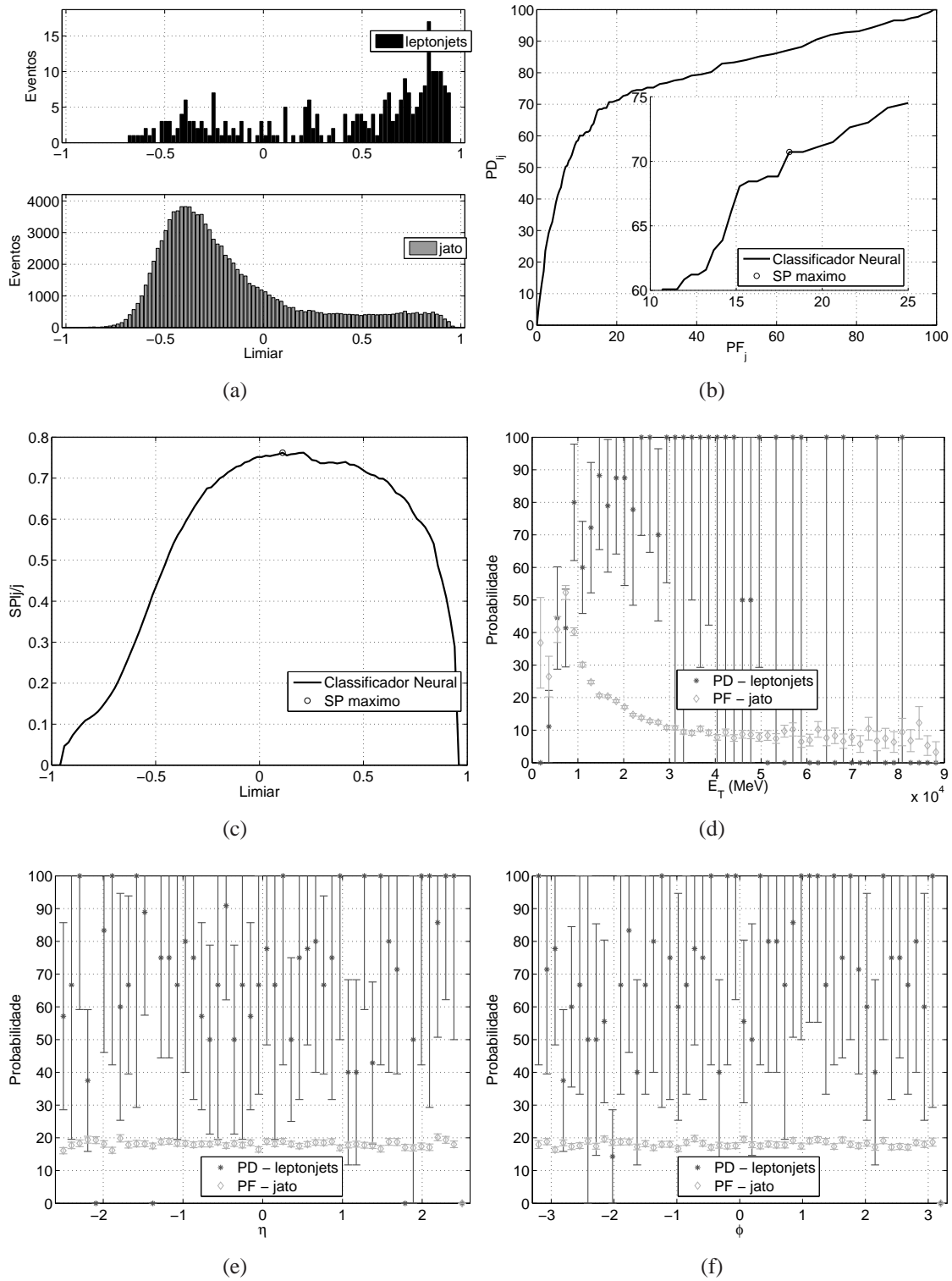


Figura 6.15: Desempenho do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia: (a) Histogramas de saída; (b) Curva ROC; (c) Produto  $SP_{ij}$ ; (d) Probabilidades de detecção e falso alarme em função da  $E_T$ ; (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector; (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

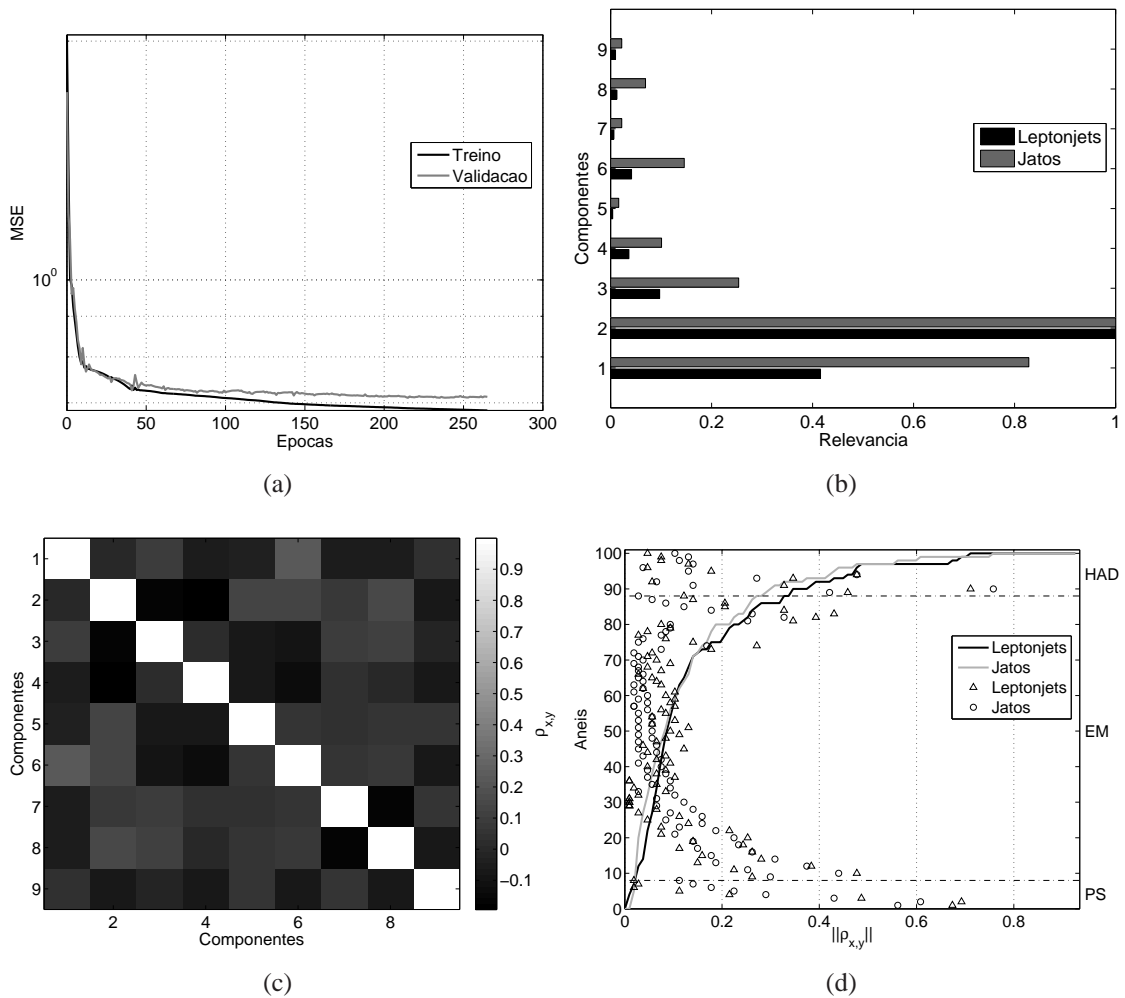


Figura 6.16: (a) Erro quadrático médio de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes principais extraídas de dados formatados em anéis e normalizados por energia: (b) Relevância das componentes principais no processo decisório implementado; (c) Correlação entre as componentes principais extraídas; (d) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes principais.

seguinte, mais detalhes dos eventos sob análise. O melhor resultado é alcançado com 9 componentes independentes do sinal quando submetido à normalização por energia. A projeção dos jatos hadrônicos nessas componentes é suficiente para uma substancial diferenciação dos eventos. Os valores obtidos para as figuras de mérito por esse classificador são: 68,82% de  $PD_{lj}$ ; 22,96% de  $PF_j$  e 0,7287 de  $SP_{lj/j}$ . Considerando esses números, podemos identificar essa estratégia como uma abordagem viável para a solução do problema tratado, porém, com desempenho ainda inferior ao do algoritmo neural operando sobre componentes principais, mostrado na seção anterior.

A Figura 6.17 mostra a evolução do desempenho obtido com um classificador linear operando sobre componentes independentes extraídas na etapa de pré-processamento. Os gráficos são exibidos em função da energia retida do sinal na etapa de branqueamento dos dados. O produto  $SP_{lj/j}$  do canal é mostrado na Figura 6.17(a) e as correspondentes taxas de eficiência na detecção de *leptonjets* e falso alarme de jatos são mostradas nas Figuras 6.17(b) e 6.17(c). Como mencionado anteriormente, o aumento no número de componentes extraídas enriquece as diferenças dos perfis de *leptonjets* e jatos hadrônicos, traduzindo-se em um valor de produto  $SP_{lj/j}$  alto.

A figura 6.18 apresenta, detalhadamente, o comportamento do classificador selecionado. Os histogramas de saída do classificador (Figura 6.18(a)) exibem a separação alcançado pelo algoritmo de Fisher entre dados da classe *leptonjets* e jatos. Nota-se que os valores de saída do classificador na régua de decisão estão afastados dos valores desejados para cada um dos eventos e não tão afastados entre si como se desejava. A moda de cada um dos histogramas de saída quase se sobrepõem, percebendo-se um número considerável de eventos da classe *leptonjets* com valor de saída praticamente sobre a moda da classe jato, evidenciando-se, mais uma vez, a complexidade do problema tratado. Apesar de as componentes independentes tratarem os dois eventos de maneira diferente, os perfis são próximos o suficiente para, mesmo sob essa deformação, guardarem características comuns fortes. A partir da variação do limiar de decisão, é possível construir a curva ROC (Figura 6.18(b)) e o produto  $SP_{lj/j}$  (Figura 6.18(c)) do classificador que refletem as considerações mencionadas. As probabilidades de detecção e de falso alarme em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  do detector ATLAS são apresentadas nas Figuras 6.18(d), 6.18(e) e 6.18(f), respectivamente.

Um outro aspecto avaliado na metodologia foi a relevância das componentes independentes extraídas e a relação delas com os anéis originais. A independência estatística, medida pela informação mútua (Figura 6.19(b)) entre as componentes, mostra o nível alcançado ao final da etapa de pré-processamento. No gráfico, o valor 0 indica nenhuma informação compartilhada. Assim, pode-se assumir as componentes extraídas como uma excelente aproximação das componentes que realmente regem o processo estudado. Po-

---

<sup>3</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.

Tabela 6.8: Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis normalizados de diferentes maneiras.

Normalização	Comp.	Energia(%) <sup>3</sup>	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ijj</sub>	limiar
Energia	1	25,91	54,75	31,41	0,6148	-0,27
	2	38,96	65,02	19,91	0,7236	0,33
	3	46,81	66,16	22,44	0,7174	0,31
	4	53,54	62,36	18,80	0,7147	0,21
	5	58,34	68,82	23,38	0,7267	0,07
	6	62,91	66,92	22,64	0,7204	-0,11
	7	66,63	68,06	23,21	0,7236	-0,07
	8	69,71	69,58	24,79	0,7237	-0,13
	9	72,58	68,06	22,97	0,7247	-0,11
	10	75,25	66,92	21,67	0,7251	-0,01
	11	77,47	66,16	21,34	0,7227	-0,05
	12	79,33	68,82	22,96	0,7287	-0,11
	13	80,95	61,98	16,44	0,7237	0,19
Sequencial	1	9,38	70,34	35,43	0,6743	-0,09
	2	16,91	62,74	27,61	0,6748	-0,05
	3	24,07	65,40	27,87	0,6873	-0,03
	4	30,39	73,38	35,71	0,6876	-0,09
	5	35,80	73,38	35,75	0,6874	-0,09
	6	40,29	72,62	31,72	0,7043	-0,13
	7	44,31	68,82	26,30	0,7124	0,03
	8	47,75	70,34	27,38	0,7148	0,01
	9	51,04	68,44	22,85	0,7273	-0,01
	10	54,07	67,68	23,12	0,7221	-0,01
	11	56,73	69,96	26,06	0,7194	-0,09
	12	59,30	69,58	25,92	0,7181	-0,09
	13	61,68	68,44	24,70	0,7183	-0,09
	14	64,01	67,68	24,25	0,7166	-0,07
	15	66,24	65,02	21,34	0,7168	-0,05
	16	68,40	66,92	22,73	0,7200	-0,03
	17	70,41	66,92	22,83	0,7195	-0,03
	18	72,42	65,40	21,44	0,7183	-0,01
	19	74,31	68,44	28,90	0,6977	0,05
20	76,11	64,64	20,53	0,7186	-0,01	
Camada	1	14,89	66,92	36,33	0,6528	-0,17
	2	25,67	68,06	38,72	0,6463	-0,19
	3	35,07	62,36	29,27	0,6648	-0,23
	4	43,32	58,94	23,47	0,6745	-0,15
	5	50,88	70,34	34,02	0,6815	-0,31
	6	57,45	63,88	27,55	0,6810	-0,25
	7	62,05	64,26	27,63	0,6825	-0,25
	8	66,31	71,86	35,82	0,6797	-0,35
	9	69,91	72,24	35,13	0,6851	-0,35
	10	73,26	70,34	37,62	0,6630	-0,29
	11	76,21	65,78	40,49	0,6261	-0,21
	12	78,63	70,72	32,73	0,6898	-0,19
	13	80,71	53,61	23,34	0,6462	0,13

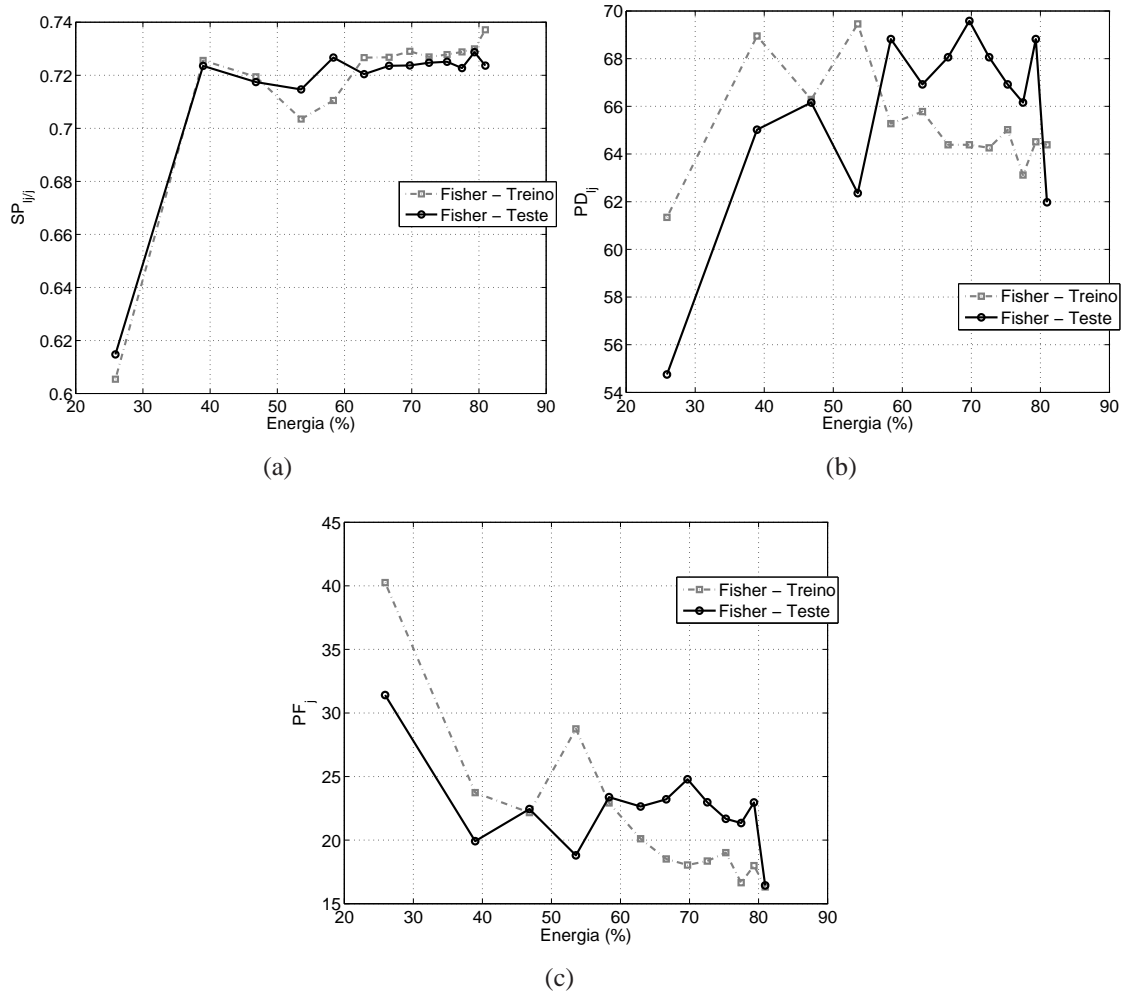


Figura 6.17: Variação de desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia. Resultados em função da energia retida do sinal: (a) Produto  $SP_{1j/j}$ ; (b) Probabilidade de detecção  $PD_{1j}$ ; (c) Probabilidade de falso alarme  $PF_{1j}$ .



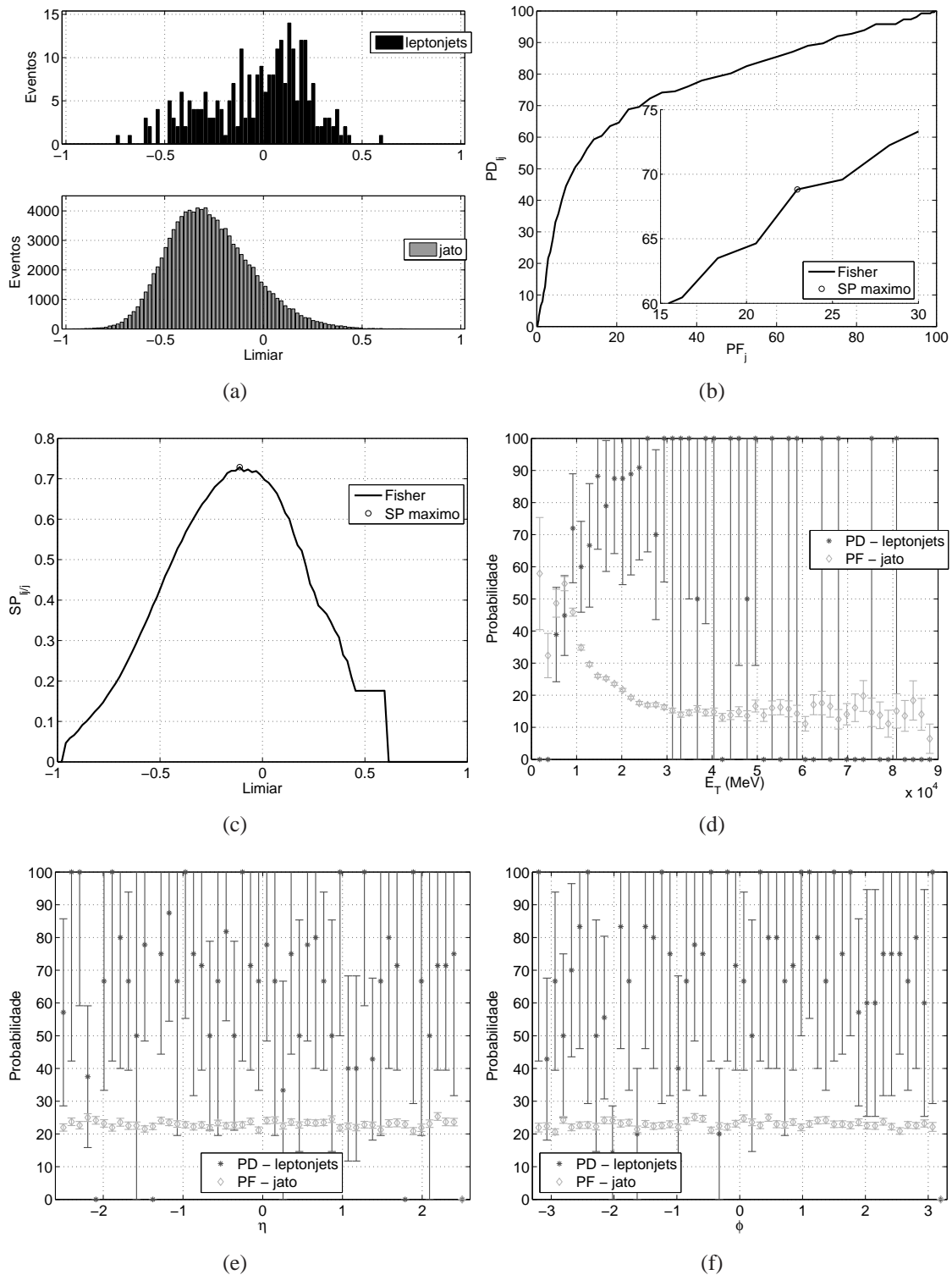


Figura 6.18: Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{l/j}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

rém, nem todas são relevantes para o processo decisório. O valor de relevância dessas variáveis (Figura 6.19(a)) mostra, principalmente, a importância de uma componente na construção da estrutura de decisão do classificador e quatro outras com um papel secundário. Não é possível estabelecer uma relação direta dessa componente independente com a componente principal de máxima relevância da seção anterior, pois a técnica ICA não guarda uma hierarquia das componentes extraídas, como a PCA. Contudo é possível verificar a correlação da componente de máxima relevância com os anéis originais (Figura 6.19(c)) e verificar quais desses e, por conseguinte, quais camadas dos calorímetros mais auxiliam o processo decisório. Essa componente mais relevante tem forte correlação com os primeiros anéis das camadas do calorímetro eletromagnético, e, portanto, uma relação mais moderada com o restante dos anéis da camada EM *front layer*. Isso indica que ela guarda informação (principalmente) da magnitude da deposição de energia da partícula no calorímetro eletromagnético e, em segundo plano, da forma do cone de deposição dessa energia, o que está alinhado com a literatura e os resultados obtidos até então.

