



UNIVERSIDADE FEDERAL DA BAHIA – UFBA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA – PPGEEL

MARTON SANDES DOS SANTOS

Dissertação de Mestrado

**CLASSIFICADORES NEURAIS DE TREINAMENTO
RÁPIDO APLICADOS NA IDENTIFICAÇÃO *ONLINE*
DE EVENTOS NO DETECTOR ATLAS**

Salvador

2018

MARTON SANDES DOS SANTOS

**CLASSIFICADORES NEURAIS DE TREINAMENTO RÁPIDO
APLICADOS NA IDENTIFICAÇÃO *ONLINE* DE EVENTOS NO
DETECTOR ATLAS**

Dissertação de Mestrado

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal da Bahia como um dos requisitos para obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Dr. Paulo César M. de A. Farias – UFBA

Coorientador: Dr. Eduardo F. de Simas Filho – UFBA

Salvador

2018

Marton Sandes dos Santos

Classificadores Neurais de Treinamento Rápido Aplicados na Identificação *Online* de Eventos no Detector ATLAS/ Marton Sandes dos Santos. – Salvador, 2018-
90 p. : il. (algumas color.) ; 30 cm.

Orientador: Dr. Paulo César M. de A. Farias – UFBA

Coorientador: Dr. Eduardo F. de Simas Filho – UFBA

Dissertação de Mestrado – Universidade Federal da Bahia – UFBA, 2018.

1. Redes Neurais Artificiais. 2. Detector ATLAS. 3. *Neural Ringer*. 4. ELM - Máquinas de Aprendizado Extremo. 5. ESN - Redes com Estados de Eco. 6. Processamento Estatístico de Sinais.

I. Dr. Paulo C. A. M de Farias II.Dr. Eduardo F. de Simas Filho.

III. UFBA – Universidade Federal da Bahia. IV. DEE – Departamento de Engenharia Elétrica. V. PPGEE – Programa de Pós-Graduação em Engenharia Elétrica

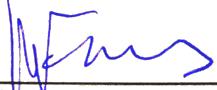
Marton Sandes dos Santos

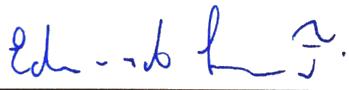
Classificadores Neurais de Treinamento Rápido Aplicados na Identificação *Online* de Eventos no Detector ATLAS

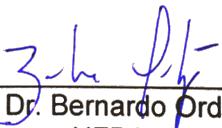
Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal da Bahia como um dos requisitos para obtenção do grau de Mestre em Engenharia Elétrica.

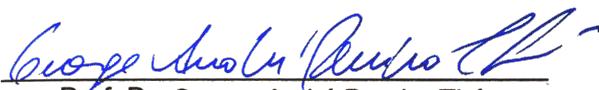
Trabalho aprovado. Salvador, 30 de agosto de 2018:

Banca Examinadora


Prof. Dr. Paulo C. Machado de A. Farias
Orientador - UFBA


Prof. Dr. Eduardo Furtado de Simas Filho
Coorientador - UFBA


Prof. Dr. Bernardo Ordoñez
UFBA


Prof. Dr. George André Pereira Thé
UFC

Salvador

2018

*Dedico esse trabalho aos meus pais,
que são meus grandes incentivadores, e orientadores.*

Agradecimentos

Agradeço ao Senhor, pelo privilégio de poder conhecer e compreender, em parte, os detalhes e beleza de sua fantástica criação, tendo a ciência como uma das ferramentas. Os agradecimentos são uma parte importante e nada fácil de fazer, pois corro o risco de esquecer de alguém.

Agradeço aos meus pais pelo apoio que sempre tive, estimulando a ser curioso para entender e compreender o que tinha curiosidade. Pois é, deu certo. Me tornei Engenheiro, e continuo curioso. Agradeço à minha esposa pela paciência e apoio, nas diversas noites em que continuei madrugada a dentro, pois o não podia perder o raciocínio que tinha iniciado.

Agradeço aos meus orientadores, Eduardo Simas e Paulo César por confiarem em mim, e permitirem que eu participasse da pesquisa desenvolvida no PPGEE-UFBA em parceria com a Colaboração ATLAS Brasil.

Ao professor Eduardo Simas, por ter sido meu orientador desde o trabalho de conclusão de curso. À Colaboração ATLAS, e o professor Seixas, coordenador da pesquisa no Brasil, pelo suporte e auxílio durante o trabalho, com críticas e sugestões, às minhas apresentações feitas à Colaboração ATLAS Brasil. A Werner, sempre disponível respondendo aos meus *emails* sobre detalhes das bases de dados e configurações das redes, muito obrigado.

A Edmar, por ter me auxiliado a iniciar a minha pesquisa, dando continuidade a parte do seu trabalho. A Tiago, pelas orientações em programação avançada em MATLAB. Minha prima Anna e minha colega, professora Perpétua, por dicas e orientações relativas à organização, língua portuguesa e trabalho acadêmico.

Aos amigos não citados, que apoiaram, intercederam e torceram, muito obrigado. Mais um degrau alcançado, sem previsão de limites.

*“Não há nada em toda a criação
que esteja escondido aos olhos de Deus;
pelo contrário tudo está patente
e a descoberto perante aquele
a quem temos de prestar contas”.
(Hebreus 4:13)*

Resumo

O ATLAS é o maior detector do *Large Hadron Collider* (LHC), maior acelerador de partículas já construído e está em operação desde 2008. Sua estrutura é altamente segmentada, sendo composta por 100.000.000 de sensores dispostos num formato cilíndrico para captar os sinais provenientes das colisões próton-próton do LHC. Entre os principais objetivos dos experimentos de física de partículas pode-se mencionar a validação de modelos teóricos e a proposição de novas teorias relacionadas aos componentes fundamentais da matéria, e suas formas de interação. Devido à sua estrutura finamente segmentada, e à natureza dos fenômenos estudados, é produzido um volume de informação considerável durante as colisões, que ocorrem a uma taxa de 40 MHz produzindo no ATLAS até 68 TB/s de informação. Assim, torna-se proibitivo o armazenamento de toda a informação produzida, para posterior processamento. Dessa forma, é necessário um sistema de seleção *online* (*Trigger*), que realize a separação dos eventos que contenham informação a respeito da física de interesse do ruído de fundo (eventos não relevantes) produzido. Neste contexto, o discriminador *online Neural Ringer* utiliza redes neurais do tipo *perceptron* de múltiplas camadas (MLP) para a separação da informação de interesse (prováveis assinaturas de elétrons) do ruído de fundo (composto em sua maioria por jatos hadrônicos). Como entradas para as redes classificadoras, o *Neural Ringer* utiliza a informação do perfil de deposição de energia do evento registrado pelo detector. Entretanto, para a obtenção do conjunto de redes neurais que compõem o *Neural Ringer*, é necessário realizar um número elevado de inicializações do processo de treinamento, o que demanda muito tempo de processamento. Neste trabalho são investigadas modificações no sistema de detecção *online* de elétrons do ATLAS (*Neural Ringer*) visando diminuir o tempo de treinamento de redes neurais artificiais e mantendo a eficiência de discriminação da física de interesse. Para isso serão utilizadas técnicas computacionalmente eficientes para treinamento dos classificadores, como as Máquinas de Aprendizado Extremo (*Extreme Learning Machine* - ELM) e as Redes com estados de Eco (*Echo State Network* - ESN). Utilizando duas bases de dados e duas técnicas de reamostragem utilizadas pela Colaboração ATLAS os classificadores foram treinados e avaliados por meio da curva ROC e índice SP na determinação da rede mais eficiente para posterior análise do tempo de treinamento. Ainda foi realizada uma análise estatística para comparação das técnicas propostas em relação ao discriminador padrão. Os resultados mostraram que as técnicas propostas produzem desempenho de classificação equivalente ao classificador em uso, e em tempo de treinamento inferior, sugerindo que as técnicas podem vir a ser utilizadas como alternativas ao classificador utilizado no detector ATLAS.

Palavras-chave: Redes Neurais. Reconhecimento de Padrões. Processamento Estatístico de Sinais. ELM. ESN. Detector ATLAS, *Neural Ringer*.

Abstract

The ATLAS is the largest detector of the Large Hadron Collider (LHC), the largest particle accelerator ever built and in operation since 2008. Its structure is highly segmented, having 100,000,000 sensors arranged in a cylindrical shape to capture proton-proton collisions occurring in the LHC. Among the main objectives of the experiments in particle physics are the validation of theoretical models and the proposition of new theories related to the fundamental components of matter and their forms of interaction. Due to its finely segmented structure and the nature of the studied phenomena, a considerable amount of information is produced during the collisions, occurring at a rate of 40 MHz producing in the ATLAS up to 68 TB/s of information. Thus, it becomes prohibitive the storage of all the information produced, for further processing. Therefore, an online selection system (Trigger) is required, which performs the separation of events that contain information about the background physics (non-relevant events) produced. In this context, the online discriminator Neural Ringer uses multilayer perceptron neural networks (MLP) to separate information of interest (likely electron signatures) from background noise (composed mainly of hadronic jets). As inputs to the classifier networks, the Neural Ringer uses the event energy deposition profile information recorded by the detector. However, to obtain the set of neural networks that make up the Neural Ringer, it is necessary to perform a large number of initializations of the training process, which requires processing time. In this work, modifications will be investigated in the ATLAS (Neural Ringer) electron detection system in order to reduce the training time of artificial neural networks and to maintain the discriminant efficiency of the physics of interest. To do this, we will use computationally efficient techniques to train classifiers, such as Extreme Learning Machines (ELM) and Echo State Networks (ESN). Using two databases and two resampling techniques used by the ATLAS Collaboration, the classifiers were trained and evaluated through the ROC curve and SP index in determining the most efficient network for later analysis of training time. A statistical analysis was also performed to compare the proposed techniques in relation to the standard discriminator. The results showed that the proposed techniques produce classification performance equivalent to the classifier in use, and at lower training time, suggesting that the techniques can be used as alternatives to the classifier used in the ATLAS detector.

Keywords: Neural Network. Pattern Recognition. Statistical Signal Processing. ELM. ESN. ATLAS Detector, Neural Ringer.

Lista de ilustrações

Figura 1 – Ilustração da localização do LHC e seus detectores.	1
Figura 2 – Agenda de atividades para o LHC em termos de nível de energia e luminosidade utilizada nas colisões.	3
Figura 3 – Ilustração do empilhamento de eventos num sensor do calorímetro do ATLAS. Os eventos em preto (curva com centro em 0) e em vermelho (curva com centro em ≈ 60).	4
Figura 4 – Exemplos de deposição de energia normalizada no canal de <i>leptons</i> (a) e em função da pseudorapidez (η) (b) com $\sqrt{s} = 500$ GeV.	5
Figura 5 – Distribuição de massa para energia $\sqrt{s} = 8$ TeV, sobreposta aos dados de medição do ano de 2012, e previsão da massa para o Bóson de Higgs segundo o modelo padrão.	6
Figura 6 – Representação do Modelo Padrão e suas partículas.	13
Figura 7 – Diagrama simplificado sobre o Modelo Padrão, contendo informações sobre as partículas básicas e as interações fundamentais.	13
Figura 8 – Diagrama simplificado dos três grupos de partículas fundamentais do Modelo Padrão.	14
Figura 9 – Representação da estrutura do LHC.	15
Figura 10 – Diagrama da estrutura do complexo acelerador do LHC.	16
Figura 11 – Diagrama do detector ATLAS, com destaque para seus subdetectores. EMB: <i>LAr Electromagnetic Calorimeter</i> ; HEC: <i>Tile Calorimeter</i> ; ID: composto por <i>Semiconductor tracker</i> , <i>Transition radiation tracker</i> e <i>Pixel detector</i> e o EMEC: <i>LAr hadronic end-cap and forward calorimeters</i> , são as tampas que fecham o detector ATLAS.	18
Figura 12 – Representação da interação das partículas no interior do detector ATLAS.	19
Figura 13 – Distribuição elétrons e jatos em função da energia total do evento.	20
Figura 14 – Representação do sistema de coordenadas do detector. O centro do sistema de coordenadas refere-se ao local da colisão, e os ângulos as possíveis trajetórias das partículas resultantes.	20
Figura 15 – Diagrama ilustrando a segmentação do detector e seus diversos Calorímetros. Barril: EMB (<i>Tile Barrel</i> e <i>LAr Electromagnetic barrel</i>); Barril Extendido: <i>Tile extended barrel</i> ; EMEC (<i>LAr forward</i> e <i>LAr hadronic end-cap</i>) e o Calorímetro de Telhas (<i>Tile Calorimeter</i>)	21
Figura 16 – Diagrama de um corte ilustrando a segmentação do barril (esquerda) e barril estendido (direita).	22
Figura 17 – Diagrama ilustrando a estrutura em acordeão do ECAL.	23
Figura 18 – Diagrama 18a de um corte transversal do detector e suas camadas e 18b representação de uma possível trajetória de um evento no interior do detector.	24
Figura 19 – Representação dos anéis do calorímetro nas RoIs.	24

Figura 20 – Exemplos típicos de assinaturas para elétrons e jatos de amostras experimentais, obtidas do calorímetro formatado em anéis.	25
Figura 21 – Esquema do <i>trigger online</i> do ATLAS após última atualização no final de 2014.	26
Figura 22 – Diagrama do modelo matemático de um neurônio artificial, o <i>Perceptron</i>	29
Figura 23 – Ilustração de um modelo de neurônio biológico.	30
Figura 24 – Diagrama de uma rede <i>feedforward</i> com o sentido de fluxo da informação.	31
Figura 25 – Exemplos de redes recorrentes	32
Figura 26 – Rede neural, na separação de classes.	33
Figura 27 – Rede <i>feedforward</i> - totalmente conectada - pesos w_n e v_m representam vetores de pesos, para simplificar o diagrama.	34
Figura 28 – Rede neural, com duas camadas ocultas, representação do algoritmo <i>backpropagation</i> - Retropropagação em representação simplificada sem pesos sinápticos.	35
Figura 29 – Exemplos de funções de ativação.	36
Figura 30 – Diagrama de uma ELM.	38
Figura 31 – Diagrama genérico de uma rede ESN, indicando os possíveis laços de realimentação	41
Figura 32 – Diagrama de uma rede ESN que possui estados de eco.	42
Figura 33 – Representação do processo de agrupamento e sorteio dos subgrupos de treinamento.	48
Figura 34 – Exemplos de curvas ROC de dois classificadores.	50
Figura 35 – Exemplos de <i>boxplot</i> para quatro possíveis distribuições.	51
Figura 36 – Representação da metodologia para avaliação dos resultados dos classificadores avaliados, ELM, ESN e MLP.	52
Figura 37 – Distribuição de assinaturas de elétrons e jatos para a base de dados experimentais.	54
Figura 38 – Distribuição de assinaturas de elétrons e jatos para a base de dados simulados.	55
Figura 39 – Número de assinaturas para elétron e jato por região na base simulada.	55
Figura 40 – <i>Boxplot</i> das melhores redes para ELM, ESN e MLP na base experimental.	59
Figura 41 – ROCs das melhores redes para ELM, ESN e MLP, utilizando as técnicas de reamostragem <i>K-fold</i> (a) e <i>Jackknife</i> (b).	60
Figura 42 – <i>Boxplot</i> ELM × ESN × MLP para cada região. Nesta, os resultados foram obtidos utilizando a técnica de reamostragem <i>Jackknife</i> . Sobre a indicação de cada região está a <i>boxplot</i> para a ELM, à esquerda a <i>boxplot</i> do MLP e à direita a <i>boxplot</i> da ESN.	63
Figura 43 – <i>Boxplot</i> ELM × ESN × MLP para cada região. Nesta, os resultados foram obtidos utilizando a técnica de reamostragem <i>K-fold</i> . Sobre a indicação de cada região está a <i>boxplot</i> para a ELM, à esquerda a <i>boxplot</i> do MLP e à direita a <i>boxplot</i> da ESN.	63
Figura 44 – ROCs para os melhores índices SP, regiões (0,0), (1,0), (2,0), (3,0) e (4,0).	64
Figura 45 – ROCs para os melhores índices SP, regiões (0,1), (1,1), (2,1), (3,1) e (4,1).	64
Figura 46 – ROCs para os melhores índices SP, regiões (0,2), (1,2), (2,2), (3,2) e (4,2).	65
Figura 47 – ROCs para os melhores índices SP, regiões (0,3), (1,3), (2,3), (3,3) e (4,3).	65

Lista de tabelas

Tabela 1 – Regiões de $ \eta $ onde estão localizados os subdetectores do ATLAS.	22
Tabela 2 – Número de anéis por camada. PS - <i>Presampler</i> ; EMB1 – EMB3: Camadas Eletromagnéticas; HEC0 – HEC2: Camadas Hadrônicas.	24
Tabela 3 – Resultados possíveis de classificação.	49
Tabela 4 – Segmentação base de dados simulados MC2015.	54
Tabela 5 – Número de assinaturas para cada corte de energia no conjunto de dados. . .	56
Tabela 6 – Número de neurônios utilizados na camada oculta do MLP, ELM e do reservatório de dinâmicas da ESN, na base experimental.	58
Tabela 7 – Dados para os melhores índices SP nas técnicas, MLP, ELM e ESN para a base Experimental.	60
Tabela 8 – Tempo de treinamento em segundos, para os melhores resultados, ELM × ESN × MLP, na base experimental.	61
Tabela 9 – Número de neurônios utilizado em cada uma das técnicas organizados por região.	62
Tabela 10 – Índice SP para os melhores resultados da ELM e ESN comparados com os valores obtidos com o MLP, em cada região. Cada coluna representa uma posição no interior do detector e em cada linha pode-se observar a melhora nos índices com a elevação da energia (Δ_{E_T} , $\Delta_{ \eta }$).	66
Tabela 11 – Probabilidade de detecção (PD) para os melhores resultados, ELM e ESN comparados com os valores obtidos com o MLP, em cada região. Cada coluna representa uma posição no interior do detector e em cada linha pode-se observar a melhora nos índices com a elevação da energia (Δ_{E_T} , $\Delta_{ \eta }$).	67
Tabela 12 – Taxa de falso alarme (FR) para os melhores resultados, ELM e ESN comparados com os valores obtidos com o MLP, em cada região. Cada coluna representa uma posição no interior do detector e em cada linha pode-se observar a melhora nos índices com a elevação da energia (Δ_{E_T} , $\Delta_{ \eta }$).	68
Tabela 13 – Tempo de treinamento em segundos, para os melhores resultados, ELM × ESN × MLP, em cada região.	69
Tabela 14 – Resultados do teste de Student, ELM × MLP, para cada uma das regiões (E_T , $ \eta $) da base.	70
Tabela 15 – Resultados do teste de Student, ESN × MLP, para cada uma das regiões (E_T , $ \eta $) da base.	70
Tabela 16 – Resultados do teste de Student, ELM × ESN, para cada uma das regiões (E_T , $ \eta $) da base.	71
Tabela 17 – Resumo dos melhores resultados obtidos com os classificadores, MLP, ELM e ESN, para as duas bases utilizadas, tendo como técnica de reamostragem o <i>Jackknife</i>	73
Tabela 18 – Segmentação base de dados utilizada.	90

Tabela 19 – Índices SP para três métodos de produzir números pseudo-aleatórios para a ELM.	90
---	----

Lista de símbolos

γ	Fóttons
η	Pseudo-rapidez
β	Matriz de pesos da camada oculta para a saída de uma rede ELM
Φ	Função de ativação
ϕ	Ângulo azimutal
Ψ	Função de Onda
\mathbf{H}	Matriz de pesos aleatórios da camada oculta
Σ	Matriz singular
λ	Autovalor
σ	Valor singular
\mathcal{L}	Luminosidade
\mathbf{W}	Matriz de pesos do Reservatório de Dinâmicas
\mathbf{W}^{back}	Matriz de pesos da Saída para Reservatório de Dinâmicas
\mathbf{W}^{in}	Matriz de pesos para a camada de entrada
\mathbf{W}^{in}	Matriz de pesos da camada de entrada para o reservatório de dinâmicas
\mathbf{W}^{inout}	Matriz de pesos da entrada para a saída
\mathbf{W}^{out}	Matriz de pesos para a camada de saída
\mathbf{W}^{outout}	Matriz de pesos de realimentação da camada de saída para a camada de saída
W^+	Bóson mediador da interação fraca
W^-	Bóson mediador da interação fraca
Z^0	Bóson mediador da interação fraca
Z'	Z prime

Lista de abreviaturas e siglas

ALICE	<i>A Large Ion Collider Experiment</i>
ATLAS	<i>A toroidal apparatus</i>
CERN	<i>Conseil Européen pour la Recherche Nucléaire</i>
CMS	<i>Compact Muon Solenoid</i>
E_T	Energia Transversa
ECAL	Calorímetro Eletrmagnético
ELM	<i>Extreme Learning Machine</i>
ESN	<i>Echo State Network</i>
FR	<i>Fake Rate</i> - Taxa de Falso alarme
H	Bóson de Higgs
HCAL	Calorímetro Hadrônicos
HEP	<i>High Energy Physics</i>
HS	<i>Hiden Sectors</i>
IA	Inteligência Artificial
ID	Detector Interno
LHC	<i>Large Hadron Colider</i>
LHCb	<i>Large Hadron Colider beauty</i>
LHCf	<i>Large Hadron Colider forward</i>
LSM	<i>Liquid State Machine</i>
MLP	<i>Multilayer Perceptron</i>
MoEDAL	<i>Monopole and Exotics Detector at the LHC</i>
NNA	<i>Neural Network Artificial</i>
PD	Probabilidade de Detecção
PS	<i>Pre sampler</i>
PSB	<i>Pre sampler booster</i>

RD	Reservatório de Dinâmicas
RNN	<i>Recurrent Neural Network</i>
NR	<i>Neural Ringer</i>
ROC	<i>Receiver Operation Curve</i>
RPROP	<i>Resilient Backpropagation</i>
SLFN	<i>Single Layer Feedforward Networks</i> - Rede em camada única em avanço
SP	Índice Soma Produto
SR	<i>Spectral Radius</i>
<i>T2Calo</i>	Discriminador do detector ATLAS
<i>TileCal</i>	Calorímetro de Telhas
TOTEM	<i>Total Elastic and diffractive cross section Measurement</i>

Sumário

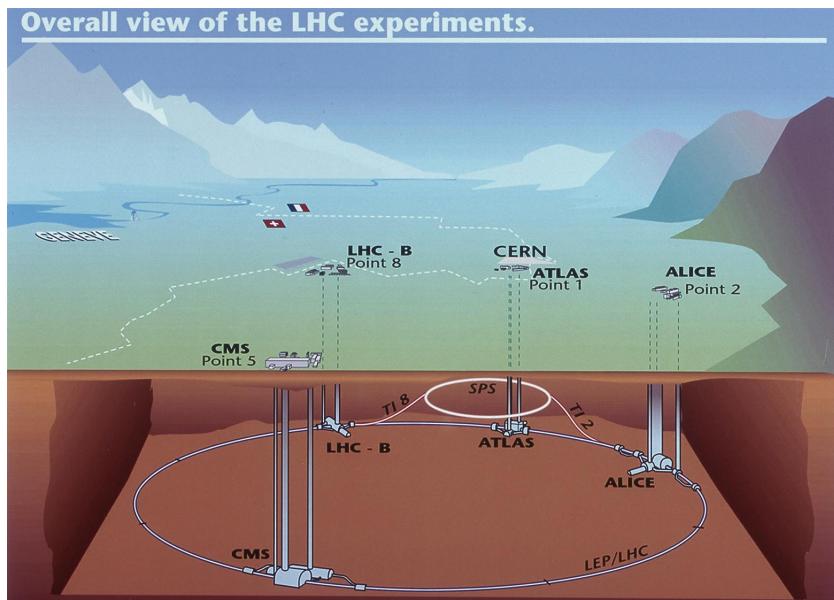
1	INTRODUÇÃO	1
1.1	Objetivo Geral	2
1.2	Objetivos Específicos	2
1.3	Justificativa	2
1.4	Organização do Documento	7
2	FÍSICA E O LHC	9
2.1	Física de Altas Energias e o LHC	9
2.1.1	Breve Histórico do Modelo Atômico	9
2.1.2	Modelo Padrão	12
2.2	O LHC	14
2.2.1	O Detector ATLAS	18
2.2.1.1	Sistema de Construção dos Anéis	23
2.2.1.2	Sistema de Seleção ou Filtragem <i>Online</i>	25
3	FUNDAMENTAÇÃO TEÓRICA	29
3.1	Redes Neurais Artificiais - RNA	29
3.1.1	Definições	29
3.1.2	Estruturas	30
3.1.2.1	Redes em Avanço	30
3.1.2.2	Redes Recorrentes	31
3.1.3	Características	32
3.1.4	Exemplos de Aplicações	36
3.2	Máquinas de Aprendizado Extremo - ELM	37
3.2.1	Exemplos de Aplicações	40
3.3	Redes com Estado de Eco - ESN	40
3.3.1	Propriedades dos Estados de Eco	43
3.3.2	Inicialização dos Pesos e Treinamento	43
3.3.3	Exemplos de Aplicações	43
4	METODOLOGIA	47
4.1	Técnicas de Reamostragem	47
4.1.1	<i>k-fold</i>	47
4.1.2	<i>Jackknife</i>	47
4.2	Determinação do número de neurônios	48
4.3	Métodos de avaliação dos resultados	49
4.3.1	Curva ROC	49
4.3.2	Índice SP	50
4.3.3	<i>Boxplot</i>	50

4.3.4	Teste de Student	51
4.4	Bases de Dados	52
4.4.1	Dados Experimentais	53
4.4.2	Dados Simulados	54
5	RESULTADOS	57
5.1	Bases de dados	57
5.1.1	Dados Experimentais	58
5.1.1.1	Informações	58
5.1.1.2	Índice SP	59
5.1.1.3	Tempo de Treinamento	61
5.1.2	Análise Estatística	61
5.1.3	Dados Simulados	62
5.1.3.1	Informações	62
5.1.3.2	Índice SP	62
5.1.3.3	Tempos de Treinamento	68
5.1.4	Análise Estatística	69
5.2	Análise dos Resultados	71
5.2.1	Resumo	73
6	CONCLUSÕES	75
6.1	Trabalhos Futuros	76
	REFERÊNCIAS	77
	APÊNDICES	85
	APÊNDICE A – TRABALHOS PUBLICADOS	87
A.1	Artigos Publicados em Anais de Congressos e Simpósios	87
A.2	Resumo Publicado em Encontro	88
	APÊNDICE B – ANÁLISE DE DESEMPENHO DA ELM	89
B.1	Sensibilidade dos pesos à distribuição de probabilidade utilizada.	89

1 Introdução

A compreensão à respeito da constituição fundamental da matéria obteve evolução significativa nos últimos anos devido à comprovações resultantes de experimentos de física de altas energias. O Grande Colisor de Hadrons (*Large Hadron Collider - LHC*) (EVANS; BRYANT, 2008) é o maior acelerador de partículas em operação atualmente e está situado no Centro Europeu para Pesquisa Nuclear (CERN)(CERN-COLLABORATION, 2016). O LHC (ver Figura 1) foi construído com o objetivo de analisar a estrutura fundamental da matéria, investigar as propriedades das partículas fundamentais propostas pelo Modelo Padrão (*Standard Model*)(MOREIRA, 2009; PIMENTA et al., 2013) e também buscar por fenômenos desconhecidos.

Figura 1 – Ilustração da localização do LHC e seus detectores.



Fonte: (CERN Collaborartion, 1999)

Para tanto, o LHC, conta com alguns experimentos: ATLAS (*A Toroidal LHC ApparatuS*), ALICE (*A Large Ion Collider Experiment*), CMS (*Compact Muon Solenoid*) e LHCb (*Large Hadron Collider beauty*), conforme mostrado na Figura 1. O túnel do acelerador tem cerca de 27 km de comprimento e as colisões podem ocorrer numa taxa de até 40×10^6 vezes por segundo (EVANS; BRYANT, 2008).

O ATLAS é um detector de propósito geral, ou seja, possui capacidade de detectar diversos tipos de partículas. Para isso, foi projetado no formato cilíndrico e conta com os seguintes subdetectores: detector de traços (identifica a trajetória das partículas carregadas eletricamente); calorímetro (medidor de energia altamente segmentado); e a câmara de múons (projeto especificamente para a detecção de múons). Como resultado dessa estrutura altamente

segmentada e da alta taxa de colisões, é produzido um volume de dados da ordem de 68 TB¹/s, o que requer a detecção (ou *trigger*) *online* das assinaturas de interesse, reduzindo o ruído de fundo produzido durante cada colisão. Essa taxa de colisões resultou somente em 2010 um total de 1 PB² somente no ATLAS (**FREUND, 2011**).

Devido ao elevado volume de informação produzido pelas colisões, o ATLAS utiliza um sistema de seleção *online* (*Trigger*), responsável por selecionar os eventos que contenham informações da física de interesse, reduzindo o volume de informação a ser armazenado em memória, para posterior análise. Este sistema possui um nível desenvolvido em *hardware* dedicado que realiza a primeira etapa de seleção dos eventos, que segue para o *Neural Ringer* (NR). Formado por um conjunto de redes neurais artificiais especialistas, baseadas em *perceptron* multicamadas, responsáveis por realizar a classificação dos eventos registrados pelo detector.

1.1 Objetivo Geral

Avaliar a utilização de técnicas de treinamento e aprendizado rápido de redes neurais artificiais, como alternativa ao classificador utilizado no *Neural Ringer* (NR), considerando aspectos como o tempo de processamento e a eficiência de detecção.

1.2 Objetivos Específicos

- Estudar a técnica Máquina de Aprendizado Extremo (*Extreme Learning Machine - ELM*) para avaliá-la como uma alternativa às técnicas utilizadas até o momento no detector;
- Desenvolver algoritmo para ELM com base na metodologia da Colaboração ATLAS-Brasil;
- Estudar a Técnica de Redes com Estados de Eco (*Echo State Network - ESN*) para avaliá-la como alternativa às técnicas utilizadas até o momento no detector.
- Desenvolver algoritmo para utilizar a ESN seguindo as especificações da Colaboração ATLAS Brasil;
- Avaliar o desempenho dos métodos propostos em comparação à versão tradicional do Neural Ringer considerando aspectos como o tempo de processamento para treinamento e a eficiência de detecção.

1.3 Justificativa

Os experimentos realizados no LHC seguem uma agenda pré-definida de estudos considerando o nível de energia a ser utilizada em suas colisões. Nesta agenda, toda a estrutura do LHC será gradativamente submetida a níveis de energia mais elevados, partindo de 450 GeV³, em cada um dos feixes, até a faixa entre 7 TeV e 8 TeV por feixe, que é a energia máxima de projeto,

¹ TeraByte - Múltiplo da unidade do *Byte* no Sistema Internacional, equivalente a 10^{12} *bytes*.

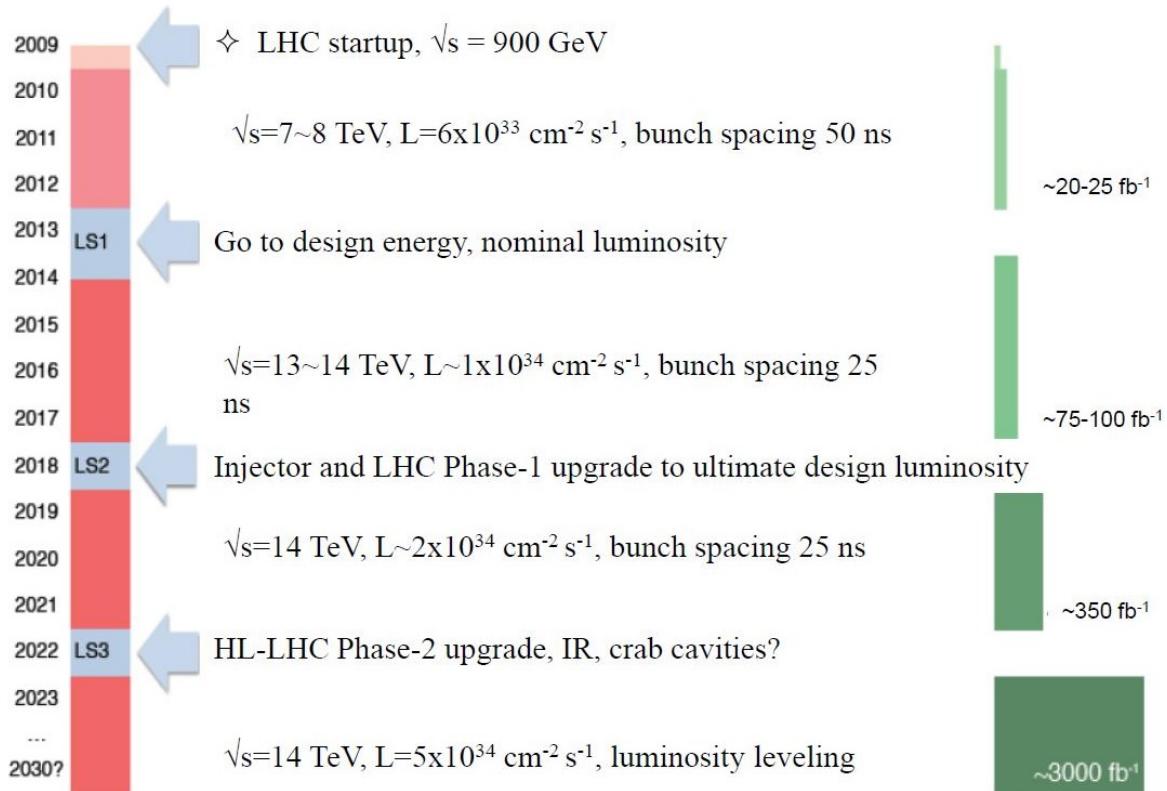
² PetaByte - Múltiplo da unidade do *Byte* no Sistema Internacional, equivalente a 10^{15} *bytes*.

³ GeV - Giga Elétron-volt. Elétron-volt energia ganha por um elétron acelerado por uma ddp de 1 volt.

prevista para as colisões. Desta forma, paradas para ajustes e atualizações de *hardware* e *software* são parte do calendário, como pode ser visualizado na [Figura 2](#), onde os eventos marcados em LS (*long shutdown*) referem-se às longas paradas para atualização e ajustes.

Desde seu início de funcionamento e primeira colisão, próton-próton, os detectores do LHC vem passando por atualizações periódicas, as quais elevam a energia, a frequência, a quantidade de partículas nos feixes de cada colisão (CERN, 2016) e luminosidade⁴. A título de curiosidade, em menos de seis anos de funcionamento (11/2009⁵ - 06/2015⁶), o LHC elevou os níveis de energia de colisão em cerca de 14 vezes, ou seja, a complexidade do tratamento dos dados obtidos desde a primeira colisão tem aumentado gradativamente. E esse cenário cria um desafio à seleção dos eventos de interesse pelos algoritmos de filtragem *online*.

Figura 2 – Agenda de atividades para o LHC em termos de nível de energia e luminosidade utilizada nas colisões.



Fonte: (HEUER, 2013)

Em 2011, o número médio esperado de colisões por feixe sofreu incremento de 5 para 15, em 2012, de 10 para quase 35 e era previsto que em 2015, quando voltasse a operar com níveis maiores de energia, chegasse a mais de 50, podendo atingir 80 (MARSHALL; ATLAS Collaboration, 2014).

Essa elevação no número de colisões produz um outro entrave na detecção das partículas,

⁴ medida do número de colisões por centímetro quadrado produzida a cada segundo [$\text{cm}^{-2} \text{s}^{-1}$].

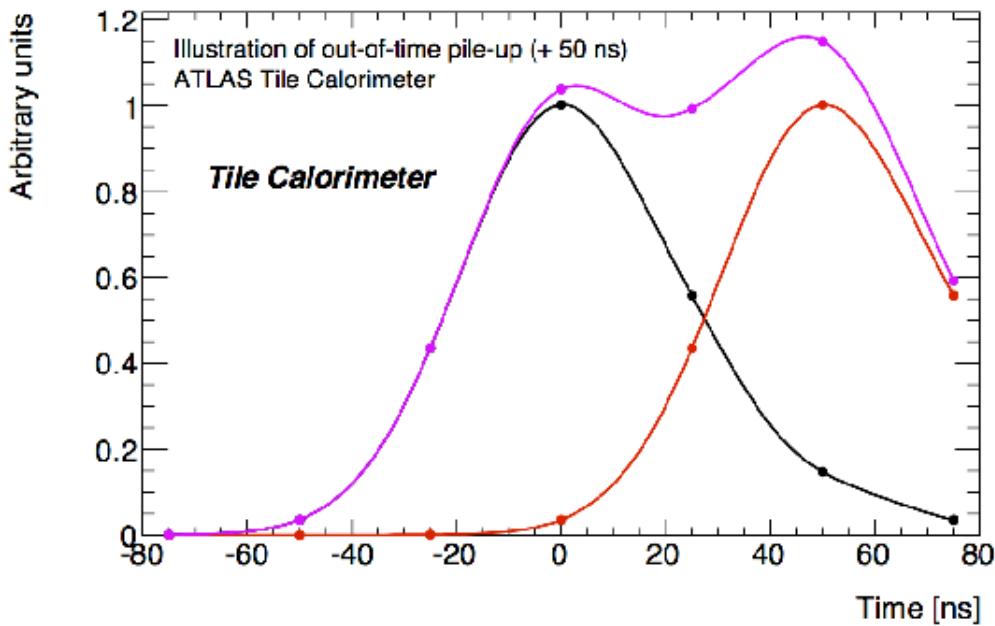
⁵ Primeira colisão a 450 GeV por feixe.

⁶ Experimentos são retomados com energia a 6,5 TeV por feixe.

que é o efeito do empilhamento (*pile-up*). Tal efeito é causado pela sobreposição de eventos num mesmo sensor, ou seja, enquanto é feita a leitura do registro de um evento anterior o sensor é sensibilizado por um novo evento. Porém, a leitura ainda não foi finalizada, e como resultado o evento registrado é a composição de dois eventos sequenciais identificados pelo sensor. Na Figura 3 é possível visualizar o efeito do empilhamento de eventos, que resulta num evento composto (em lilás) que mascara a real informação dos eventos originais, as curvas à esquerda (em preto) e à direita (em vermelho) (LUZ et al., 2016).

Devido à sua estrutura altamente segmentada, 100.000.000 de canais (ATLAS Collaboration, 2010), 187.652 somente nos calorímetros (ATLAS Collaboration, 2008a), o detector ATLAS produz um volume de dados da ordem de 68 TB/s. Tal fato, torna proibitivo utilizar um sistema para o armazenamento de toda a informação produzida. Como solução, o detector ATLAS realiza a detecção *online* dos eventos que possuam prováveis assinaturas de interesse antes de armazená-las em mídia permanente, para análise futura. Esse processo de detecção (*trigger*) *online*, visa reduzir o ruído de fundo produzido nas colisões.

Figura 3 – Ilustração do empilhamento de eventos num sensor do calorímetro do ATLAS. Os eventos em preto (curva com centro em 0) e em vermelho (curva com centro em ≈ 60).



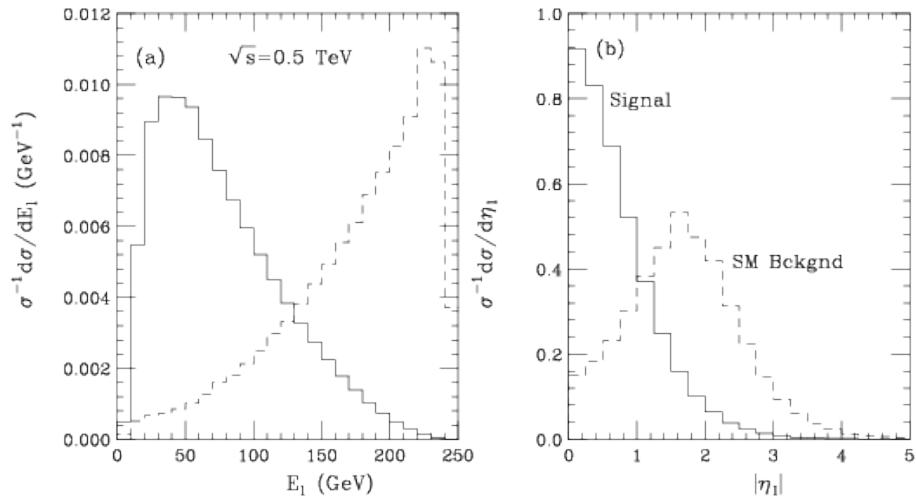
Fonte: Peralva e SEIXAS (2015)

Por apresentar formato cilíndrico, as camadas do calorímetro são sobrepostas, o que naturalmente produz correlação entre os anéis, ou seja, redundância (BRUMFIEL, 2012). E além disso, é esperado um comportamento, ligeiramente, não linear dos sensores do calorímetro (há saturação para valores elevados de energia - *pile-up* (FREUND; PINTO; SEIXAS, 2016; WIGMANS, 2008)).

Dentre os eventos de interesse, a detecção de elétrons é muito importante no detector ATLAS, pois estão envolvidos em decaimentos raros como do bóson *prime* (Z') e o bóson de

Higgs (FREUND; PINTO; SEIXAS, 2016). A identificação de elétrons se baseia fortemente na informação dos calorímetros e é dificultada devido à ocorrência de uma elevada taxa de ruído de fundo, basicamente compostos de partículas hadrônicas, conhecidas como jatos. Sua identificação utiliza informações do canal e/γ , o qual busca identificar assinaturas de elétrons, pósitrons ou fóton (FREUND, 2011, p 76). Na Figura 4 são exibidos possíveis perfis de deposição de energia para o sinal de interesse (linha sólida) e o ruído de fundo (*background*, linha tracejada).

Figura 4 – Exemplos de deposição de energia normalizada no canal de *leptons* (a) e em função da pseudorapidez (η) (b) com $\sqrt{s} = 500$ GeV.



Fonte: Han e Hewett (1999)

O sistema de seleção ou filtragem *online* (*trigger*) do ATLAS (ANJOS et al., 2006) é responsável pela seleção dos eventos interessantes para o experimento e, também pela redução do ruído de fundo (assinaturas não relevantes) produzido nas colisões. Esse sistema, opera sob grandes restrições temporais de processamento, pois deve selecionar sinais⁷ que indiquem assinaturas de eventos físicos de interesse em meio a um volume de informação significativo num tempo reduzido.

A estrutura do sistema de filtragem possui três etapas, em cascata, responsáveis por reduzir o volume de dados proveniente das colisões que produzem informação numa taxa próxima de 1 GHz para uma taxa de 300 Hz, pois, a física de interesse é de rara ocorrência e está envolta num chuveiro de partículas hadrônicas com elevada taxa de produção.

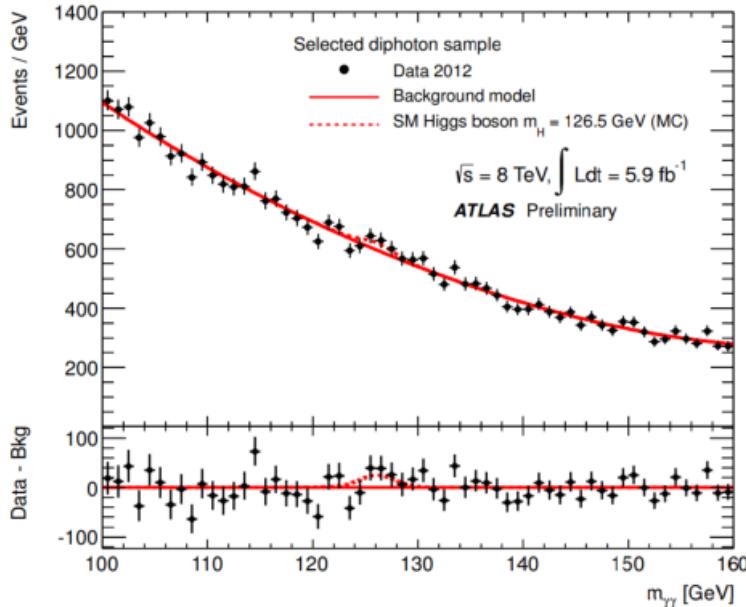
Na Figura 5 é possível visualizar o registro do bóson de Higgs, com massa⁸ de 126,5 GeV (ATLAS Collaboration, 2012a) conforme previsto no modelo padrão, em meio aos dados de medição no ano de 2012. Tais eventos necessitam de um sistema de *trigger* de alta eficiência e seletividade, dada a rara ocorrência e estreita faixa possível de registro de ocorrência. Para o bóson de Higgs, a faixa esperada para a massa é, $106 \text{ GeV} < m_{\gamma\gamma} < 160 \text{ GeV}$ tendo o valor da

⁷ É definido matematicamente como uma função de uma ou mais variáveis, a qual veicula informação sobre a natureza de um fenômeno físico (HAYKIN, 2001).

⁸ A referência à massa utilizando a unidade de energia eV, é comum na literatura de físicas de altas energias, visto que a velocidade na qual as partículas são aceleradas é c , velocidade da luz. Explicitamente, $m = E_c/c^2$.

massa ajustado para $125,09 \pm 0,24$ GeV (AAD et al., 2015, p. 5).

Figura 5 – Distribuição de massa para energia $\sqrt{s} = 8$ TeV, sobreposta aos dados de medição do ano de 2012, e previsão da massa para o Bóson de Higgs segundo o modelo padrão.



Fonte: ATLAS Collaboration (2012a)

Um outro ponto, relevante, associado ao *Neural Ringer* é o elevado tempo de treinamento do sistema. Pois o processo precisa ser repetido para as diferentes configurações de operação e também para diferentes regiões do detector. Seguindo a metodologia, atual⁹, adotada pela Colaboração ATLAS, é necessário realizar o número de treinamentos de ordem superior a 10^4 redes neurais para o projeto de cada discriminador do sistema.

Logo, a atualização e aprimoramento nas técnicas utilizadas para a seleção e detecção de eventos no detector ATLAS tornam-se de relevância, e um desafio para a colaboração ATLAS frente aos saltos no nível de informação produzidos após cada etapa de atualizações previstas na agenda.

Este trabalho propõe a utilização de duas técnicas como alternativas ao classificador MLP no discriminador *Neural Ringer*. A primeira técnica, uma Máquina de Aprendizado Extremo (do inglês: *Extreme Learning Machine - ELM*) (HUANG; ZHU; SIEW, 2006; HUANG; WANG; LAN, 2011; HUANG et al., 2015; HUANG, 2015), que é uma rede neural em avanço de camada única que não possui um processo iterativo para treino. A segunda técnica é a Rede com Estado de Eco (do inglês: *Echo State Network - ESN*) (JAEGER, 2003; JAEGER; HAAS, 2004; JAEGER, 2010), uma rede neural de estrutura recorrente, na qual a saída saída da rede é obtida por meio de um combinador linear.

As máquinas de aprendizado extremo (ELM) foram propostas inicialmente em Huang,

⁹ Na qual a base de dados é segmentada em regiões (E_T , η), e cada uma treinada 5.000 vezes. Mais detalhes no Capítulo 4.

Zhu e Siew (2006) e têm sido aplicadas em problemas nos quais deseja-se obter uma rede otimizada a partir de repetidos ensaios com grande volume de dados em tempo de treinamento reduzido, como em Termenon et al. (2016).

A ELM possui uma estrutura semelhante à de uma rede neural MLP, com uma única camada oculta (*Single Layer Feedforward Network - SLFN*). Para o processo de treinamento da ELM assume-se que é possível gerar aleatoriamente os pesos da camada de entrada, e determinar, analiticamente, os melhores pesos para a camada oculta (HUANG, 2015) com base nos pesos da camada de entrada e as entradas e saídas alvo. Deste modo, o tempo de treinamento de uma ELM é consideravelmente reduzido, pois não existe um procedimento iterativo de retropropagação de erro para o ajuste dos pesos do modelo.

As redes com estado de eco (ESN) são estruturas compostas de três partes: uma camada de entrada, um reservatório de dinâmicas uma camada de saída. O reservatório de dinâmicas, compõe a camada oculta, nele tem-se uma rede neural recorrente, totalmente conectada, com elementos de processamento não-linear. Sua camada de saída é determinada pela combinação linear entre os elementos de saída do reservatório de dinâmicas (JAEGER, 2002; BOCCATO, 2013; SIMEÓN, 2015) e os valores alvo.

A ESN faz parte dos métodos de Computação com Reservatórios (*Reservoir Computing*), propostos por Jaeger (2001) e foram chamadas de ESN. As redes ESN além de aproveitar as vantagens de uma estrutura recorrente, conseguem simplificar significativamente o processo de treinamento.

1.4 Organização do Documento

O Capítulo 2 apresenta uma breve pesquisa bibliográfica sobre o históricos de alguns dos principais fatos das descobertas da física que contribuíram para a estrutura do modelo padrão de interação entre as partículas elementares em uso atualmente, o LHC, detector ATLAS. O Capítulo 3 apresenta as técnicas propostas neste trabalho de mestrado. No Capítulo 4, é descrita a metodologia adotada para a execução do trabalho, no que se refere à análise quantitativa e qualitativa dos resultados para cada uma das técnicas propostas. O Capítulo 5, apresenta os resultados obtidos com dados simulados e experimentais. O Capítulo 6 apresenta as conclusões e perspectivas de trabalhos futuros.

No Apêndice A é apresentado o resumo dos trabalhos publicados em anais de congressos e simpósios. No Apêndice B é apresentada uma breve análise de sensibilidade da técnica ELM quanto ao tipo de distribuição utilizada para a geração dos números pseudo-aleatórios da camada interna, utilizando uma das bases de dados simulados.

2 Física e o LHC

Este capítulo é estruturado em duas partes. Na primeira é apresentada uma breve pesquisa bibliográfica sobre a história do modelo atômico, a qual abordará a evolução do modelo até os dias atuais citando algumas das principais descobertas no entendimento da estrutura fundamental da matéria. A segunda parte contemplará a apresentação do Grande Colisor de Hadrons (*Large Hadron Collider*) e do detector ATLAS.

2.1 Física de Altas Energias e o LHC

2.1.1 Breve Histórico do Modelo Atômico

A discussão sobre a estrutura da matéria e de como seria o átomo¹⁰ vem desde o século V A.C. Alguns filósofos defendiam a ideia de indivisibilidade do átomo, outros acreditavam que a matéria era infinitamente divisível e contínua (ROCHA, 2002).

Em 1807, Dalton¹¹ publica a sua teoria atômica que teve impulso quando leu sobre a pesquisa de Lavoisier¹² identificando que o ar é composto por, pelo menos, dois gases de pesos diferentes. Nesta publicação, Novo Sistema de Filosofia Química (*New System of Chemical Philosophy*) estabelece as leis básicas da nova química (PINHEIRO; COSTA; MOREIRA, 2011).

Em 1904, Nagaoka¹³ propôs um modelo em que o núcleo era uma esfera, grande e rígida, e os elétrons estariam distribuídos em anéis tipo saturnianos (INAMURA, 2016). Thomson¹⁴ em seus experimentos chega à descoberta do elétron, fato que o levou propor um modelo diferente para o átomo, e nesse, levando em consideração a presença dos elétrons. Tal modelo ficou conhecido como pudim de passas, no qual a carga atômica estaria distribuída num volume preenchido por cargas negativas (elétrons) distribuídas uniformemente resultando em equilíbrio elétrico. Tal modelo apresentava inconsistências pois cargas opostas se neutralizam ao interagirem (ROCHA, 2002; PINHEIRO; COSTA; MOREIRA, 2011).

Alguns anos mais tarde, em 1910, Rutherford¹⁵ conduziu experimentos com radioatividade nos quais bombardeou uma fina placa metálica, com feixe de partículas alfa¹⁶. Nesse experimento ele concluiu que os núcleos eram muito pequenos, com raios entre 10^{-12} e 10^{-13} cm, e observou que parte do feixe partículas atravessava a folha metálica sem nenhum desvio, enquanto outras sofriam desvio. Tais observações o levaram a propor o modelo atômico planetário, com o núcleo, pequeno e de carga positiva ao centro, os elétrons distribuídos ao redor em órbitas circulares e

¹⁰ Partícula que se considerava o último grau da divisão da matéria (PRIBERAM, 2016).

¹¹ John Dalton (1766 – 1844) químico, físico e meteorologista inglês.

¹² Antoine Laurent de Lavoisier (1743 – 1794), químico Francês.

¹³ Hantaro Nagaoka (1865–1950), físico Japonês.