Para os testes com o classificador neural, diferentes valores de energia retida e diferentes topologias de classificadores (todas com variação no número de neurônios na camada intermediária da rede) foram implementados, de forma que a variação desses dois parâmetros resultou em diferentes valores finais de desempenho. O número de neurônios da camada escondida da rede variou de 2 a 10 neurônios, enquanto o percentual de energia retida alcançou valores em torno de 80%, como nos testes com o classificador linear.

Seguindo o raciocínio até aqui exposto, selecionamos os melhores resultados em função da nossa figura de mérito, o produto  $SP_{ij/j}$ , como mostra a Tabela 6.9.

Tabela 6.9: Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Neurônios	Comp.	Energia(%) <sup>4</sup>	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ij/j</sub>	limiar
Energia	9	4	53,54	69,96	19,25	0,7526	0,13
Sequencial	9	19	74,31	69,58	24,20	0,7266	0,11
Camada	5	12	78,63	65,78	22,55	0,7150	0,17

As virtudes oferecidas pelo algoritmo de classificação não-linear resultaram no aumento da eficiência em relação ao obtido com o classificador de Fisher. O aumento não tão expressivo quanto o esperado, indica que o mérito nesse caso é compartilhado com o pré-processamento realizado, pois com o uso do algoritmo linear - como mostrado anteriormente - a tarefa foi também bem desempenhada. Ainda assim, a evolução na detecção de *leptonjets* e o decréscimo no falso alarme são sempre bem vindos na difícil tarefa de separação dos padrões estudados. A abordagem se apresenta como a segunda melhor dentre todos os resultados obtidos até aqui. Escolhendo a normalização por energia, pois

<sup>4</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.

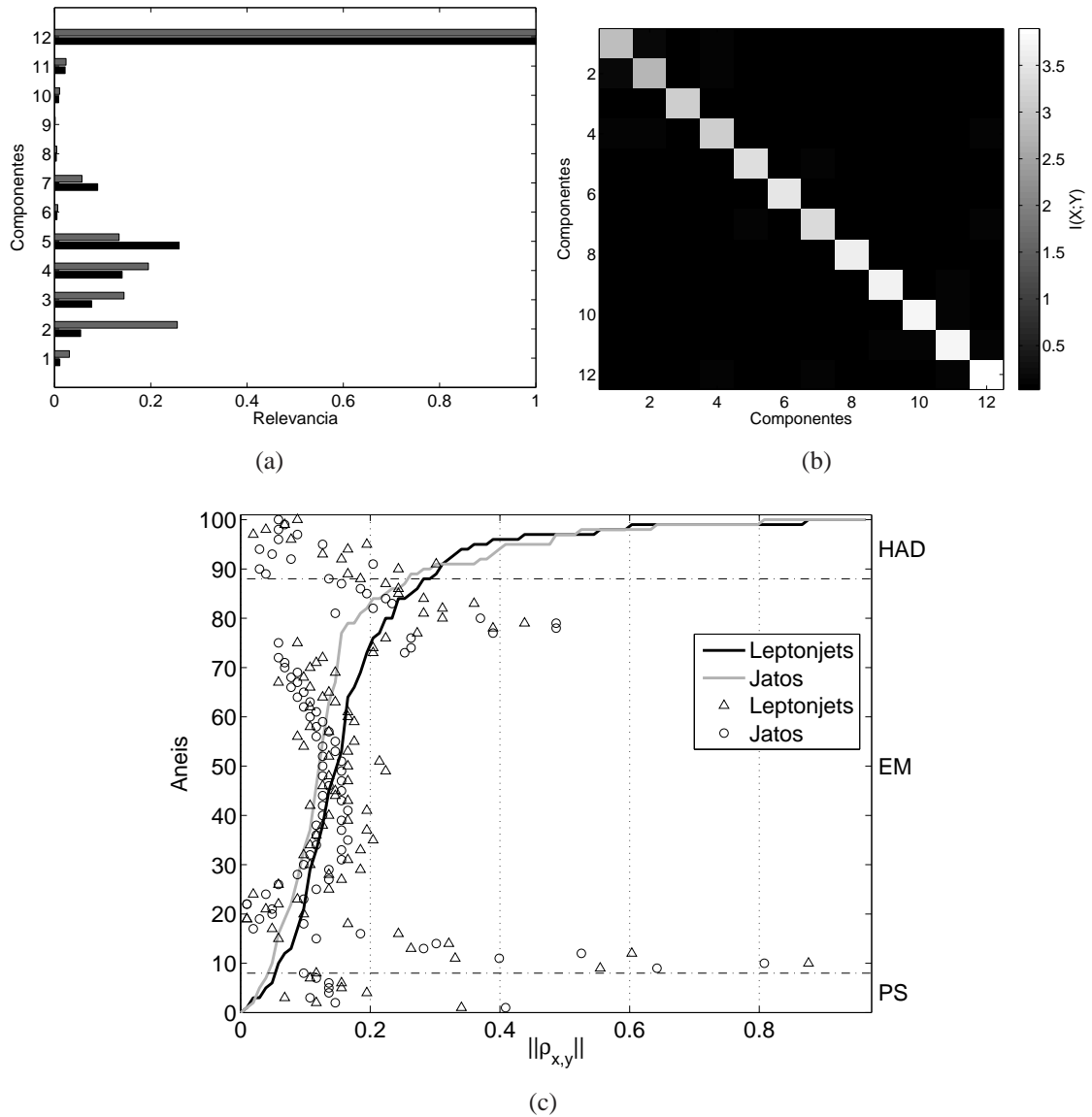


Figura 6.19: Medidas qualitativas das variáveis de entrada do classificador linear de Fisher operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia: (a) Relevância das componentes independentes no processo decisório implementado; (b) Informação mútua entre as componentes independentes extraídas; (c) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes.

é a que oferece o melhor resultado final, verificamos o comportamento dos diferentes classificadores testados através das curvas obtidas para cada uma das figuras de mérito (Figura 6.20).

A Figura 6.20(b) mostra o gráfico de comportamento do produto  $SP_{l_j/j}$  do canal. A variação do número de neurônios intermediários da rede, assim como do percentual de energia retido do sinal se aproxima do observado com a PCA. O número de componentes extraídas é mais determinante no resultado final do que a complexidade do classificador usado. Para um mesmo número de componentes extraídas, a curva em função do número de neurônios intermediários sofre uma leve variação, principalmente com baixa energia, haja visto a pouca informação disponível para modelar a estrutura interna de decisão. Quando o percentual de energia retido aumenta, o número elevado de parâmetros livres do classificador neural ganha certo protagonismo no desempenho final, contudo não é nesse cenário que os valores máximos são alcançados, dado que informação não relevante - inclusive informação que confunde o classificador - é inserida. O desempenho máximo é alcançado, provavelmente, quando as componente de máxima relevância para o processo decisório são extraídas, em torno de 55% da energia do sinal.

Os diferentes valores de produto  $SP_{l_j/j}$  obtidos estão associados, respectivamente, às probabilidades de detecção e falso alarme mostradas nas Figuras 6.20(c) e 6.20(d).

O melhor classificador apresenta a topologia 4 – 9 – 1. Os 9 neurônios intermediários são alimentados por 4 componentes independentes extraídas (ou 53,54% da energia do sinal). Com esse classificador é possível discriminar 69,96% dos *leptonjets* para uma taxa de falso alarme de 19,25%. A abordagem apresentou duas características interessantes: manteve o elevado índice de aceitação de *leptonjets* já alcançado com a PCA (melhor resultado até aqui) e obteve uma taxa de compactação dos dados de entrada das mais expressivas.

Apesar de a configuração final obtida pelo classificador (topologia 4-9-1) inflar a dimensão dos dados de entrada na camada escondida da rede, a compactação implementada (ICA) implica uma operação bastante simples do ponto de vista computacional, além de uma taxa final de compactação de 96%. Os detalhes do desempenho do classificador e as interpretações de seus resultados são apresentados na Figura 6.21 e discutidos adiante. Assim como na PCA, acredita-se que com a compactação realizada (e sua transformação de base), *leptonjets* e *jatos* tenham as características que os distinguem ressaltadas. Ao eliminar as componentes menos energéticas na etapa de branqueamento dos dados e implementar o algoritmo de busca pela independência estatística, as informações que causariam confusão no discernimento entre *leptonjets* e *jatos* hadrônicos são reduzidas.

É possível verificar pelos histogramas de saída da rede que os padrões, principalmente os *jatos* hadrônico, apresentam dificuldades de se aproximarem de seu valor desejado de saída, como mostra a Figura 6.21(a) e isso se reflete na curva ROC (Figura 6.21(b)) e produto  $SP_{l_j/j}$  (Figura 6.21(c)). A maior confusão - eventos perdidos e falsos positivos

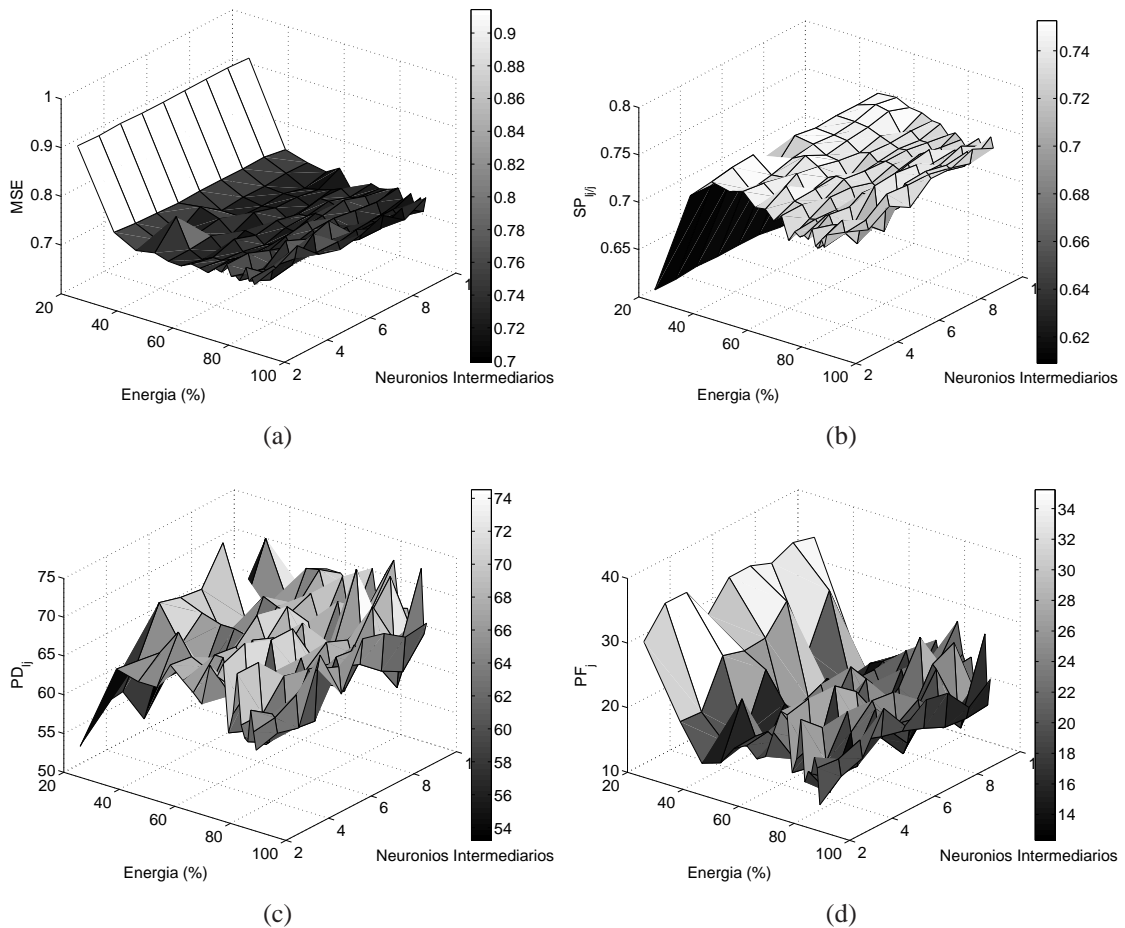


Figura 6.20: Variação de desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis normalizados por energia em função do número de neurônios na camada intermediária da rede e da energia retida do sinal: (a) Erro quadrático médio (MSE) final de treinamento; (b) Produto  $SP_{ij}$  associado à melhor rede; (c) Probabilidade de detecção  $PD_{ij}$  associada à melhor rede; (d) Probabilidade de falso alarme  $PF_j$  associada à melhor rede.

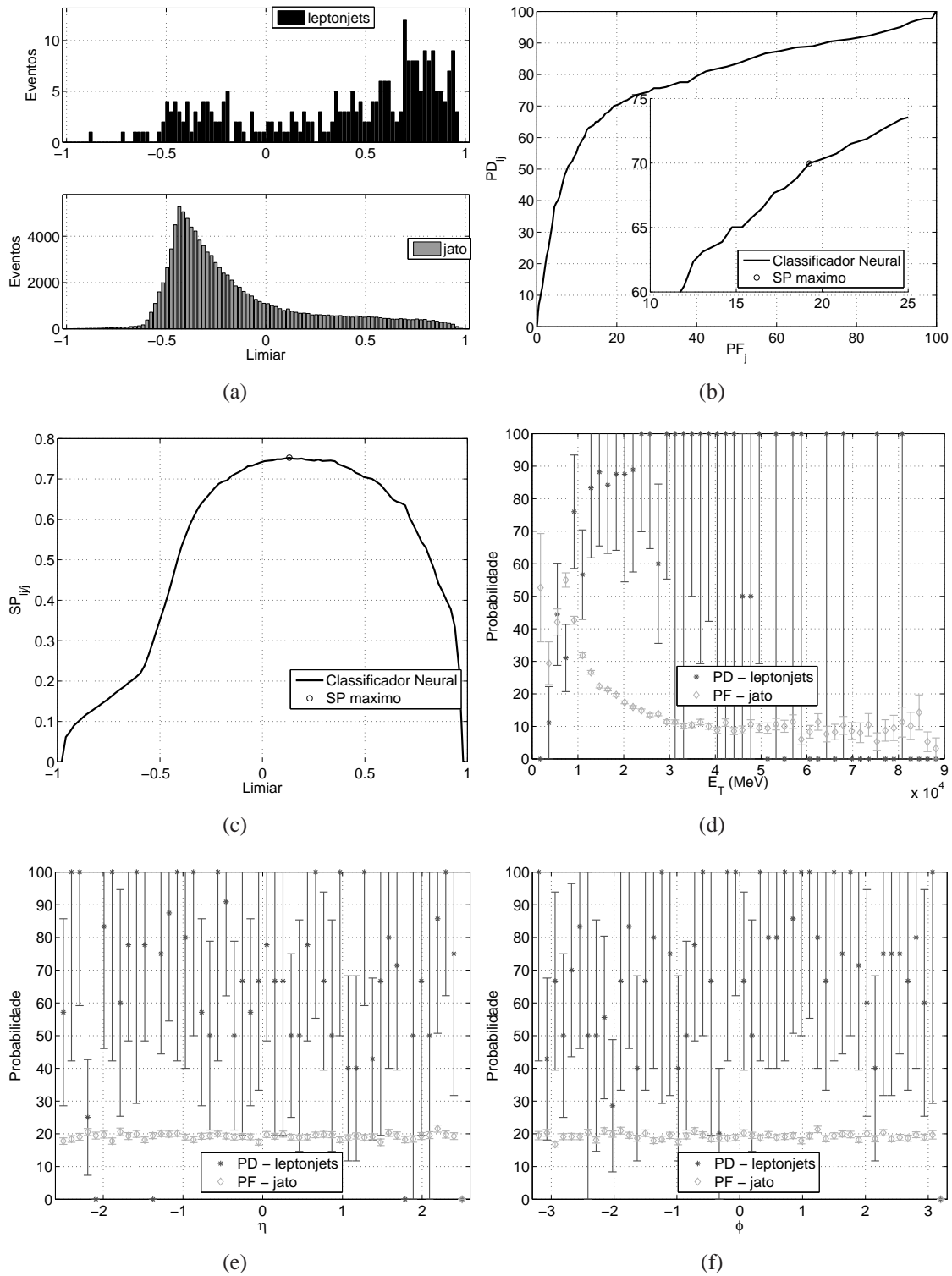


Figura 6.21: Desempenho do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{II}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

- é observada em eventos de baixa energia. Os *leptonjets* nessas condições são menos colimados o que os aproximaria, ainda mais, do perfil dos jatos hadrônicos. Apesar do esforço do algoritmo de melhorar os resultados nessa faixa de energia ( $E_T < 20\text{GeV}$ , Figura 6.21(d)), fica evidente que um outro tipo de processamento teria de ser despendido, especificamente, para eventos nessa região sem, é claro, uma deteriorização no restante da faixa de energia coberta. As Figuras 6.21(e) e 6.21(f) apresentam o comportamento do classificador em função das coordenadas  $\eta$  e  $\phi$  do detector ATLAS.

A análise da abordagem se completa com a análise dos dados que alimentam o sistema. Para isso, um conjunto de medidas são apresentadas na Figura 6.22. A qualidade da independência estatística alcançada na extração das componentes é medida pela informação mútua entre elas, como mostra a Figura 6.22(c). Os valores alcançados indicam que elas apresentam um baixo valor de informação compartilhada, i.e., uma boa aproximação das componentes independentes. Todas as componentes extraídas apresentam uma relevância significativa para o processo decisório (Figura 6.22(b)) e que a mais significativa delas congrega, principalmente, informação dos primeiros anéis das camadas eletromagnéticas e hadrônicas e que, de maneira secundária, dos anéis adjacentes ao primeiro anel da camada EM *front layer*.

## 6.4 Pré-processamento Segmentado

A redução de dimensionalidade dos dados de entrada na abordagem segmentada considera a divisão entre as diferentes camadas dos calorímetros como diferentes processos aleatórios (sete no total). A classificação é realizada através dos dados transformados em cada uma das camadas.

As componentes retidas por valor de energia especificado se referem ao número mínimo de componentes necessária para alcançar tal patamar. Esse procedimento foi escolhido devido à heterogeneidade da deposição de energia dos eventos em cada camada dos calorímetros. A extração de características de modo segmentado tem a vantagem de explorar as diferentes potencialidades de cada camada do calorímetro de maneira independente. Espera-se uma maior robustez a falhas nesse procedimento, uma vez que é de conhecimento que os sensores podem parar de funcionar ao longo do tempo e a abordagem segmentada retém essa perda de informação nas características extraídas na camada em questão. Em contraponto, uma taxa de compactação menor em relação a alcançada pela abordagem não segmentada é previsível, dado que as componentes finais obtidas fazem uso apenas de informação local e não consideram a relação dessa com a informação de outras camadas.

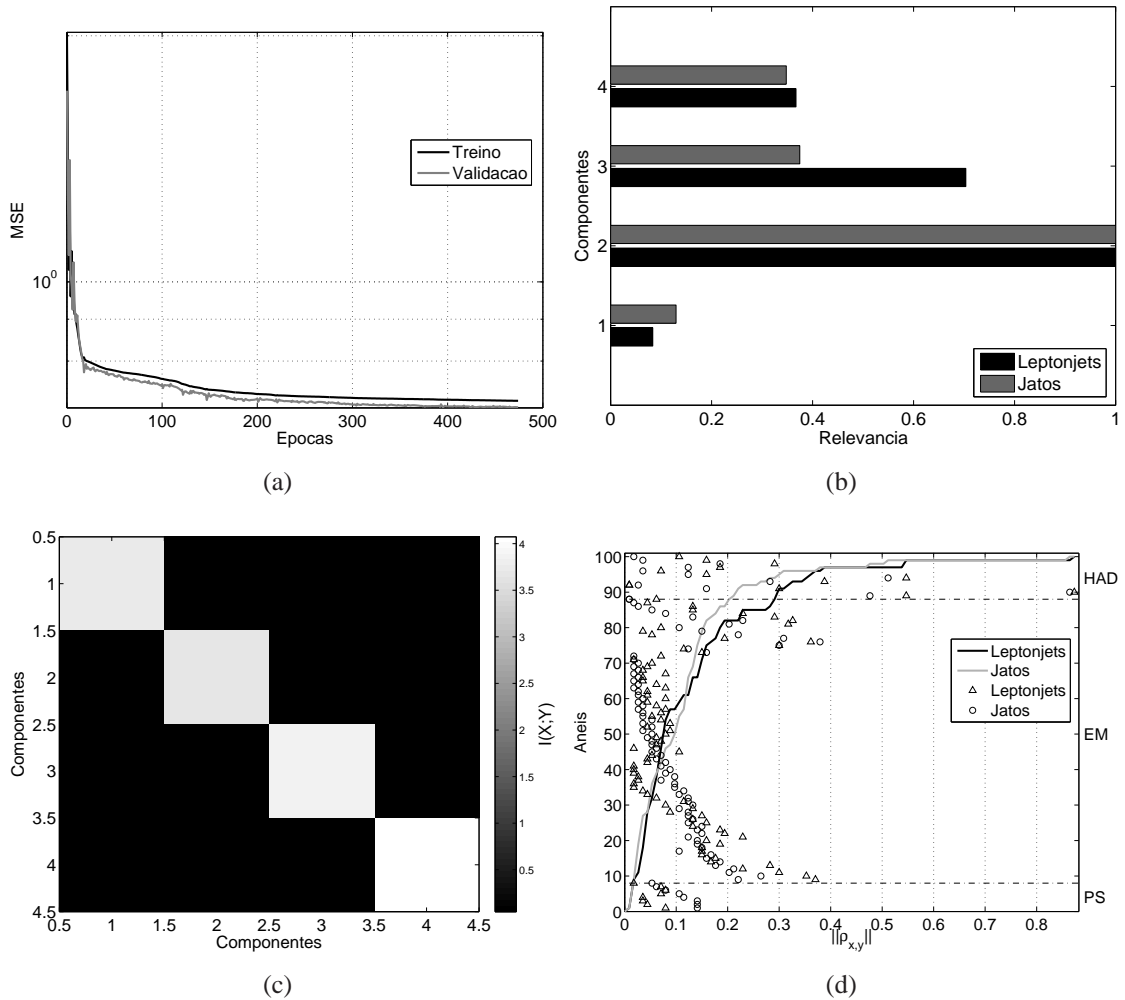


Figura 6.22: (a) Erro médio quadrático de treinamento do algoritmo neural. Medidas qualitativas das variáveis de entrada do classificador neural operando sobre componentes independentes extraídas de dados formatados em anéis e normalizados por energia: (b) Relevância das componentes independentes no processo decisório implementado; (c) Informação Mútua entre as componentes independentes extraídas; (d) Correlação entre a componente de máxima relevância no processo decisório e os anéis usados na extração das componentes independentes.