¹⁴ Joseph John Thomson (1856 – 1940) físico britânico, Nobel de Física de 1906 pelas investigações sobre a condução de eletricidade nos gases.

¹⁵ Ernest Rutherford (1871–1937), físico e químico neozelandês.

¹⁶ Partículas formadas por dois prótons e dois nêutrons de carga +2, formadas a partir da ionização do núcleo de He.

um espaço vazio entre o núcleo e os elétrons (OLIVEIRA; FERNANDES, 2006).

O modelo proposto por Rutherford ainda continha uma inconsistência física. Ele não era estável, pois os elétrons desenvolveriam trajetórias elípticas em direção ao núcleo o que geraria uma possível *catástrofe atômica*. Mais tarde, Bohr¹⁷ (1913) traz contribuições ao modelo proposto por Rutherford, o que ficou conhecido como modelo de Rutherford-Bohr, incorporando teorias sobre distribuição e movimentos dos elétrons; teorias que se baseiam na teoria quântica de Plank¹⁸ (PINHEIRO; COSTA; MOREIRA, 2011).

Apesar das contribuições feitas por Bohr, o modelo ainda possuía questões em aberto, como por exemplo, não consegue explicar a energia constante do elétron. Werner Heisenberg¹⁹, um dos assistentes de Bohr²⁰ ao ler sobre a teoria apresentada por Bohr para o modelo atômico, verifica que não é possível observar os orbitais definidos por Bohr em seu modelo, porém, é possível verificar a transição entre os orbitais (ROCHA, 2002). Como consequência dos resultados ele estabelece as relações de incerteza (1927), as quais dizem ser impossível determinarmos simultaneamente a posição e a quantidade de movimento (mv) de uma partícula, em um certo instante. Essa relação indica não ser possível observar um fenômeno sem causar interferência durante o processo (PINHEIRO; COSTA; MOREIRA, 2011; OLIVEIRA; FERNANDES, 2006; MELZER; AIRES, 2015).

Erwin Schrödinger²¹ atuava como professor na universidade de Zurique quando tomou conhecimento da tese de doutorado de De Broglie²² (1926) o que lhe interessou e motivou na busca de uma equação de movimento para as ondas de matéria. Tal busca culminou na função de onda Ψ , a qual é um objeto matemático que apresenta o mesmo caráter de um campo estendido no espaço. Os trabalhos desses dois últimos físicos citados, Heisenberg e Schrödinger, atualizam o modelo Rutherford-Bohr e formam a base do modelo atômico atual, no qual a posição do elétron é definida em uma probabilidade calculada pela função Ψ , e define-se uma nuvem eletrônica na qual o elétron ocupa uma posição desconhecida (ROCHA, 2002).

Paralelo aos avanços descobertas no modelo atômico, descobertas importantes sobre a estrutura do núcleo atômico ocorriam, as quais culminam para a formulação do modelo padrão utilizado, atualmente, e em constante aperfeiçoamento (PINHEIRO; COSTA; MOREIRA, 2011).

A seguir são listadas algumas descobertas relevantes citadas por ano, reunidas em Rocha (2002, p. 363 – 372):

- 1897 - Descoberta do elétron e sua carga negativa por Thomson;
- 1903 - Nobel de Física pela descoberta da radioatividade: Antonie Henri Becquerel, Pierre Curie e Marie Skłodowska-Curie²³;

¹⁷ Niels Henrick David Bohr (1885 – 1962), físico dinamarquês.

¹⁸ Max Karl Ernest Plank (1858 – 1947), físico alemão, Nobel de Física em 1918 pela descoberta dos quanta de energia.

¹⁹ Werner Karl Heisenberg (1901 – 1976), físico teórico alemão.

²⁰ Max Born (1882 – 1970), físico e matemático alemão.

²¹ Erwin Rudolf Josef Alexander Schrödinger (1887 – 1961), físico teórico austríaco.

²² Louis-Victor-Pierre-Raymond, 7.º duque de Broglie (1892 – 1987), físico francês, Nobel de Física de 1929 pela descoberta da natureza ondulatória dos elétrons.

²³ Becquerel (1852 – 1908), físico Francês; Pierre Curie (1859 – 1906), físico Francês; Marie Curie (1867 – 1934),

- 1917 - Nobel de Física pela descoberta dos Raio-X: Charles Glover Barkla²⁴;
- 1918 - Nobel de Física pela descoberta dos quanta de energia: Max Planck;
- 1929 - Nobel de Física pela descoberta da natureza ondulatória do elétron: De Broglie;
- 1932 - Nobel de Física pela criação da Mecânica Quântica: Werner Heisenberg;
- 1933 - Nobel de Física pela descoberta de novas formas para a teoria atômica: Erwin Schrödinger;
- 1935 - Nobel de Física pela descoberta do Nêutron: Sir James Chadwick²⁵;
- 1949 - Nobel de Física pela descoberta do Méson: Hideki Yukawa;
- 1954 - Nobel de Física pela pesquisa fundamental sobre a Mecânica Quântica: Max Born e Walther Bothe²⁶;
- 1958 - Descoberta do Antinêutron: Pavel Aleksejetic Cherenkov, Il'já Michajlovic Frank e Igor' Evgen'evic Tamm²⁷;
- 1959 - Nobel de Física pela Descoberta do Antípróton: Emilio Gino Segrè e Owen Chamberlain²⁸;
- 1983 - Tevatron, acelerador de partículas com energia de colisão de até 1,8 TeV é construído no Fermilab;
- 1984 - Nobel de Física pelas contribuições que permitiram a descoberta das partículas de campo W e Z: Carlo Rubbia e Simon van der Meer²⁹;
- 1992 - Nobel de Física pela invenção e desenvolvimento de detectores de partículas - a câmara proporcional de multifios: Russell A. Hulse³⁰;
- 1995 - Nobel de Física pelas contribuições à descoberta do lépton tau a Martin L. Perl e a Frederick Reines pela detecção do neutrino³¹;
- 2008 - LHC entra em operação.
- 2012 - Bóson de Higgs, primeiro registro de detecção (ATLAS Collaboration, 2012b)
- 2018 - Bóson de Higgs, segundo registro de detecção. Desta vez, associado ao *top quark* (ATLAS Collaboration, 2018).

²⁴ física Polonesa; 1^a mulher a ganhar dois Prêmios Nobel; em 1911 - Nobel em Química pela descoberta do Rádio e Polônio

²⁵ (1877 – 1944), físico Inglês.

²⁶ (1891 – 1974), físico Inglês.

²⁷ (1891 – 1957), físico Alemão.

²⁸ Cherenkov (1904 – 1990), físico Russo; Frank (1908 – 1990) físico Russo; Tamm (1895 – 1971), físico Russo.

²⁹ Segrè (1905 – 1989) físico Italiano, Chamberlain (1920 – 2006) físico dos EUA.

³⁰ Rubbia (1930 –), físico Italiano; Meer (1925 – 2011), físico Holandês.

³¹ (1950 –), físico dos EUA.

³¹ Perl (1927 – 2014) físico dos EUA, Reines (1918 – 1998) físico dos EUA

2.1.2 Modelo Padrão

Em 1960 se iniciam as discussões sobre o modelo padrão, o qual é uma das teorias mais completa sobre a natureza da matéria em uso atualmente. Segundo Gondon Kane, um físico teórico da Universidade de Michigan:

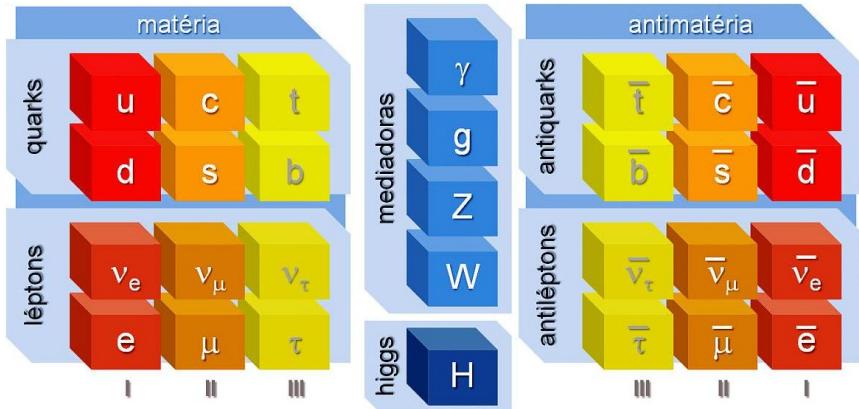
[...] Rather it is a conclusion embodied in the most sophisticated mathematical theory of nature in history, the Standard Model of particle physics. Despite the word “model” in its name, the Standard Model is a comprehensive theory that specifies what are the basic particles and how they interact. Everything that happens in our world (except for the effects of gravity) results from Standard Model particles interacting according to its rules and equations [...] (KANE, 2003).

Em tradução livre:

[...] Pelo contrário, o Modelo Padrão é, na história, a mais sofisticada teoria matemática sobre a natureza. Apesar da palavra “modelo” em seu nome, o Modelo Padrão é uma teoria abrangente que identifica as partículas básicas e especifica como interagem. Tudo o que acontece em nosso mundo (exceto os efeitos da gravidade) resulta das partículas do Modelo Padrão interagindo de acordo com suas regras e equações [...]

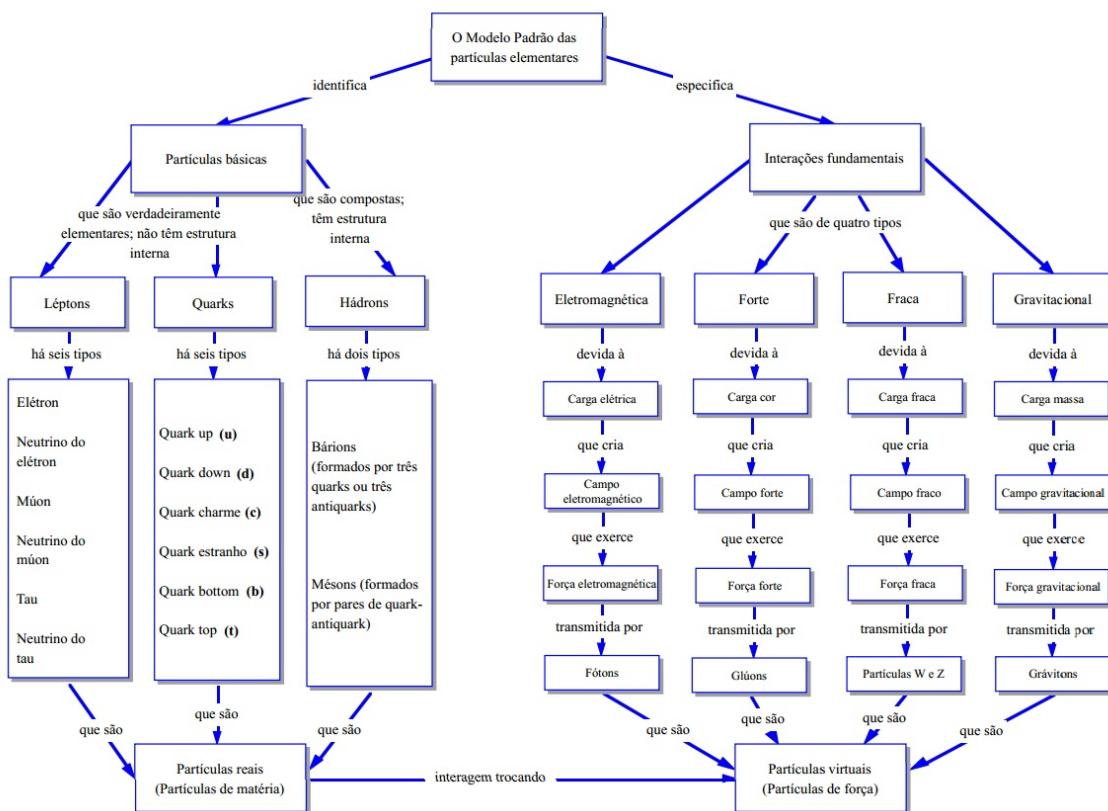
Na Figura 6 são exibidas as principais partículas responsáveis pelos quatro campos fundamentais segundo o Modelo Padrão (MP) atual, a saber, o campo de fótons (eletromagnético), o campo de glúons (forte), o campo de partículas W e Z (fraco) e o campo de grávitons (gravitacional), esse último ainda não foi observado. No MP os constituintes básicos da matéria, são as partículas elementares: quarks, léptons e bósons mediadores. Sendo que para cada uma dessas partículas existe a correspondente antipartícula, com mesma massa, *spin* e paridade da correspondente partícula, porém com números quânticos opostos (PIMENTA et al., 2013; BRAIBANT; GIACOMELLI; SPURIO, 2012; ELLWANGER, 2012).

Na Figura 6 tem a presença do bóson de Higgs (ELLWANGER, 2012, Cap. 7), a qual seria a partícula criada pelo campo de Higgs no momento em que esse recebe energia suficiente. Em contrapartida, quando a partícula de Higgs interage com as demais partículas elementares (léptons e quarks, por exemplo) ela transfere energia na forma de massa, do campo de Higgs para a partícula elementar (PIMENTA et al., 2013).

Figura 6 – Representação do Modelo Padrão e suas partículas.

Fonte: (GROSSMANN, 2013)

É possível compreender o papel de tais partículas na Figura 7, na qual, em resumo, são apresentadas as partículas elementares e suas interações fundamentais (MOREIRA, 2009).

Figura 7 – Diagrama simplificado sobre o Modelo Padrão, contendo informações sobre as partículas básicas e as interações fundamentais.

Fonte: Extraído de Moreira (2009)

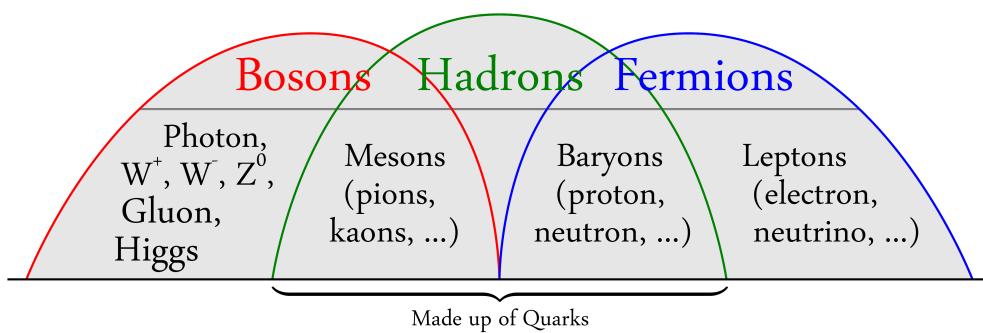
Os quarks são partículas elementares fermiônicas³², as quais podem interagir através de todas as interações fundamentais. São seis os tipos de quarks, também chamados sabores, os quais são o quark *u* (*up*), o quark *d* (*down*), o quark *s* (*strange*), o quark *c* (*charm*), o quark *b* (*bottom*) e o quark *t* (*top*), sendo cada um possuidor de uma carga cor *R* (*red*), *G* (*green*) e *B* (*blue*). A carga cor é a responsável pelo confinamento dos quarks, pois somente os estados (hádrons) sem cor são os observados (PIMENTA et al., 2013).

O segundo grupo é o dos léptons³³, também são seis: *elétron* (*e*), *múon* (μ), *tau* (τ), *neutrino do elétron* (ν_e), *neutrino do múon* (ν_μ) e *neutrino do tau* (ν_τ). Não sujeitos à interação forte nem constituídos por quarks. Os três neutrinos não possuem cor ou carga, dessa forma só interagem via força fraca e gravitacional, por isso são de difícil observação (ELLWANGER, 2012, Cap. 6, 7).

O último grupo é o grupo dos bóson mediadores, partículas de *spin* inteiro e que intermedeiam as interações entre os férmons. Os bósons W^+ , W^- e Z^0 são mediadores da interação fraca, os fótons (γ) da interação eletromagnética e os glúons (g) a interação forte (BRAIBANT; GIACOMELLI; SPURIO, 2012, Cap. 8).

As partículas constituintes do MP podem ser organizadas num diagrama exibido na Figura 8, a qual contempla os bóson, os hádrons e os férmons. Os férmons compostos por léptons e barions³⁴. Os hádrons³⁵ é grupo formado por partículas de *spin* inteiro, os barions, e partículas com *spin* não inteiro, os mesons³⁶, e o último, o grupo dos bósons

Figura 8 – Diagrama simplificado dos três grupos de partículas fundamentais do Modelo Padrão.



Fonte: Extraído de Belle (2017)

2.2 O LHC

O grande colisor de hádrons (*Large Hadron Collider - LHC*) (CERN, 2017) é o maior acelerador de partículas em operação atualmente. Possui formato circular, um perímetro de 27 km aproximadamente, está localizado no Centro Europeu para Pesquisa Nuclear (CERN), no

³² Partículas com *spin* semi-inteiro. *spin* - característica intrínseca das partículas elementares; um dos quatro números quânticos que definem uma partícula.

³³ Do grego *lépton* que significa leve ou pequeno, foi por esse motivo usada para identificar as partículas menores.

³⁴ Do grego *baros* que significa pesado, foi por esse motivo usada para identificar as partículas maiores.

³⁵ Do grego *hadros* - forte. Partículas compostas por quarks, sujeitos a força nuclear forte.

³⁶ Do grego *mesos* que significa intermediário ou médio, usada para identificar partículas com massa mediana.

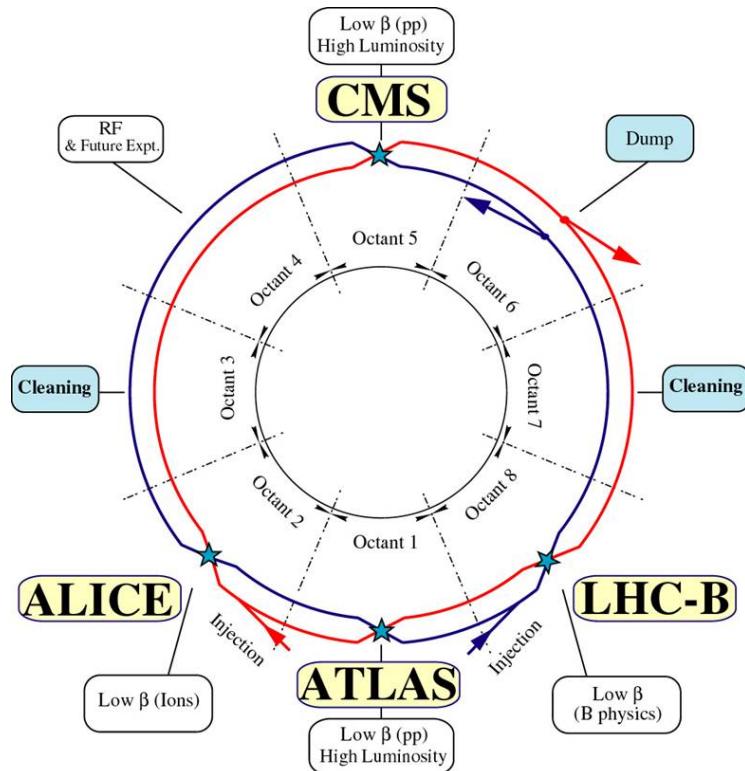
subsolo (entre 50 m e 175 m de profundidade) na fronteira franco-suíça, próximo a Genebra, Suíça.

Teve um custo de construção total de 4332 Mi CHF³⁷, aproximadamente €3,76 Bi. O custo total, somente com os seus detectores foi de 1500 Mi CHF, aproximadamente €1,3 Bi (CERN, 2017).

É composto por sete experimentos, CMS (*Compact Muon Solenoid*), ATLAS (*A Toroidal ApparatuS*), LHCb (*Large Hadron Collider beauty experiment*), ALICE (*A Large Ion Collider Experiment*), TOTEM (*Total Elastic and diffractive cross section Measurement*), LHCf (*Large Hadron Collider forward*) e MoEDAL (*Monopole and Exotics Detector At the LHC*), esses três últimos, respectivamente, de menor escala (FREUND, 2011).

Na Figura 9 é possível observar como a estrutura do LHC é dividida. São oito octantes: no primeiro, ficam o ATLAS e LHCf; no quinto, diametralmente oposto ao primeiro, ficam o CMS e TOTEM; nos 2º e 8º octantes, são os pontos onde os feixes de prótons são inseridos, um no sentido horário e o outro no sentido anti-horário, e também é a localização dos detectores ALICE e LHCb/MoEDAL, respectivamente.

Figura 9 – Representação da estrutura do LHC.



Fonte: (EVANS; BRYANT, 2008)

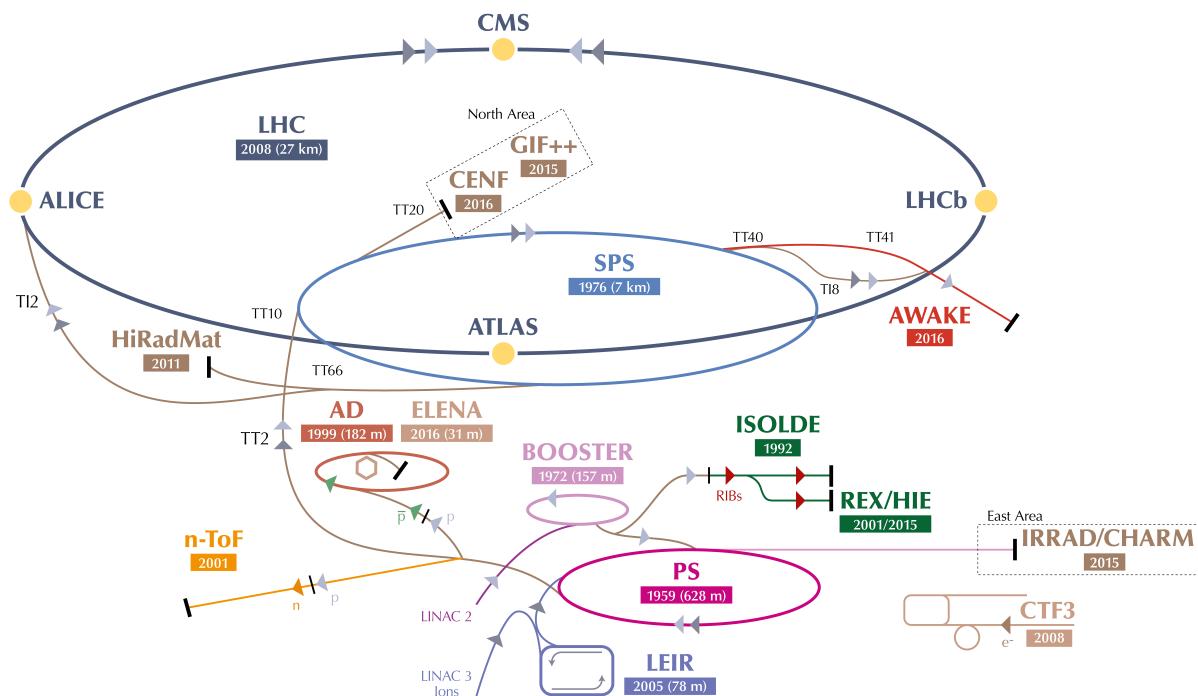
O complexo acelerador do LHC, conta com três estágios de aceleração nos quais os feixes de prótons, provenientes de átomos de hidrogênio, são acelerados até 99,9999991% da velocidade

³⁷ CHF - Francos suíços.

da luz, e atingem a energia de até 14 TeV (CERN, 2017, p 5).

Na Figura 10 é possível visualizar esses estágios onde os feixes são acelerados: PSB³⁸, PS³⁹ e SPS⁴⁰. No estágio do PSB (indicado como *BOOSTER* na Figura 10) os feixes de prótons são injetados com energia de 50 MeV vindos do Linac2⁴¹, e saem desse estágio com energia de 1,4 GeV. Em seguida chegam ao PS e ficam até atingirem a energia de 25 GeV, e são direcionados ao terceiro estágio antes do LHC, passando pelo SPS e são acelerados até atingirem 450 GeV. Nesse momento, os feixes são inseridos nos octantes 2 e 8 do LHC, ver Figura 9, para acelerarem durante 20 min até obterem valores de energia de 7 TeV, por feixe, e colidirem nos detectores ATLAS e CMS, octantes 1 e 5, respectivamente (CERN, 2017).

Figura 10 – Diagrama da estrutura do complexo acelerador do LHC.



Fonte: Adaptado de CERN (2017)

O ATLAS e o CMS, são detectores de propósito geral, porém, projetados de maneira diferente. O objetivo de ter dois detectores de estruturas distintas, operando de forma isolada para o mesmo propósito, está no fato de obter confirmação dos resultados eliminando respostas tendenciosas, ou seja, os resultados obtidos no LHC passam pelo registro e confirmação desses dois detectores de maneira independente.

O LHCb é um detector especializado no estudo do méson B e para compreensão da Violação CP e a diferença matéria e antimateria. O subdetector ALICE, especialista na detecção

³⁸ Proton Synchrotron Booster.

³⁹ Proton Synchrotron.

⁴⁰ Super Proton Synchrotron.

⁴¹ Linac 2 - Linear Accelerator 2.

de íons pesados, é destinado a explorar eventos na interação núcleo-núcleo. O TOTEM, mede a seção de choque total de colisões $p-p$, e estuda colisões elásticas⁴² e difrativas⁴³. O LHCf, é o detector responsável pelo estudo da influência de raios cósmicos nos experimentos, e o MoEDAL, tem o objetivo de buscar evidências de partículas hipotéticas, estáveis, de monópole magnético e estáveis supersimétricas.