Tabela 6.10: Melhor desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Camada	Energia(%) <sup>5</sup>	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ij/j</sub>	limiar
Energia	<i>Pre Sampler</i>	75,56	55,51	31,53	0,6182	-0,68
	<i>EM front layer</i>	79,60	65,40	24,66	0,7028	-0,35
	<i>EM middle layer</i>	98,73	58,17	31,69	0,6314	0,25
	<i>EM back layer</i>	98,66	74,52	36,75	0,6877	0,76
	<i>HAD layer 0</i>	67,97	72,62	30,51	0,7105	0,58
	<i>HAD layer 1</i>	63,40	61,60	42,89	0,5933	0,31
	<i>HAD layer 2</i>	63,40	61,60	42,89	0,5933	0,31
Sequencial	<i>Pre Sampler</i>	20,12	25,48	35,28	0,4279	-0,03
	<i>EM front layer</i>	24,97	66,16	22,97	0,7149	0,05
	<i>EM middle layer</i>	87,63	60,08	34,32	0,6285	-0,76
	<i>EM back layer</i>	98,28	67,30	36,84	0,6521	-0,23
	<i>HAD layer 0</i>	93,41	59,32	24,52	0,6715	-0,05
	<i>HAD layer 1</i>	93,10	61,98	34,92	0,6352	-0,19
	<i>HAD layer 2</i>	89,43	51,33	36,82	0,5710	-0,37
Camada	<i>Pre Sampler</i>	32,17	54,75	27,44	0,6334	0,21
	<i>EM front layer</i>	87,38	69,96	23,41	0,7324	0,39
	<i>EM middle layer</i>	93,97	60,08	33,79	0,6310	0,41
	<i>EM back layer</i>	55,06	58,56	40,90	0,5883	-0,43
	<i>HAD layer 0</i>	90,00	64,26	35,09	0,6459	-0,47
	<i>HAD layer 1</i>	91,10	57,41	30,63	0,6325	-0,29
	<i>HAD layer 2</i>	72,70	60,46	44,74	0,5783	-0,25

### 6.4.1 PCA

O uso da PCA como extrator local de característica foi implementado em cada uma das 7 camadas dos calorímetros. Um classificador foi responsável por avaliar a separabilidade dos padrões sob a ótica das variáveis transformadas. A partir daí, obteve-se o número ideal de componentes retidas que maximizava o produto  $SP_{ij/j}$ . Aproveitando esses resultados, implementou-se um classificador cujas entradas são as componentes selecionadas anteriormente. O intuito foi verificar se os resultados produzidos pontualmente com essas componentes, poderiam ser melhorados com a junção da informação de outras camadas.

A estratégia da abordagem segmentada foi realizada com o uso de classificadores lineares e neurais. A Tabela 6.10 apresenta os resultados do melhor desempenho do classificador linear de Fisher ao lidar com as componentes extraídas em cada camada. Esse primeiro teste é importante pois serve para determinar o piso de eficiência por camada dos calorímetros na separação dos padrões. O termo piso foi escolhido devido ao classificador linear de Fisher representar a estrutura de decisão mais simples usada. A partir desses valores, pode-se verificar o ganho alcançado com outras estruturas de decisão mais complexas ou pelo uso combinado da informação das camadas em um classificador, também, linear.

O desempenho alcançado pela maioria das camadas, independentemente da normali-

<sup>5</sup>O percentual de energia se refere ao acumulado nas componentes principais.

zação, não é tão elevado, salvo algumas exceções que se aproximam dos melhores valores atingidos em testes anteriores. Isoladamente a primeira camada eletromagnética é a que fornece um melhor resultado final na diferenciação dos eventos, independentemente da normalização. Essa informação já tinha sido observada em alguns dos testes anteriores e mostra que a informação espacial, já que a camada EM *front layer* é a de maior resolução, tem bastante impacto na separação dos padrões. A magnitude da energia depositada também é notada nessa camada. Apesar de não ser a mais profunda, torna-se relevante por ser a primeira camada do calorímetro eletromagnético.

A partir dos resultados mostrados na Tabela 6.10, implementou-se um classificador linear cujas entradas são as componentes nela descritas. O intuito foi verificar se essas componentes conjuntamente poderiam produzir resultados ainda melhores. Parte-se das componentes pertencentes a camada com melhor desempenho e gradativamente se agrega as componentes das camadas subsequentes. A Tabela 6.11 mostra o melhor desempenho para cada uma das normalizações. A agregação da informação de algumas das camadas no processo decisório, implicou a melhora do resultado final para todas as normalizações estudadas. O resultado aglomerativo foi melhor do que o melhor dos resultados de cada camada isoladamente. Isso sinaliza que as potencialidades locais podem ser combinadas para maximizar os resultados.

O melhor classificador foi obtido com as informações das componentes principais extraídas da cada uma das camadas submetidas à normalização por energia. A configuração final obtida, indica um classificador simples do ponto de vista computacional. As 49 variáveis de entrada representam quase a metade do espaço de dimensão original. A Figura 6.23 ilustra o comportamento do classificador final obtido. Os histogramas de saída (Figura 6.23(a)) sugerem que a tarefa de classificação nesse espaço ainda é bastante complexa. Os padrões de *leptonjets* e jatos hadrônicos ficam muito próximos na régua de decisão, o que resulta em um classificador com margem de decisão (limiar de funcionamento) muito estreita. Mesmo com essa peculiaridade, o resultado obtido é comparável com o das outras abordagens até então testadas. A Figura 6.23(b) mostra a curva ROC do classificador. A Figura 6.23(c) apresenta o comportamento do produto  $SP_{ij}$  em função do limiar de decisão. As Figuras 6.23(d), 6.23(e) e 6.23(f) exibem, respectivamente, a eficiência do classificador em função da  $E_T$ , das coordenadas  $\eta$  e  $\phi$  do evento. Nota-se que para o limiar de trabalho escolhido não há qualquer tipo de tendência nos resultados do classificador.

Para o classificador neural, a Tabela 6.12 resume os melhores resultados por camada. No estudo do classificador neural foi variado o número de neurônios na camada escondida da rede. Um mínimo de 2 e um máximo de 5 neurônios foi estipulado para todas as camadas, com exceção das camadas hadrônicas, em que a dimensão é menor. Nessas camadas a rede neural foi fixada com 2 neurônios em sua camada escondida.

---

<sup>6</sup>O percentual de energia se refere ao acumulado nas componentes principais.

Tabela 6.11: Desempenho do classificador linear de Fisher operando sobre as componentes principais extraídas de forma segmentada mostradas na Tabela 6.10.

Normalização	Componentes							PD <sub>lj</sub>	PF <sub>j</sub>	SP <sub>lj/j</sub>	limiar
	Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2				
Energia	2	27	7	7	2	3	1	61,22	11,77	0,7411	0,41
Sequencial	0	4	0	7	3	3	0	67,30	19,81	0,7360	0,03
Camada	1	18	5	3	3	3	0	68,82	21,06	0,7380	0,29

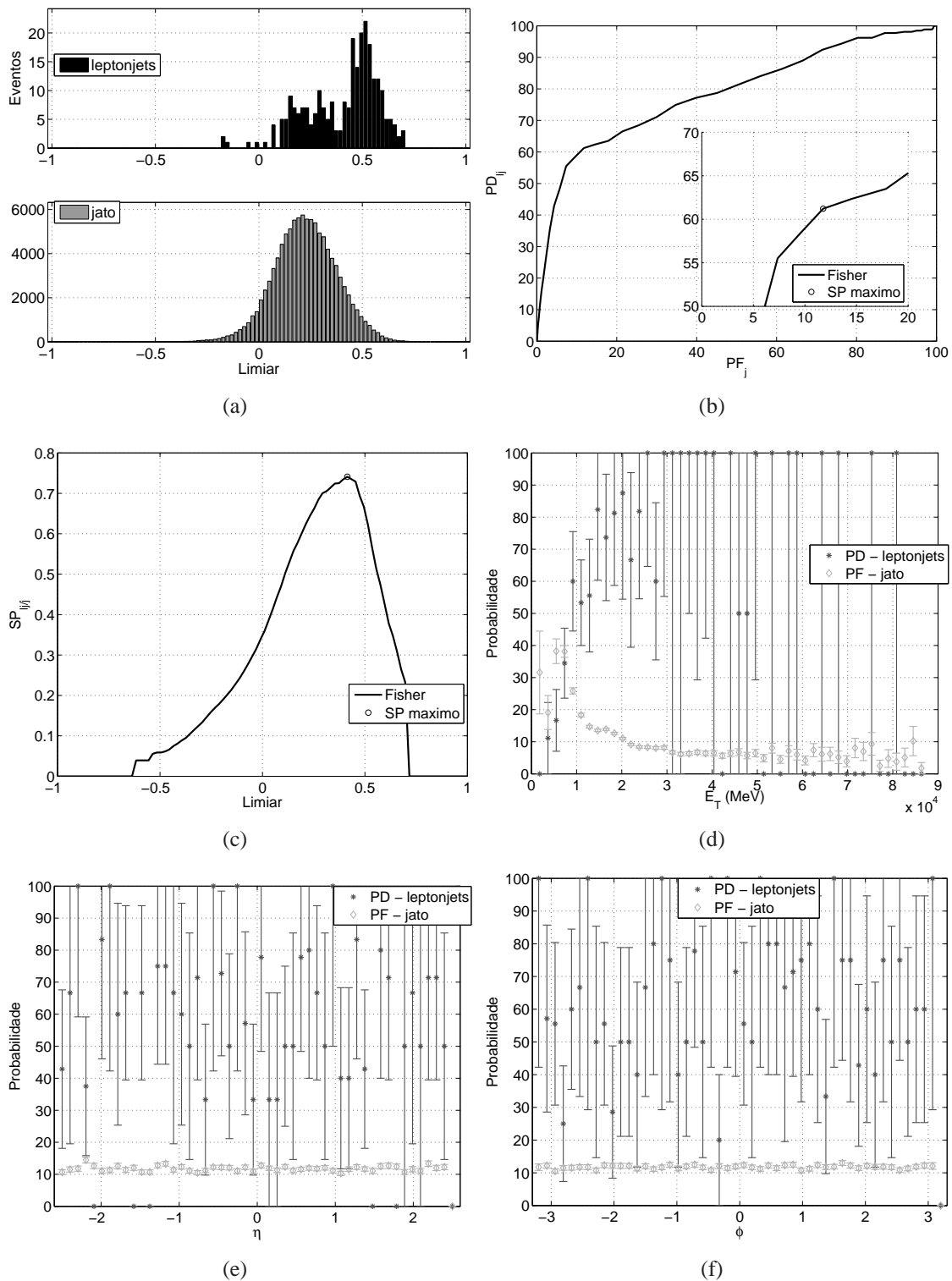


Figura 6.23: Desempenho do classificador linear de Fisher operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{l/j}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

Tabela 6.12: Desempenho do classificador neural operando sobre componentes principais extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Camada	Neurônios	Energia(%) <sup>6</sup>	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ij/j</sub>	limiar
Energia	<i>Pre Sampler</i>	4	94,77	74,14	46,35	0,6348	0,09
	<i>EM front layer</i>	4	87,57	61,60	13,90	0,7333	0,27
	<i>EM middle layer</i>	2	98,73	61,60	30,61	0,6543	0,23
	<i>EM back layer</i>	3	83,34	73,76	33,55	0,7006	0,21
	<i>HAD layer 0</i>	2	67,97	75,67	31,97	0,7180	0,25
	<i>HAD layer 1</i>	2	82,33	68,82	37,50	0,6562	0,27
	<i>HAD layer 2</i>	2	84,21	64,26	45,10	0,5949	0,19
Sequencial	<i>Pre Sampler</i>	3	97,28	68,44	42,35	0,6293	0,17
	<i>EM front layer</i>	5	54,60	68,82	22,96	0,7287	0,09
	<i>EM middle layer</i>	3	93,28	54,75	27,89	0,6313	0,23
	<i>EM back layer</i>	3	98,28	68,82	36,20	0,6629	0,21
	<i>HAD layer 0</i>	2	93,41	70,34	35,12	0,6758	0,09
	<i>HAD layer 1</i>	2	93,10	62,74	36,29	0,6322	0,11
	<i>HAD layer 2</i>	2	89,43	58,17	41,72	0,5823	0,17
Camada	<i>Pre Sampler</i>	5	79,97	71,48	39,40	0,6593	0,17
	<i>EM front layer</i>	2	84,66	69,20	19,87	0,7457	0,17
	<i>EM middle layer</i>	4	99,92	66,16	34,33	0,6591	0,19
	<i>EM back layer</i>	5	67,56	66,54	44,02	0,6114	0,13
	<i>HAD layer 0</i>	2	90,00	57,79	25,13	0,6606	0,19
	<i>HAD layer 1</i>	2	91,10	65,40	39,12	0,6312	0,11
	<i>HAD layer 2</i>	2	88,85	52,47	36,53	0,5784	0,21

Os resultados acompanharam as características positivas apresentadas pelo classificador linear e agregaram certas peculiaridades inerentes à não linearidade inserida. As componentes que produziram o melhor resultado, foram apresentadas a um único classificador neural. A variação na topologia deste classificador neural foi feita em relação ao número de neurônios na camada escondida da rede, de 2 a 5 neurônios. Os resultados são mostrados na Tabela 6.13.

A evolução dos resultados localmente também foi alcançada com a agregação da informação de diferentes camadas, independentemente da normalização aplicada. O melhor classificador, obtido com dados submetidos a normalização por energia, opera com dados de todas as camadas dos calorímetros porém com um número de componentes menor do que o requerido pelo classificador linear. Credita-se esses ganhos às características não-lineares trazidas pela rede neural usada. A topologia final do classificador 32 – 5 – 1 (5 neurônios na camada escondida) é suficiente para realizar a tarefa desejada com uma baixa complexidade computacional.

A Figura 6.24 mostra o comportamento do classificador para as diferentes figuras de mérito avaliadas. Os histogramas de saída (Figura 6.24(a)) mostram a separação alcançada entre *leptonjets* e jatos hadrônicos. Neles, fica visível o esforço da rede neural para isolar esses padrões. Ainda que mais espalhados na régua de decisão, os valores de saída indicam que para o patamar ótimo de decisão o desempenho do classificador é, apenas, levemente superior ao seu concorrente linear. Esse desempenho ainda não consegue di-

Tabela 6.13: Desempenho do classificador neural operando sobre as componentes principais extraída de forma segmentada mostradas na Tabela 6.12.

Normalização	Neurônios	Componentes							PD <sub>lj</sub>	PF <sub>j</sub>	SP <sub>lj/j</sub>	limiar
		Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2				
Energia	5	5	12	7	3	1	2	2	63,88	11,50	0,7569	0,29
Sequencial	2	7	15	5	7	3	3	0	69,20	17,63	0,7564	0,03
Camada	5	5	15	7	0	3	0	0	72,62	22,23	0,7518	0,07

minuir um número considerável de eventos perdidos, i.e., rotulados como ruído de fundo. Isso se refletiu na grande quantidade de *leptonjets* associados a uma valor de saída, situado sobre a régua de decisão, praticamente sobre a moda dos valores de saída dos jatos hadrônicos. Todas essas características se refletiram na curva ROC (Figura 6.24(b)). O regime de operação (intervalo seguro do limiar de decisão) mais abrangente, como mostra o produto  $SP_{l_j/j}$  (Figura 6.24(c)) está diretamente ligado ao regime de treinamento do classificador neural usado.

Considerando a eficiência em energia (Figura 6.24(d)), observa-se que, a abordagem apresenta alguns pontos positivos: a curva de detecção de *leptonjets* cresce com o aumento da energia do evento, enquanto o falso alarme apresenta uma lógica inversa. Essa característica permite prever e calibrar o comportamento do classificador em função da faixa de energia dos eventos analisados. Para energias muito baixas ( $E_T < 10$  GeV), a eficiência na detecção de *leptonjets* é baixa e o falso alarme é considerado alto. Na faixa de energia seguinte ( $10 \text{ GeV} < E_T < 20 \text{ GeV}$ ), as curvas de eficiência mostram, de maneira acentuada, valores mais próximos dos desejados para a operação de filtragem de nível 2. A partir desse limiar ( $E_T > 20 \text{ GeV}$ ), o comportamento do classificador é melhor e está em linha com o observado para a maioria das abordagens bem sucedidas testadas. Esses resultados são respaldados também pela eficiência obtida em função das coordenadas  $\eta$  e  $\phi$  dos eventos apresentando um comportamento livre de tendências.

A implementação resultou em uma significativa redução de dimensionalidade dos dados a serem classificados, uma separação entre os padrões que resolve o problema tratado, porém não é o melhor classificador obtido até aqui.

## 6.4.2 ICA

Seguindo a metodologia usada com uso da PCA, esta seção apresenta os resultados pertinentes ao uso da ICA como extrator de características. Mais uma vez, os percentuais de energia retida a que se referem os resultados, relacionam-se com a energia retida na etapa de branqueamento dos dados de cada camada. Como já dito, essa etapa é inerente à técnica de análise de componentes independentes escolhida.

De maneira análoga ao procedido com a PCA, um classificador linear de Fisher foi usado para determinar o piso de eficiência fornecido pelas componentes independentes extraídas de cada uma das camadas dos calorímetros eletromagnético e hadrônico. A Tabela 6.14, guiada pelo produto  $SP_{l_j/j}$ , mostra o desempenho obtido.

Os valores não apresentam grandes surpresas, seguindo, em boa parte, o observado com a PCA. A maior variação foi verificada com dados normalizados por energia. Por exemplo, observa-se que com o uso dessa normalização, a primeira camada do calorí-

---

<sup>7</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.

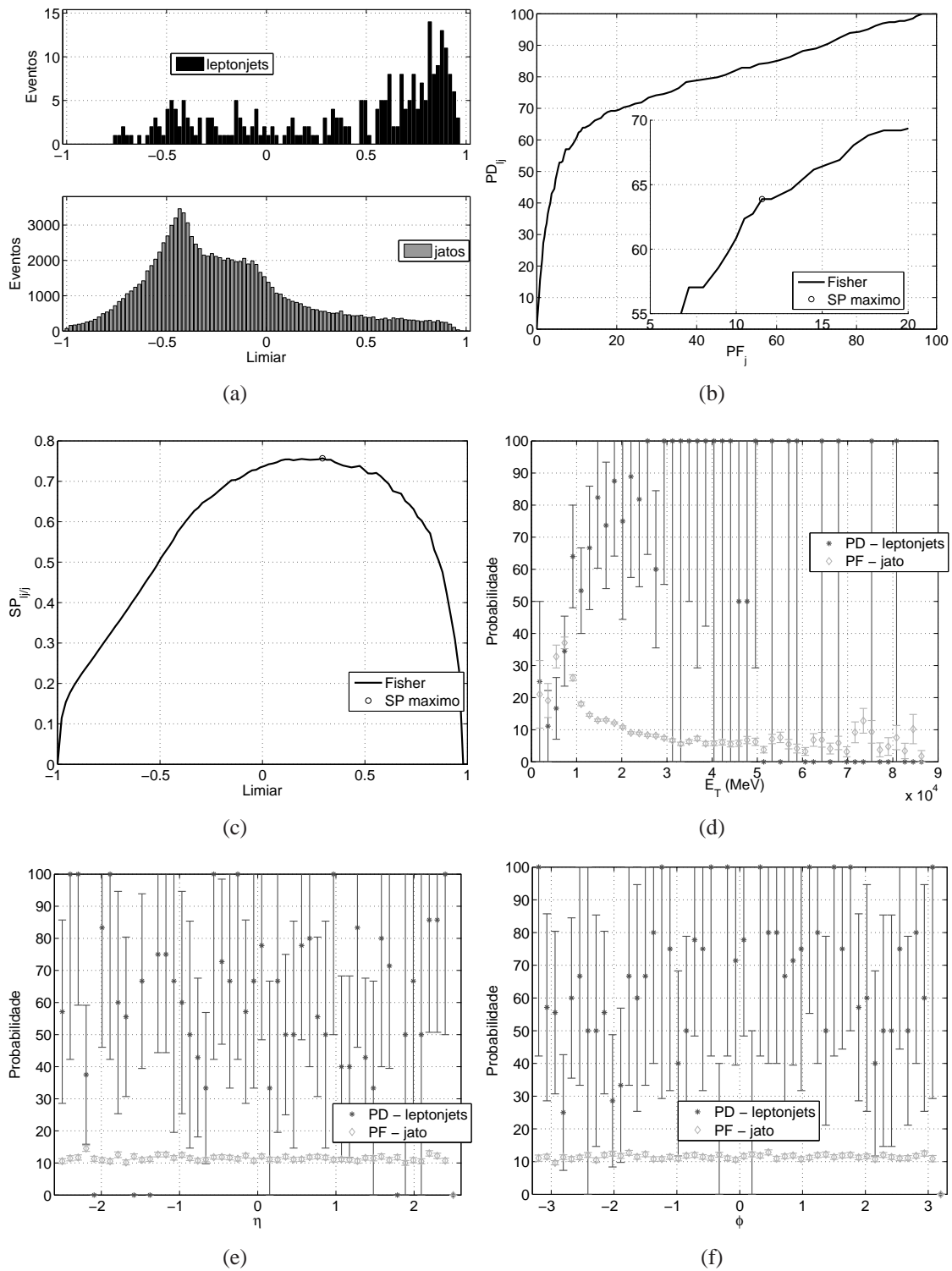


Figura 6.24: Desempenho do classificador neural operando sobre componentes principais extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{l_j/j}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.



Tabela 6.14: Melhor desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Camada	Energia(%) <sup>7</sup>	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ijj</sub>	limiar
Energia	<i>Pre Sampler</i>	90,97	50,19	25,02	0,6196	0,23
	<i>EM front layer</i>	96,35	64,26	18,92	0,7242	0,23
	<i>EM middle layer</i>	98,73	58,17	31,69	0,6314	0,25
	<i>EM back layer</i>	93,29	72,62	33,90	0,6932	0,80
	<i>HAD layer 0</i>	97,40	66,54	27,36	0,6955	0,60
	<i>HAD layer 1</i>	97,39	68,82	38,84	0,6494	0,39
	<i>HAD layer 2</i>	63,40	61,60	42,89	0,5933	0,31
Sequencial	<i>Pre Sampler</i>	97,28	57,79	33,77	0,6194	-0,15
	<i>EM front layer</i>	24,97	66,16	22,97	0,7149	0,05
	<i>EM middle layer</i>	87,63	60,08	34,32	0,6285	-0,76
	<i>EM back layer</i>	98,28	69,58	40,90	0,6423	-0,25
	<i>HAD layer 0</i>	93,41	59,32	24,52	0,6715	-0,05
	<i>HAD layer 1</i>	93,10	61,98	34,92	0,6352	-0,19
	<i>HAD layer 2</i>	89,43	51,33	36,82	0,5710	-0,37
Camada	<i>Pre Sampler</i>	79,97	47,15	18,02	0,6336	0,64
	<i>EM front layer</i>	87,38	69,96	23,41	0,7324	0,39
	<i>EM middle layer</i>	93,97	60,08	33,79	0,6310	0,41
	<i>EM back layer</i>	95,78	68,32	53,18	0,5706	-0,54
	<i>HAD layer 0</i>	90,00	64,26	35,09	0,6459	-0,47
	<i>HAD layer 1</i>	91,10	57,41	30,63	0,6325	-0,29
	<i>HAD layer 2</i>	72,70	60,46	44,74	0,5783	-0,25

metro apresenta o melhor desempenho, enquanto que com a PCA esse título tinha sido dado a primeira camada hadrônica. Uma característica geral observada é que a variação do desempenho, quando obtida, se deu majoritariamente para um número maior, i.e., uma variação positiva. Ao realizar a busca pela independência das componentes, a técnica ICA produz resultados levemente superiores aos obtido com a PCA. Outro ponto observado foi a convergência das camadas com melhor poder de diferenciação de eventos em cada uma das normalizações aplicadas. Em todas as normalizações a informação mais importante está na camada *EM front layer*, seguida pela informação que provém da camada *HAD layer 0*. De maneira geral, as primeiras camadas de cada um dos calorímetros (cada uma com suas especificidades) são importante para o processo decisório.

Aproveitando os resultados mostrados na Tabela 6.14, implementou-se um classificador linear cujas entradas são as componentes nela descritas. O intuito foi verificar se essas componentes, que alguns casos apresentaram a capacidade de separar os padrões em estudo, conjuntamente poderiam produzir resultados ainda melhores. Novamente, parte-se das componentes pertencentes a camada com melhor desempenho e gradativamente se agrega as componentes das camadas com eficiência subsequente. A Tabela 6.15 mostra o resultado do melhor desempenho para cada uma das normalizações.

Em todas as abordagens, a aglomeração da informação gerou um classificador com eficiência melhor do que a obtida por cada uma das camadas isoladamente. O melhor classificador foi obtido com a informação das componentes independentes extraídas de

Tabela 6.15: Desempenho do classificador linear de Fisher operando sobre as componentes independentes extraídas de forma segmentada apresentadas na Tabela 6.14.

Normalização	Componentes							PD <sub>lj</sub>	PF <sub>j</sub>	SP <sub>lj/j</sub>	limiar
	Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2				
Energia	0	27	7	5	3	3	0	65,02	15,90	0,7425	0,21
Sequencial	0	4	0	7	3	3	0	67,30	19,81	0,7360	0,03
Camada	5	18	5	7	3	3	2	69,58	22,34	0,7357	0,27

dados normalizados por energia de todas as camadas com exceção da *Pre Sampler* e *HAD layer 2*. As componentes independentes extraídas dessas duas camadas não contribuíam para o processo decisório. A configuração final obtida aponta para classificador simples do ponto de vista computacional. As 45 variáveis de entrada representam um espaço de dimensão bem inferior ao original e inferior inclusive ao obtido com a técnica PCA (seção anterior). A Figura 6.25 ilustra o comportamento do classificador obtido. A saída do classificador para os dados de teste é mostrada na Figura 6.25(a). Os padrões de *leptonjets* e jatos hadrônicos ficam muito próximos na régua de decisão, o que resulta em um classificador com margem de decisão (limiar de funcionamento) muito estreita, como observado na maioria das respostas das abordagens implementadas com classificadores lineares. A Figura 6.25(b) mostra a curva ROC do classificador e a Figura 6.25(c), o comportamento do produto  $SP_{l_j/j}$  em função do limiar de decisão. As Figuras 6.25(d), 6.25(e) e 6.25(f) exibem, respectivamente, a eficiência do classificador em função da  $E_T$ , das coordenadas  $\eta$  e  $\phi$  do evento. O classificador obtido também é considerado uma solução viável para o problema tratado.

Para o classificador neural, a Tabela 6.16 resume os melhores resultados por camada. O estudo do classificador neural envolveu a variação do número de neurônios na camada escondida da rede. Um mínimo de 2 e um máximo de 5 neurônios na camada escondida da rede foi estipulado para os classificadores, com exceção das camadas hadrônicas, em que a dimensão é menor. Nessas camadas a rede neural foi fixada com 2 neurônios escondidos.

Tabela 6.16: Desempenho do classificador neural operando sobre componentes independentes extraídas de forma segmentada de dados formatados em anéis e normalizados de diferentes maneiras.

Normalização	Camada	Neurônios	Energia(%) <sup>8</sup>	PD <sub>l<sub>j</sub></sub>	PF <sub>j</sub>	SP <sub>l<sub>j</sub>/j</sub>	limiar
Energia	<i>Pre Sampler</i>	5	98,98	51,71	25,27	0,6269	0,17
	<i>EM front layer</i>	2	94,25	62,36	14,48	0,7348	0,27
	<i>EM middle layer</i>	2	98,73	67,30	38,35	0,6444	0,21
	<i>EM back layer</i>	5	83,34	74,52	34,22	0,7009	0,25
	<i>HAD layer 0</i>	2	97,40	69,20	26,09	0,7153	0,29
	<i>HAD layer 1</i>	2	82,33	70,34	39,11	0,6553	0,23
	<i>HAD layer 2</i>	2	84,21	66,92	47,55	0,5946	0,21
	Sequencial	<i>Pre Sampler</i>	4	97,28	56,65	30,66	0,6284
<i>EM front layer</i>		3	20,61	69,20	24,31	0,7241	0,11
<i>EM middle layer</i>		3	87,63	70,72	43,84	0,6323	0,07
<i>EM back layer</i>		5	88,80	65,78	33,24	0,6627	0,19
<i>HAD layer 0</i>		2	93,41	59,70	24,58	0,6733	0,17
<i>HAD layer 1</i>		2	93,10	61,60	34,76	0,6340	0,11
<i>HAD layer 2</i>		2	89,43	55,13	38,58	0,5823	0,15
Camada		<i>Pre Sampler</i>	5	70,89	69,96	39,18	0,6531
	<i>EM front layer</i>	4	83,63	69,96	21,93	0,7396	0,13
	<i>EM middle layer</i>	4	93,97	65,78	36,98	0,6439	0,19
	<i>EM back layer</i>	4	79,51	67,68	47,97	0,5960	0,11
	<i>HAD layer 0</i>	2	74,02	57,03	25,26	0,6559	0,25
	<i>HAD layer 1</i>	2	91,10	64,64	37,84	0,6339	0,13
	<i>HAD layer 2</i>	2	88,85	64,26	47,48	0,5824	0,09

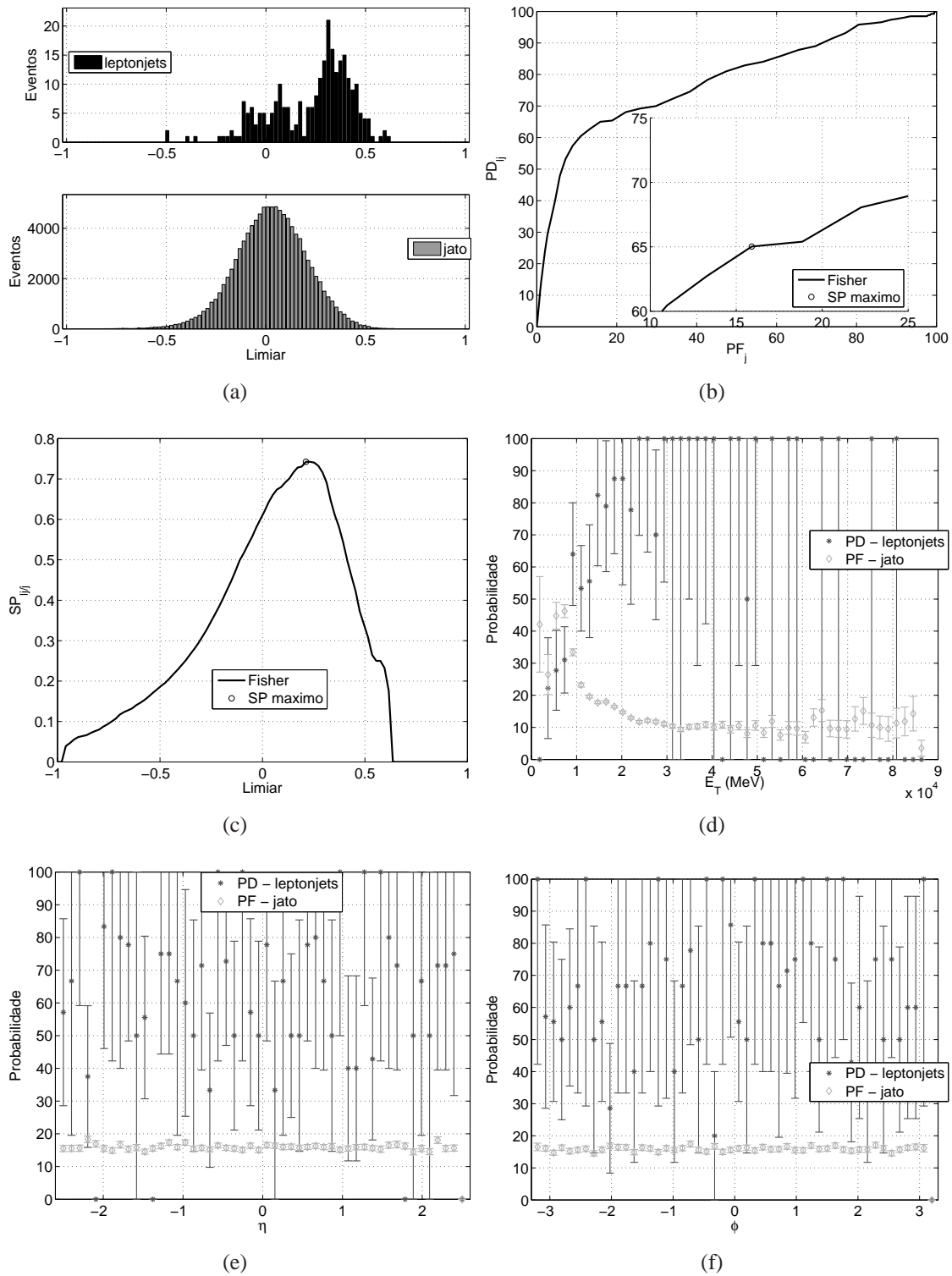


Figura 6.25: Desempenho do classificador linear de Fisher operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por energia. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{II}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

As camadas que conseguem captar as principais diferenças entre o evento desejado e o ruído de fundo são as mesmas observadas com o uso do classificador linear. Contudo, as demais camadas apresentam uma evolução em seus números finais, impulsionadas, evidentemente, pela não linearidade trazida pela rede neural empregada. Isso se aplica para todas as normalizações. As componentes que produziram o melhor resultado foram, gradualmente apresentadas a um único classificador neural, cujos resultados são mostrados na Tabela 6.17.

O classificador obtido aproxima-se bastante da melhor abordagem até aqui testada (classificador neural sobre pré-processamento não segmentado com o uso da PCA). Acredita-se que a eficiência final poderia ser aumentada, pois os testes dos classificadores neurais aglomerativos na abordagem segmentada foram realizados para um número máximo de neurônios na camada escondida da rede de 5, enquanto na abordagem não segmentada alcançou 10. Novos estudos que buscassem essa linha de investigação podem apresentar números finais ainda mais expressivos.

O melhor classificador opera com dados normalizados por camada e engloba informação de cada uma, exceto da última camada hadrônica. A topologia do classificador obtido também é uma característica positiva da metodologia aplicada. Com apenas dois neurônios na camada escondida da rede e alimentado por 33 componentes ( $33 - 2 - 1$ ), a complexidade computacional é similar a dos classificadores neurais testados até então e suficiente para realizar a tarefa desejada.

A Figura 6.26 mostra o comportamento do classificador para as diferentes figuras de mérito avaliadas. Os histogramas de saída (Figura 6.26(a)) exibem a separação obtida e deixam em evidência um conjunto de *leptonjets* que definitivamente compartilham grande parte de suas características de deposição com os jatos hadrônicos. Esses eventos podem ser entendidos como muito difíceis de serem corretamente rotulados e tratados, inclusive como um problema dentro do problema. Essas características são refletidas na curva ROC (Figura 6.26(b)) e no produto  $SP_{lj/j}$  (Figura 6.26(c)). Já a eficiência em função da  $E_T$  e das coordenadas  $\eta$  e  $\phi$  dos eventos são mostradas nas Figuras 6.26(d), 6.26(e) e 6.26(f), respectivamente.

A implementação realizada implica uma expressiva redução de dimensionalidade dos dados, uma separação de padrões que acompanha de perto a maioria das abordagens testadas, além de uma topologia final de decisão considerada computacionalmente apta a operar no nível 2 de filtragem do detector.

---

<sup>8</sup>O percentual de energia se refere ao acumulado nas componentes principais, pré-processadas pelo algoritmo ICA.

Tabela 6.17: Desempenho do classificador neural operando sobre as componentes independentes extraída de forma segmentada apresentadas na Tabela 6.16.

Normalização	Neurônios	Componentes							PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ij/j</sub>	limiar
		Pre S.	EM f.	EM m.	EM b.	HAD 0	HAD 1	HAD 2				
Energia	3	0	21	0	3	3	2	0	65,78	13,00	0,7602	0,23
Sequencial	5	0	3	0	0	3	0	0	67,68	18,40	0,7448	0,09
Camada	2	4	14	5	5	2	3	0	69,96	17,31	0,7619	0,01

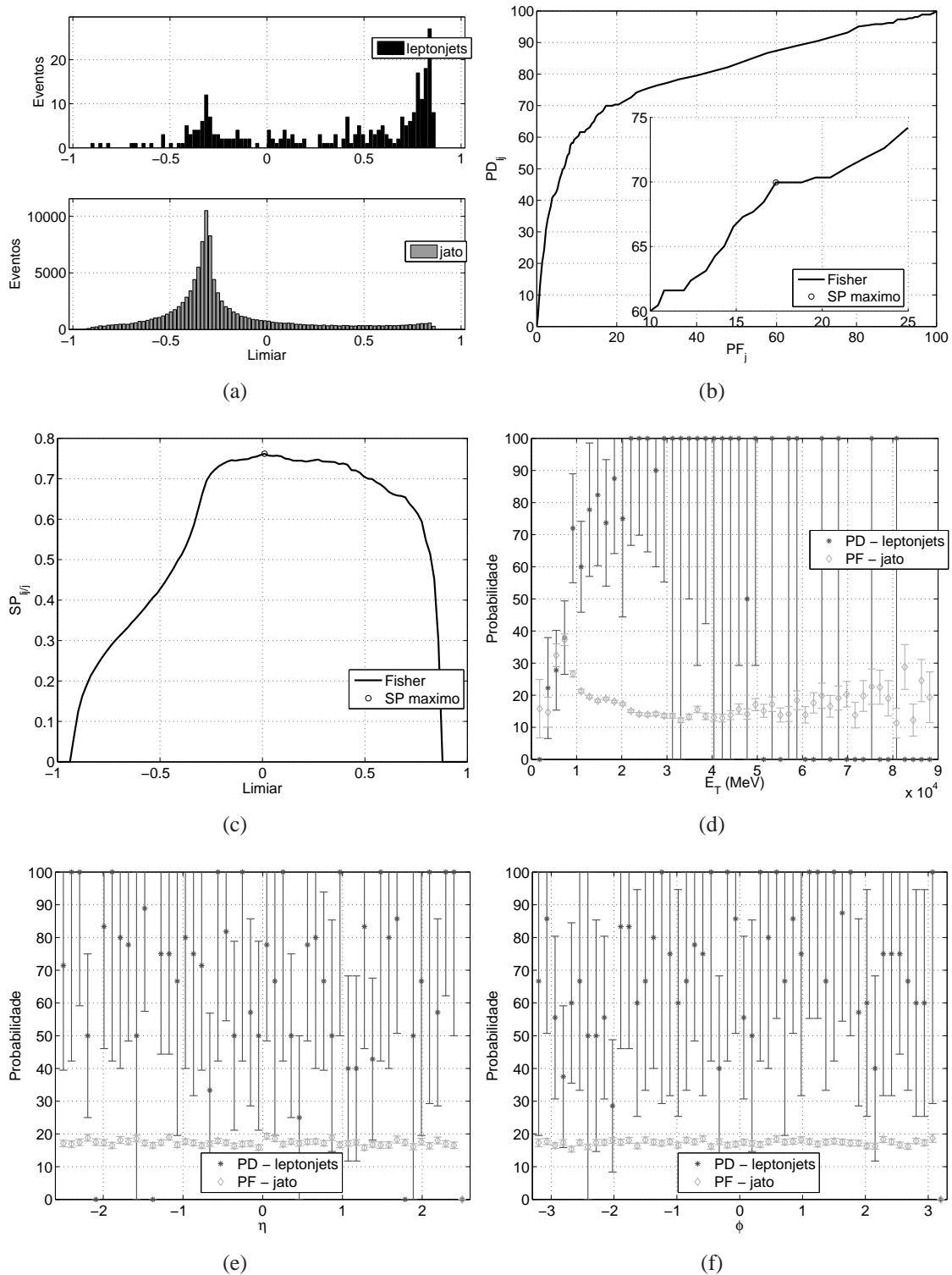


Figura 6.26: Desempenho do classificador neural operando sobre componentes independentes extraídas de maneira segmentada de dados formatados em anéis e normalizados por camada. (a) Histogramas de saída. (b) Curva ROC. (c) Produto  $SP_{ljj}$ . (d) Probabilidades de detecção e falso alarme em função da  $E_T$ . (e) Probabilidades de detecção e falso alarme em função da coordenada  $\eta$  do detector. (f) Probabilidades de detecção e falso alarme em função da coordenada  $\phi$  do detector.

## 6.5 Comparação

Os testes implementados neste capítulo visaram apontar um caminho a ser seguido na construção de um futuro canal de detecção e seleção de eventos *leptonjets*/jatos. Os resultados com diferentes abordagens identificaram interessantes caminhos a se seguir na pesquisa, cada um com potencialidades e limitações. Esses pontos positivos e negativos foram observados e destacados na eficiência alcançada (detecção de eventos de interesse do canal), na facilidade de interpretação dos resultados, na taxa de compactação obtida e na complexidade computacional requerida. A Tabela 6.18 resume o desempenho de cada uma delas, enquanto as Figuras 6.27 e 6.28 apresentam a curva ROC e o produto  $SP_{lj}$  dos respectivos classificadores. A partir das curvas algumas comparações podem ser feitas.

De maneira geral, todas as abordagens apresentaram desempenho próximos. Observa-se que os classificadores que operam alimentados apenas com a informação dos anéis, tanto o linear quanto o neural, servem como referências para as demais abordagens ao indicar que conseguem realizar a tarefa de detectar *leptonjets* e rejeitar jatos hadrônicos lidando com toda a informação de calorimetria disponível. Apesar de operarem em um ambiente de alta dimensionalidade (100 anéis originais), as medidas qualitativas usadas forneceram uma estimativa pontual, detalhada (não apenas camadas, mas anéis), da informação que realmente é importante para o processo decisório. Esse dado é pertinente pois, além de servir de base para o confronto entre o especialista na área e a interpretação que o algoritmo faz dos dados fornecidos, também ajuda nas abordagens seguintes, quando as técnicas de compactação e extração de características usam a informação de certos anéis para a construção de novas variáveis de entrada. Aqui, o algoritmo linear identificou a informação de alguns anéis das camadas eletromagnéticas como relevantes. Já o classificador neural conseguiu explorar, além delas, informações de alguns anéis das camadas *Pre Sampler* e *HAD layer 0*. Por exemplo, a identificação dos anéis da camada EM *middle layer*, ora com mais ora com menos peso na construção da estrutura interna de decisão, faz todo sentido pois, sendo essa a camada eletromagnética mais profunda, ela armazena a maior parte da energia eletromagnética dos eventos de interesse e do ruído de fundo, servindo perfeitamente para diferenciá-los. O classificador linear é considerado simples computacionalmente e o algoritmo neural escolhido (com apenas 7 neurônios na camada escondida) não representam um problema para as limitação do nível 2 de filtragem. Essas características credenciam qualquer um dos classificadores a operarem em um futuro canal do detector ATLAS.

A seleção de regiões concêntricas ( $\Delta R$ ) ao ponto de penetração da partícula também apresentou aspectos positivos. É possível afirmar que melhores resultados podem ser produzidos com um estudo mais detalhado da normalização a ser aplicada. Em todo caso, com as técnicas propostas foi possível reduzir 75% da dimensão de entrada, implementar classificadores simples (tanto o linear quanto neural) e produzir resultados que se aproxi-



Tabela 6.18: Comparação dos resultados das diferentes abordagens propostas.