Com o LHC em funcionamento a existe a possibilidade de verificar algumas questões fundamentais da física de partículas elementares previstas, e outras questões que vão além do Modelo Padrão (ELLWANGER, 2012, Cap. 8). A seguir, breve citação das teorias com as quais o LHC pode contribuir para sua verificação (FREUND, 2011), e maiores detalhes podem ser obtidos em Nath et al. (2010):

- **Descoberta do Bóson de Higgs** - Partícula responsável por transferir energia, na forma de massa, do campo de Higgs, para as partículas com as quais ele interage (MOREIRA, 2009);
- **Busca pela SUSY** - Teoria da supersimetria, na qual cada partícula deve ter uma contraparte supersimétrica. Algumas partículas previstas nesse modelo devem ser detectadas na região de TeV;
- **Violiação CP** - Estudo da violação Carga Paridade, um subtópico da SUSY;
- **Matéria Escura** - Partículas ou conglomerados maciços de partículas que não brilham ou dissemelam luz;
- **Física do quark top** - Busca uma melhor compreensão da física dessa partícula;
- **Física do Z prime (Z')** - Possíveis bósons Z adicionais. Tais bósons ocorrem em extensões do MP, e caso eles sejam detectados na região com massa de TeV, o LHC pode identificá-los;
- **Assinaturas visíveis do Setor Escuro (HS)** - Modelos baseados em cordas e membranas. Interações entre a região visível e escura podem ocorrer e produzir bósons Z' , os quais poderiam estar na faixa de frações de GeV, e nessas situações o LHC poderia detectá-los;
- **Provar a origem da massa dos léptons neutrinos** - Como o LHC tem capacidade de detecção de massas na região de TeV, tal mecanismo gerador de massas pode ser detectado se estiver nessa região;
- **Busca por dimensões extras** - Modelos de ordens superiores de dimensão. Os modelos de dimensões extras são uma alternativa à supersimetria, os quais permitem a produção de um rico conjunto de assinaturas, incluindo buracos negros, os quais podem ser testados no LHC;

⁴² São colisões nas quais não há perda líquida de energia cinética como resultado da colisão.

⁴³ São colisões $p-p$, nas quais há dissociação parcial de um dos prótons, em apenas alguns novos prótons, sendo que o outro próton permanece intacto. Essas, estão no limiar entre as colisões (elásticas) que não resultam na produção de novas partículas, e as colisões (rígidas) desejadas (ELLWANGER, 2012, Cap. 8), nas quais há possibilidade de produção de física de alta energia (FREUND, 2018).

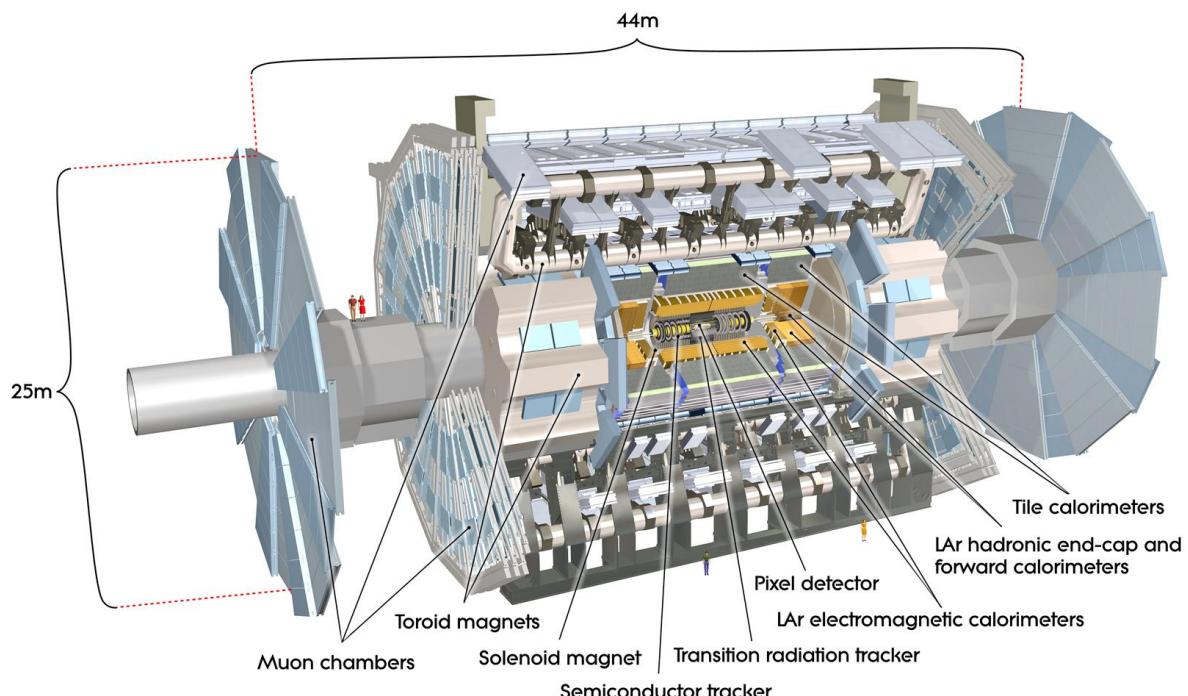
- **Busca por cordas no LHC** - Teoria que pode possibilitar a unificação das quatro⁴⁴ interações conhecidas na natureza incluindo a gravidade. Modelos independentes preveem cordas na escala de TeV, as quais podem ser testadas no LHC .

2.2.1 O Detector ATLAS

O ATLAS, é um dos principais detectores do LHC. Construído em formato cilíndrico, com raio de 12,5 m, comprimento de 44 m e com massa de aproximadamente 7.000 toneladas (ATLAS Collaboration, 2016).

Sua estrutura, ver Figura 11, é composta dos seguintes detectores: Detector Interno (ID), Calorímetro Eletromagnético (EMB), Calorímetro Hadrônico (HEC) e Espectrômetro de Múons. Cada detector possui uma função específica de detecção: o ID é responsável pelas partículas carregadas eletricamente, o ECAL (*Electromagnetic Calorimeter*) responsável por detectar e absorver elétrons, fótons e pósitrons, o HEC detectar e absorve partículas com componentes hadrônicas, como nêutrons, prótons e outros mésons. Os Múons devido a sua energia devem atravessar os calorímetros e serem detectados somente pelo Espectrômetro de Múons. Os Neutrinos não são detectados pelos subdetectores do ATLAS (FREUND, 2018).

Figura 11 – Diagrama do detector ATLAS, com destaque para seus subdetectores. EMB: *LAr Electromagnetic Calorimeter*; HEC: *Tile Calorimeter*; ID: composto por *Semiconductor tracker*, *Transition radiation tracker* e *Pixel detector* e o EMEC: *LAr hadronic end-cap and forward calorimeters*, são as tampas que fecham o detector ATLAS.

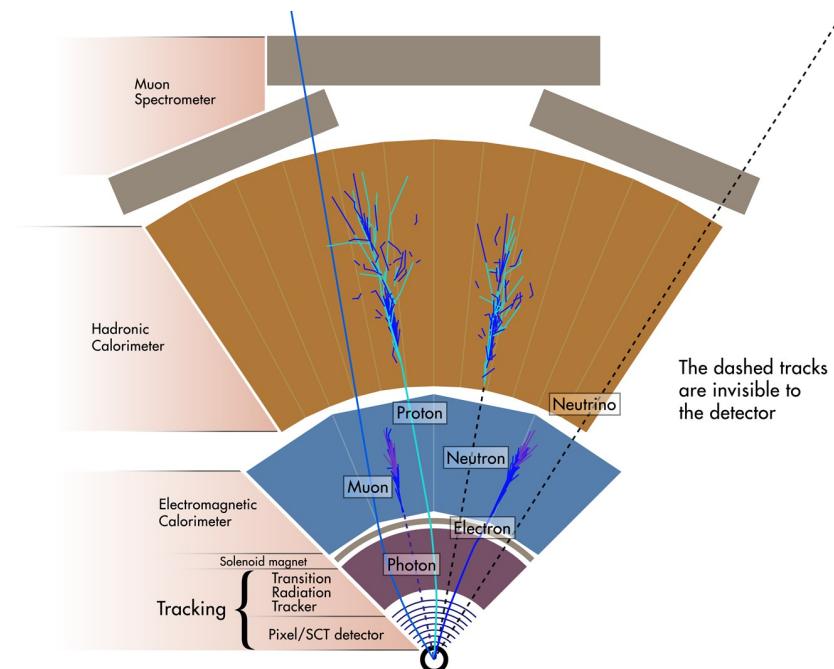


Fonte: (CERN, 2015)

⁴⁴ Força Nuclear Forte, Força Nuclear Fraca, Força Eletromagnética e Gravidade.

Na Figura 12, é possível observar o diagrama de um corte transversal contendo um setor do detector. Nesse corte é indicada a interação das partículas ao longo das camadas constituintes do detector. No círculo preto, no vértice do setor, está localizado o túnel onde os feixes de prótons são acelerados e colidem. Em seguida as regiões dos detectores ID, EMB, HEC e Espectrômetro de Muons, camada mais externa. Partindo do ponto de colisão são ilustradas possíveis trajetórias das partículas ao longo da interação com as camadas do detector, incluindo as partículas que não são visíveis para o detector, os neutrinos.

Figura 12 – Representação da interação das partículas no interior do detector ATLAS.



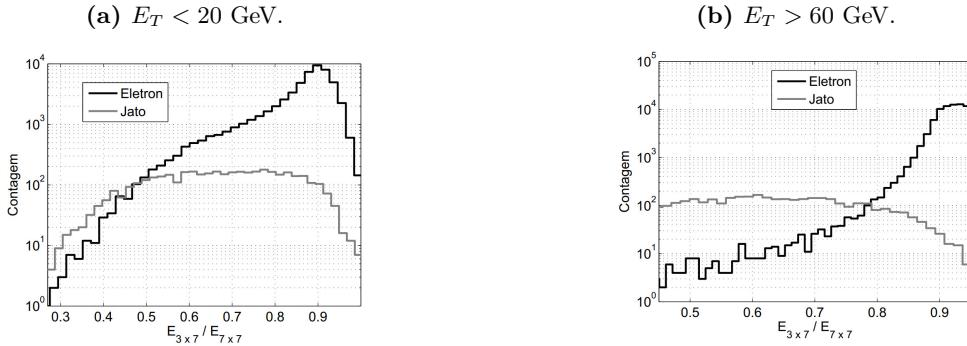
Fonte: Adaptado de ATLAS Collaborartion (2013)

A identificação de elétrons é muito importante para o desempenho do detector, pois a busca por assinaturas de interesse⁴⁵ podem estar relacionadas aos elétrons, e, para isso, são utilizadas informações dos calorímetros. O sistema de calorímetros do ATLAS é composto por 187.652 sensores (ATLAS Collaboration, 2017) com o objetivo de medir a energia depositada pelas partículas produzidas nas colisões do LHC. Um dos discriminadores utilizados atualmente no ATLAS para a identificação online de elétrons é o *Neural Ringer* (SEIXAS et al., 1996), no qual o perfil de deposição de energia é utilizado como entrada para uma rede neural tipo *perceptron* de múltiplas camadas, que opera como classificador.

Em SIMAS FILHO (2010, p 52–53) foi feita uma análise da distribuição de elétrons e jatos em função da energia total do evento, e apresentada na Figura 13. Nota-se que a variação da distribuição em função da energia envolvida no evento, favorece a separação de classes para energia mais alta ($E_T > 60$ GeV).

⁴⁵ O primeiro registro do Bóson de Higgs (2012) ocorreu com o auxílio de informações contidas em canais de elétrons, múons e fôtons isolados (FREUND; PINTO; SEIXAS, 2016).

Figura 13 – Distribuição elétrons e jatos em função da energia total do evento.

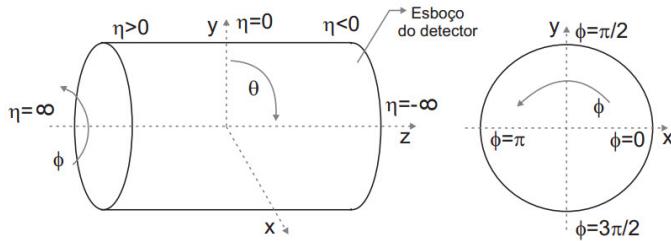


Fonte: SIMAS FILHO (2010)

O detector ATLAS possui um sistema de coordenadas cilíndricas representado na Figura 14a. Na Figura 14b detalhe para a representação dos ângulos θ e ϕ e a pseudo-rapidez. O eixo z indica a direção de propagação do feixe de partículas, x e y descrevem o plano transversal ao feixe.

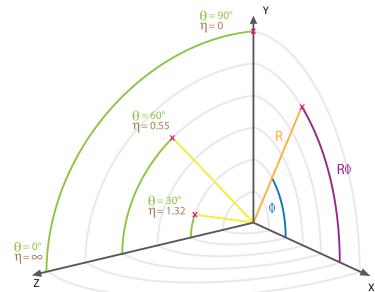
Figura 14 – Representação do sistema de coordenadas do detector. O centro do sistema de coordenadas refere-se ao local da colisão, e os ângulos as possíveis trajetórias das partículas resultantes.

(a) Diagrama do sistema de coordenadas do detector.



Fonte: Extraído de Souza (2015)

(b) Pseudo-rapidez (η) e ângulos θ e ϕ .



Fonte: Extraído de Lenzi (2013)

Baseado no sistema de coordenadas, as regiões dentro do detector que indicam a posição onde ocorrem as colisões podem ser definidas com os parâmetros: ϕ definido pela Equação 2.1, que indica a rotação em torno do eixo de colisão; o ângulo polar θ (Equação 2.2) define a pseudo-rapidez η na Equação 2.3 a qual representa a direção de propagação das partículas após a colisão, e a energia transversa E_T , definida pela Equação 2.4.

$$\phi = \arctan\left(\frac{x}{y}\right), \quad (2.1)$$

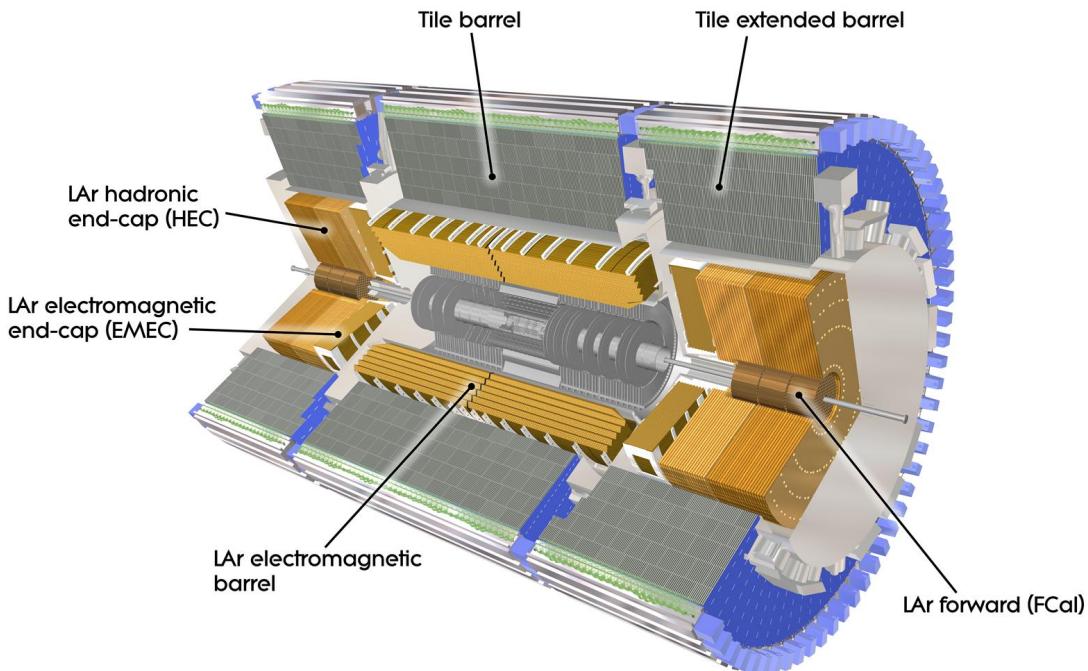
$$\theta = \arctan\left(\frac{x}{z}\right), \quad (2.2)$$

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

$$E_T = E \sin(\theta). \quad (2.4)$$

Na Figura 15 pode-se observar os calorímetros e as regiões internas do detector. Tomando o centro do detector como referência e se afastando do centro sobre o eixo z (Figura 14a) até as extremidades, são quatro partes: o Barril (EMB⁴⁶), ao centro, o barril estendido, as tampas (EMEC⁴⁷) e os calorímetros FCAL⁴⁸. Cada região dessas possui segmentação específica. O barril abrange a faixa de $|\eta| < 1,52$ e possui a maior granularidade, 112.448 canais de leitura dos 187.652. Pois é nessa região que os feixes de prótons são postos em colisão e espera-se que o maior número de colisões frontais ocorra, liberando a energia máxima, produzindo grande número de partículas. A segunda região, é a do barril estendido, e abrange a região de $0,8 < |\eta| < 1,7$ com 2.304 canais de leitura (em ambos os lados). As tampas abrangem a faixa de $1,375 < |\eta| < 3,2$, e o FCAL faixa de $3,1 < |\eta| < 4,9$ (ATLAS Collaboration, 2017).

Figura 15 – Diagrama ilustrando a segmentação do detector e seus diversos Calorímetros. Barril: EMB (*Tile Barrel* e *LAr Electromagnetic barrel*); Barril Extendido: *Tile extended barrel*; EMEC (*LAr forward* e *LAr hadronic end-cap*) e o Calorímetro de Telhas (*Tile Calorimeter*)



Fonte: ATLAS Collaboration (2008b)

Devido a estrutura altamente segmentada e complexa do detector, existem três regiões simétricas em ϕ onde existe um número reduzido de sensores. Isso produz degradação no registro das leituras das colisões para $|\eta| < 0,02$; $1,34 < |\eta| < 1,54$, e $2,47 < |\eta| < 2,5$. A causa dessa degradação é uma fissura (*crack*) na junção das partes do barril para passagem de cabos e outra fissura próxima às tampas, respectivamente (FREUND, 2018).

⁴⁶ *Electromagnetic Barrel*

⁴⁷ *Electromagnetic Endcap*

⁴⁸ *Forward Calorimeter*

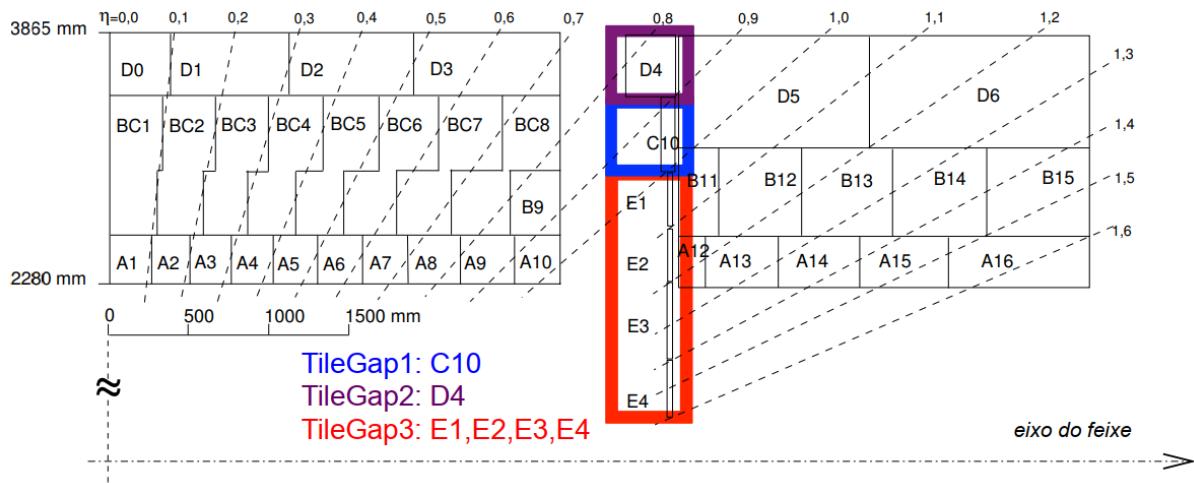
A seguir, na Tabela 1, é apresentada a localização de alguns módulos dos calorímetros do detector ATLAS, indicando e a região de $|\eta|$ correspondente (ATLAS Collaboration, 2017).

Tabela 1 – Regiões de $|\eta|$ onde estão localizados os subdetectores do ATLAS.

Calorímetros e região de $ \eta $.	
Calorímetro	faixa de η
PS	$ \eta < 1,52$
EMB1	$ \eta < 1,4$
EMB2	$ \eta < 1,4$
EMB3	$ \eta < 1,35$
EMEC	$1,375 < \eta < 3,2$
FCAL	$3,1 < \eta < 4,9$
HEC	$1,5 < \eta < 3,2$

A seguir, na Figura 16, um corte com detalhe da segmentação na estrutura do calorímetro do detector é apresentado. Na região mais próxima de $|\eta| = 0$ a granularidade aumenta e a medida que $|\eta|$ cresce o número de células vai reduzindo. As linhas tracejadas referem-se aos valores de η constante. A parte colorida em destaque representa a região de *crack* do detector. À direita da região do *crack* é o barril estendido, com segmentação reduzida em relação ao barril, lado esquerdo (FREUND, 2018).

Figura 16 – Diagrama de um corte ilustrando a segmentação do barril (esquerda) e barril estendido (direita).

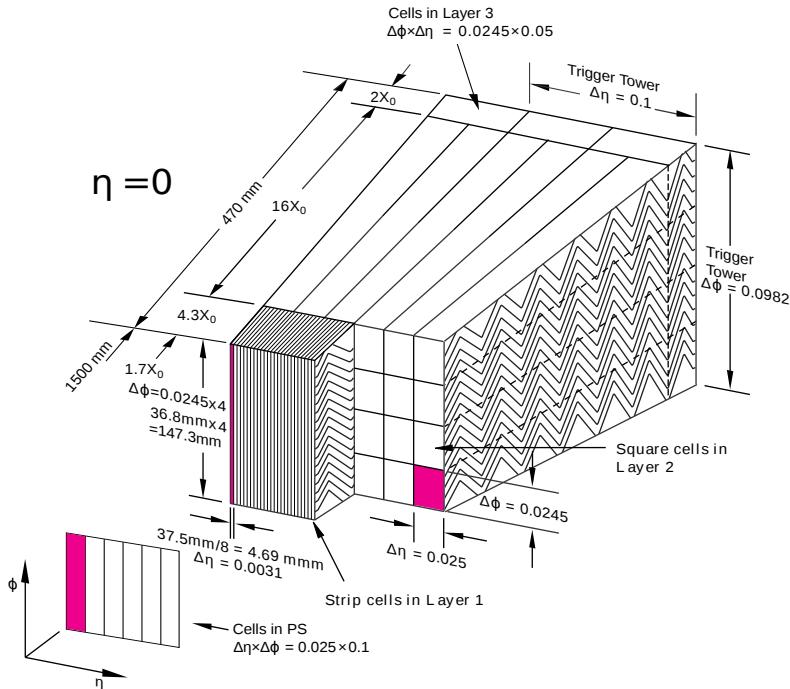


Fonte: Freund (2018)

Na região mais próxima ao feixe de prótons está localizado o calorímetro eletromagnético (ECAL⁴⁹), um dos constituintes do barril. Sua estrutura é em formato de acordeão, o que permite cobrir toda a região em ϕ , obter uma estrutura de camadas com granularidade diferente e evitar fissuras (*gaps*) as quais degradam a resposta do detector, ver Figura 17.

⁴⁹ *Electromagnetic Calorimeter*

Figura 17 – Diagrama ilustrando a estrutura em acordeão do ECAL.



Fonte: ATLAS Collaboration (2016)

Os experimentos observados no detector ATLAS são de alta energia e associado a essa alta energia é necessário definir e obter o parâmetro chamado luminosidade, a qual é a medida do número de colisões por centímetro quadrado produzida a cada segundo, definida pela Equação 2.5 (NÓBREGA; MACKEDANZ, 2013).

$$\mathcal{L} = n \left(\frac{N_1 N_2}{A} \right) f [cm^{-2}s^{-1}]. \quad (2.5)$$

na qual n é o número de feixes de partículas, N_i o número de partículas por feixe, A , a área de secção transversal do feixe e f a frequência de colisão.

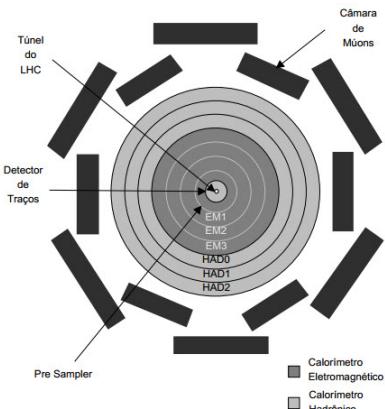
2.2.1.1 Sistema de Construção dos Anéis

Essa técnica realiza um pré-processamento organizando a informação do evento registrado pelos calorímetros em 100 anéis concêntricos distribuídos ao longo das 7 camadas do detector. Na Figura 18a detalhe para as camadas internas do detector e mais externamente a câmara de mísulas. Na Figura 18b é ilustrada uma possível trajetória das partículas produzidas em uma colisão no interior do detector ATLAS.

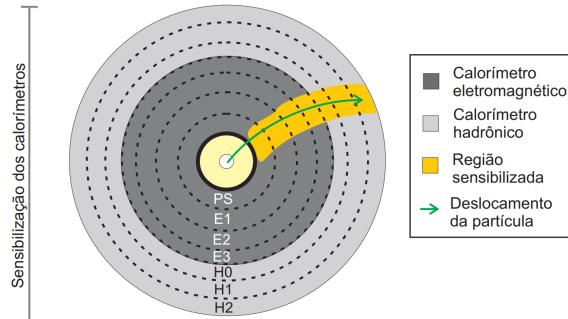
Dentre os sensores da região sensibilizada, o algoritmo detecta a célula de maior deposição de energia, pois essa célula dará origem ao primeiro anel, e consequentemente o centro da ROI. Em seguida, as células adjacentes à ROI formam o segundo anel e assim sucessivamente até completar 100 anéis concêntricos ao longo das camadas do detector. A informação de energia registrada em cada anel é somada, concatenada num vetor de 100 posições e normalizada. É esse vetor, contendo a representação do perfil de energia depositado durante a colisão que será utilizado como entrada do classificador neural.

Figura 18 – Diagrama 18a de um corte transversal do detector e suas camadas e 18b representação de uma possível trajetória de um evento no interior do detector.

(a) Camadas do detector ATLAS.



(b) Possível trajetória de um evento.



Fonte: SIMAS FILHO (2010)

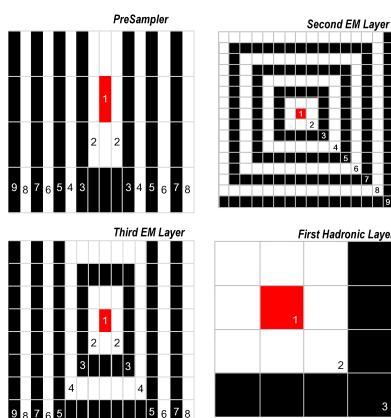
A seguir, na **Tabela 2** é exibido o número de anéis em cada uma das camadas do detector. O número de anéis diferente para cada camada é justificado pela diferente granularidade da estrutura do detector.

Tabela 2 – Número de anéis por camada. PS - *Presampler*; EMB1 – EMB3: Camadas Eletromagnéticas; HEC0 – HEC2: Camadas Hadrônicas.

Camada	PS	EMB1	EMB2	EMB3	HEC0	HEC1	HEC2
Anéis	08	64	08	08	04	04	04

Na **Figura 19** é possível visualizar a representação dos anéis produzidos com base das informações provenientes da trajetória registrada pelos calorímetros ao longo das camadas do detector.

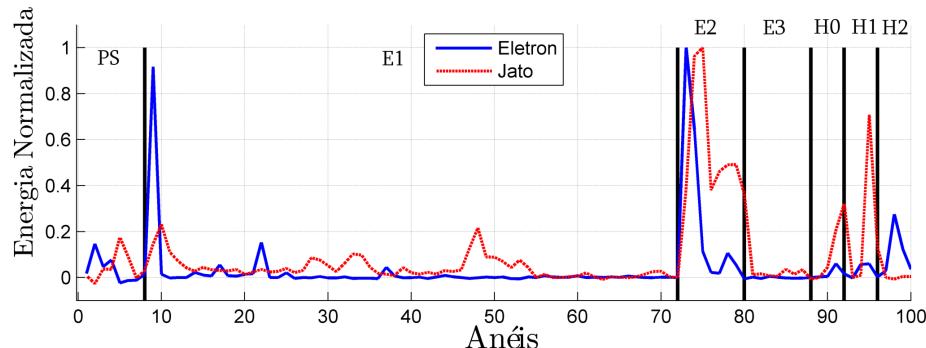
Figura 19 – Representação dos anéis do calorímetro nas RoIs.



Fonte: Adaptado de Freund (2011)

Na Figura 20 são exibidos exemplos de perfis de deposição de energia para elétrons e jatos provenientes do processo de anelamento da informação de uma colisão.

Figura 20 – Exemplos típicos de assinaturas para elétrons e jatos de amostras experimentais, obtidas do calorímetro formatado em anéis.



Fonte: Dados experimentais.

2.2.1.2 Sistema de Seleção ou Filtragem *Online*

O sistema de seleção ou filtragem *online* (*trigger*) do ATLAS (ACHENBACH et al., 2008) é responsável pela seleção dos eventos interessantes para o experimento e, também pela redução do ruído de fundo (assinaturas não relevantes) produzido nas colisões. Sua estrutura é composta de uma camada (*Level 1*) de *hardware* dedicado, a qual é responsável pelo primeiro estágio de filtragem, e outra camada de *software*, na qual operam os discriminadores.

O primeiro discriminador (elétron/jato) utilizado, anteriormente, para a filtragem *online*, no ATLAS, foi o T2Calo (ATLAS Collaboration, 2003). Nesse discriminador são realizados cortes lineares dos níveis de energia agrupadas nas variáveis R_{CORE} ⁵⁰, E_{RATIO} ⁵¹, E_{HAD} ⁵² e R_{TEM} ⁵³ (CIODARO, 2012).

O segundo discriminador utilizado para aumentar a eficiência na identificação de elétrons no ATLAS é o *Neural Ringer* (NR) (ANJOS et al., 2006). Neste discriminador, a informação de uma colisão registrada pelos sensores dos calorímetros é organizada num vetor de 100 posições, as quais representam a intensidade de energia depositada nas camadas do detector ao longo da trajetória percorrida pela partícula. Conforme apresentado na Figura 12. Essa informação organizada em um vetor é utilizada como entrada de um sistema de classificação baseado numa rede neural artificial tipo perceptron de múltiplas camadas (*multi-layer perceptron* - MLP) totalmente conectada (HAYKIN, 2008).

Na configuração atual do NR, utiliza-se um conjunto (*ensemble*) de redes neurais artificiais especialistas . Para o projeto de tais redes, o detector é subdividido em regiões de $|\eta|$, associadas a níveis de energia transversa (E_T), sendo que para cada par ($|\eta|$, E_T), uma rede neural é projetada.

⁵⁰ R_{CORE} : razão de núcleo, calculada na EMB2.

⁵¹ E_{RATIO} : Razão de energia, calculada na EMB1.

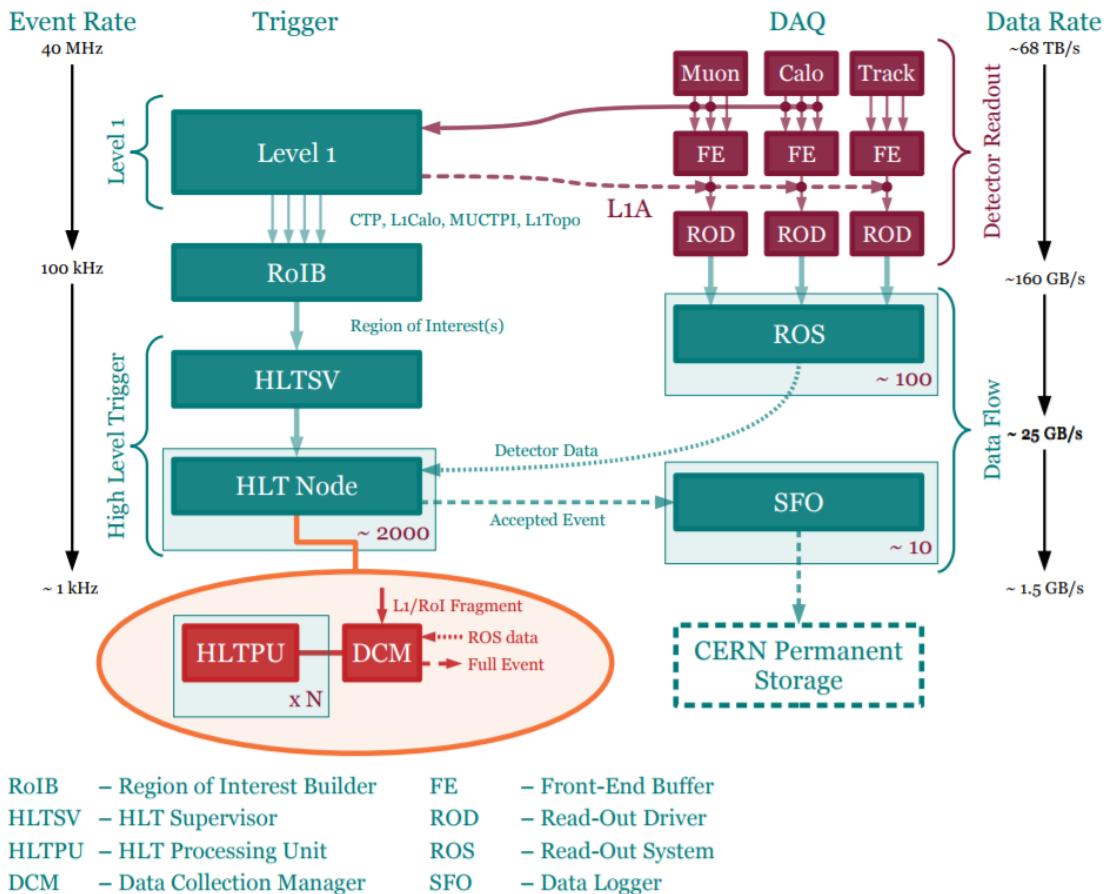
⁵² E_{HAD} : fração entre a quantidade de energia depositada nas três primeiras camadas do HCAL e a quantidade nas três camadas do EMB.

⁵³ R_{TEM} : soma de E_T em todas as células das três camadas do EMB.

Dessa forma, o projeto do classificador neural torna-se denso, visto que são necessárias diversas redes neurais para o processo de classificação (FREUND, 2018).

Após a parada para atualizações entre os anos de 2013 e 2014, o sistema de *trigger* do detector ATLAS passou por atualizações significativas em seus subsistemas tanto em *hardware*, quanto em *software*, ver Figura 21. Essas atualizações elevaram o número de colisões por feixe, nível de luminosidade e as taxas aplicadas ao *trigger*, elevando a frequência de entrada aceita do L1 de 75 kHz para 100 KHz e frequência de saída de 300 Hz para 1 kHz, além disso, o EF que antes era separado do L2 passa a integrar uma única etapa do HLT (*High Level Trigger*), o que reduz a complexidade e melhora a dinâmica dos algoritmos (GALSTER, 2015; KILBY; ATLAS Collaboration, 2016; MARTÍNEZ; ATLAS Collaboration, 2016; VAZQUEZ, 2016).

Figura 21 – Esquema do *trigger online* do ATLAS após última atualização no final de 2014.



Fonte: (GALSTER, 2015)

Uma questão associada ao *Neural Ringer* é o elevado tempo de treinamento do sistema. Tal fato decorre do número elevado de inicializações necessárias para a obtenção da melhor rede. Cada estrutura de rede é inicializada k^{54} centena de vezes. Esse método visa reduzir problemas associados a mínimos locais e oscilações na estatística da base de dado utilizada. Também é

⁵⁴ O valor de k depende da técnica de reamostragem utilizada. Para k -fold, utiliza-se $k = 50$; para *Jackknife*, utiliza-se $k = 10$.

importante citar, que as bases de dados utilizadas no processo de projeto dos classificadores, possuem dimensão elevada, com número de assinaturas com ordem de grandeza acima de 10^5 .

Soma-se a esse fato as atualizações com incremento nos níveis de energia envolvidos em cada colisão no LHC, que implicam em atualizações do *trigger online*, necessárias ao processo de identificação dos decaimentos de interesse para a verificação de fenômenos previstos na física teórica (MOREIRA, 2009; PIMENTA et al., 2013).

3 Fundamentação Teórica

Este capítulo apresenta as técnicas utilizadas como classificadores neurais artificiais neste trabalho. O *perceptron* multicamadas (MLP), as máquinas de extremo (ELM) e as redes com estado de eco (ESN). Serão abordados as definições, suas características e exemplos de aplicação na literatura recente.

3.1 Redes Neurais Artificiais - RNA

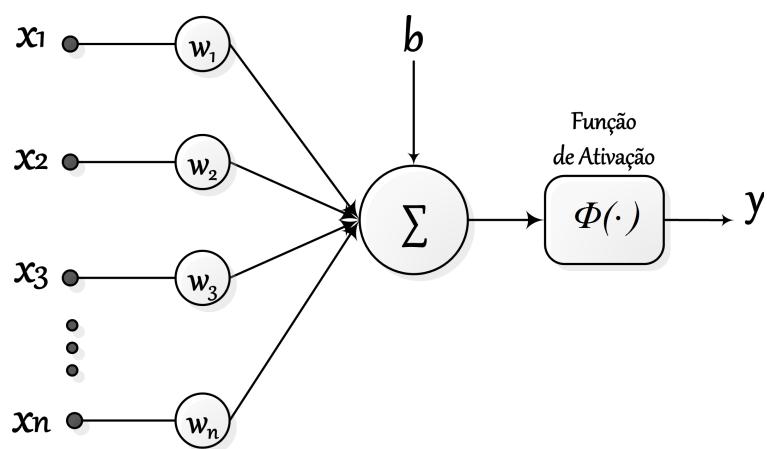
3.1.1 Definições

São sistemas paralelos, distribuídos, compostos por unidades de processamento simples (neurônios artificiais) que calculam determinadas funções matemáticas (normalmente não-lineares) (BRAGA; CARVALHO; LUDERMIR, 2007).

Uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que tem a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso (HAYKIN, 2008, p. 2).

Um neurônio é uma unidade de processamento de informação que é fundamental às operações de uma RNA (HAYKIN, 2008). Na Figura 22 é apresentada a representação de um neurônio artificial.

Figura 22 – Diagrama do modelo matemático de um neurônio artificial, o *Perceptron*.



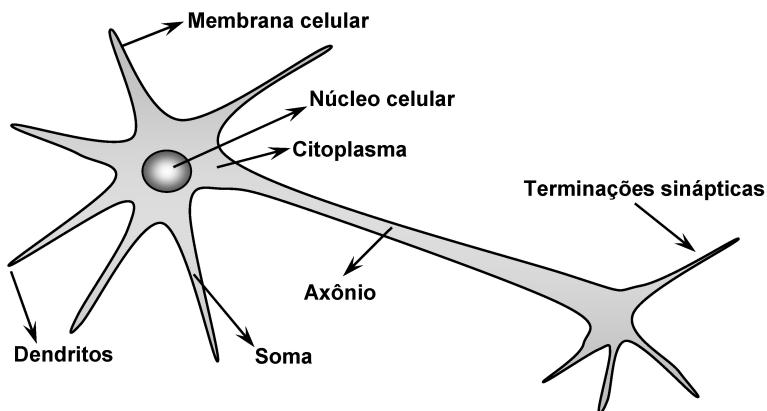
Uma rede neural é constituída de um conjunto de neurônios artificiais que podem ter seu modelo matemático dado pela Equação 3.1. O sinal b (*bias* - viés) é um parâmetro livre de ajuste da rede; Φ é a função de ativação; w_i é o vetor de pesos e x_i é o vetor de sinais de entrada da rede. Um diagrama representativo é apresentado na Figura 22. Esse modelo busca se aproximar do modelo de um neurônio biológico Figura 23, sendo as sinapses representadas pelos pesos

atribuídos à cada entrada, informações vindas de outros neurônios ou dos neurotransmissores espalhados pelo corpo.

$$y[n] = \Phi\left(\sum_{i=1}^n w_i x[i] + b\right). \quad (3.1)$$

Logo, uma RNA nada mais é mais do que o encadeamento de neurônios artificiais, de maneira análoga ao modelo de rede neural utilizado para o cérebro, Figura 23, em escala reduzida, mas mantendo o mesmo princípio, de processamento paralelo e distribuído.

Figura 23 – Ilustração de um modelo de neurônio biológico.



Fonte: (BARRA, 2013)

3.1.2 Estruturas

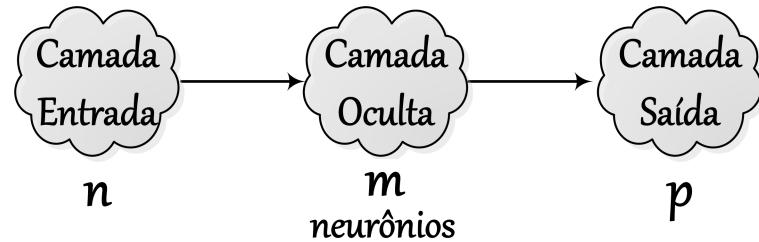
Desde os primeiros estudos sobre redes neurais, e o primeiro neurônio artificial desenvolvido, o *perceptron*⁵⁵ as estruturas de uma rede neural podem ser classificadas em dois tipos (BOCCATO, 2013; HAYKIN, 2008), as redes do tipo em avanço (do inglês: *feedforward*) e as redes recorrentes, cada uma dessas estruturas com suas variantes.

3.1.2.1 Redes em Avanço

Nas redes do tipo em avanço (do inglês: *feedforward*), ver Figura 24, o sinal proveniente das entradas percorre a estrutura da rede num único sentido. Seguem da entrada para a saída sem nenhuma etapa de realimentação, ou seja, as saídas de uma camada não interferem em suas entradas, ou camadas imediatamente anteriores.

⁵⁵ O tipo de classificador neural *feedforward*, linear, mais simples desenvolvido por Frank Rosenblatt (1928-1971) em 1957.

Figura 24 – Diagrama de uma rede *feedforward* com o sentido de fluxo da informação.



Algumas variações para as redes em avanço:

- Uma ou mais camadas ocultas;
- Ser totalmente conectada, ou seja, a saída de cada neurônio da camada imediatamente anterior será entrada de todos os neurônios da camada imediatamente posterior;
- Parcialmente conectada, alguns neurônios não recebem o sinal de saída da camada imediatamente posterior;

Duas das estruturas que serão utilizadas neste trabalho a MLP (do inglês: *MLP - Perceptron Multilayer* - Perceptron Multicamadas) e a ELM, são estruturas de rede em avanço.

3.1.2.2 Redes Recorrentes

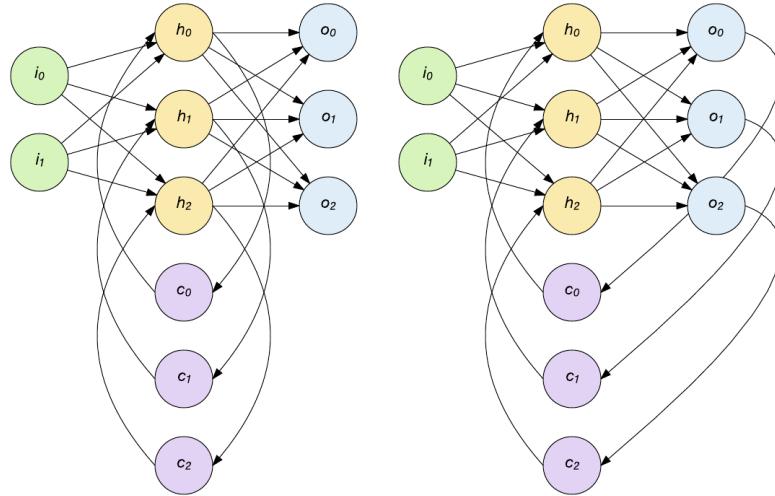
As redes recorrentes são estruturas de redes que possuem pelo menos um laço de realimentação em sua topologia (HAYKIN, 2008). Essa estrutura se assemelha ao modelo das conexões entre os neurônios biológicos, e esse fato possibilita à rede ter uma capacidade de memória. Isso decorre do fato de que, em cada novo sinal fornecido aos neurônios da rede, existe a informação que foi processada no instante imediatamente anterior. O que se deve aos laços de realimentação e capacidade de aproximação universal. Características que as tornam ferramentas eficientes no processamento de sinais e tratamento de problemas dinâmicos (BOCCATO, 2013).

Algumas variações para as redes recorrentes (JONES, 2017):

- Redes de Hopfield, estrutura com laços de realimentação entre todos os neurônios, [Figura 25b](#);
- Redes de Elman, não possui laços de realimentação da saída para o resto da rede, [Figura 25a](#);
- Redes de Jordan, existem laços de realimentação da camada de saída somente para a camada de oculta, [Figura 25a](#);
- Redes com Estados de Eco, [Figura 25c](#).

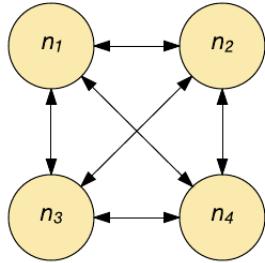
Figura 25 – Exemplos de redes recurrentes

(a) Elman (E) e Jordan (D)



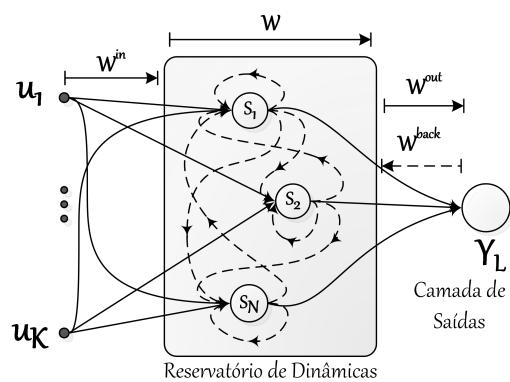
Fonte: Adaptado de Jones (2017)

(b) Hopfield



Fonte: Adaptado de Jones (2017)

(c) rede ESN



3.1.3 Características

As RNA possuem propriedades úteis, dentre as quais podemos destacar (HAYKIN, 2008):

- Não-linearidade - Podem trabalhar tanto com funções lineares, quanto não-lineares;
- Capacidade de generalização - Produz saídas adequadas para sinais que não estavam presentes no momento do treinamento;
- Capacidade de adaptação - Uma RNA treinada para uma determinada tarefa pode ter seus pesos sinápticos atualizados com esforço reduzido;
- Tolerância a falhas - Devido à sua característica distribuída uma RNA só terá seu desempenho degradado significativamente caso ocorra uma falha relevante, no sinal de entrada ou em seus ramos de conexão entre camadas.

Sua aplicação é de grande valia onde não se conhece o modelo dinâmico do sistema, ou quando não é possível obtê-lo.

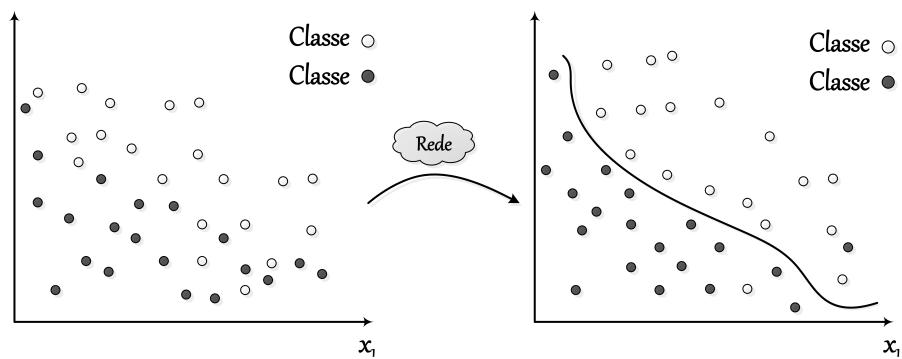
As redes neurais têm aplicações em sistemas onde se deseja obter o reconhecimento/identificação de padrões, onde o processamento de sinal torna-se complexo, no que se refere à capacidade de separação das características de interesse.

As principais tarefas que uma RNA pode executar, segundo Braga, Carvalho e Ludermir (2007) são:

- Classificação - separar classes ou atribuir uma classe a um padrão desconhecido (Figura 26).
Ex: Reconhecimento de caracteres;
- Categorização (*clustering*) - típico de aprendizado não-supervisionado, visa identificar as classes/categorias dentro do conjunto de dados. Ex: Agrupamento de clientes;
- Previsão - estimativa de funções, tomando por base o estado atual e anteriores. Ex: Previsão do tempo;
- Regressão - Ferramenta estatística para obtenção de um modelo representativo (aproximado) das relações existentes entre as variáveis de um sistema.

Na Figura 26, há uma representação do resultado após aplicação de amostras contendo características de duas classes a serem separadas por uma rede neural. A rede neural age como um operador matemático realizando uma transformação, de forma a organizar os sinais de tal maneira que seja possível gerar um hiperplano que separe cada classe do problema em questão. Esse mesmo princípio é aplicado para problemas de complexidade elevada, com número de classes superior a dois. E o processo de ajuste do número de neurônios, bem como a arquitetura da rede utilizada são determinados de forma experimental, ajustando cada parâmetro até o atendimento das especificações mínimas do problema.

Figura 26 – Rede neural, na separação de classes.



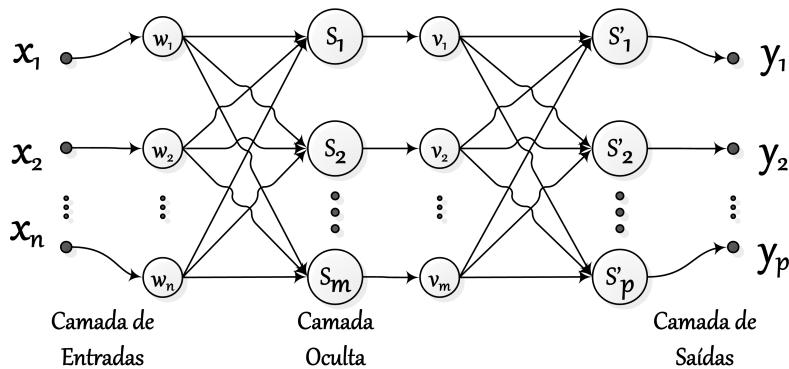
Para a aplicação de uma RNA em qualquer tipo de problema, é necessário treiná-la. Ou seja, deve-se fornecer exemplos com características relevantes das classes que a RNA deve identificar. E para essa tarefa existem dois métodos de treinamento: o supervisionado ou não-supervisionado. No supervisionado, são apresentadas à rede amostras com características relevantes do padrão/classe a ser identificado, bem como qual classificação amostra deve receber,

ou seja, são fornecidos os padrões de entrada e saída (HAYKIN, 2008; BRAGA; CARVALHO; LUDERMIR, 2007).

No treinamento não-supervisionado, não é fornecida à RNA uma tabela de entradas e saídas. O treinamento envolve o processo iterativo de atualização dos pesos sinápticos, com base na informação apresentada à rede (HAYKIN, 2008; BRAGA; CARVALHO; LUDERMIR, 2007).

Na rede ilustrada na Figura 27, cada neurônio das camadas oculta e de saída possuem modelo matemático descrito pela Equação 3.2 e pela Equação 3.3, respectivamente. Logo, para a camada oculta são necessárias $m \times n$ operações de soma e $m \times n$ operações produto, e de igual modo, na camada de saída $p \times m$ operações de soma e $p \times m$ operações de produtos.

Figura 27 – Rede *feedforward* - totalmente conectada - pesos w_n e v_m representam vetores de pesos, para simplificar o diagrama.



Nessa estrutura de rede, o número de operações de soma e operações de produto realizadas em cada uma de suas camadas pode ser determinado observando-se o número de neurônios de suas camadas. Os parâmetros livres (*bias*) foram omitidos para simplificação do diagrama.