Técnica	Anéis (Comp.)	Energia(%)	Neurônios	PD <sub>ij</sub>	PF <sub>j</sub>	SP <sub>ij/j</sub>	limiar
Fisher - Anéis	100			63,88	16,76	0,7324	0,52
Neural - Anéis	100		7	68,44	17,28	0,7541	0,07
Fisher - $\Delta R$ Anéis	25			58,56	12,22	0,7243	0,29
Neural - $\Delta R$ Anéis	25		5	63,50	12,47	0,7503	0,31
Fisher - PCA (Não Seg.)	(12)	79,33		68,82	22,96	0,7287	-0,11
Neural - PCA (Não Seg.)	(9)	72,58	9	70,72	18,09	0,7621	0,11
Fisher - ICA (Não Seg.)	(12)	59,30		68,82	79,33	0,7287	-0,11
Neural - ICA (Não Seg.)	(4)	30,39	9	69,96	53,54	0,7526	0,13
Fisher - PCA (Seg.)	(49)			61,22	11,77	0,7411	0,41
Neural - PCA (Seg.)	(32)		5	63,88	11,50	0,7569	0,29
Fisher - ICA (Seg.)	(45)			65,02	15,90	0,7425	0,21
Neural - ICA (Seg.)	(33)		2	69,96	17,31	0,7619	0,01

maram bastante do obtido com os concorrentes operando sobre a dimensão completa da RoI e até mesmo das melhores abordagens que envolveram compactação. Foi possível realizar uma redução significativa da dimensão dos dados baseada somente em conhecimento do especialista.

Nas abordagens realizadas com o auxílio das diferentes técnicas de análise de componentes, verificou-se que a extração de característica empregada pelos algoritmos PCA e ICA são eficientes no fornecimento da informação relevante para o algoritmo de decisão. A transformação dos dados de alta dimensionalidade dos sinais dos calorímetros do ATLAS em variáveis com poder de descrição do processo analisado, foi realizada tanto na abordagem não segmentada quanto na segmentada. Ambas com sucesso.

Na abordagem não segmentada, foi possível atingir um nível de compactação bastante elevado. Um outro benefício produzido, foi a diferenciação das características de *leptonjets* e jatos hadrônicos nas variáveis resultantes. Isso fica explícito pelos significativos valores finais obtidos, inclusive com o melhor dos resultados sendo produzido nessa metodologia. O que se nota, é que, independentemente da técnica de análise de componentes empregada, o custo final nos resultados em relação ao ganho obtido em eficiência são justificáveis e administráveis. Além disso, pela análise de relevância e correlação entre as variáveis de máxima relevância e os anéis originais (medidas qualitativas do estágio de pré-processamento), ficou explícita a contribuição de cada região dos calorímetros na formação desses descritores. Com a PCA, a informação presente nos primeiros anéis das camadas eletromagnéticas, além da *Pre Sampler* e *HAD layer 0*, estava bem representada. No uso da ICA, isso se repetiu, com um aumento de representação dos anéis secundários da camada *EM front layer*. Essa conclusão está em linha com o que fora verificado na abordagem com anéis. Por fim, as estruturas de decisão criadas são simples e também se acomodam perfeitamente no ambiente de filtragem de nível 2 de um possível canal a ser criado.

No caso da abordagem segmentada, os valores de desempenho são tão expressivos quanto os obtidos de maneira não segmentada. A aplicação local das técnicas de análise de componentes (extração de características em cada camada dos calorímetros) permitiu potencializar a transformação da informação disponível. A alimentação do classificador pela informação realmente discriminante de cada camada ou de apenas algumas camadas, eliminou outras variáveis consideradas nos testes anteriores que complicavam (mais do que auxiliavam) a tarefa de identificação do evento de interesse. Aqui, também as taxas de compactação obtidas foram consideradas altas. O principal diferencial em relação a abordagem não segmentada é uma maior robustez a falhas, dada a característica local empregada na extração de características. Essa abordagem produziu o segundo melhor resultado dentre todas testadas, mas que poderia substituir a primeira sem perdas significativas de desempenho, mostrando uma convergência nos resultados das melhores técnicas avaliadas. Nesse contexto segmentado, a contribuição de cada camada no processo decisório

fica explicitada não só graficamente como quantificada. Novamente as camadas EM *front layer* e HAD *layer 0* impactam fortemente no desempenho dos classificadores. O que já tinha sido observado nas abordagens anteriores encontra eco aqui. De tudo isso, pode-se concluir que a diferenciação de *leptonjets* e jatos hadrônicos passa, de alguma maneira, pela informação dessas camadas.

Falando especificamente do melhor resultado (abordagem não segmentada com o uso da PCA e do classificador neural) investigou-se o tempo requerido para a sua execução no ambiente de filtragem do detector ATLAS, o Athena. Na cadeia de processamento do RingerNeuralHypo - um dos algoritmos de referência do canal elétron/jato - o tempo de execução, calculado e detalhado na Seção 3.4, indicou a sua viabilidade técnica. Comparando as duas abordagens, verifica-se que a única diferença entre elas é o bloco de pré-processamento PCA aqui inserido. O tempo calculado para a sua execução foi calculado em  $0,0314 \pm 0,0021$  ms. Assim, ao tempo de execução do RingerNeuralHypo (tabela 3.2), que é de aproximadamente  $1,2352 \pm 0,1288$ , acrescenta-se o tempo de execução do bloco PCA. Chega-se, então, a um valor final que ainda está dentro do permitido para um algoritmo de nível 2 de filtragem do detector ATLAS.

A investigação realizada e tudo que nela está envolvido (metodologia, formatação dos dados, normalização dos dados, tipos de classificadores, medidas de desempenho etc), compreende um pequeno tutorial sobre o problema tratado. A partir dela, é possível traçar um caminho a seguir na construção de algoritmos de filtragem para o problema de separação de *leptonjets* e jatos hadrônicos no contexto da busca de um candidato a matéria escura.

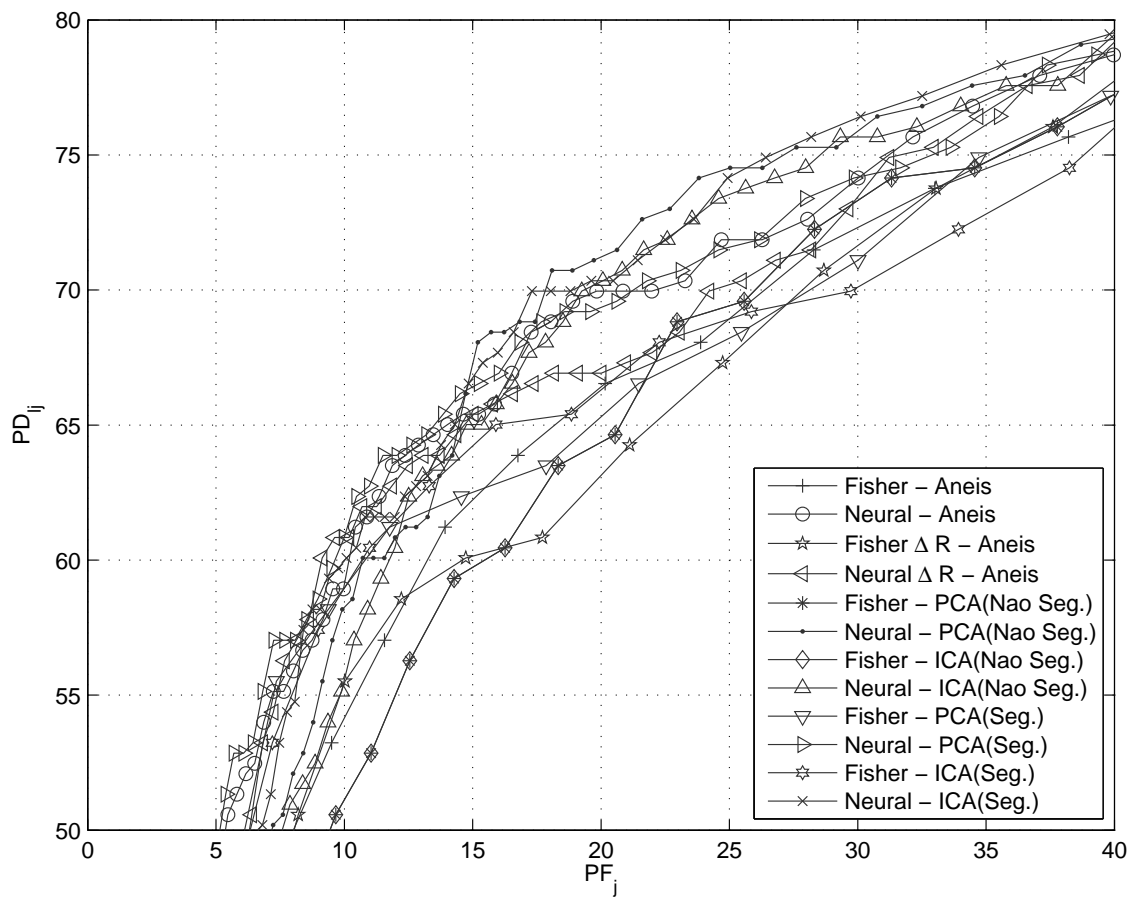


Figura 6.27: Curvas ROC para o canal  $lj/j$  obtidas com cada uma das abordagens propostas.

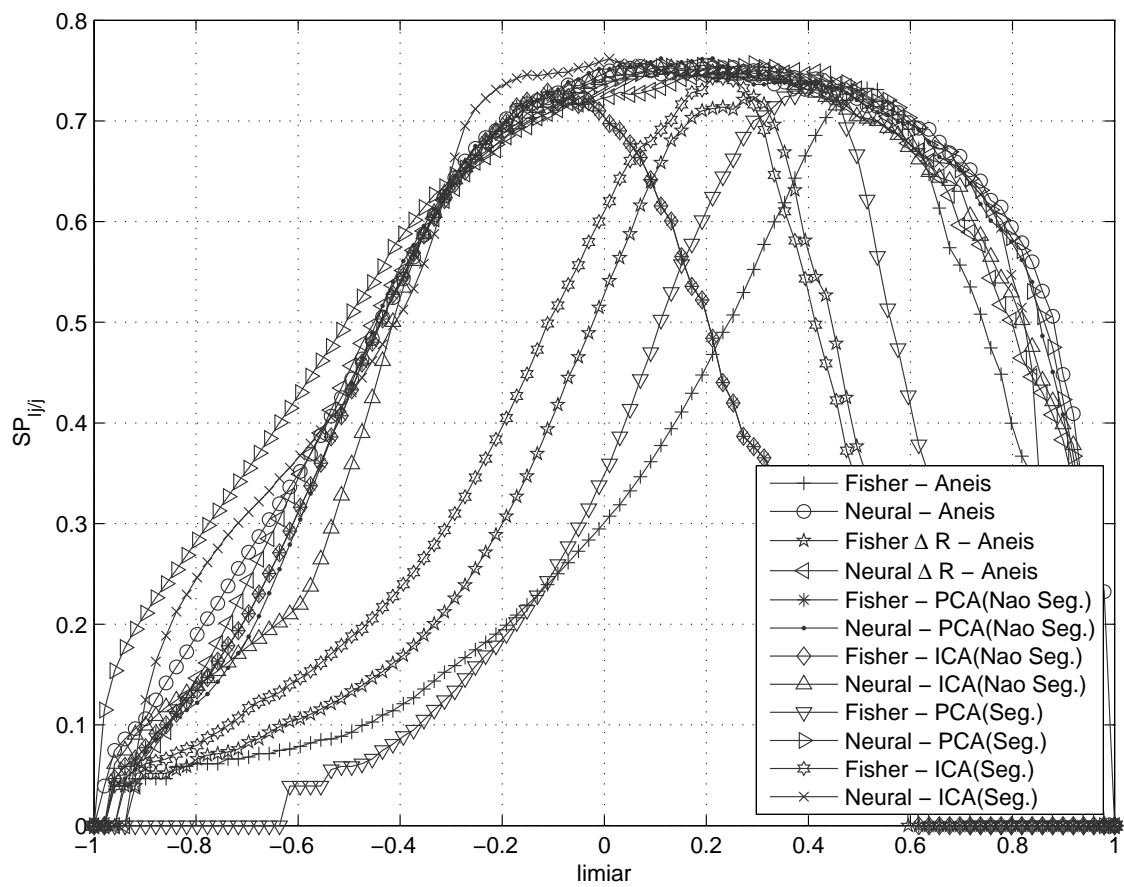


Figura 6.28: Produto  $SP_{l_j/j}$  obtido com cada uma das abordagens propostas.

# Capítulo 7

## Conclusões e Trabalhos Futuros

O grande colisionador hadrônico LHC, no CERN, iniciou uma nova era de estudos e investigações na área de física de partículas e altas energias. Com o objetivo de entender as características fundamentais da matéria, o aparato tecnológico experimental estabeleceu alguns recordes: energia e luminosidade alcançadas; colaboradores científicos envolvidos; produtos difundidos na sociedade e valores monetários despendidos. Para isso, ao longo dos anos, se envolveram profissionais de diferentes áreas do conhecimento, como: engenharia, computação, matemática e física.

Para lidar com a enorme massa de dados gerada em cada uma das colisões que ocorreram e ocorrerão, robustos sistemas de filtragem *online* foram criados para satisfazer cada uma das exigências requeridas: a alta eficiência na detecção dos eventos de interesse; uma baixa aceitação do ruído de fundo; robustez a falhas e operação dentro do tempo limite.

O adequado funcionamento do sistema de filtragem é fundamental para o desempenho do detector; porém, atender a todas as demandas inerentes à aplicação não é uma tarefa trivial. Em geral, os sistemas de filtragem são hierarquizados e a informação não relevante é gradualmente reduzida.

O trabalho apresentado investigou o problema de identificação e filtragem de eventos nesse ambiente de física de partículas. Um conjunto de técnicas de análise de componentes e máquinas de aprendizado atuando como classificadores de padrão foram usados nos diferentes procedimentos propostos. Esses procedimentos, quando possível, concorrem com os atuais algoritmos de referência do canal estudado, como, por exemplo, no caso do canal elétron/jato, em que os algoritmos de referência se chamam T2Ca10 e RingerNeuralHypo.

Uma introdução ao ambiente de física de partículas foi apresentada, assim como os respectivos eventos de interesse em cada canal estudado. Visando a otimização do desempenho de discriminação dos eventos no canal elétron/jato e *leptonjets*/jato, três abordagens foram seguidas: o processamento dos dados de calorimetria formatados em anéis normalizados, o pré-processamento segmentado desses dados e o pré-processamento não segmentado dos dados. Os objetivos em questão eram a maximização de elétrons e *leptonjets*

detectados com a minimização do falso alarme para o canal elétron/jato, enquanto para o canal *leptonjets*/jato a maximização da detecção dos *leptonjets* se contrapõe a minimização dos falsos positivos, i.e., jatos hadrônicos identificados como *leptonjets*.

Na abordagem em anéis, esses foram apresentados aos algoritmos de teste de hipótese (classificadores lineares e não-lineares) responsáveis pela correta rotulação dos eventos sob as condições de filtragem pertinentes ao ambiente analisado. Nas demais abordagens, houve a inserção de uma etapa de extração de características. No procedimento segmentado, a aplicação de técnicas de análise de componentes foi feita localmente, em cada camada dos calorímetros, e só então os dados foram apresentados para o algoritmo de classificação. No procedimento não segmentado, a aplicação das técnicas de análise de componentes foi feita sobre os dados de calorimetria como um todo, sem a separação de camadas.

Testes com diferentes paradigmas e algoritmos de processamento de sinais foram realizados e seus resultados apresentados. A classificação alimentada com dados formatados em anéis produziu os melhores valores de detecção de elétrons (98% para os classificadores linear e não-linear) e falso alarme de jatos (3% para ambos classificadores) no canal elétron/jato, porém com uma baixa eficiência na detecção de *leptonjets*, 14%. A aplicação segmentada é interessante no que diz respeito a interpretação física do processo decisório do classificador, como mostrou o potencial local para cada uma das componentes retidas por camada do calorímetro. Além disso, o procedimento segmentado é mais robusto a falhas. É fato que ao longo do funcionamento do detector ATLAS alguns de seus sensores serão prejudicados pelo estresse ao qual o equipamento é submetido, assim, a informação perdida tem um efeito negativo local, o que, dependendo da configuração final alcançada, pode representar um efeito limitado, uma virtude (ou dependendo da região afetada um ponto negativo). O uso combinado da PCA na extração de características e do classificador neural resultou no melhor desempenho geral, resumido num produto  $SP_{e/j/l_j}$  de 0,7966 ( $PD_e = 94\%$ ,  $PF_j = 15\%$  e  $PD_{l_j} = 61\%$ ). Já a aplicação não segmentada é interessante no que diz respeito a eliminação de redundância entre camadas, o que proporciona uma maior taxa de compactação dos dados, como mostraram os resultados (redução até 95% da dimensão dos dados com  $PD_e$  em torno de 91%,  $PF_j$  em torno de 14% e  $PD_{l_j}$  em torno de 50%).

Como os sinais formatados em anéis formam um espaço de dimensão elevada, além de possuírem algum grau de redundância na informação, fez-se uso de técnicas de processamento estatístico de sinais na busca de um espaço de dimensão reduzida cuja informação fosse tão representativa quanto a original.

Os testes e resultados em função da energia retida no processo, assim como as medidas de desempenho usadas, exprimiram detalhes do complexo processo de classificação tratado. Os resultados (inéditos para dados exóticos, *leptonjets*), evidenciaram um caminho promissor a ser seguido no pré-processamento dos dados.

Resultados com dados de simulação de Monte Carlo para colisões próton-próton compreendendo aproximadamente 470000 assinaturas de elétrons e 310000 assinaturas de jatos, além de aproximadamente 1000 assinaturas de *leptonjets* ( $4e$ ) foram mostrados para cada uma das abordagens testadas, assim como para as figuras de mérito existentes no caso do canal elétron/jato. Para elétrons e jatos podemos dizer que a estatística disponível era rica e suficiente para fazer afirmações mais conclusivas sobre os valores finais alcançados. Já a base de dados para *leptonjets* precisa de um incremento no número de eventos para melhorar o detalhamento já alcançado com elétrons e jatos.

Quanto aos algoritmos de classificação, foram testados algoritmos lineares e não-lineares. O algoritmo linear escolhido foi classificador linear de Fisher, com vasta literatura publicada e muito simples computacionalmente de ser implementado. Sobre o algoritmo não-linear escolhido, optou-se por redes neurais do tipo percéptron de múltiplas camadas, também com vasta literatura publicada e de complexidade computacional administrável para o nível 2 de filtragem. Os testes foram confrontados (quando possível) com os algoritmos de referência do canal de seleção, como, por exemplo, os algoritmos de referência do canal elétron/jato, T2Calo e RingerNeuralHypo. Este último já fazia uso de uma rede neural como algoritmo de teste de hipótese. Alguns dos procedimentos propostos seguiram essa linha de pesquisa, o que proporcionou uma comparação final satisfatória.

Além disso, os sistemas apresentados possibilitam resposta a evidências. No contexto de classificação de padrões, o algoritmo de decisão pode ser projetado para fornecer informação não somente sobre qual padrão particular selecionar, mas também sobre a confiança ou crença na decisão tomada. Esta última informação é de grande valor para o analista do sistema.

Baseado neste estudo, conclui-se que os procedimentos apresentados podem corretamente classificar eventos pontuais de elétrons, mesmo em condições de alto ruído de fundo (jatos), além de aumentar significativamente o percentual de aceitação de eventos exóticos (*leptonjets*) no contexto do canal elétron/jato. No que se refere ao canal *leptonjets*/jato proposto, foram realizados basicamente os mesmos testes empregados para o canal elétron/jato, com algumas novidades no que diz respeito a redução de dimensionalidade baseada em conhecimento de especialista (regiões  $\Delta R$ ). Os resultados compreendem uma outra solução para o problema de detecção de eventos exóticos. No contexto desse canal, pôde-se fornecer uma substancial separação dos eventos (considerando-se a semelhança na deposição de energia por parte dos padrões em análise), além de propor um novo instrumento de estudo dedicado a questões que não são focadas em outros canais de seleção de eventos. O melhor resultado foi obtido na abordagem não segmentada com o uso da PCA como extrator de características e da rede neural como classificador,  $SP_{ij} = 0,7621$  ( $PD_{ij} = 70\%$  e  $PF_j = 18\%$ ).

Todas essas conclusões foram enriquecidas pelas análises feitas tanto da qualidade da



resposta do classificador, quanto da informação fornecida para esse mesmo classificador. O que só fortaleceu a análise e conclusões finais.

## 7.1 Trabalhos Futuros

Como trabalhos futuros, pode-se destacar alguns pontos a serem abordados.

- A análise de relevância das componentes estudadas revelou a importância distinta das variáveis no processo decisório entre elétrons, jatos e *leptonjets*. Em alguns casos há indícios de que entradas pouco relevantes podem piorar o processo decisório de maneira que uma análise criteriosa da relevância de cada variável é um bom caminho na busca de uma redução ainda maior da dimensão dos dados de entrada.
- A base de dados para *leptonjets* representa uma fração dos possíveis decaimentos esperados para dados exóticos, dentre eles, poderíamos destacar o decaimento final em outros pares de léptons, como múons ( $\mu$ ) ou mesmo uma mistura de elétrons e múons, além de um número final maior de léptons por RoI, o que elevaria o nível de dificuldade da tarefa de classificação desses eventos. A criação de estratégias de filtragem para esses cenários, fortaleceria, ainda mais, o peso de um possível canal *leptonjets/jato*.
- A realização de testes usando dados simulados com as últimas incorporações do simulador de Monte Carlo (este vem sofrendo modificações desde o início das análises de colisão no LHC), assim como testes usando dados reais torna-se um passo muito importante na avaliação de cada uma das técnicas propostas. É de conhecimento do especialista que ajustes em alguns dos algoritmos são necessários, devido a pequenas diferenças que ocorrem entre o simulador de calorimetria e os sensores reais.
- A incorporação da abordagem vencedora do canal e/j no ambiente Athena.
- Mesmo a geração de eventos exóticos (*leptonjets*) em simuladores não sendo trivial, o enriquecimento quantitativo da base de dados usada possibilitaria um melhor estudo desses eventos, principalmente nos resultados do canal proposto *leptonjets/jato*.
- O estudo de outras técnicas de análise de componentes (PCD, NLPCA, NLICA - Nonlinear Independent Component Analysis - etc.) aplicadas ao canal *leptonjets/jato* pode seguir o mesmo caminho que a literatura já apresenta para o canal elétron/jato.