A de se observar que, a cada neurônio adicionado à rede visando a elevação da taxa de acerto, são adicionadas n operações de soma e n operações de produto, realizadas na camada oculta. Na camada de saída, m operações de soma e m operações de produto. Essa elevação do número de neurônios implica em aumento da complexidade da RNA (OLIVEIRA; BUTKEWITSCH, 2000; REYES; VELLASCO; TANSCHEIT, 2012) que resulta em elevação do custo computacional. Outro fator relevante diz respeito à capacidade de generalização da rede, que pode ser comprometida com o aumento indiscriminado do número de neurônios, ocasionando resultados indesejáveis conhecidos como *overfitting*.

$$S_m = \Phi\left(\left[\sum_{i=1}^n w_i x[i] = w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots + w_n x_n\right] + b_m\right) \quad (3.2)$$

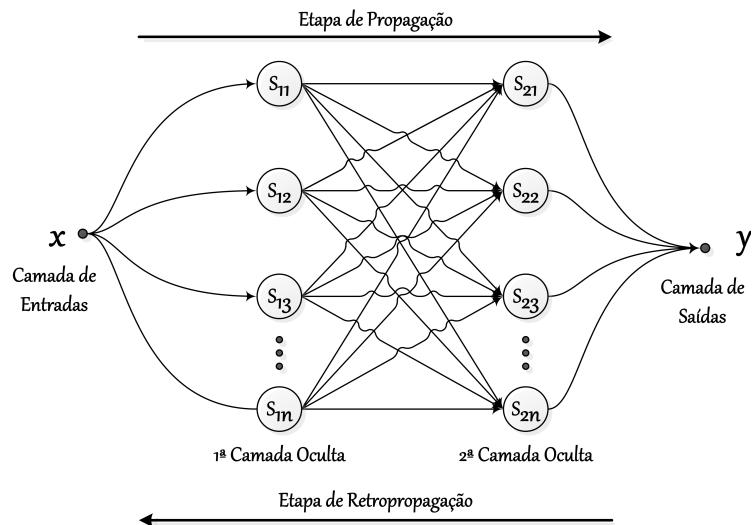
$$S'_p = \Phi\left(\left[\sum_{j=1}^m v_j S[j] = v_1 S_1 + v_2 S_2 + v_3 S_3 + \dots + v_m S_m\right] + b_p\right) \quad (3.3)$$

Para que uma RNA seja utilizada é necessário que essa esteja treinada, e atendendo a critérios pré-estabelecidos, relativos à cada situação onde uma RNA é utilizada. O critério

de treinamento mais utilizado é o de critério de erro de saída. O sinal de saída de uma RNA é comparado com o resultado desejado, e caso a tolerância para o erro não seja atendida, o algoritmo ajusta os pesos sinápticos até que o critério de erro seja satisfeito.

Para uma RNA multicamadas o algoritmo de treinamento supervisionado mais utilizado é o *Backpropagation*. Esse algoritmo é dividido em duas etapas, uma chamada propagação, e uma retropropagação. A etapa de propagação consiste em aplicar um padrão à entrada da RNA, até obter o sinal de saída respectivo. A etapa de retropropagação, consiste no ajuste dos pesos sinápticos começando da última camada da RNA, em direção à camada de entrada, conforme indicado na [Figura 28](#). Após essas duas etapas estarem completas, o segundo padrão é apresentado à RNA e a partir desse instante o processo se repete até que o critério de erro seja atendido.

Figura 28 – Rede neural, com duas camadas ocultas, representação do algoritmo *backpropagation* - Retropropagação em representação simplificada sem pesos sinápticos.



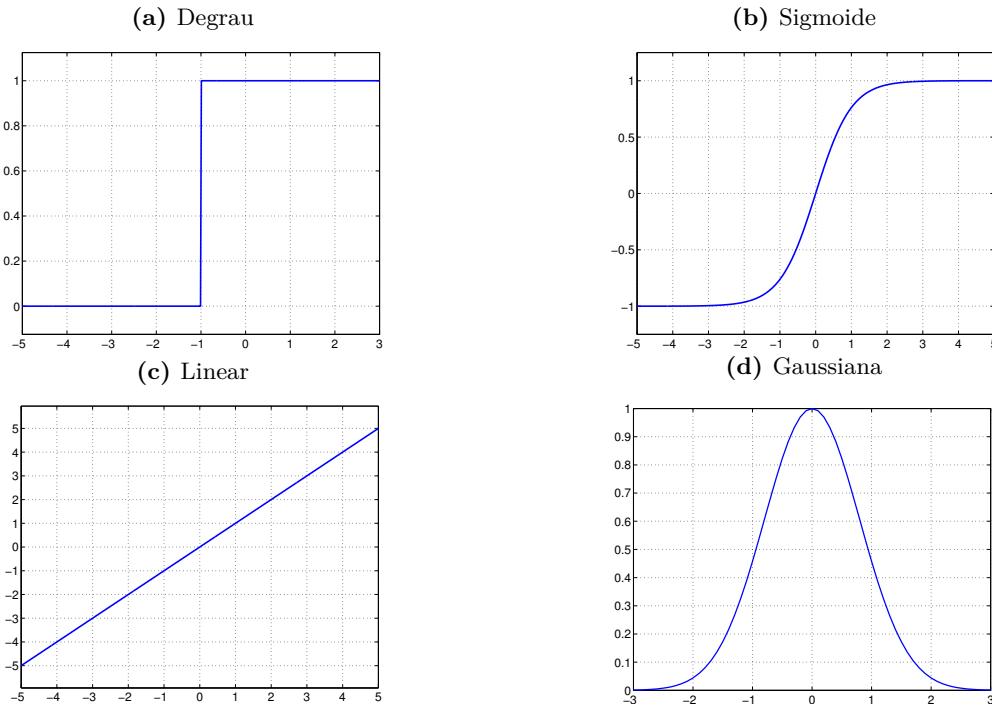
As funções de ativação são responsáveis por gerar a saída y de cada neurônio, a partir dos valores dos pesos $w = (w_1, w_2, w_3, \dots, w_n)^T$ e as entradas $x = (x_1, x_2, x_3, \dots, x_n)$ (BRAGA; CARVALHO; LUDERMIR, 2007). Na [Figura 29](#), exemplos de algumas funções de ativação utilizadas em neurônios artificiais. As expressões analíticas correspondentes às funções de ativação são apresentadas nas Equações 3.4 a 3.7, respectivamente.

$$f(u) = \begin{cases} 1 & \sum_{i=1}^n x_i w_i \geq \theta \\ 0 & \sum_{i=1}^n x_i w_i < \theta. \end{cases} \quad (3.4)$$

$$f(u) = \frac{1}{1 + e^{-\beta u}}. \quad (3.5)$$

$$f(u) = u. \quad (3.6)$$

$$f(u) = e^{\frac{-(u-\mu)^2}{\sigma^2}}. \quad (3.7)$$

Figura 29 – Exemplos de funções de ativação.

Na Equação 3.5, β representa a inclinação da curva. Na Equação 3.7, μ é o centro, e σ , o desvio padrão.

3.1.4 Exemplos de Aplicações

Em Dvorkin, Goldschmit e Storti (2010) foi aplicada, para reconhecimento de acordes, uma RNA perceptron de multicamadas, com uma camada oculta contendo 61 neurônios e uma de saída. Foi utilizado um teclado Yamaha® PSR-E4313, que foi configurado para reproduzir o som de um piano, um cravo, um órgão e um violão. Com esses timbres foi montado um banco de acordes com 144 amostras gravadas.

Em Soares et al. (2011) uma rede MLP foi utilizada para predição e estimação do diâmetros de árvores de eucalipto para a extração de madeira de qualidade no momento em que as árvores estão prontas para a colheita.

Em Freund (2011) um classificador neural numa rede com estrutura MLP, com uma camada oculta, totalmente conectada, foi desenvolvido para a classificação de elétrons/jatos e utilizada no sistema de *trigger* do detector ATLAS. O desempenho obtido pelo classificador proposto superou o algoritmo padrão utilizado pela colaboração ATLAS em três bases de dados utilizadas para o seu desenvolvimento.

Em Santos (2014) uma rede neural do tipo MLP com uma camada oculta foi utilizada para classificação de acordes naturais de guitarra. Nesse sistema foi utilizado como pré-processamento a *chroma feature* para obtenção de um vetor característico para cada acorde, o qual continha a contribuição de cada uma das doze componentes (notas) constituintes na escala cromática. Os melhores resultados obtidos foram utilizando 16 neurônios na camada oculta com desempenho global de 94,32%.

Em Souza et al. (2014a) foi proposto um discriminador neural para realizar a detecção de partículas eletromagnéticas (elétrons e fótons) no segundo nível de *trigger online* de eventos do detector ATLAS. Para tanto, foi utilizada uma combinação de técnicas de extração de características, tais como DWT (*Discret Wavelet Transform* - Transformada Discreta de Wavelet), PCA (*Principal Analysis Component* - Análise de Componentes Principais) e ICA (*Independent Component Analysis* - Análise de Componentes Independentes) com classificadores neurais. Os resultados obtidos foram semelhantes ao classificador *Neural Ringer* sem pré-processamento, possibilitando a redução do número de componentes utilizados em até 80%.

Em Souza et al. (2014b) foi proposta uma arquitetura de classificação via rede neural segmentada também para o problema de detecção *online* de elétrons no ATLAS. A informação proveniente de cada camada do calorímetro é processada separadamente e utilizada para alimentar classificadores neurais (num total de sete, um para cada camada). As saídas de cada classificador segmentado são utilizadas para alimentar uma outra rede neural (formando um segundo estágio de classificação), que combina as características segmentadas para produzir a decisão final.

Em Fernandes Jr. e Damiani (2014), uma RNA foi utilizada num trabalho cujo objetivo foi a extração de tempo musical utilizando transformada *Wavelet* e rede neural artificial. Foi desenvolvido um método para detecção de tempo, batidas por minuto (bpm) de uma música, onde a transformada *Wavelet* foi utilizada para a construção de funções de detecção de *onsets*⁵⁶. E uma rede neural de uma camada oculta, do tipo *feedforward*, foi utilizada para mapear os descritores multiresolucionais, no tempo musical correspondente.

Em Freund, Pinto e SEIXAS (2016) é descrita uma arquitetura em redes neurais, do tipo MLP, utilizada para seleção dos eventos no canal eletromagnético do detector ATLAS, utilizando a informação anelada de calorimetria. Utilizando dados provenientes da simulação Monte Carlo ⁵⁷ e validação cruzada, as redes obtiveram desempenho semelhante no final da cadeia de detecção, porém, atingiram uma redução de ~ 2 na taxa de Falso Alarme (FA).

3.2 Máquinas de Aprendizado Extremo - ELM

As máquinas de aprendizado extremo (*Extreme Learning Machines* - ELM) foram propostas inicialmente em Huang, Zhu e Siew (2004). Utilizando uma estrutura semelhante à de uma rede neural MLP com uma única camada oculta⁵⁸, ver Figura 30, o treinamento da ELM assume que é possível gerar aleatoriamente os pesos da camada de entradas e determinar analiticamente os melhores pesos para a camada oculta. Deste modo, o tempo de treinamento de uma ELM é consideravelmente reduzido, pois não existe um procedimento iterativo de retro-propagação de erro para o ajuste dos pesos do modelo.

Foi demonstrado que uma rede ELM, assim como uma rede MLP é um aproximador universal e possui capacidade de interpolação nos trabalhos de Huang, Zhu e Siew (2006), Huang,

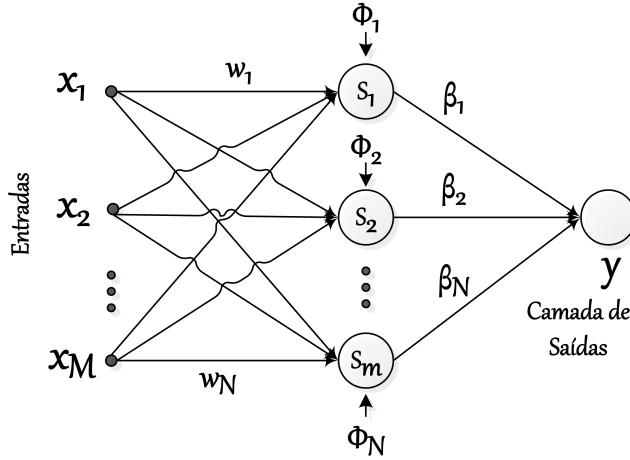
⁵⁶ Momento de início de uma nota, quando sua amplitude sai de zero a um valor de pico.

⁵⁷ Método estatístico de simulações baseadas no uso de sequências de números pseudo-aleatórios para resolução de problemas, em particular para estimar os parâmetros de uma distribuição desconhecida. Utilizado especialmente quando a complexidade do problema torna inviável a obtenção de uma solução analítica ou com métodos numéricos tradicionais (BRAIBANT; GIACOMELLI; SPURIO, 2012, p. 27).

⁵⁸ SLFN - Single Layer feedforward Networks

Wang e Lan (2011), Huang (2015), nos quais também são apresentadas variações nos modelos das redes ELM. Entretanto, em alguns casos, redes ELM comparadas com redes MLP requerem um número maior de neurônios na camada oculta para resolver, com desempenho equivalente, o mesmo problema (WANG; HUANG, 2005).

Figura 30 – Diagrama de uma ELM.



Para um conjunto de M pares entrada-saída $(\mathbf{x}_i, \mathbf{y}_i)$ com $\mathbf{x}_i \in \mathbb{R}^{d_1}$ e $\mathbf{y}_i \in \mathbb{R}^{d_2}$, a saída de uma SLFN com N neurônios na camada oculta é modelada pela Equação 3.8.

$$\mathbf{y}_j = \sum_{i=1}^N \beta_i \Phi(\mathbf{w}_i \mathbf{x}_j + b_i), \quad j \in [1, M] \quad (3.8)$$

sendo Φ a função de ativação, \mathbf{w}_i e b_i os pesos e o *bias* da camada de entrada, respectivamente, e β_i os pesos da camada de saída.

A equação 3.8 pode ser reescrita como $\mathbf{H}\beta = \mathbf{Y}$, sendo,

$$\mathbf{H} = \begin{pmatrix} \Phi(\mathbf{w}_1 \mathbf{x}_1 + b_1) & \dots & \Phi(\mathbf{w}_N \mathbf{x}_1 + b_N) \\ \vdots & \ddots & \vdots \\ \Phi(\mathbf{w}_1 \mathbf{x}_M + b_1) & \dots & \Phi(\mathbf{w}_N \mathbf{x}_M + b_N) \end{pmatrix}, \quad (3.9)$$

e $\beta = (\beta_1^T \dots \beta_N^T)^T$ e $\mathbf{Y} = (y_1^T \dots y_M^T)^T$.

Como função de ativação, as redes ELM podem utilizar as mesmas funções aplicáveis às redes MLP, como por exemplo, linear, sigmoide, gaussiana, funções de base radial (do inglês: *Radial Basis Functions*-RBF).

A solução baseia-se em determinar a matriz inversa generalizada de Moore-Penrose de \mathbf{H} , definida como $\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$, que pode ser obtida por mínimos quadrados ordinários (do inglês: *Ordinary Least Squares* - OLS) ou via decomposição em valores singulares (do inglês: *Singular Value Decomposition* - SVD) (SOUTO, 2000; COLIBORO, 2008).

Na SVD uma matriz $\mathbf{A}_{m \times n}$ é decomposta da seguinte forma (SOUTO, 2000)

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \quad (3.10)$$

sendo $\mathbf{U}_{m \times m}$, $\Sigma_{m \times n}$ e $\mathbf{V}_{n \times n}$. A matriz Σ é da forma

$$\Sigma = \begin{pmatrix} \mathbf{D} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}, \quad (3.11)$$

com $\mathbf{D}_{p \times p}$ uma matriz diagonal formada pelos valores singulares da decomposição de \mathbf{A} , determinados por meio dos autovalores associados a matriz $\mathbf{A}^T\mathbf{A}$, tais que $\sigma_p = \sqrt{\lambda_p} \geq 0$, sendo σ_p o valor singular e λ_p o autovalor associado.

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \sigma_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & 0 \\ 0 & 0 & 0 & \dots & \sigma_p & 0 \end{pmatrix}, \quad p = \min\{m, n\}. \quad (3.12)$$

A inversa generalizada de Moore-Penrose, \mathbf{A}^\dagger , a partir de seus valores singulares é determinada da seguinte forma (MACAUSLAND, 2014):

$$\mathbf{A}^\dagger = \mathbf{V}\Sigma^+\mathbf{U}^T, \quad (3.13)$$

$$\Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & 0 \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_p} & 0 \end{pmatrix}^T. \quad (3.14)$$

A seguir, o resumo do processo de decomposição de valores singulares descrito em Coliboro (2008):

1. Calcular a matriz $\mathbf{A}^T\mathbf{A}$, seus autovalores e autovetores associados;
2. Montar a matriz $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_m]$ a partir dos autovetores de $\mathbf{A}^T\mathbf{A}$;
3. Calcular os valores singulares $\sigma_p = \sqrt{\lambda_p}$ e montar a matriz Σ ;
4. Calcular os vetores $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sigma_i}$, com $i \in \{1, 2, \dots, n\}$, e montar a matriz $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_n]$.

3.2.1 Exemplos de Aplicações

Em Wang e Huang (2005) redes ELM foram comparadas a redes MLP como classificadores de sequência de proteínas, neste trabalho o desempenho das redes ELM foi semelhante às redes MLP, tendo um tempo de treinamento pelo menos 180 vezes menor, porém com um número de neurônios na camada oculta (160) superior ao utilizado pelas redes MLP (35).

Em Zhang e Luo (2015) quatro modelos de redes ELM foram testadas quanto à robustez a *outliers*⁵⁹ no conjunto de dados. Uma rede em estrutura ELM clássica, e as outras três utilizando os multiplicadores de Lagrange para definição de um parâmetro de otimização: uma baseada no erro da rede (RELM⁶⁰); outro baseado na relação erro e pesos da rede (WRELM⁶¹) e a última associando a saída de referência e o erro (ORELM⁶²). Os testes de regressão, mostraram que a rede ORELM obteve o menor erro médio quadrático nos testes com contaminação por *outliers*. Nos problemas de classificação, a contaminação por *outliers* avaliada, foi de 0%, 10%, 20% e 40%. Nos conjuntos sem contaminação a rede que obteve o melhor desempenho foi a RELM. Nos testes com contaminação por *outliers* a ORLEM obteve o melhor desempenho em relação as demais.

Em Huang et al. (2015) pode-se verificar as variações da ELM, assim como a fundamentação matemática e a demonstração de algumas propriedades relevantes, como a capacidade de aproximação universal da ELM.

A ELM vem sendo utilizada em diferentes aplicações como, por exemplo, em Termenon et al. (2016), para desenvolver uma ferramenta de apoio à extração de características de imagens de ressonância magnética no diagnóstico de mal de Alzheimer. Em Horata, Chiewchanwattana e Sunat (2013), Barreto e Barros (2016), foi associada a Estimadores-M (RUCKSTUHL, 2014) como classificador robusto com baixa sensibilidade a *outliers*.

Em Qu et al. (2016) uma estrutura com duas camadas foi avaliada e comparada em problemas de regressão e classificação sendo observado que a estrutura torna-se interessante para problemas complexos na presença de recursos computacionais de armazenamento limitados.

Outros trabalhos já foram desenvolvidos onde apresentam estudos para melhoria da ELM quanto a robustez a *outliers* e problemas computacionais quando a matriz de saída da camada oculta não possui posto completo (HORATA; CHIEWCHANWATTANA; SUNAT, 2013) baseados em estimadores M⁶³.

3.3 Redes com Estado de Eco - ESN

As redes com estados de eco (ESN) são redes neurais compostas por: uma camada de entradas; uma camada interna denominada reservatório de dinâmicas (RD), constituída de neurônios organizados numa estrutura recorrente totalmente conectados utilizando funções de ativação não-linear; e uma camada de saídas de característica linear a qual tem seu resultado obtido de maneira semelhante ao que ocorre com a ELM, por meio da inversa generalizada de

⁵⁹ Amostras de valores discrepantes em relação ao conjunto de dados analisados.

⁶⁰ Regularized ELM.

⁶¹ Weighted Regularized ELM.

⁶² Outliers-robust ELM.

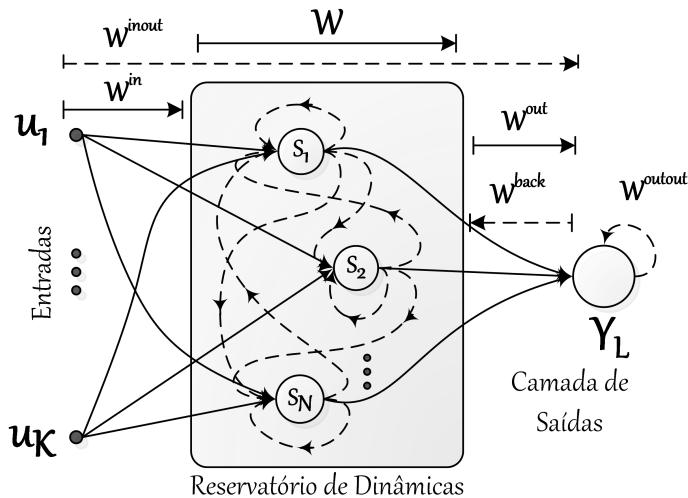
⁶³ Maximum likelihood estimator - Estimador de Máxima Verossimilhança

Moore Penrose, ou método de regressão linear dos mínimos quadrados, por exemplo, (JAEGER, 2001).

Na Figura 31 é exibido um diagrama genérico de uma rede ESN, indicando todas as possíveis conexões entre as camadas da rede, a saber:

- \mathbf{W}^{in} - matriz de pesos da camada de entrada para o RDs;
- \mathbf{W}^{inout} - matriz de pesos da camada de entrada para a camada de saída;
- \mathbf{W} - matriz de pesos do RD;
- \mathbf{W}^{out} - matriz de pesos da camada de entrada para o RD;
- \mathbf{W}^{back} - matriz de pesos (realimentação) da camada de saída para o RD;
- \mathbf{W}^{outout} - matriz de pesos da camada de saída para a camada de saída.

Figura 31 – Diagrama genérico de uma rede ESN, indicando os possíveis laços de realimentação .



Para a rede genérica da Figura 31, na qual o RD é uma camada totalmente conectada formada de elementos de função de ativação não-linear, a atualização dos estados é definida segundo as Equações 3.15 e 3.16.

Os sinais de entrada da rede, $\mathbf{u}(\mathbf{n}) = [u_1(n), u_2(n), \dots, u_K(n)]^T$, são combinados linearmente gerando o vetor de entradas do reservatório de dinâmicas, $\mathbf{x}(\mathbf{n}) = [x_1(n), x_2(n), \dots, x_N(n)]^T$, $f(\cdot)$ é a função de ativação, as matrizes de pesos $\mathbf{W}^{in} \in \mathcal{R}^{N \times K}$ e $\mathbf{W} \in \mathcal{R}^{N \times N}$ são geradas aleatoriamente, e o vetor de saídas $\mathbf{y}(\mathbf{n}) = [y_1(n), y_2(n), \dots, y_L(n)]^T$ que representa o conjunto de estados da rede em cada instante n , pode ser determinado por um método de regressão linear. Na ESN apenas as conexões entre o RD e a camada de saída são treinada (SIMEÓN, 2015).

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n) + \mathbf{W}^{back}\mathbf{y}(n) + \mathbf{W}^{bias}) \quad (3.15)$$

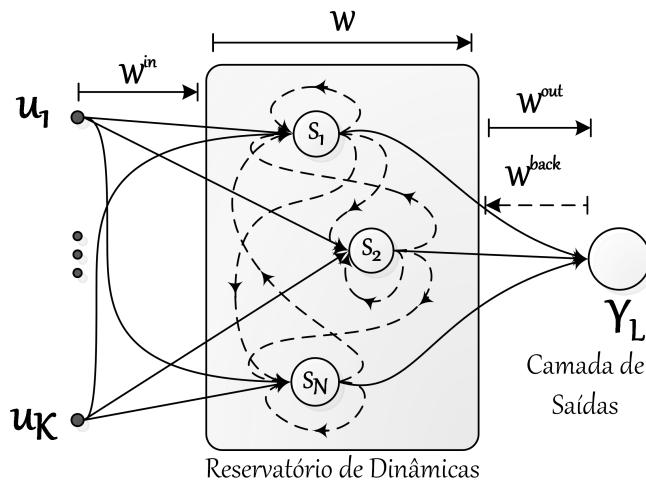
$$\mathbf{y}(n+1) = \mathbf{f}^{out}(\mathbf{W}^{inout}\mathbf{u}(n+1) + \mathbf{W}^{out}\mathbf{x}(n+1) + \mathbf{W}^{outout}\mathbf{y}(n+1) + \mathbf{W}^{biasout}) \quad (3.16)$$

Já na Figura 32, é exibida uma rede ESN que possui estados de eco. E seus estados são atualizados conforme as Equações 3.17 e 3.18.

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n)). \quad (3.17)$$

$$\mathbf{y}(n+1) = \mathbf{W}^{out}\mathbf{x}(n+1). \quad (3.18)$$

Figura 32 – Diagrama de uma rede ESN que possui estados de eco.



Com base nos padrões disponíveis para o treinamento e resposta esperada, \mathbf{Y} , é possível determinar os coeficientes da matriz \mathbf{W} , Equação 3.19 por meio da inversa generalizada expressa na Equação 3.20.

$$\mathbf{W}^{out} = \mathbf{X}^\dagger \mathbf{Y}. \quad (3.19)$$

$$\mathbf{X}^\dagger = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T. \quad (3.20)$$

Adicionalmente pode ser acrescido o parâmetro α (*leak rate*) na Equação 3.15 o que resulta na Equação 3.21, e a escolha adequada do valor parâmetro permite a melhora no ajuste da dinâmica do reservatório da ESN (SIMEÓN, 2015). O valor ótimo para o parâmetro α pode ser definido empiricamente, ou por busca num conjunto de valores por uma função de otimização. No trabalho de Antonelo, Schrauwen e Stroobandt (2008) um pequeno robô é treinado no contexto de computação de reservatórios, utilizando redes ESN e aborda métodos de busca do valor adequado para o parâmetro α .

$$\mathbf{x}(n+1) = \mathbf{f}((1 - \alpha)\mathbf{x}(n) + \alpha(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n))) \quad (3.21)$$

3.3.1 Propriedades dos Estados de Eco

Jaeger (2010), numa revisão de um trabalho anterior (JAEGER, 2001), apresenta os requisitos necessários à existência dos estados de eco em uma rede neural de estrutura recorrente. A seguir tais requisitos são apresentados:

- $|\sigma_{max}(\mathbf{W})| < 1$, no qual σ é o valor singular de \mathbf{W} .
- $|\lambda_{max}(\mathbf{W})| < 1$, sendo λ o autovalor de \mathbf{W} é chamado como raio espectral da ESN (JAEGER, 2010).

Levando em consideração os critérios demonstrados em Jaeger (2010), basta criar uma matriz \mathbf{W} que atenda a esses critérios, definir uma matriz \mathbf{W}^{in} de maneira arbitrária, que o treinamento da camada de saída de uma rede ESN é realizado por meio da solução de um problema de regressão linear.