- O estudo do potencial de outros classificadores, i.e., que englobem em sua construção algum grau de comprometimento com o erro de classificação, como a SVM, pode ser uma linha de pesquisa e ser seguida para os dois canais estudados.

# Referências Bibliográficas

- [1] EVANS, L., BRYANT, P. “LHC Machine”, *Journal of Instrumentation*, v. 3, n. 08, pp. S08001, 2008. Disponível em: <<http://stacks.iop.org/1748-0221/3/i=08/a=S08001>>.
- [2] COLLABORATION, T. A. “The ATLAS Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation*, v. 3, n. 08, pp. S08003, 2008. Disponível em: <<http://stacks.iop.org/1748-0221/3/i=08/a=S08003>>.
- [3] ALICE COLLABORATION - CERN. “A Large Ion Collider Experiment”. <http://aliceinfo.cern.ch/Collaboration/index.html>, 2009. Última visita em 10/09/2009.
- [4] CMS EXPERIMENT - CERN. “The Compact Muon Solenoid Experiment”. <http://cms.web.cern.ch/cms/index.html>, 2008. Última visita em 10/09/2009.
- [5] LHCb EXPERIMENT - CERN. “Large Hadron Collider beauty Experiment”. <http://lhcb-public.web.cern.ch/lhcb-public/>, 2008. Última visita em 10/09/2009.
- [6] JOLLIFFE, I. T. *Principal Component Analysis*. 2nd ed. USA, Springer, October 2002.
- [7] HYVÄRINEN, A., KARHUNEN, J., OJA, E. *Independent Component Analysis*. USA, John Wiley & Sons, Inc., 2001.
- [8] DUDA, R. O., HART, P. E., STORK, D. G. *Pattern Classification*. 2 ed. USA, Wiley-Interscience, 2000.
- [9] HAYKIN, S. *Redes Neurais, Princípios e prática*. 2 ed. Brasil, Bookman, 2001.
- [10] ORSI, S. “PAMELA: A payload for antimatter matter exploration and light nuclei astrophysics”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 580, n. 2, pp. 880 – 883, 2007. Imaging 2006 - Proceedings of the 3rd International Conference on Imaging Techniques in Subatomic Physics, Astrophysics, Medicine, Biology and Industry.

- [11] BARWICK, S. W., BEATTY, J. J., BOWER, C. R., et al. “The High-Energy Antimatter Telescope (HEAT): An instrument for the study of cosmic-ray positrons”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 400, n. 1, pp. 34–52, 1997.
- [12] SEO, E.-S., ADAMS, J. H., BASHINDZHAGYAN, G. L., et al. “Advanced thin ionization calorimeter to measure ultrahigh energy cosmic rays”, *Advances in Space Research*, v. 19, n. 5, pp. 711–718, 1997. Cosmic radiation: Spectra and Composition.
- [13] TEEGARDEN, B. J., STURNER, S. J. “INTEGRAL Observations of Gamma-Ray Bursts”. In: *Bulletin of the American Astronomical Society*, v. 31, *Bulletin of the American Astronomical Society*, pp. 717–+, apr 1999.
- [14] BERNABEI, R., BELLI, P., BUSSOLOTTI, A., et al. “The DAMA/LIBRA apparatus”, *Nuclear Instruments and Methods in Physics Research A*, v. 592, pp. 297–315, jul. 2008. doi: 10.1016/j.nima.2008.04.082.
- [15] LÖNNBLAD, L., PETERSON, C., RÖGNVALDSSON, T. “Finding gluon jets with a neural trigger”, *Phys. Rev. Lett.*, v. 65, n. 11, pp. 1321–1324, Sep 1990.
- [16] LÖNNBLAD, L., PETERSON, C., RÖGNVALDSSON, T. “Using neural networks to identify jets”, *Nuclear Physics B*, v. 349, n. 3, pp. 675–702, 1991.
- [17] A.GRAHAM, M., JONES, L. M., HERBIN, S. “Neural network classification of quark and gluon jets”, *Phys. Rev. D*, v. 51, n. 9, pp. 4789–4807, May 1995. doi: 10.1103/PhysRevD.51.4789.
- [18] HOLMSTROM, L., SAIN, S. R., MIETTINEN, H. E. “A new multivariate technique for top quark search”, *Computer Physics Communications*, v. 88, n. 2–3, pp. 195–210, 1995.
- [19] HULTQVIST, K., JACOBSSON, R., JOHANSSON, K. “Using a neural network in the search for the Higgs boson”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 364, n. 1, pp. 193–200, 1995.
- [20] AALTONEN, T., ADELMAN, J., AKIMOTO, T., et al. “Search for  $WW$  and  $WZ$  production in lepton plus jets final state at CDF”, *Phys. Rev. D*, v. 79, n. 11, pp. 112011, Jun 2009. doi: 10.1103/PhysRevD.79.112011.
- [21] AALTONEN, T., ADELMAN, J., AKIMOTO, T., et al. “Search for standard model Higgs boson production in association with a  $W$  boson using a neural network

discriminant at CDF”, *Phys. Rev. D*, v. 80, n. 1, pp. 012002, Jul 2009. doi: 10.1103/PhysRevD.80.012002.

- [22] DELPHI EXPERIMENT. “DEtector with Lepton, Photon and Hadron Identification”. <http://delphiwww.cern.ch/>. Última visita em 20/01/2011.
- [23] CERN. “LEP - Large Electron-Positron Collider”. <http://public.web.cern.ch/public/en/research/lep-en.html>, . Última visita em 20/01/2011.
- [24] FERMI NATIONAL ACCELERATOR LABORATORY. “The Collider Detector at Fermilab”. <http://www-cdf.fnal.gov/>, . Última visita em 20/01/2011.
- [25] FERMI NATIONAL ACCELERATOR LABORATORY. “Tevatron Department - Beams Division”. <http://www-bdnew.fnal.gov/tevatron/>, . Última visita em 20/01/2011.
- [26] US DEPARTMENT OF ENERGY. “Fermi National Accelerator Laboratory”. <http://www.fnal.gov/>. Última visita em 20/01/2011.
- [27] PRICE, D. D. “Search for Standard Model Higgs in  $\mathbf{WH} \rightarrow l\nu b\bar{b}$  at the Tevatron”, 2009.
- [28] PENNING, B. “Search for  $\mathbf{H} \rightarrow \mathbf{WW} \rightarrow ll$  at D0”, 2008.
- [29] FERMI NATIONAL ACCELERATOR LABORATORY. “The D0 Experiment”. <http://www-d0.fnal.gov/>, . Última visita em 20/01/2011.
- [30] MAGGIPINTO, T., NARDULLI, G., DUSINI, S., et al. “Role of neural networks in the search of the Higgs boson at LHC”, *Physics Letters B*, v. 409, n. 1–4, pp. 517–522, 1997.
- [31] DENBY, B. “Neural networks and cellular automata in experimental high energy physics”, *Computer Physics Communications*, v. 49, n. 3, pp. 429–448, 1988.
- [32] PETERSON, C. “Track finding with neural networks”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 279, n. 3, pp. 537–545, 1989.
- [33] GYULASSY, M., HARLANDER, M. “Elastic tracking and neural network algorithms for complex pattern recognition”, *Computer Physics Communications*, v. 66, n. 1, pp. 31–46, 1991.
- [34] LINDSEY, C. S., DENBY, B., HAGGERTY, H., et al. “Real time track finding in a drift chamber with a VLSI neural network”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 317, n. 1–2, pp. 346–356, 1992.

- [35] DENBY, B., LINDSEY, C. S., DICKSON, M., et al. “Performance of the CDF neural network electron isolation trigger at Fermilab”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 356, n. 2–3, pp. 485–492, 1995.
- [36] SEIXAS, J. M., CALÔBA, L. P., SOUZA, M. N., et al. “Neural second-level trigger system based on calorimetry”, *Computer Physics Communications*, v. 95, n. 2–3, pp. 143–157, 1996.
- [37] ABRAMOWICZ, H., CALDWELL, A., SINKUS, R. “Neural network based electron identification in the ZEUS calorimeter”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 365, n. 2–3, pp. 508 – 517, 1995.
- [38] DESY. “The ZEUS Experiment”. <http://www-zeus.desy.de/index.html>, . Última visita em 20/01/2011.
- [39] DESY. “The Hadron Electron Ring Accelerator”. <http://www-zeus.desy.de/public/hera.php3>, . Última visita em 20/01/2011.
- [40] DESY. “DESY - Deutsches Elektronen-Synchrotron”. <http://www.desy.de>, . Última visita em 20/01/2011.
- [41] ABT, I., AHMED, T., AID, S., et al. “The H1 detector at HERA”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 386, n. 2–3, pp. 310–347, 1997.
- [42] KIESLING, C. “The H1 neural network trigger”. In: *Neural Networks, 2003. Proceedings of the International Joint Conference on*, v. 1, pp. 393–398, july 2003. doi: 10.1109/IJCNN.2003.1223378.
- [43] DENBY, B., GARDA, P., GRANADO, B., et al. “Fast triggering in high-energy physics experiments using hardware neural networks”, *Neural Networks, IEEE Transactions on*, v. 14, n. 5, pp. 1010–1027, sept 2003. ISSN: 1045-9227. doi: 10.1109/TNN.2003.816903.
- [44] TAUSCHER, L., VLACHOS, S., WADHWA, M. “A charged particle trigger based on neural networks for the L3 experiment using the inner-TEC detector”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 389, n. 1-2, pp. 138 – 140, 1997. New Computing Techniques in Physics Research V.
- [45] CERN. “The L3 Experiment at CERN”. <http://l3.web.cern.ch/l3/>, . Última visita em 20/01/2011.

- [46] HORN, D. “Neural computation methods and applications: Summary talk of the AI session”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 389, n. 1–2, pp. 381–387, 1997. New Computing Techniques in Physics Research V.
- [47] DENBY, B. “Neural networks in high energy physics: A ten year perspective”, *Computer Physics Communications*, v. 119, n. 2–3, pp. 219–231, 1999.
- [48] ANJOS, A. D., TORRES, R. C., SEIXAS, J. M., et al. “Neural triggering system operating on high resolution calorimetry information”, *Nuclear Instruments and Methods in Physics Research A*, v. 559, pp. 134–138, abril 2006. doi: 10.1016/j.nima.2005.11.132.
- [49] DOS ANJOS, A. R. *Sistema Online de Filtragem em um Ambiente com Alta Taxa de Eventos*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, Dezembro 2006.
- [50] STRANDLIE, A. “Track reconstruction - from bubble chambers to the LHC”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 535, n. 1–2, pp. 57–64, 2004. Proceedings of the 10th International Vienna Conference on Instrumentation.
- [51] LIMA, H. P., SEIXAS, J. M. “A segmented principal component analysis applied to calorimetry information at ATLAS”, *Nuclear Instruments and Methods in Physics Research A*, v. 559, pp. 129–133, abril 2006. doi: 10.1016/j.nima.2005.11.131.
- [52] TORRES, R. C. *Sistema online de filtragem em um ambiente com alta taxa de eventos e fina granularidade*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, março 2010.
- [53] DE SIMAS FILHO, E. F. *Análise não-linear de componentes independentes para uma filtragem online baseada em calorimetria de alta energia e com fina segmentação*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, dezembro 2010.
- [54] ABDALA, M. C. B. *O discreto charme das partículas elementares*. 1 ed. São Paulo, Editora UNESP, 2004.
- [55] LATTES, C. M. G., MUIRHEAD, H., OCCHIALINI, G. P. S., et al. “Process Involving Charged Mesons”, *Nature*, v. 159, pp. 694–697, maio 1947. doi: 10.1038/159694a0.
- [56] MOREIRA, M. A. “O Modelo Padrão da Física de Partículas”, *Revista Brasileira de Ensino de Física*, v. 31, n. 1, pp. 1306, 2009.



- [57] ZWICKY, F. “On the Masses of Nebulae and of Clusters of Nebulae”, v. 86, out. 1937. doi: 10.1086/143864.
- [58] BABCOCK, H. W. “The rotation of the Andromeda Nebula”, *Lick Observatory Bulletin*, v. 19, pp. 41–51, 1939.
- [59] SCOTT, D. “The standard cosmological model”, 2005. 16 pages, 2 figures, invited talk at Theory Canada 1, June 2005, Vancouver.
- [60] LAHAV, O., LIDDLE, A. R. “The cosmological parameters 2005”, *Review of Particle Physics 2006*, 2006. doi: 10.1088/0954-3899/33/1/001. On page 224-232 of the Review of Particle Physics, please cite the entire review.
- [61] ARKANI-HAMED, N., FINKBEINER, D. P., SLATYER, T. R., et al. “A Theory of Dark Matter”, *Physical Review D*, v. 79, pp. 015014, 2009. Disponível em: <doi:10.1103/PhysRevD.79.015014>.
- [62] ARKANI-HAMED, N., WEINER, N. “LHC Signals for a SuperUnified Theory of Dark Matter”, *JHEP*, v. 0812, pp. 104, 2008. Disponível em: <doi:10.1088/1126-6708/2008/12/104>.
- [63] BAUMGART, M., CHEUNG, C., RUDERMAN, J. T., et al. “Non-Abelian Dark Sectors and Their Collider Signatures”, *JHEP*, v. 0904, pp. 014, 2009. Disponível em: <doi:10.1088/1126-6708/2009/04/014>.
- [64] STRASSLER, M. J., ZUREK, K. M. “Echoes of a Hidden Valley at Hadron Colliders”, *Physics Letters B*, v. 651, pp. 374, 2007. Disponível em: <doi:10.1016/j.physletb.2007.06.055>.
- [65] RUDERMAN, J. T., VOLANSKY, T. “Decaying into the Hidden Sector”. 2009. Disponível em: <http://www.citebase.org/abstract?id=oai:arXiv.org:0908.1570extgreater .
- [66] ABAZOV, D. V. “Search for NMSSM Higgs bosons in the  $h \rightarrow aa \rightarrow \mu\mu\mu, \mu\mu\tau$  channels using  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96\text{TeV}$ ”, *Physical Review Letters*, v. 103, pp. 061801, 2009. Disponível em: <doi:10.1103/PhysRevLett.103.061801>.
- [67] MARTIN, S. “A Supersymmetry Primer”. In: Kane, G. L. (Ed.), *Perspectives on Supersymmetry*, pp. 1–+, 1998.
- [68] FALKOWSKI, A., RUDERMAN, J. T., VOLANSKY, T., et al. “Hidden Higgs Decaying to Lepton Jets”, *JHEP*, v. 05, pp. 077, 2010. doi: 10.1007/JHEP05(2010)077.



- [69] FALKOWSKI, A., RUDERMAN, J. T., VOLANSKY, T., et al. “Discovering Higgs Boson Decays to Lepton Jets at Hadron Colliders”, *Phys. Rev. Lett.*, v. 105, n. 24, pp. 241801, Dec 2010. doi: 10.1103/PhysRevLett.105.241801.
- [70] ABAZOV, V. M., OTHERS. “Search for events with leptonic jets and missing transverse energy in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV”, *Phys. Rev. Lett.*, v. 105, pp. 211802, 2010. doi: 10.1103/PhysRevLett.105.211802.
- [71] LHC. “Large Hadron Collider”. <http://lhc.web.cern.ch/lhc/>, 2009. Última visita em 10/09/2009.
- [72] CBPF. “LHC - O Gigante criador da materia”. <http://mesonpi.cat.cbpf.br/desafios/>, 2007. Última visita em 07/08/2009.
- [73] CERN. “The accelerator complex”. <http://public.web.cern.ch/public/en/Research/AccelComplex-en.html>, . Última visita em 20/01/2011.
- [74] CHECCHIA, P., EREDITATO, A. “Particle detectors for future collider experiments: a description of basic principles and general characteristics”, *Instrumentation Measurement Magazine, IEEE*, v. 8, n. 2, pp. 27 – 32, jun 2005. ISSN: 1094-6969. doi: 10.1109/MIM.2005.1405921.
- [75] ATLAS COLLABORATION. *ATLAS detector and physics performance, technical design report*. Relatório Técnico I, CERN, Maio 1999.
- [76] CERN. “The ATLAS Experiment, Mapping the Secrets of the Universe”. <http://atlasexperiment.org/>, 2007. Última visita em 07/04/2008.
- [77] PEQUENÃO, J., MAIO, A. “A virtual reality image of ATLAS”, ATLAS Collection., 2002.
- [78] THE ATLAS TDAQ COLLABORATION. *ATLAS High-Level Trigger Data Acquisition and Controls Technical Design Report*. Relatório técnico, CERN/LHCC/2003-022, 2003.
- [79] THE ATLAS COLLABORATION. “The ATLAS Inner Detector commissioning and calibration”, *ArXiv e-prints*, abr. 2010.
- [80] GOLUTVIN, A. “Review of calorimeters”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 453, n. 1–2, pp. 192–198, 2000.
- [81] THE ATLAS COLLABORATION. “Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions”, *ArXiv e-prints*, dez. 2009.

- [82] PEQUENAO, J. “Event Cross Section in a computer generated image of the ATLAS detector.” Mar 2008.
- [83] TAREM, S., PANIKASHVILI, N. “Muon identification and reconstruction in the ATLAS detector at the LHC”. In: *Nuclear Science Symposium Conference Record, 2004 IEEE*, v. 4, pp. 2186–2190 Vol. 4, Outubro 2004. doi: 10.1109/NSSMIC.2004.1462696.
- [84] THE ATLAS COLLABORATION. “Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays”, *ArXiv e-prints*, jun. 2010.
- [85] MELLO, A., DOS ANJOS, A., ARMSTRONG, S., et al. “Overview of the High-Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC”, *Nuclear Science, IEEE Transactions on*, v. 53, n. 5, pp. 2839–2843, outubro 2006. ISSN: 0018-9499. doi: 10.1109/TNS.2006.882049.
- [86] ATLAS COLLABORATION. *Athena Developer Guide*. CERN, Agosto 2001.
- [87] DAMAZIO, D., LIMA, D. E. F. D. “TrigT2CaloRinger”. <https://twiki.cern.ch/twiki/bin/view/Atlas/TrigT2CaloRinger>, 2009. Última visita em 20/09/2009.
- [88] GONZALEZ, S., HANSL-KOZANECKA, T., WIELERS, M. *Selection of high-p<sub>T</sub> electromagnetic clusters by the Second Level Trigger of ATLAS*. Relatório técnico, ATL-COM-DAQ-99-009, 1999.
- [89] GONZALEZ, S., HANS-KOZANECKA, T., WIELERS, M. “Selection of high-p<sub>T</sub> electromagnetic clusters by the level-2 trigger of ATLAS”. janeiro 2000.
- [90] ARMSTRONG, S., BAINES, J., BEE, C., et al. “Studies for a common selection software environment in ATLAS: from the level-2 trigger to the offline reconstruction”, *Nuclear Science, IEEE Transactions on*, v. 51, n. 3, pp. 915–920, Junho 2004. ISSN: 0018-9499. doi: 10.1109/TNS.2004.829548.
- [91] SCHIAVI, C. “Implementation and performance of the high level trigger electron and photon selection for the ATLAS experiment at the LHC”, *Nuclear Science, IEEE Transactions on*, v. 53, n. 3, pp. 1424–1429, Junho 2006. ISSN: 0018-9499. doi: 10.1109/TNS.2006.871902.
- [92] LIMA, D. E. F. D. “TrigMultiVarHypo”. <https://twiki.cern.ch/twiki/bin/view/Atlas/TrigMultiVarHypo>, 2009. Última visita em 20/09/2009.
- [93] ROBERT, C. P., CASELLA, G. *Monte Carlo Statistical Methods*. 2 ed. USA, Springer Verlag, 2004.

- [94] TREES, H. L. V. *Detection, Estimation, and Modulation Theory, Part I*. John Wiley and Sons, 2001.
- [95] TREES, H. L. V. *Detection, Estimation, and Modulation Theory, Part III*. John Wiley and Sons, 2001.
- [96] ANJOS, A. D., TORRES, R., FERREIRA, B., et al. “Otimização do Segundo Nível de Trigger do ATLAS Baseado em Calorimetria”. In: *XXVI Encontro Brasileiro de Física de Partículas e Campos*, 2005.
- [97] ANJOS, A. D. *Sistema Neuronal de decisão baseado em calorimetria de altas energias*. Tese de Mestrado, COPPE/UFRJ, 2001.
- [98] JÚNIOR, H. P. L. *Monitoração online e processamento distribuído para um ambiente com alta taxa de eventos e fina segmentação*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro, Outubro 2006.
- [99] DE SOUZA, D. L., DE SEIXAS, J. M. “Análise de Componentes Principais Não-Lineares aplicada num Detector Finamente Segmentado”. In: *XXVIII Encontro Nacional de Física de Partículas e Campos*, setembro 2007.
- [100] DE SOUZA, D. L., DE SEIXAS, J. M. “Redução de Dimensionalidade num Detector Finamente Segmentado Utilizando a Análise de Componentes Principais Não-Lineares”. In: *Congresso Brasileiro de Redes Neurais (CBRN 2007)*, pp. 8–11, outubro 2007.
- [101] TORRES, R. C., SIMAS, E., DE SOUZA, D. L., et al. “Segmented Online Neural Filtering System Based On Independent Components Of Pre-Processed Information”. In: *Signal Processing*, v. 1, pp. 337–358, India, India: In-Tech, 2010. ISBN: 0097895376196.
- [102] CHEUNG, C., RUDERMAN, J., WANG, L.-T., et al. “Lepton jets in (supersymmetric) electroweak processes”, *Journal of High Energy Physics*, v. 2010, pp. 1–23, 2010. Disponível em: <[http://dx.doi.org/10.1007/JHEP04\(2010\)116](http://dx.doi.org/10.1007/JHEP04(2010)116)>. 10.1007/JHEP04(2010)116.
- [103] SOUZA, D., DE SEIXAS, J. M. “Statistical Processing on High Resolution Calorimetry Information”. In: *WCI 2010*, oct 2010. Disponível em: <<http://XXXXXX/74870.pdf>>.
- [104] DE SOUZA, D. L., DE SEIXAS, J. M. “Neural leptonjets/jet discrimination in highly segmented calorimeters”. In: *X Congresso Brasileiro de Inteligência Computacional (CBIC 2011)*, v. 1, pp. 49–54, 2011.