3.3.2 Inicialização dos Pesos e Treinamento

Uma vez que as propriedades de estados de eco foram atendidas, o treinamento pode ser realizado seguindo as etapas (JAEGER, 2001; SIMEÓN, 2015; BOCCATO, 2013):

- Gerar uma matriz de pesos aleatórios (com média zero e variância 1) \mathbf{W} com certo grau de esparsividade, em torno de 20%;
- Normalizar \mathbf{W} com base no raio espectral;
- Definir uma matriz de pesos de entrada \mathbf{W}^{in} arbitrária;
- Calcular a matriz de pesos de saída \mathbf{W}^{out} por meio de um algoritmo de regressão linear. Neste trabalho será utilizada a inversa generalizada de Moore-Penrose.

3.3.3 Exemplos de Aplicações

Em Antonelo, Schrauwen e Stroobandt (2008) um pequeno robô é treinado no contexto de computação de reservatórios, utilizando redes ESN e aborda métodos de busca do valor adequado para o parâmetro α .

Boccato (2013) apresenta novas abordagens para as partes fundamentais de uma rede ESN, o RD, a camada de saída, e uma unificação entre a ESN e a ELM, essa última aplicada como camada de saída rede. Neste trabalho é proposta uma arquitetura que utiliza um filtro de Volterna em alternativa ao combinador linear de saída, que permite explorar as características estatísticas produzidas no RD, porém, sem afetar a simplicidade do processo de treinamento.

Em Siqueira (2013) a ESN é avaliada como alternativa e aperfeiçoamento à previsão de vazões médias mensais de usinas hidroelétricas brasileiras. O trabalho foi desenvolvido com dados das séries históricas das usinas de Furnas, Emborcação e de Sobradinho. Foram avaliadas três estrutura de redes neurais, MLP, ELM e ESN em alternativa ao método PAR (Periódicos auto-regressivos) (REIS, 2013), e em todas os resultados superaram o PAR. Das técnicas avaliadas

duas estruturas com ESN foram as que apresentaram os melhores resultados na predição, sendo a primeira com combinador linear proposta por Jaeger (2001) e a mesma rede, porém utilizando um filtro de Voltera.

Em Ganjefar e Ghasemi (2014), uma ESN foi utilizada no sistema de controle de turbinas eólicas de baixa potência (1 – 100 kW). O objetivo do trabalho era manter o sistema “rastreando” o ponto de operação de máxima geração de potência, algoritmo conhecido como MPPT⁶⁴. No algoritmo é necessário conhecer as características da turbina utilizada bem como monitorar as condições de vento, o que se torna um problema de complexidade elevada, devido às características de dinâmica não-lineares do sistema de geração eólica. Três métodos foram propostos: No 1º, o controlador foi projetado conhecendo-se a velocidade do vento. No 2º, o controlador baseado na ESN (com 100 neurônios), não tinha a informação da velocidade do vento. E 3º, foi adicionado um estimador da velocidade do vento utilizando a ESN. Os resultados, simulados, foram comparados com o resultados de um controlador PID e ABPC⁶⁵. Os métodos 2 e 3 foram comparados com o método 1 e a eficiência para a potência média alcançada foi de 99,9986% e 99,8843% respectivamente.

No trabalho de Wen, Li e Li (2015), um conjunto de redes ESN (*Ensemble Convolutional Echo State Network* - EC-ESN) foi utilizado para o reconhecimento de padrões de expressões faciais. Utilizando imagens de duas bases de dados sem nenhuma técnica de extração de características, as imagens foram apresentadas as redes SVM, SRC⁶⁶, Softmax, ESN e EC-ESN. Os resultados indicaram a que a ESN tem capacidade de separação de classes em problemas de reconhecimento de expressões faciais.

SIMEÓN (2015) propõe uma abordagem utilizando a ESN para o prognóstico de vida útil remanescente de equipamentos baseada em dados históricos utilizando o algoritmo de colônia de abelhas (ESN-ABC). A aplicação do método ABC⁶⁷ junto com a ESN possibilitou o ajuste dos parâmetros da rede, tendo o RD de tamanho fixo, resultando no menor erro quadrático médio quando comparada com o método clássico e o método de treinamento com filtro de Kalman (AIUBE, 2005, Cap 4).

Já em Trentin, Scherer e Schwenker (2015), uma variação da ESN, π -ESN (*Probabilistic ESN*) foi aplicada num problema de reconhecimento de cinco expressões de fala de mulheres. Os sinais utilizados tinham duração entre 0,7 s e 1,7s, e os resultados foram comparados com outros quatro classificadores, 1-NN, SVM, MLP, e AdaBoost, e os resultados foram muito significativos tendo a π -ESN como a maior percentual médio de classificação.

No trabalho de Schaetti, Salomon e Couturier (2016) as redes ESN foram aplicadas no reconhecimento de dígitos manuscritos, e seus resultados comparados a estruturas de redes neurais convolucionais⁶⁸ que são o estado da arte na classificação de imagens. Foi utilizada a base de dados MNIST⁶⁹, a qual contém 60.000 amostras para treino e 10.000 amostras para teste.

⁶⁴ Maximum Power Point Tracking.

⁶⁵ Adaptive Passivity-Based Control.

⁶⁶ Sparse representation classifier

⁶⁷ Artificial Bee Colony

⁶⁸ CNN - Convolutional Neural Networks.

⁶⁹ Modified National Institute of Standards and Technology.

Os resultados obtidos para a ESN apresentaram variação de 0,93% a 1,68% para a taxa de erro de classificação, enquanto que o SVM obteve 1,1% e redes convolucionais chegaram a um erro máximo de 0,35%.

Redes ESN em conjunto⁷⁰, foi utilizadas numa estrutura convolucional no trabalho de Wang, Wang e Liu (2016). Neste trabalho é proposta uma nova abordagem para tratamento em problemas com séries temporais multivariadas⁷¹ no reconhecimento de expressões faciais⁷².

Em Prater (2017) uma estrutura baseada em ESN, a ESMVE (*Echo State Mean-Variance Estimation*) foi desenvolvida e comparada comparada com K-ELM (*Kernel Extreme Learning Machine*) e SVM (*Support Vector Machine*), os resultados obtidos com a estrutura ESMVE mostraram-se mais eficientes.

Em ARAÚJO (2017) a ESN foi utilizada como estimador de vazões médias mensais no reservatório da Usina Hidrelétrica de Furnas (UHE Furnas). Duas versões foram aplicadas, a ESN clássica e a ESN com *leaky rate*. A base de dados utilizada era composta de informações das séries de vazões naturais mensais de janeiro de 1931 a dezembro de 2015 da UHE Furnas. Para teste foram utilizados 03 períodos de 5 anos, num total de 60 amostras cada. Período de seca, intervalo nos anos de 1952 – 1956, período de cheias, anos de 1979 – 1983 e período de vazões medianas, anos de 2006 – 2010. O período restante foi utilizado para treinamento. Os resultados indicaram que as redes ESN com *leaky rate* geraram o menor erro no processo de estimação das vazões para a UHE Furnas. Também indicaram que os modelos produziram resultados consistentes com as observações, o que permite prever as vazões com eficiência, sendo uma alternativa eficaz.

⁷⁰ EC-ESN – Ensemble Echo State Network.

⁷¹ MTS – Multivariate Time Series.

⁷² FER – Facil Expressions Recognition.

4 Metodologia

Introdução

Neste capítulo serão abordados procedimentos utilizados durante a pesquisa no atendimento aos objetivos propostos.

4.1 Técnicas de Reamostragem

A técnica de reamostragem (COSTA, 2006, p. 41) baseia-se em subdividir o conjunto de dados em subconjuntos menores, nos quais haja representação das características do conjunto total de dados. No processo de divisão dos subconjuntos, sorteios aleatórios são realizados com o intuito de evitar possíveis tendências na seleção das amostras a serem utilizadas para treinamento e teste. Todas as redes foram treinadas, na nuvem, em máquinas virtuais, com processadores de arquitetura Intel[©] Xeon E5 v4 2,2 GHz com 8 núcleos e 24 GB de RAM.

4.1.1 *k-fold*

No treinamento da redes utilizadas como classificadores, foi realizado o método de validação cruzada *k-fold* (FACELLI et al., 2011; XU et al., 2018), ver Figura 33. Neste trabalho, a base de dados foi subdividida em 10 subconjuntos de tamanhos idênticos, em seguida, foi realizado o sorteio de 6 subconjuntos (60%) para o treino e 4 (40%) para teste e validação, então, essa configuração de rede foi treinada e testada 100 vezes.

Com o auxílio da Equação 4.1 é possível verificar que existem 210 combinações distintas para realizar o sorteio, tanto para treino, quanto para o teste ($n = 10$, $p = 6$). Com o primeiro ciclo realizado o sorteio repete-se até que sejam feitos 50 sorteios, e cada um treinado 100 vezes. Essa é uma das metodologias para treino das RNA utilizada pela Colaboração ATLAS, e será aplicada neste trabalho permitindo realizar comparações com os resultados da colaboração como referência de desempenho.

$$C_{n,p} = \frac{n!}{p!(n-p)!}. \quad (4.1)$$

4.1.2 *Jackknife*

É um método de validação cruzada não paramétrico utilizado para estimar o enviesamento de uma amostra, ou um parâmetro de interesse de uma amostra aleatória de uma população. Esse método também é conhecido como *leave-one-out*⁷³, por dividir a base de dados em **n** subconjuntos de igual tamanho ($\mathbf{x} = [x_1, x_2, x_3, \dots, x_n]$), em seguida processar a análise em $n-1$ subconjuntos até completar o processamento de toda a base de dados, n vezes. Dessa forma cada um das n etapas de processamento, chamadas amostras *jackknife*, só diferem entre si por

⁷³ Deixe um fora - em tradução livre.

um dos n subconjuntos da base, mantendo a característica de interesse (COSTA, 2006; ABDI; WILLIAMS, 2010).

Cada amostra *jackknife* tem o seguinte formato:

$$\mathbf{x}^1 = [x_2, x_3, x_4, \dots, x_{n-1}, x_n] \quad (4.2)$$

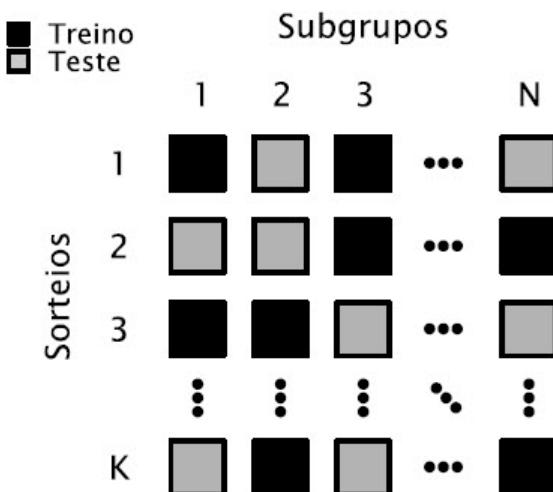
$$\mathbf{x}^2 = [x_3, x_4, x_5, \dots, x_{n-1}, x_n] \quad (4.3)$$

⋮

$$\mathbf{x}^n = [x_1, x_2, x_3, \dots, x_{n-2}, x_{n-1}] \quad (4.4)$$

Cada uma das amostras, Equação 4.2, Equação 4.3 e Equação 4.4, é utilizada para determinação do parâmetro $\hat{\theta} = s(\mathbf{x})$, na qual $s(\mathbf{x})$ é uma estatística de interesse para a população em estudo (ALVES, 2013).

Figura 33 – Representação do processo de agrupamento e sorteio dos subgrupos de treinamento.



4.2 Determinação do número de neurônios

Para cada uma das três técnicas utilizadas nesse trabalho foi adotado o seguinte procedimento de determinação:

- MLP
 - Base Experimental: Número de neurônios utilizados no trabalho de Souza (2015);
 - Dados Simulados: Número de neurônios utilizados pela Colaboração ATLAS.
- ELM: Rede que obteve o melhor índice SP, nos testes com número de neurônios na camada oculta avaliados de 5 até 100;
- ESN: Rede que obteve o melhor índice SP, nos testes com número de neurônios no reservatório de dinâmicas avaliados de 5 até 60.

4.3 Métodos de avaliação dos resultados

Neste trabalho serão utilizadas três bases de dados. Uma com dados experimentais obtidos no ano de 2011 e uma com dados obtidos pela técnica de Monte Carlo (**YORIYAZ, 2009**) no ano de 2015. O classificador utilizado pela colaboração ATLAS, baseado em *perceptron multilayer* (MLP) será a referência utilizada para validação das técnicas ELM e ESN, propostas para uso como classificadores alternativos ao MLP. Para isso os parâmetros de ajuste do MLP serão os utilizados pela Colaboração.

4.3.1 Curva ROC

Para avaliação de desempenho do discriminador binário (elétron/jato) será utilizada a curva ROC⁷⁴ (**FAWCETT, 2006**). A curva ROC auxilia na análise dos resultados provenientes de classificadores distintos, em problemas de classificação binária⁷⁵ quando se deseja avaliar a qualidade da separação efetuada pelos classificadores em análise.

Na [Tabela 3](#), é apresentado o resumo das respostas possíveis de um classificador. O significado para cada termo da tabela é:

- Verdadeiros Positivos **VP**: Valores da classe positiva classificados corretamente.
- Falsos Negativos **FN**: Valores da classe positiva classificados como negativos.
- Falsos Positivos **FP**: Valores da classe negativa classificados como positivos.
- Verdadeiros Negativos **VN**: Valores da classe negativa classificados corretamente

Tabela 3 – Resultados possíveis de classificação.

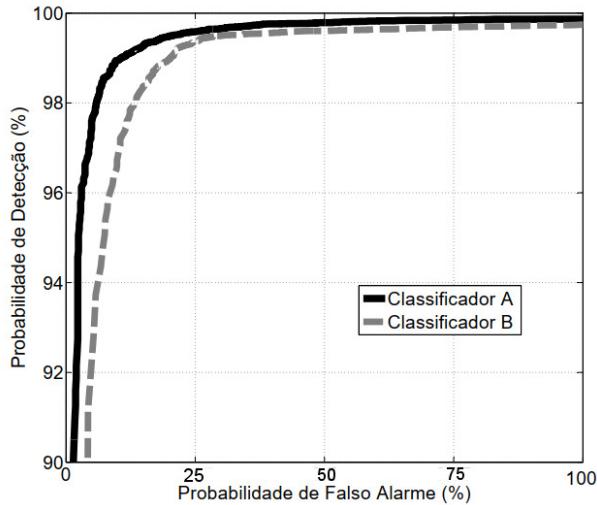
		Valor Observado	
		Positivos	Negativos
Valor Predito	Positivos	VP	FP
	Negativos	FN	VN

Neste trabalho, as classes em análise são: elétrons e jatos hadrônicos. Os elétrons são a classe de interesse, dessa forma, serão associados aos valores VP, que neste trabalho foram chamados de probabilidade de detecção (PD), enquanto que os jatos, foram chamados de taxa de falso alarme (FR). A curva ROC consiste de um gráfico (x, y), no qual o eixo das abscissas corresponde à taxa de falsos alarmes (FR), e o eixo das ordenadas à probabilidade de detecção (PD). Na [Figura 34](#) é exibido um exemplo de curvas ROC para dois classificadores.

⁷⁴ *Receiver Operating Characteristic Curve*.

⁷⁵ Existe uma classe de interesse, que deseja-se obter a separação numa base dados contaminada por outra classe, tomada como ruído de fundo.

Figura 34 – Exemplos de curvas ROC de dois classificadores.



Fonte: Souza (2015)

Na Figura 34 observa-se que o classificador A possui um melhor desempenho de classificação em relação ao classificador B, pois sua curva alcança valores de PD elevados com menores FR, em comparação com o classificador B. Um classificador ideal, seria aquele no qual a sua curva ROC atinge o máximo valor de PD (100%) para FR igual a zero, e se mantém no máximo ao longo de toda a faixa de FR.

4.3.2 Índice SP

É um parâmetro utilizado para auxiliar na definição do ponto de operação ótimo de um determinado classificador (TORRES et al., 2009). É definido conforme Equação 4.5.

$$SP = \sqrt{\frac{(Ef_e + Ef_j)}{2} \times \sqrt{Ef_e \times Ef_j}}. \quad (4.5)$$

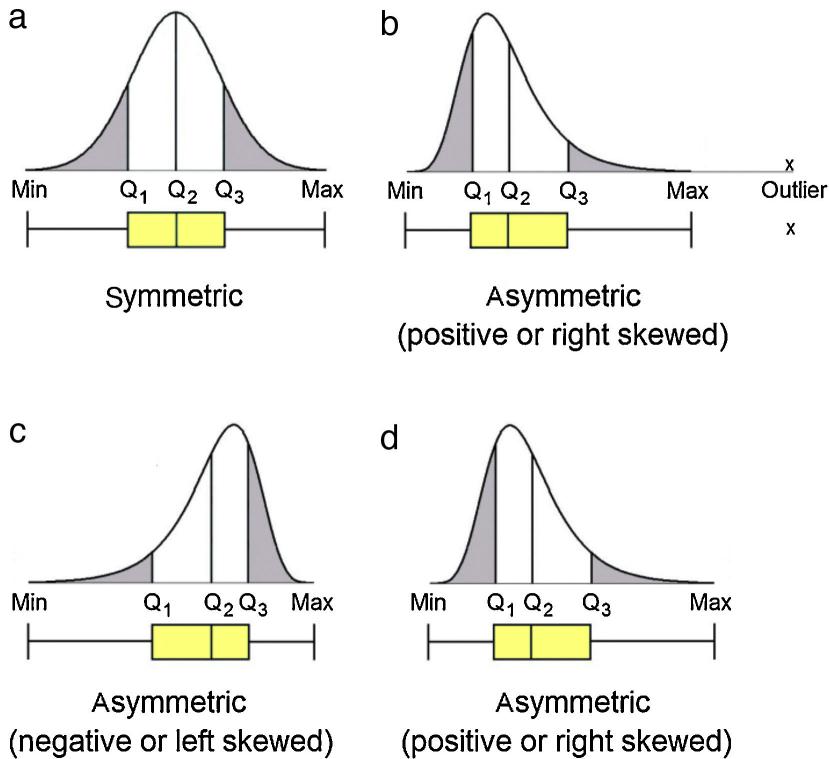
onde $Ef_e = PD$ e $Ef_j = 1 - FR$ são as eficiências obtidas, respectivamente, para elétrons e jatos (sendo PD a probabilidade de detecção de elétrons e FR probabilidade de classificar um jato hadrônico incorretamente). A eficiência de um classificador está associada ao maior valor para o índice SP. Um índice $SP = 1$ (classificador ideal), indica máxima taxa de probabilidade de detecção (PD) para erro (FR) zero.

4.3.3 Boxplot

A apresentação dos resultados obtidos nos ensaios realizados será feita em gráficos utilizando a *boxplot* (gráfico de caixa). Esse tipo de gráfico é utilizado para representar a distribuição empírica dos dados (Portal Action, 2016) de uma série de eventos (FERREIRA et al., 2016). Na Figura 35, Q_1 e Q_3 , referem-se ao 1º e 3º quartis respectivamente, IQR, é a faixa entre quartis e indica o grau de dispersão dos dados. Os limites superior e inferior são definidos por segmentos chamados *Whisker*, ou “fio de bigode”. Os quais são calculados por $Q_1 - 1,5 \times IQR$

para o inferior e $Q_3 + 1,5 \times IQR$ para o superior. Os pontos que por ventura fiquem fora destes limites são chamados *outliers* e representam valores discrepantes.

Figura 35 – Exemplos de *boxplot* para quatro possíveis distribuições.



Fonte: Ferreira et al. (2016)

Os Quartis, Q_1 , Q_2 e Q_3 , respectivamente, são os valores que marcam os limites onde estão situados os 25%, 50%, e 75% das observações obtidas numa amostra, sendo essas organizadas de maneira crescente.

A definição de cada quartil j segue a expressão da Equação 4.6, sendo n o número de elementos da amostra.

$$Q_j = X_k + \left(\frac{j(n+1)}{4} - k \right) (X_{k+1} - X_k), \quad (4.6)$$

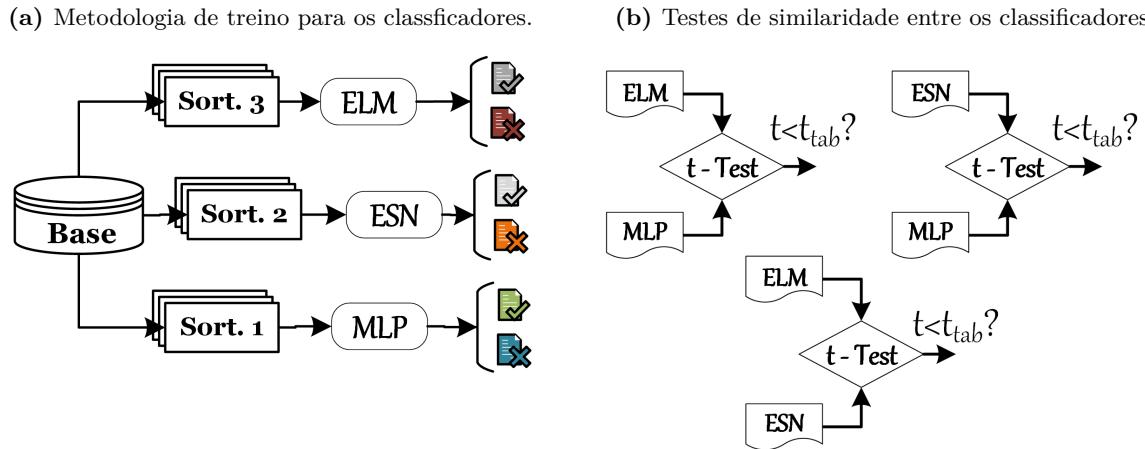
calcula-se k como a parte inteira de $\frac{j(n+1)}{4}$, para $j = \{1, 2, 3\}$ e X_k é a posição da observação k da amostra organizada de maneira crescente.

4.3.4 Teste de Student

Além da avaliação por meio do índice SP, um teste estatístico será utilizado, o teste de Student. Para isso, o treinamento das redes será realizado, ou seja, serão realizados três sorteios, um para cada classificador, ver Figura 36a, e os resultados serão comparados por meio do teste de Student⁷⁶ conforme Figura 36b (KUNCHEVA, 2004).

⁷⁶ pseudônimo de William Sealy Gosset (1876-1937), químico e estatístico inglês que desenvolveu o teste de hipóteses conhecido como t-student.

Figura 36 – Representação da metodologia para avaliação dos resultados dos classificadores avaliados, ELM, ESN e MLP.



Este teste é utilizado como parâmetro estatístico de análise de desempenho conforme descrito em Kim (2015) e em Kuncheva (2004, p. 18,19). Nele, dois classificadores A e B são avaliados (comparados) quanto a diferença de desempenho de classificação no treino por validação cruzada, em conjuntos de treino independentes, com k treinos em cada sorteio, ver Figura 33, e seus resultados organizados em \mathbf{P}_A e \mathbf{P}_B . Em seguida, um conjunto de diferenças entre os resultados dos classificadores para cada um dos k sorteios é obtido, Equação 4.7. Na Equação 4.8 esse conjunto de diferenças determina um valor na tabela de distribuição de Student, com $k-1$ graus de liberdade. A hipótese nula nesse teste, H_0 , é de que os classificadores possuem mesmo desempenho se $t_{calculado} < t_{tabelado}$ para um nível de significância de 0,05 bilateral⁷⁷.

Neste trabalho os vetores \mathbf{P}_A e \mathbf{P}_B , referem-se aos índices SP apresentados nos dados das melhores boxplot em cada uma das três técnicas: MLP, ELM e ESN. Três comparações foram realizadas: $\mathbf{P}_{MLP} \times \mathbf{P}_{ELM}$, $\mathbf{P}_{MLP} \times \mathbf{P}_{ESN}$ e $\mathbf{P}_{ESN} \times \mathbf{P}_{ELM}$.

$$\mathbf{P}^{(k)} = \mathbf{P}_A^{(k)} - \mathbf{P}_B^{(k)}. \quad (4.7)$$

$$t = \frac{\bar{\mathbf{P}}\sqrt{k}}{\sqrt{\sum_{i=1}^k \frac{(\mathbf{P}^{(i)} - \bar{\mathbf{P}})^2}{k-1}}}, \quad (4.8)$$

sendo que

$$\bar{\mathbf{P}} = \frac{1}{k} \cdot \sum_{i=1}^k \mathbf{P}^{(i)}. \quad (4.9)$$

4.4 Bases de Dados

Nesta dissertação as bases de dados utilizadas foram fornecidas pela Colaboração ATLAS. Estão disponíveis bases de dados simulados, obtidos através de técnicas de Monte Carlo e dados

⁷⁷ Os valores de referência são de uma tabela de probabilidade bicaudal.

experimentais de colisões do LHC. Todas as bases, simuladas e experimentais, são validadas pelo ambiente *offline*, o qual possui três níveis de aceitação selecionáveis (COSTA, 2016):

- *Loose* - Critério de aceitação de sinal elevado, porém baixa rejeição ao ruído de fundo;
- *Medium* - Nível de aceitação intermediário entre o *loose* e o *tight*;
- *Tight* - Neste ajuste a redução do ruído de fundo é elevada, entretanto, o critério de aceitação para sinal é mais criterioso, resultando na menor eficiência entre os três níveis.

O ambiente *offline* é composto por um conjunto de algoritmos com elevado poder computacional, o qual possui menos restrições temporais de processamento. Sua resposta é utilizada como referência pelos físicos para análise e é utilizada para o desenvolvimento de classificadores neurais para separação elétron/jato (SOUZA, 2015).

Os sinais simulados por técnicas de Monte Carlo (YORIYAZ, 2009) são utilizados no desenvolvimento dos experimentos físicas de altas energias na tentativa de representar as condições de operação do detector (sejam com os parâmetros atuais, sejam em condições futuras).

Para os dados simulados a energia do centro de massa foi ajustada para 14 TeV, com luminosidade máxima de 10^{34} (SOUZA, 2015). Com esses níveis de energia e luminosidade é possível que ocorra empilhamento de eventos, ou seja, o sistema de instrumentação ainda está sensibilizado por um evento anterior quando o evento atual se desenvolve. Tais condições, são interessantes no que diz respeito à robustez do algoritmo proposto, uma vez que deve ser capaz de separar as classes, elétron/jato, mesmo com a ocorrência do empilhamento (*pile-up*). Em cada base de dados simulados, a Colaboração ATLAS busca representar as características físicas do detector, assim como as condições de operação do mesmo no momento das colisões.