- [105] CHEN, L.-H., CHANG, S. “An adaptive learning algorithm for principal component analysis”, *Neural Networks, IEEE Transactions on*, v. 6, n. 5, pp. 1255–1263, setembro 1995.
- [106] KARHUNEN, J., JOUTSENSALO, J. “Generalizations of principal component analysis, optimization problems, and neural networks”, *Neural Networks*, v. 8, n. 4, pp. 549–562, 1995. ISSN: 0893-6080. doi: [http://dx.doi.org/10.1016/0893-6080\(94\)00098-7](http://dx.doi.org/10.1016/0893-6080(94)00098-7).
- [107] CHATTERJEE, C., KANG, Z., ROYCHOWDHURY, V. “Algorithms for accelerated convergence of adaptive PCA”, *Neural Networks, IEEE Transactions on*, v. 11, n. 2, pp. 338–355, março 2000.
- [108] COMON, P., GOLUB, G. “Tracking a few extreme singular values and vectors in signal processing”, *Proceedings of the IEEE*, v. 78, n. 8, pp. 1327–1343, aug 1990.
- [109] OWSLEY, N. “Adaptive data orthogonalization”. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78.*, v. 3, pp. 109 – 112, apr 1978.
- [110] TUFTS, D., MELISSINOS, C. “Simple, effective computation of principal eigenvectors and their eigenvalues and application to high-resolution estimation of frequencies”, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 34, n. 5, pp. 1046 – 1053, oct 1986. ISSN: 0096-3518.
- [111] SHARMAN, K. “Adaptive algorithms for estimating the complete covariance eigenstructure”. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, v. 11, pp. 1401 – 1404, apr 1986.
- [112] BUNCH, J. R., NIELSEN, C. P., SORENSEN, D. C. “Rank-one modification of the symmetric eigenproblem”, *Numerische Mathematik*, v. 31, pp. 31–48, 1978. ISSN: 0029-599X. Disponível em: <http://dx.doi.org/10.1007/BF01396012>.
- [113] KARASALO, I. “Estimating the covariance matrix by signal subspace averaging”, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 34, n. 1, pp. 8 – 12, fev 1986. ISSN: 0096-3518.
- [114] OJA, E. “Simplified neuron model as a principal component analyzer”, *Journal of Mathematical Biology*, v. 15, pp. 267–273, 1982. ISSN: 0303-6812. Disponível em: <http://dx.doi.org/10.1007/BF00275687>.

- [115] KARHUNEN, J., OJA, E. “New methods for stochastic approximation of truncated Karhunen-Loève expansions”. In: *International Conference on Pattern Recognition - ICPR*, pp. 550–553, 1982.
- [116] KARHUNEN, J. “Adaptive algorithms for estimating eigenvectors of correlation type matrices”. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, v. 9, pp. 592 – 595, mar 1984.
- [117] YANG, J. F., KAVEH, M. “Adaptive eigensubspace algorithms for direction or frequency estimation and tracking”, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 36, n. 2, pp. 241 –251, feb 1988. ISSN: 0096-3518. doi: 10.1109/29.1516.
- [118] REDDY, V., EGARDT, B., KAILATH, T. “Least squares type algorithm for adaptive implementation of Pisarenko’s harmonic retrieval method”, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 30, n. 3, pp. 399 – 405, jun 1982. ISSN: 0096-3518.
- [119] BANNOUR, S., AZIMI-SADJADI, M. R. “An adaptive approach for optimal data reduction using recursive least squares learning method”. In: *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, v. 2, pp. 297 –300 vol.2, 23-26 1992. doi: 10.1109/ICASSP.1992.226061.
- [120] K. JAIN, A. *Fundamentals of Digital Image Processing*. New Jersey, Prantice-Hall Inc., 1989.
- [121] YANG, B. “Projection Approximation Subspace Tracking”, *IEEE Transactions on Signal Processing*, v. 43, n. 1, pp. 95–107, Janeiro 1995.
- [122] YANG, B. “An extension of the PASTd algorithm to both rank and subspace tracking”, *Signal Processing Letters, IEEE*, v. 2, n. 9, pp. 179 –182, sep 1995. ISSN: 1070-9908. doi: 10.1109/97.410547.
- [123] ABED-MERAIM, K., CHKEIF, A., HUA, Y. “Fast orthonormal PAST algorithm”, *Signal Processing Letters, IEEE*, v. 7, n. 3, pp. 60 –62, mar 2000. ISSN: 1070-9908. doi: 10.1109/97.823526.
- [124] ZHANG, H., REN, G., ZHANG, H., et al. “An improved OPAST algorithm for spatio-temporal multiuser detection technique based on subspace tracking”. In: *Communications Systems, 2004. ICCS 2004. The Ninth International Conference on*, pp. 401 –404, 7-7 2004. doi: 10.1109/ICCS.2004.1359407.

- [125] OJA, E. “The nonlinear PCA learning rule in independent component analysis”, *Neurocomputing*, v. 17, n. 1, pp. 25–45, Setembro 1997.
- [126] ZHU, X.-L., ZHANG, X.-D. “Adaptive RLS Algorithm for Blind Source Separation Using a Natural Gradient”, *IEEE Signal Processing Letters*, v. 9, n. 12, pp. 432–435, Dezembro 2002.
- [127] KRAMER, M. A. “Nonlinear principal component analysis using autoassociative neural networks”, *AIChE Journal*, v. 34, n. 2, pp. 161–310, Fevereiro 1991.
- [128] OJA, E. “From neural learning to independent components”, *Neurocomputing*, v. 22, n. 1–3, pp. 187–199, Novembro 1998.
- [129] HASTIE, T., STUETZLE, W. “Principal Curves”, *Journal of the American Statistical Association*, v. 84, n. 406, pp. 502–516, 1989.
- [130] LEBLANC, M., TIBSHIRANI, R. “Adaptive principal surfaces”, *Journal of the American Statistical Association*, v. 89, n. 425, pp. 53–64, 1994.
- [131] SCHÖLKOPF, B., SMOLA, A., MÜLLER, K.-R. “Nonlinear Component Analysis as a Kernel Eigenvalue Problem”, *Neural Computation*, v. 10, n. 5, pp. 1299–1319, julho 1998.
- [132] KAMBHATLA, N., LEEN, T. K. “Dimension Reduction by Local Principal Component Analysis”, *Neural Computation*, v. 9, n. 7, pp. 1493–1516, outubro 1997.
- [133] MALTHOUSE, E. C. “Limitations of Nonlinear PCA as Performed with Generic Neural Networks”, *IEEE Transactions on Neural Networks*, v. 9, n. 1, pp. 165–173, Janeiro 1998.
- [134] KARHUNEN, J., PAJUNEN, P., OJA, E. “The Nonlinear PCA Criterion in Blind Source Separation: Relations with Other Approaches”, *Neurocomputing*, v. 22, n. 5/6, pp. 5–20, Novembro 1998.
- [135] KARHUNEN, J., PAJUNEN, P. “Blind source separation and tracking using nonlinear PCA criterion: a least-squares approach”, *Neural Networks International Conference*, v. 4, pp. 2147–2152, Setembro 1997.
- [136] ZHU, X.-L., ZHANG, X.-D., DING, Z.-Z., et al. “Adaptive Nonlinear PCA Algorithms for Blind Source Separation Without Prewhitening”, *IEEE Transactions on Circuits and Systems*, v. 53, n. 3, pp. 745–753, Março 2006.



- [137] PAJUNEN, P., KARHUNEN, J. “Least Square Methods for Blind Source Separation Based on Nonlinear PCA”, *International Journal of Neural Systems*, v. 8, n. 5/6, pp. 601–612, Outubro–Dezembro 1998.
- [138] GOLUB, G., LOAN, C. V. *Matrix Computations*. 3 ed. USA, Johns Hopkins University Press, 1996.
- [139] ANS, B., HÉRAULT, J., JUTTEN, C. “Adaptive neural architectures: detection of primitives”. In: *Proc. of COGNITIVA '85*, pp. 593–597, Paris, France, 1985.
- [140] COMON, P. “Independent component analysis, a new concept?” *Signal Process.*, v. 36, n. 3, pp. 287–314, 1994. ISSN: 0165-1684. doi: [http://dx.doi.org/10.1016/0165-1684\(94\)90029-9](http://dx.doi.org/10.1016/0165-1684(94)90029-9).
- [141] CARDOSO, J.-F., SOULOUMIAC, A. “Blind beamforming for non Gaussian signals”, *IEE Proceedings-F*, v. 140, n. 6, pp. 362–370, dec 1993.
- [142] BELOUCHRANI, A., ABED MERAÏM, K., COIS CARDOSO, J.-F., et al. “A blind source separation technique based on second order statistics”, *ieeSP*, v. 45, n. 2, pp. 434–44, feb 1997.
- [143] TONG, L., SOON, V., HUANG, Y., et al. “AMUSE: a new blind identification algorithm”. In: *Circuits and Systems, 1990., IEEE International Symposium on*, pp. 1784 –1787 vol.3, 1-3 1990. doi: 10.1109/ISCAS.1990.111981.
- [144] ZIEHE, A., MÜLLER, K. R. “TDSEP - an efficient algorithm for blind separation using time structure”. In: *Proceedings of the 8th International Conference on Artificial Neural Networks, ICANN 98, Perspectives in Neural Computing*, pp. 675 – 680, Berlin, Jan 1998. Springer Verlag.
- [145] FERREOL, A., ALBERA, L., CHEVALIER, P. “Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM)”, *Signal Processing, IEEE Transactions on*, v. 53, n. 5, pp. 1640 – 1653, may 2005. ISSN: 1053-587X. doi: 10.1109/TSP.2005.845431.
- [146] DELFOSSE, N., LOUBATON, P. “Adaptive separation of independent sources: a deflation approach”. In: *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, v. iv, pp. IV/41 –IV/44 vol.4, 19-22 1994. doi: 10.1109/ICASSP.1994.389881.
- [147] HYVÄRINEN, A. “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”, *IEEE Transactions on Neural Networks*, v. 10, n. 3, pp. 626–634, Maio 1999.

- [148] MOREAU, E. “Criteria for Complex Sources Separation”. In: *In Proc. EUSIPCO*, pp. 931–934, Set 1996.
- [149] LEE, T.-W., GIROLAMI, M., SEJNOWSKI, T. J. “Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources”, *Neural Computation*, v. 11, n. 2, pp. 417–441, 1999. doi: 10.1162/089976699300016719. Disponível em: <\url{http://www.mitpressjournals.org/doi/abs/10.1162/089976699300016719}>.
- [150] ICHI AMARI, S. “Natural Gradient Works Efficiently in Learning”, *Neural Computation*, v. 10, n. 2, pp. 251–276, 1998. doi: 10.1162/089976698300017746. Disponível em: <\url{http://www.mitpressjournals.org/doi/abs/10.1162/089976698300017746}>.



# Apêndice A

## Extração de Características e Teste de Hipótese

Considerando problemas onde o espaço de entrada de dados tem dimensão bastante elevada a compactação da informação e extração de características pode se mostrar atraente. Tal processo, em geral, é possível pois há dois tipos de dimensionalidade: a superficial e a intrínseca. A dimensionalidade superficial dos dados pode ser definida como o número de variáveis aparentes. A dimensionalidade intrínseca pode ser definida como o número de variáveis independentes que regem as variações não-aleatórias nas observações, i.e., é o número de variáveis que estão realmente por trás do processo que se deseja estudar, lembrando-se que essas variáveis nem sempre estão visíveis como aquelas que foram descritas como superficiais.

As técnicas de análise de componentes, sejam principais (lineares ou não) ou independentes, são abordagens consagradas no tratamento de: análise estatística, extração de características, compressão e visualização de dados.

O ponto de início das técnicas de análise de componentes é um vetor de observação  $\mathbf{x} \in \mathbb{R}^l$ , onde  $n$  observações compõem a amostra de dados  $\mathbf{X}$ . Tipicamente, os vetores  $\mathbf{x}$  são valores de um sinal adquirido em diferentes instantes de tempo ou espaço. O que é essencial para a aplicação dessas técnicas é que os dados sejam mutuamente correlacionados e que haja, então, um certo grau de redundância das amostras.

O vetor de observação  $\mathbf{x}$  é o resultado da captação dos sinais fonte, independentes, alterados pelo estatística do canal. Na forma matricial pode-se escrever como

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{A.1}$$

onde  $\mathbf{A} \in \mathbb{R}^{l \times f}$  é uma matriz determinística de mistura que caracteriza o meio ou canal,  $\mathbf{s} \in \mathbb{R}^f$  é o vetor de sinais fonte com matriz de correlação não-singular  $\mathbf{C}_s = E[\mathbf{s}\mathbf{s}^T]$  [7].

A identificação e a separação de múltiplos sinais é alcançada sem se conhecer as características do canal ou meio onde os sinais trafegam. O exemplo clássico é o problema, ma-

tematicamente tratável, da identificação cega de fontes - *Blind Source Separation* (BSS). A identificação cega se compromete a revelar tanto  $\mathbf{A}$  quanto  $\mathbf{s}$  a partir de  $\mathbf{x}$ .

Se os fatores  $a_{ij}$  de  $\mathbf{A}$  são conhecidos, o problema é resolvido trivialmente a partir de

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (\text{A.2})$$

onde  $\mathbf{W} = \mathbf{A}^{-1}$ .

Dependendo do problema tratado diferentes abordagens podem ser usadas com sucesso, dentre elas as análise lineares e não-lineares de componentes.

A Análise de Componentes Principais (*Principal Component Analysis* - PCA) tem sido extensivamente utilizada na compactação e representação de sinais a partir de componentes de maior energia. Essas componentes principais correspondem aos autovetores da matriz de correlação relativa ao conjunto de dados.

A Análise de Componentes Principais Não-Lineares (*Nonlinear Principal Component Analysis* - NLPCA) é utilizada para identificar e remover correlações, assim como a PCA. Enquanto a PCA identifica somente correlação linear entre as variáveis, a NLPCA abrange tanto a correlação linear quanto a relação não-linear, sem restrição quanto ao tipo de não-linearidade presente nos dados.

A Análise de Componentes Independentes (*Independent Component Analysis* - ICA) é uma transformação que busca a independência estatística das componentes de saída. Nessa tarefa, é comum o uso de estatística de ordem superior no processamento. A busca pela minimização da dependência estatística pode ser vista como uma extensão da PCA e NLPCA.

## A.1 Análise de Componentes Principais

É vasta a literatura que mostra a PCA como uma solução ótima para diferentes problemas de representação da informação [6, 105–107]. Isso inclui:

- Maximização das variâncias linearmente transformadas.
- Minimização do erro quadrático médio de reconstrução, quando os dados de entrada são aproximados usando um sub-espço linear dimensionalmente menor (Figura A.1).
- Descorrelação das saídas.
- Minimização da representação da entropia.

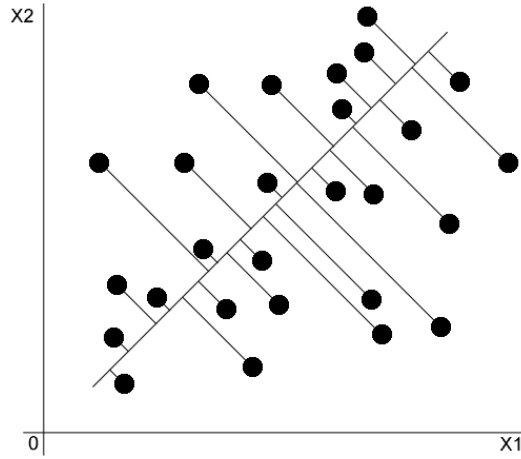


Figura A.1: Minimização do desvio ortogonal.

Considerando um conjunto de vetores do tipo

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_l \end{bmatrix} \quad (\text{A.3})$$

uma matriz de covariância é extraída do conjunto de dados iniciais

$$\mathbf{C}_x = E \left\{ (\mathbf{x} - \mathbf{m}_x) (\mathbf{x} - \mathbf{m}_x)^T \right\} \quad (\text{A.4})$$

onde

$$\mathbf{m}_x = E \{ \mathbf{x} \} \quad (\text{A.5})$$

é o vetor médio e  $E \{ \cdot \}$  é o valor esperado do argumento. Para  $n$  observações, pode-se aproximar discretamente as equações (A.4) e (A.5) por

$$\mathbf{C}_x = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \mathbf{x}_k^T - \mathbf{m}_x \mathbf{m}_x^T \quad (\text{A.6})$$

e

$$\mathbf{m}_x = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \quad (\text{A.7})$$

de tal forma que o elemento  $c_{ii}$  de  $\mathbf{C}_x$  é a variância de  $x_i$ , e os elementos  $c_{ij}$  são as covariâncias entre os elementos  $x_i$  e  $x_j$  desses vetores. Se os elementos  $x_i$  e  $x_j$  não estão correlacionados, suas covariâncias são igual a zero,  $c_{ij} = c_{ji} = 0$  [7].

Como  $\mathbf{C}_x$  é real e simétrica, é sempre possível encontrar um conjunto de  $l$  autovetores ortonormais [6, 106].

Seja  $\mathbf{T}$  uma matriz cujas colunas são formadas a partir dos autovetores de  $\mathbf{C}_x$ , ordenada de modo que sua primeira coluna é o autovetor obtido a partir do maior autovalor e a última coluna é o autovetor obtido a partir do menor autovalor de  $\mathbf{C}_x$ . A transformada é obtida através de

$$\mathbf{y} = \mathbf{T}^T (\mathbf{x} - \mathbf{m}_x) \quad (\text{A.8})$$

onde a matriz de transformação  $\mathbf{T}$  mapeia os vetores  $\mathbf{x}$  em vetores  $\mathbf{y}$  [6].

A média dos vetores  $\mathbf{y}$  resultantes é zero e a matriz de covariância do plano de componentes principais pode ser obtida em termos de  $\mathbf{T}$  e  $\mathbf{C}_x$  através de

$$\mathbf{C}_y = \mathbf{T}^T \mathbf{C}_x \mathbf{T} \quad (\text{A.9})$$

Além disso,  $\mathbf{C}_y$  é uma matriz diagonal cujos elementos são os autovalores de  $\mathbf{C}_x$ , i.e.,

$$\mathbf{C}_y = \begin{bmatrix} \alpha_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_n \end{bmatrix} \quad (\text{A.10})$$

Os elementos fora da diagonal de  $\mathbf{C}_y$  possuem valor zero, comprovando que os componentes dos vetores  $\mathbf{y}$  estão descorrelacionados.

A implementação das técnicas PCA tem sido baseada em algoritmos “batch”, que realizam a decomposição em autovalores - *Eigenvalue Decomposition* (ED) - da matriz de correlação ou na decomposição em valor singular - *Singular Value Decomposition* (SVD) - da matriz de dados. Essa abordagem não é apropriada para processamento adaptativo pois requer repetidas operações ED/SVD, que são bastante custosas tanto computacionalmente quanto em tempo.

Para contornar esta dificuldade um número considerável de algoritmos adaptativos tem sido desenvolvido. A maior parte dessas técnicas pode ser agrupada em três famílias. Na primeira, métodos clássicos dos algoritmos “batch” ED/SVD têm sido modificados para o uso em processamento adaptativo [108–111]. Na segunda família, variações do algoritmo de *rank* um [112], como [113], são aplicados na busca de vetores ortogonais entre si a partir de uma base obtida dos dados originais. E a terceira família de algoritmos considera a decomposição ED/SVD como um problema de otimização com ou sem restrição. Métodos baseados no gradiente [114–117], Gauss-Newton [118, 119], gradiente conjugado [120] e RLS [121–124] podem ser aplicados na busca de maiores e menores autovalores e seus correspondentes autovetores.

## A.2 Análise de Componentes Principais Não-lineares

A extensão não-linear da PCA tem sido proposta há, pelo menos, duas décadas e hoje é considerada uma importante área de pesquisa.

Em muitos casos as variáveis que compõem o vetor de informação de entrada apresentam relação não-linear entre si, de tal modo que uma análise de componentes principais não-linear é mais eficiente do que uma técnica linear padrão. A NLPCA é análoga à PCA do ponto de vista da minimização do erro quadrático, acrescida de mapeamentos não-lineares entre os espaços de dimensão original  $\mathbb{R}^l$  e de dimensão reduzida  $\mathbb{R}^f$ . Dessa forma, se houver relação não-linear entre as variáveis, a NLPCA descreverá os dados com maior acurácia e/ou com um número menor de componentes do que a PCA, a menos que não haja dados suficientes para permitir a formulação de mapeamentos mais complexas [125, 126].

Uma outra motivação para usar não-linearidades, é que elas introduzem de uma maneira implícita estatística de ordem elevada na computação. Estatística de ordem elevada, definida por cumulantes iguais ou maiores do que segundos momentos, são necessários para uma boa caracterização de dados não gaussianos [106, 127]. Estatística de ordem elevada, definida por cumulantes e momentos superiores ao de segunda ordem são necessários para uma boa caracterização de dados não-gaussianos. Além disso as saídas da PCA padrão são, usualmente, no máximo mutuamente descorrelacionadas, mas não independentes (a exceção de dados gaussianos), o que seria desejável em muitos casos.

Naturalmente, a NLPCA tem algumas desvantagens em relação à PCA padrão. A análise matemática dos algoritmos de aprendizagem é freqüentemente difícil, fazendo suas propriedades menos compreendidas. Os algoritmos de aprendizagem são mais complicados e podem algumas vezes cair mais facilmente em mínimos locais [106, 128].

Em NLPCA o mapeamento no espaço de componentes é generalizado para permitir não-linearidades arbitrárias

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \tag{A.11}$$

onde  $\mathbf{g}$  é um vetor de funções não-lineares, isto é,  $\mathbf{g} = \{g_1, g_2, \dots, g_f\}$  [127].

O termo NLPCA não é único e dentre as técnicas de análise não-linear de dados pode-se destacar: Curvas Principais [129, 130], redes neurais auto-associativas [127] e descorrelação não-linear [7].

Os métodos da primeira categoria são baseados em transformação não-linear. Kernel PCA [131], proposto por Scholkopf, é um método que realiza PCA para dados de entrada mapeados por uma função não-linear. Conseqüentemente, o método constrói as componentes principais ordenadas como a PCA. Contudo, uma maneira adequada para determinar uma função não-linear de mapeamento para um dado conjunto de dados não é bem estabelecida.

Os métodos da segunda categoria são baseados na aproximação local linear. Curvas Principais [129], proposta por Hestie et al., consiste na representação de dados de alta dimensionalidade com linhas poligonais. O PCA local proposto por Kambhalta et al. [132] divide o espaço de entrada em domínios locais e realiza a PCA para cada domínio. Como

as bases dos domínios locais adjacentes são descontínuas, isso provê um entendimento global da distribuição dos dados.

Os métodos na terceira categoria são baseados em redes neurais, tais como a rede neural auto-associativa proposta por Kramer [127]. O método pode obter as funções de mapeamento global não-linear para reduzir a dimensionalidade. Contudo, esses métodos não produzem uma maneira para decidir o número de componentes principais fora o conhecido "tentativa e erro". Além disso, o número de componentes principais tem de ser especificado *a priori*. Muitas abordagens empregam o erro quadrático médio de reconstrução associado ao número de componentes principais extraídas. O paralelo entre o erro de reconstrução da PCA e da NLPCA pode fornecer uma estimativa da contribuição daquela componente para a descrição do sinal original. Algumas das limitações do método pode ser verificada em [133].

Cada um desses métodos possui suas peculiaridades, seja em tempo de treinamento, complexidade computacional, precisão e qualidade das componentes extraídas.

## A.2.1 Descorrelação não-linear

A correlação entre duas variáveis aleatórias está relacionada com a independência estatística, de maneira que duas variáveis independentes são sempre descorrelacionadas, porém o oposto não é verdade.

Estendendo o conceito de correlação linear pode-se definir a correlação não-linear entre duas variáveis  $y_1$  e  $y_2$  como  $E\{g(y_1)h(y_2)\} = 0$ , onde  $g(\cdot)$  e  $h(\cdot)$  são duas funções, e que pelo menos uma é não-linear. Isso significa que primeiro as variáveis são transformadas não-linearmente e então o cálculo usual é realizado. Ao atingir a descorrelação linear e não-linear entre as variáveis estudadas, uma aproximação suficiente o bastante das componentes independentes é alcançada [134].

Enquanto o objetivo da PCA é minimizar o erro quadrático médio de reconstrução do sinal, projetando as componentes em uma base ortonormal, a NLPCA pode ser definida de modo simples através da função objetivo a ser minimizada

$$J(\mathbf{W}) = E\{\|\mathbf{x} - \mathbf{W}^T g(\mathbf{W}\mathbf{x})\|^2\} \quad (\text{A.12})$$

onde  $\mathbf{W}$  é a base do sub-espço onde serão projetadas as entradas  $\mathbf{x}$ . No mínimo  $J$ , o produto  $\mathbf{W}\mathbf{x}$  fornece as componentes principais não-lineares [135, 136]. A única diferença em relação ao PCA, é a inserção da função não-linear  $g(\cdot)$  no mapeamento dos fatores  $\mathbf{W}\mathbf{x}$ .

Interessantemente, para dados branqueados, é mostrado [134] que o critério dado pela equação (A.12) tem uma relação exata com outras funções contraste como: a maximização/minimização da *kurtosis* e máxima verossimilhança; o que o aproxima das abordagens ICA.

Para dados branqueados, ao invés de vetores de observação  $\mathbf{x}$ , usa-se vetores projetados ortogonalmente  $\mathbf{v}$ . É assumido, também, que no branqueamento, a dimensão de  $\mathbf{v}$  pode ser reduzida à de  $\mathbf{s}$ .

Note que minimizar o critério (A.12) não significa um erro quadrático médio menor do que na PCA padrão. Ao invés disso, a virtude desse critério é que ele introduz estatística de ordem elevada através da não-linearidade  $g(\cdot)$  [136, 137].

### Branqueamento

Um vetor aleatório  $\mathbf{v}$  com média nula é dito branco se suas suas componentes são descorrelacionadas e têm variância unitária [7]. Isso implica que a matriz de covariância (e a de correlação) de  $\mathbf{v}$  são iguais a matriz identidade.

O branqueamento é algumas vezes chamado *sphering* e é mais forte do que a descorrelação, dado o seu compromisso com a média e a variância dos sinais transformados.

Um método de branqueamento bastante aplicado é através da decomposição SVD [138] da matriz de covariância do conjunto de dados. Nessa abordagem, considerando  $\mathbf{U}$  a matriz cujas colunas são os autovetores unitários da matriz de covariância de  $\mathbf{x}$ ,  $\mathbf{C}_x$ , e  $\mathbf{D}$  uma matriz diagonal formada pelos autovalores de  $\mathbf{C}_x$ , a matriz de transformação linear  $\mathbf{V}$  que branqueia os dados é obtida por

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{U}^T \quad (\text{A.13})$$

## A.3 Análise de Componentes Independentes

Hérault e Jutten parecem ser os primeiros (em 1983) a, informalmente, usarem o conceito de ICA na solução do problema de separação cega de fontes [139]. Anos mais tarde, Common apresentou uma formulação matemática da ICA e mostrou que cumulantes de ordem elevada podem ser usados para resolver problemas desse tipo [140]. Em paralelo, Cardoso e Souloumiac desenvolveram o algoritmo JADE [141], baseado na diagonalização aproximadamente conjunta. Enquanto essas duas abordagens usam estatística de segunda e quarta ordem, outras abordagens exploram apenas estatística de segunda ordem.

A exploração da informação matrizes de covariância datam do fim dos anos 1980, mas a teoria completa só é apresentada anos mais tarde [142]. O mesmo tipo de abordagem é desenvolvido de maneira independente por diferentes pesquisadores [142–144] que resultaram nos métodos AMUSE, SOBI e TDSEP, respectivamente. Em 1999 Müller et al. propõem uma versão modificada do JADE, que usa tanto estatística de segunda ordem quanto de quarta ordem. Mais recentemente Alberta et al. apresentaram uma extensão

do SOBI para estatística de quarta ordem, FOBIUM, que trata ICA sob determinadas restrições (mais componentes que observações). Os autores também propõem um método algébrico chamado ICAR [145], que utiliza a matriz de covariância de quarta ordem, chamada matriz de quadricovariância na solução do problema.

Enquanto os métodos anteriores identificam simultaneamente as componentes independentes, Delfosse e Loubaton [146] propuseram a extração de uma componente por vez, o que também é conhecido como técnica de deflação. Alguns anos depois Hyvarinen et al. propuseram o método FastICA [147] que iterativamente maximiza um funcional de quarta ordem. Enquanto a primeira versão usa a deflação, Hyvarinen et al. [7] propõem uma versão simultânea do FastICA cujo passo de ortonormalização conjunta é similar ao apresentado por Moreau [148].

Ao invés de explorar, explícita ou implicitamente, estatística de segunda e quarta ordens para resolver o problema ICA, algumas abordagens usam diretamente o conceito de independência. De fato, Lee et al. [149] apresentam uma abordagem que maximiza a informação baseada na parametrização das distribuições de probabilidade, sejam sub ou super gaussianas, para criar uma regra de aprendizado geral, que é otimizada pelo algoritmo do gradiente natural, proposta por Amari et al. [150]. Pham propõe o uso de estimativas não-paramétricas da verossimilhança ou da informação mútua.

A lista de métodos ICA citados pode ainda ser acrescida de alguns outros, o que mostra que a literatura sobre o assunto é vasta e crescente.

Na ICA, a identificação e a separação de múltiplos sinais é alcançada sem o conhecimento das características do canal onde o sinal trafega, o que de certa maneira associa diretamente com os problemas de separação cega de fontes. Nesse tipo de problema, almeja-se que as variáveis transformadas sejam estatisticamente independentes.

Dentre os diferenciais da análise de componentes independentes há algumas indeterminâncias intrínsecas ao modelo: a extração das componentes independentes não obedece a qualquer ordem de importância, assim como cada um dos sinais estimados podem estar escalados por fatores positivos ou negativos [140]. Felizmente essas limitações são insignificantes em muitas aplicações.

O que poderia guiar a PCA, a energia do sinal transformado (variância), não se aplica diretamente a ICA. Na minimização da dependência estatística, que é o principal interesse para qualquer algoritmo ICA, é comum o uso de cumulantes de ordem elevada (*kurtosis*), assim como medidas teóricas de informação (negentropia)

Contudo, é comum o uso de algum pré-processamento para a compactação dos dados, especialmente para aplicações com dimensão elevada dos dados. Nesse caso, a PCA pode ser aplicada como um passo anterior a ICA, assim como o branqueamento dos dados de entrada.



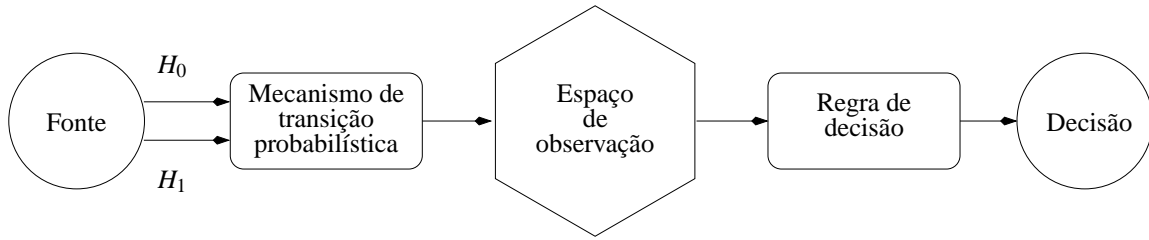


Figura A.2: Componentes de um problema clássico de detecção de sinais.

## A.4 Teste de Hipótese

Considerando-se o caso clássico de detecção e estimação de sinais com duas hipóteses, o problema de classificação pode ser resumido da seguinte maneira: os sinais fonte originais alterados pela estatística do canal devem ser detectados a partir das observações do processo. As regras de decisão, que formam o sistema classificador, são projetadas para maximizar a probabilidade de detecção e minimização do falso alarme [94, 95].

No caso da decisão binária, cada vez que uma observação é efetuada 4 situações podem ocorrer: decisão pela hipótese 1, sendo a hipótese 2 verdadeira; decisão pela hipótese 2, sendo a hipótese 1 verdadeira; decisão pela hipótese 1, sendo a hipótese 1 verdadeira; decisão pela hipótese 2, sendo a hipótese 2 verdadeira. As duas primeiras alternativas são consideradas erros de decisão e as duas últimas definidas como classificações corretas. Cada uma das hipóteses é associada a uma saída da fonte, que é mapeada em uma região do espaço de observação.

O mecanismo de transição probabilística (a estatística do canal) gera pontos no espaço de observação de acordo com as densidades de probabilidade condicionais anteriores. Quando essas probabilidades são conhecidas ou podem ser estimadas de alguma forma, o projeto do sistema de classificação pode ser simplificado e critérios probabilísticos usados para a escolha da regra de decisão [94, 95]. Quando essa informação não está disponível, alternativas são aplicadas.

### A.4.1 Classificador Linear de Fisher

O algoritmo linear de Fisher enquadra-se na categoria de algoritmos supervisionados. O potencial da abordagem vem da resposta à pergunta: como usar a informação que rotula os dados para encontrar projeções que otimizem essa informação. Com esse propósito o algoritmo propõe maximizar a seguinte função objetivo:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (\text{A.14})$$

onde  $\mathbf{S}_B$  é a matriz de espalhamento entre classes e  $\mathbf{S}_W$  é a matriz de espalhamento intra-classes [8]. Devido ao fato de as matrizes de espalhamento serem proporcionais às

matrizes de covariância, pode-se definir o funcional  $J$  por meio destas - a constante de proporcionalidade não teria efeito na solução final. As definições das matrizes de espalhamento são:

$$\mathbf{S}_B = \sum_c (\mu_c - \mu)(\mu_c - \mu)^T \quad (\text{A.15})$$

$$\mathbf{S}_W = \sum_c \sum_{i \in c} (\mathbf{x}_i - \mu_c)(\mathbf{x}_i - \mu_c)^T \quad (\text{A.16})$$

onde  $\mu$  é o vetor média dos dados,  $\mu_c$  é o vetor média dos dados da classe  $c$  e  $\mathbf{x}_i$  é o  $i$ -ésimo vetor pertencente a classe  $c$ .

A interpretação da solução da função objetivo é vista quando o vetor média das classes estão distantes, i.e., a variância entre as classes é grande e, ao mesmo tempo, a variância intraclasse é pequena, ou seja, os dados de cada classe são bem comportados.

Uma propriedade importante da função objetivo  $J$  é a sua invariância com respeito ao escalonamento do vetor  $\mathbf{w} \rightarrow \alpha \mathbf{w}$  [8].

## A.4.2 Classificador Neural

A construção de máquinas capazes de aprender através da experiência tem sido objeto de pesquisas por vários anos. O aspecto técnico tem recebido um enorme impulso com o avanço da computação. Esse avanço tem demonstrado que máquinas podem apresentar um nível significativo de habilidade de aprendizagem, embora os limites desta aprendizagem não estejam totalmente definidos. Já a disponibilidade de sistemas de aprendizado eficientes é de importância estratégica em diferentes campos de aplicação, principalmente onde nenhum modelo matemático do problema seja conhecido. Dentre eles, a classificação de padrões em um ambiente de estatística desconhecida.

### Definição e Características

Uma Rede Neural Artificial (RNA) pode ser vista como um modelo matemático composto de elementos computacionais lineares ou não, chamados neurônios, trabalhando em paralelo e massivamente conectados por ligações caracterizadas por diferentes pesos denominadas sinapses [9]. Os modelos neurais são principalmente especificados pela topologia da rede, características dos neurônios e regras de treinamento e aprendizado.

Segundo Haykin [9], as redes neurais artificiais podem fornecer as seguintes propriedades úteis:

- Mapeamento de Entrada-Saída.
- Adaptabilidade.
- Resposta a Evidências.
- Tolerância a Falhas.

## Aprendizado

Existem tarefas, problemas, de complicada solução por uma abordagem matemática tradicional em que os modelos são complexos ou a modelagem torna-se difícil devido a aspectos de não-linearidade ou por o processo incorporar estatísticas presentes, porém desconhecidas.

Algumas abordagens para a solução desse tipo de problema podem fazer uso de uma máquina de aprendizado. Os pares entrada/saída tipicamente refletem uma relação funcional, embora isso nem sempre seja o caso para exemplos em que as saídas estão corrompidas por ruído. Quando uma função básica de mapeamento entrada/saída existe, ela é referida como função alvo. A estimativa da função alvo que é aprendida pelo algoritmo é conhecida como a solução do problema de aprendizagem. No caso de classificação de padrões essa função é algumas vezes referenciada função decisão. A solução é escolhida de um conjunto de funções candidatas que realizam o mapeamento do espaço de entrada para espaço de saída. Os algoritmos que utilizam os dados de treinamento e selecionam uma função do espaço de funções são referenciados como algoritmos de aprendizagem.

No caso particular em que os exemplos são pares rotulados de entrada/saída, isso é chamado aprendizado supervisionado. Os exemplos de entrada/saída são funcionalmente chamados de dados de treinamento [9]. A Figura A.3 exemplifica o caso de um ambiente de aprendizado supervisionado, o bloco denominado “professor” é responsável por informar a resposta desejada para o respectivo dado de entrada.

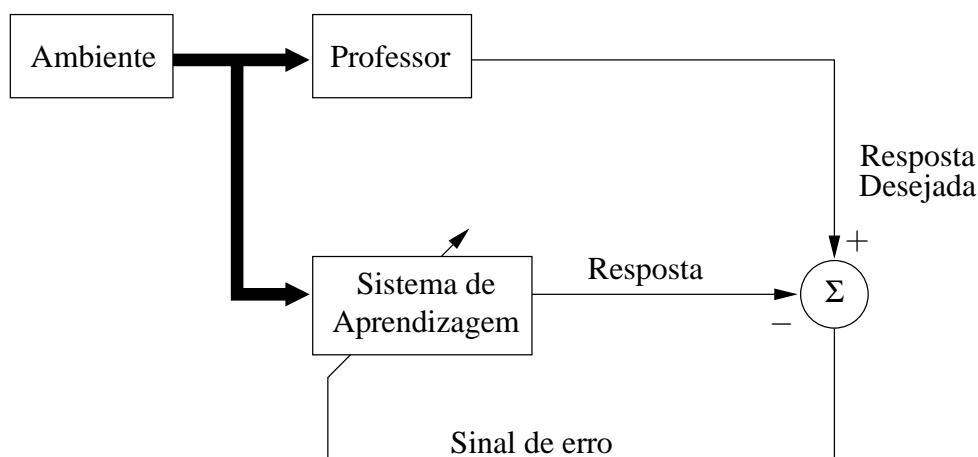


Figura A.3: Aprendizado supervisionado.

Variações nos modelos de aprendizado também são encontradas quanto ao modo como os dados de treinamento são gerados e apresentados ao algoritmo.

## Validação

A validação de um algoritmo de aprendizagem é a etapa de constatação do conhecimento adquirido no treinamento, geralmente realizada através de um teste de gene-

realização. Nesse processo, um conjunto de dados, denominado conjunto de validação, é apresentado ao algoritmo de aprendizagem que estima a saída correspondente, realizando assim uma inferência sobre dados não observados anteriormente. Valores de saída próximos da saída desejada indicam adequada generalização, enquanto valores de saída divergentes indicam má generalização.

Pode-se usar também uma subdivisão do conjunto de treinamento, criando uma espécie de conjunto de validação, para verificar a eficiência da rede quanto à sua capacidade de generalização durante o treinamento, que seria empregado como critério de parada do processo.

### **Classificador de Padrões**

Os seres humanos são bons no reconhecimento de padrões. Recebemos dados do mundo à nossa volta através dos nossos sentidos e somos capazes de reconhecer a fonte dos dados. Frequentemente, somos capazes de fazer isso quase que imediatamente e praticamente sem esforço. Podemos, por exemplo, reconhecer um rosto familiar de uma pessoa, muito embora esta pessoa tenha envelhecido desde o último encontro entre ambos, podemos identificar uma pessoa familiar pela sua voz ao telefone, apesar de uma comunicação ruim, dentre outros exemplos. Os humanos realizam o reconhecimento de padrões através de um processo de aprendizagem; e assim acontece com as redes neurais artificiais [9].

O reconhecimento de padrões é formalmente definido como o processo pelo qual um padrão/sinal recebido é atribuído a uma classe dentre um número predeterminado de classes (categorias). Uma rede neural supervisionada realiza o reconhecimento de padrões passando inicialmente por uma seção de treinamento, durante a qual se apresenta à rede um conjunto de padrões de entrada junto com a categoria a qual cada padrão particular pertence. Mais tarde, apresenta-se à rede um novo padrão que não foi visto antes, mas que pertence a mesma população de padrões utilizada para treiná-la. A rede é capaz de identificar a classe daquele padrão particular por causa da informação que ela extraiu dos dados de treinamento. O reconhecimento de padrões realizado por uma rede neural é de natureza estatística, com os padrões sendo representados por pontos em um espaço de decisão multidimensional. O espaço de decisão é dividido em regiões, cada uma das quais associada a uma classe. As fronteiras de decisão são determinadas pelo processo de treinamento. A construção dessas fronteiras é tornada estatística pela variabilidade inerente que existe dentro das classes e entre as classes [9].

Para que o desempenho do classificador para novos padrões de entrada (diferentes dos usados na fase de treinamento) possa ser avaliado, o classificador passa por uma fase de teste. Durante essa fase, deve ser observado o número de acertos e erros realizados pelo classificador. Uma estimativa do potencial do classificador é obtida com a apresentação das porcentagens de padrões corretamente e incorretamente classificados.

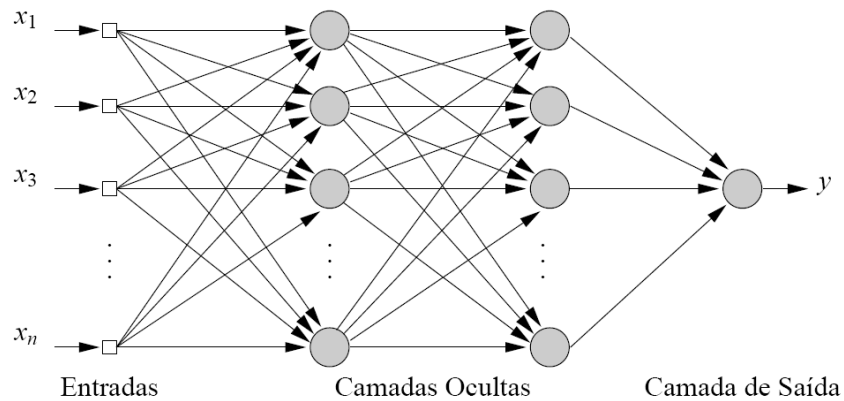


Figura A.4: Rede Percéptron de Múltiplas Camadas.

### Rede Perceptrons de Múltiplas Camadas

As redes Perceptrons de Múltiplas Camadas, do inglês *Multilayer Perceptrons* (MLP), surgiram como alternativa para a solução de problemas de complexidade não-linear. A solução dessa classe de problemas passa pelo uso de redes com uma ou mais camadas intermediárias alimentadas adiante (*feedforward*), onde cada camada é composta por um conjunto de neurônios em que nas camadas intermediárias (ocultas) é comum a utilização de funções de ativação não-linear. A Figura A.4 apresenta a estrutura de uma MLP típica.

Para treinar as redes com mais de uma camada, as técnicas mais comuns são aquelas baseadas no gradiente descendente. O principal algoritmo utilizado para o treinamento dessas redes é o algoritmo supervisionado da retropropagação (*backpropagation*) [9] e algumas de suas variações. Esses algoritmos são baseados na regra de aprendizagem por correção de erro. Para que esse método possa ser utilizado, a função de ativação precisa ser contínua e diferenciável. A função de ativação deve possibilitar a inferência dos erros cometidos pela rede para as camadas interiores com a maior precisão possível.