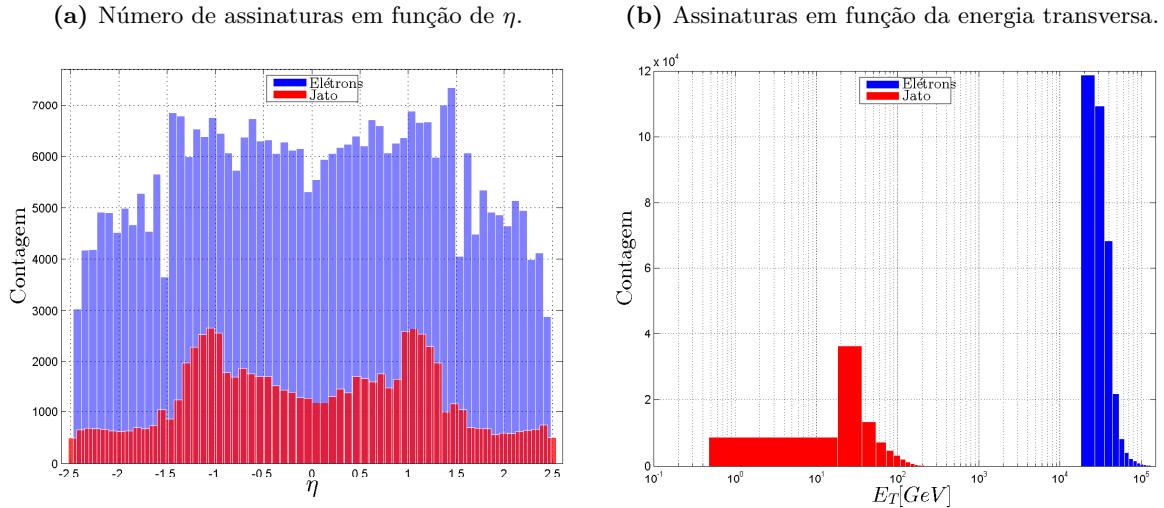
4.4.1 Dados Experimentais

Os dados disponíveis nesta base foram obtidos de dois eventos de colisões registradas pelo detector ATLAS e validados pelo ambiente *offline*, no projeto referenciado como data11_7TeV com as seguintes características:

- Assinaturas de elétrons:
 - Energia por feixe de 3,48 TeV;
 - Luminosidade de $1,34 \cdot 10^{30} \text{ cm}^{-2}\text{s}^{-1}$;
 - Número de assinaturas: 337.658;
- Assinaturas de jatos:
 - Energia por feixe de 3,48 TeV;
 - Luminosidade de $1,45 \cdot 10^{30} \text{ cm}^{-2}\text{s}^{-1}$;
 - Número de assinaturas: 78.353.

Na Figura 37a é apresentado o número de assinaturas para elétrons e jatos em função da posição η de interação com o detector, enquanto que na Figura 37b é exibido o número de assinaturas em função da energia transversa. É possível observar que o número de assinaturas de elétrons é da ordem de 3 vezes o número de assinaturas para os jatos, também nota-se a separação entre os níveis de energia para elétrons e jatos, da ordem de 10^2 aproximadamente.

Figura 37 – Distribuição de assinaturas de elétrons e jatos para a base de dados experimentais.



4.4.2 Dados Simulados

As assinaturas disponíveis nesta base foram obtidas por meio da técnica de Monte-Carlo. Nesta, os dados foram segmentados em intervalos de $\Delta_{|\eta|}$ e Δ_{E_T} . Na Tabela 4, são apresentados os cinco intervalos nos níveis de energia transversa (Δ_{E_T}) e os quatro intervalos de posições dentro do detector ($\Delta_{|\eta|}$), totalizando 20 regiões. Nesta base de dados a energia das colisões foi ajustada para 13 TeV, contendo número médio de colisões ($\langle \mu \rangle$) entre 0 e 60.

Tabela 4 – Segmentação base de dados simulados MC2015.

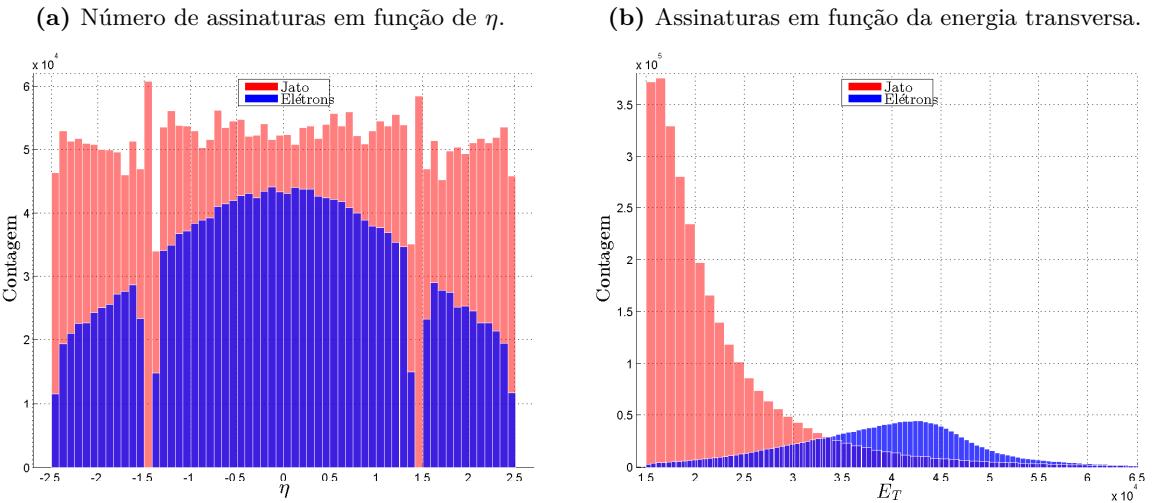
Intervalos	
Δ_{E_T} [GeV]	$\Delta_{ \eta }$
[15; 20]	[0, 0; 0, 8]
[20; 30]	[0, 8; 1, 37]
[30; 40]	[1, 37; 1, 54]
[40; 50]	[1, 54; 2, 5]
[50; ∞ [–

Para simplificação da representação de cada uma das regiões da base de dados simulados, será adotada a representação em par ordenado (x, y) . Nesta representação os valores de Δ_{E_T} referem-se à coordenada $x = [0, 1, 2, 3, 4]$, e os intervalos de $|\eta|$, referem-se à coordenada $y = [0, 1, 2, 3]$. Dessa forma, a representação da região com energia na faixa [15; 20] GeV, no intervalo de $\Delta_{|\eta|} = [0, 0; 0, 8]$, será representada como a região (0,0). Procedendo dessa maneira, obtém-se

20 regiões, provenientes da combinação dos cinco intervalos para a energia transversa (Δ_{ET}) e os quatro intervalos para pseudo-rapidez ($\Delta_{|\eta|}$), indo de (0,0) até (4,3).

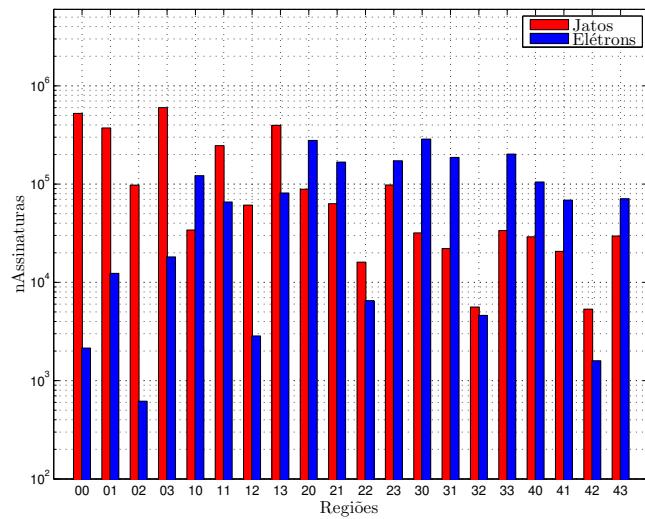
Na Figura 38a é apresentado o número de assinaturas para elétrons e jatos em função da posição η de interação com o detector, enquanto que na Figura 38b o número de assinaturas em função da energia transversa. Observa-se que para essa base, a região de $1,4 < \Delta_{|\eta|} < 1,5$, região de *crack*, o registro de assinaturas de elétrons é quase zero (17 para $-1,5 < \Delta_\eta < -1,4$ e 11, para $1,4 < \Delta_\eta < 1,5$) em comparação com as demais regiões.

Figura 38 – Distribuição de assinaturas de elétrons e jatos para a base de dados simulados.



Na Figura 39 o número de assinaturas de elétrons e jatos para cada uma das 20 regiões da base de dados simulados, proveniente da combinação dos intervalos para Δ_{ET} e $\Delta_{|\eta|}$. Nessa base, novos parâmetros de operação para o detector fazem com que o número de assinaturas geradas para elétrons e jatos estejam na mesma ordem de grandeza.

Figura 39 – Número de assinaturas para elétron e jato por região na base simulada.



Na Tabela 5 são exibidos o número de assinaturas para elétrons e jatos presentes na base de dados, agrupados por região. O número de assinaturas está representado em pares ordenados,

$(\Delta_{E_T}, \Delta_{|\eta|})$, no total de 20 pares para essa base, referentes aos intervalos de variação da energia transversa e intervalos de variação pseudo-rapidez apresentados na Tabela 4.

Tabela 5 – Número de assinaturas para cada corte de energia no conjunto de dados.

Número de assinaturas por região										
	Regiões									
	(0,0)	(0,1)	(0,2)	(0,3)	(1,0)	(1,1)	(1,2)	(1,3)	(2,0)	(2,1)
Elétrons	21.490	12.362	618	18.154	121.811	65.684	2.854	81.125	277.833	167.207
Jatos	523.981	372.257	97.614	600.198	346.143	246.076	61.155	396.224	88.659	63.230
Total	526.130	384.619	98.232	618.352	155.954	311.760	64.009	477.349	366.492	230.437
	(2,2)	(2,3)	(3,0)	(3,1)	(3,2)	(3,3)	(4,0)	(4,1)	(4,2)	(4,3)
Elétrons	6.159	172.515	286.840	186.662	4.618	202.108	105.114	68.736	1.597	71.013
Jatos	16.062	97.835	31.857	22.087	5.607	33.642	29.068	20.689	5.224	29.627
Total	22.581	270.350	318.697	208.749	10.225	235.750	134.182	89.425	6.931	100.640

Fonte: Colaboração ATLAS.

5 Resultados

Introdução

Neste capítulo serão apresentados os resultados dos treinamentos realizados nas bases de dados Experimentais, simulados por meio da técnica de Monte Carlo, utilizando três classificadores neurais, um baseado em *Perceptron Multilayer* (MLP), o segundo em Máquinas de Aprendizado Extremo (ELM) e o terceiro em Redes com Estados de Eco (ESN).

5.1 Bases de dados

A seguir são apresentados os principais ajustes para as redes MLP, ELM e ESN nas três bases de dados utilizadas. O número de neurônios utilizado em cada base, assim como variações de parâmetros, serão apresentados em suas respectivas seções.

- MLP

As redes MLP foram a referência para comparação das técnicas propostas, dessa forma, os parâmetros utilizados foram os mesmos já utilizados pela Colaboração ATLAS. Com isso, é possível validar a metodologia e os resultados obtidos com as técnicas alternativas, e sugerir testes no ambiente do detector ATLAS.

- Número de Épocas: 300;
- Função de ativação: tangente hiperbólica;
- mínimo gradiente: 0;
- Algoritmo de treino: *Resilient Backpropagation*;
- Máx. número de falhas: 150⁷⁸;

- ELM

Para a ELM, a escolha dos ajustes que foram aplicados durante o processo de treinamento das redes, foi orientada pela pesquisa apresentada no [Capítulo 2, seção 3.2](#), e testes de sensibilidade⁷⁹.

- Pesos na camada oculta: pseudo-aleatórios com distribuição normal;
- Função na camada oculta: tangente hiperbólica;
- Determinação da matriz de saída: Inversa generalizada de Moore-Penrose;
- Função na camada de saída: tangente hiperbólica.

⁷⁸ Número máximo de épocas as quais o erro alcançado no grupo de dados de teste cresce. Atingido esse número o treinamento é interrompido.

⁷⁹ O resultado do teste de sensibilidade da ELM ao método de geração dos números pseudo-aleatórios é apresentado no [Apêndice B](#).

- ESN

O procedimento adotado para a escolha dos ajustes para a ESN, foi o mesmo utilizado para a ELM. A escolha dos ajustes que foram aplicados durante o processo de treinamento das redes, foi orientada pela pesquisa apresentada no [Capítulo 2](#), seção 3.3.

- Pesos na camada oculta: pseudo-aleatórios com distribuição normal;
- Grau de esparsividade: 30%;
- Função na camada oculta: tangente hiperbólica;
- Determinação da matriz de saída: Inversa generalizada de Moore-Penrose;
- Função na camada de saída: tangente hiperbólica.

Para as três técnicas foram utilizados duas técnicas de reamostragem para obtenção dos resultados. Esse procedimento é necessário para validação da metodologia junto à Colaboração ATLAS.

- Técnica de reamostragem:

- *K-fold*: 50 sorteios, 100 inicializações;
 - * Amostras para treino: 60%;
 - * Amostras para teste e validação: 40%;
- *Jackknife*: 10 sorteios, 100 inicializações;
 - * Amostras para treino: 90%;
 - * Amostras para teste e validação: 10%;

5.1.1 Dados Experimentais

5.1.1.1 Informações

A seguir os resultados na base experimental cujo número de assinaturas de elétrons é de 337.658, e o número de assinaturas de jatos é de 78.353, tendo no total 416.011. O número de neurônios utilizado em cada uma das três técnicas para os melhores desempenho de classificação é exibido na [Tabela 6](#), de acordo com a metodologia descrita na [seção 4.2](#).

Tabela 6 – Número de neurônios utilizados na camada oculta do MLP, ELM e do reservatório de dinâmicas da ESN, na base experimental.

Número de neurônios		
ELM	MLP	ESN
100	23	15

5.1.1.2 Índice SP

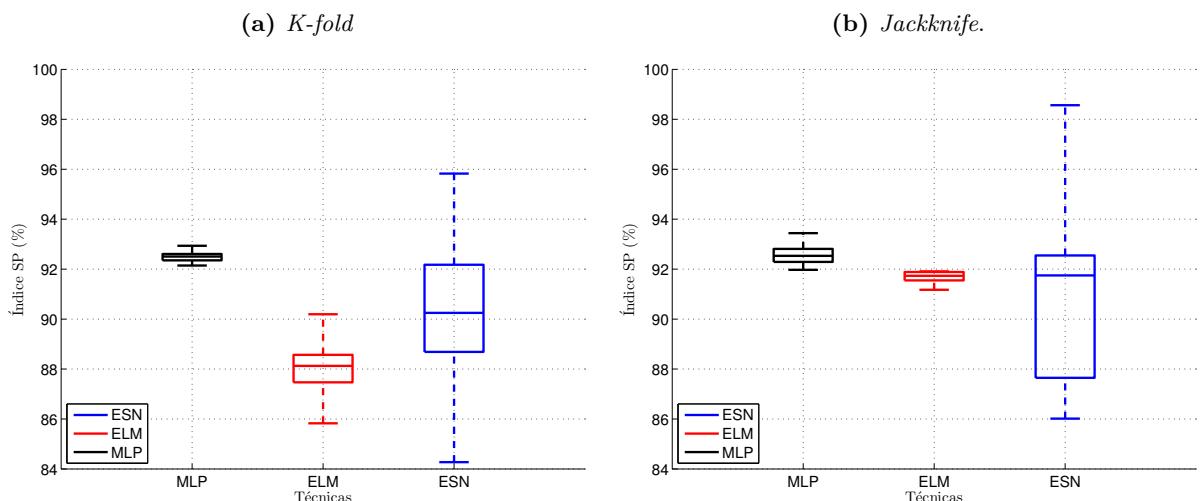
Na Figura 40a são exibidas as *boxplot* de cada um dos classificadores, MLP, ELM e ESN nessa ordem, para o melhor das 100 inicializações realizados em cada uma das técnicas, utilizando a técnica de reamostragem *K-fold*. Já na Figura 40b são exibidas as *boxplot* de cada um dos classificadores, MLP, ELM e ESN nessa ordem, utilizando a técnica de reamostragem *Jackknife*. Cada *boxplot* apresenta o melhor resultado dentre as 100 inicializações realizadas em cada subconjunto disjunto da base de dados.

É possível observar que os resultados alcançados com a técnica de reamostragem *Jackknife*, na Figura 40b, foram superiores aos resultados alcançados com o *k-fold*, Figura 40a. Nota-se que para o MLP houve uma pequena melhora para o máximo valor atingido, apesar de uma pequena elevação na dispersão entre os índices SPs, mínimo e máximo. Para a ELM houve melhora na dispersão dos valores do mínimo e máximo SPs alcançados, e elevação do valor médio, que saiu de 88% com a técnica *K-fold* para próximo de 92% com a técnica *Jackknife*. Já na ESN, nota-se que o diferença entre o maior e o menor valor do índice SP da *boxplot* se manteve aproximadamente constante, entretanto, houve melhora no valor médio, que subiu de próximo de 90% para próximo de 92%, e a *boxplot* sofreu um deslocamento, no qual seu início sai de 84% para 88%, no menor SP, e de 95% para acima de 98%, para o máximo SP.

Os resultados utilizando a técnica de reamostragem *Jackknife*, sugerem que o método foi capaz de obter grupos de treino, teste e validação que contivessem características relevantes da base de dados. Desta forma, os classificadores produziram uma separação de classes elétron jato com melhor eficiência.

Em preto é exibido o resultado para o MLP. Os resultados apresentam pequena dispersão para o melhor subconjunto. Sendo o mínimo valor para o índice SP de 91,972% e o máximo de 93,441%. Em vermelho, são os resultados obtidos com a ELM. Em relação ao MLP a ELM apresentou um resultado com dispersão similar, entre 91,174% e 92,845%. Por fim, em azul, são os resultados obtidos com a ESN. Nestes a dispersão é superior tanto ao MLP quanto à ELM, com resultados para o índice SP entre 86,017% e 98,564%.

Figura 40 – *Boxplot* das melhores redes para ELM, ESN e MLP na base experimental.

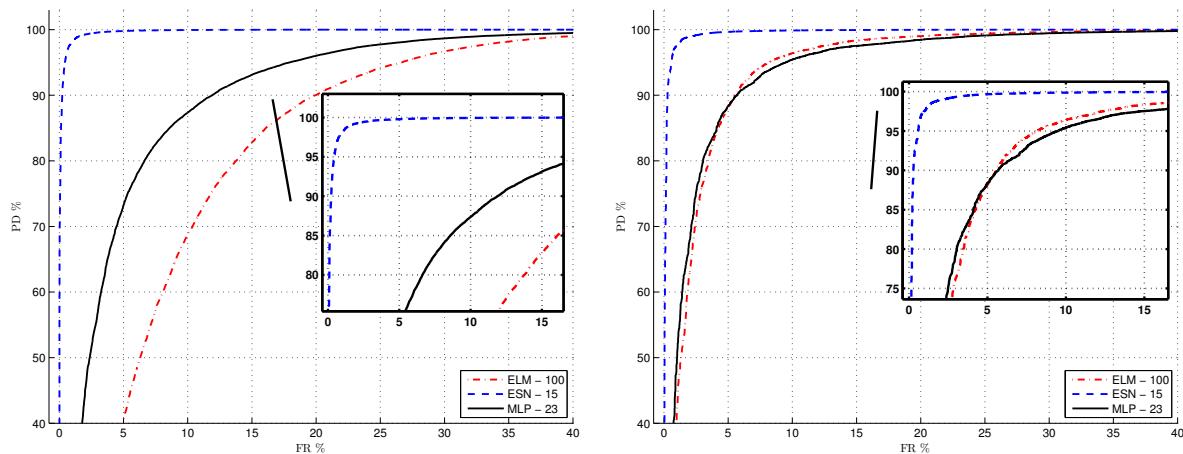


Nas Figuras 41a e 41b, são exibidas as curvas ROC para cada uma das técnicas em seus melhores resultados para o índice SP, utilizando *K-fold* e *Jackknife*. Observa-se o que a curva ROC, Figura 41b, para a rede ELM sugere uma similaridade de desempenho em relação à rede MLP, pois as curvas estão quase sobrepostas. Já em relação à ESN, o MLP, apresenta seu melhor resultado para uma taxa de falso alarme de 8,359%, enquanto que a ESN tem seu melhor resultado com taxa de falso alarme de 1,519% aproximadamente. Observa-se que a ESN obteve uma melhor separação para as classes elétron e jato em relação à ELM e ao MLP.

Figura 41 – ROCs das melhores redes para ELM, ESN e MLP, utilizando as técnicas de reamostragem *K-fold* (a) e *Jackknife* (b).

(a) *K-fold*.

(b) *Jackknife*.



Na Tabela 7 é possível observar os dados para o máximo índice SP alcançado em cada uma das técnicas utilizando *Jackknife*. Observa-se que as redes ELM, tiveram um desempenho equivalente ao das redes MLP, pois o seu índice SP ficou 0,59% abaixo, e as incertezas também próximas as alcançadas pelo MLP.

Ao observar os valores para as redes ESN, nota-se a superioridade nos três parâmetros, SP, PD e FR. Para o PD 3,39 % maior, para o SP 5,12% maior e para o FR 6,84 % menor. Porém, é na ESN, que encontram-se os maiores níveis de incerteza para os valores, pelo menos 8 vezes maior que o MLP. Esse fato advém da grande variabilidade para os índices SP alcançados, de 86,017% a 98,564% o que produz um $\Delta_{SP_{ESN}} = 12,547\%$. Em comparação com o MLP, a faixa de valores foi de: 91,972% a 93.441% resultando num $\Delta_{SP_{MLP}} = 1,469\%$, aproximadamente 10 vezes menor.

Tabela 7 – Dados para os melhores índices SP nas técnicas, MLP, ELM e ESN para a base Experimental.

	FR (%)	SP (%)	PD (%)
ELM	$7,848 \pm 0.9383$	$92,845 \pm 0,460$	$93,541 \pm 0,7954$
MLP	$8,359 \pm 0,540$	$93,441 \pm 0,399$	$95,258 \pm 0,446$
ESN	$1,519 \pm 4,674$	$98,564 \pm 3,924$	$98,646 \pm 3,219$

5.1.1.3 Tempo de Treinamento

A Tabela 8 mostra os tempos de treino para as redes com o maior índice SP, que foram apresentados na Tabela 7. Nota-se uma redução no tempo de treinamento significativa tanto com a ELM quanto com a ESN. Ambas foram mais rápidas que o MLP. A ELM foi 12,65 vezes, e a ESN 11,18 vezes, o que representa redução de 92,09% e 91,06% do tempo do MLP, respectivamente. Esses resultados indicam que com as redes ELM e ESN seria possível realizar um número maior de reconfigurações e treinos das redes no mesmo tempo utilizado para o treino de redes MLP.

Tabela 8 – Tempo de treinamento em segundos, para os melhores resultados, ELM × ESN × MLP, na base experimental.

	Técnicas		
	ELM	MLP	ESN
	tempo (s)	$22,770 \pm 4,182$	$288,130 \pm 42,064$

5.1.2 Análise Estatística

A seguir os resultados obtidos com o teste de Student para os melhores resultados obtidos com a base de dados experimentais, utilizando as expressões das Equações 4.7, 4.8 e 4.9. Os dados de entrada das equações foram os índices SP das melhores inicializações para cada uma das técnicas. Três análises serão apresentadas: ELM × MLP, ESN × MLP e ELM × ESN. Em cada uma será avaliada a seguinte hipótese nula, H_0 : os classificadores possuem o mesmo desempenho de classificação. Caso $t_{calc} > t_{tab}$, a hipótese deve ser descartada, do contrário, a hipótese não pode ser descartada, pois os classificadores possuem desempenho semelhante.

O valor de referência tabelado para uma distribuição de Student com nove graus ($k-1 = 9$) de liberdade é $t_{tab} = 2,262^{80}$ para um nível de significância de 95%. Os dados utilizados são os dados que produziram as melhores *boxplot* da Figura 40b.

- ELM × MLP: O resultado do teste na comparação entre os classificadores foi de $t = 5,191$, que é maior que o $t_{tab} = 2,262$, logo, a hipótese nula pode ser descartada, pois os classificadores possuem diferenças significativas.
- ESN × MLP: Na comparação entre esses classificadores o resultado do teste foi de $t = 1,146$, menor do que o $t_{tab} = 2,262$, o que indica desempenho de classificação semelhante, pois as diferenças não são significativas
- ELM × ESN: Nessa comparação o resultado foi de $t = 0,500$, significativamente inferior ao valor tabulado $t_{tab} = 2,262$. Dessa forma conclui-se que os classificadores não possuem diferenças significativas, ou seja, possuem desempenho de classificação similar.

Os testes realizados considerando as combinações possíveis entre os classificadores indicam que os classificadores, ELM e MLP possuem diferenças significativas, ou seja, os classificadores

⁸⁰ Valor para uma probabilidade bicaudal.

projetados não possuem resultados similares. A comparação entre ESN e MLP, indicou que os resultados são similares, pois os classificadores possuem mesmo desempenho de classificação. Na comparação entre ESN e ELM, o teste indicou que os classificadores são semelhantes em desempenho, visto o valor obtido ser significativamente inferior ao tabelado.

5.1.3 Dados Simulados

5.1.3.1 Informações

Nesta base com dados simulados foram utilizadas duas técnicas de reamostragem, a *K-fold* e *Jackknife*, e serão apresentados os resultados para a técnica, *Jackknife*, que produziu os melhores resultados. Três técnicas de classificadores neurais foram utilizadas: MLP, ELM e ESN. E o número de neurônios utilizado em cada região, obtido pela metodologia descrita na seção 4.2, é exibido na Tabela 9.

Tabela 9 – Número de neurônios utilizado em cada uma das técnicas organizados por região.

Número de neurônios por Região na base simulada										
	(0,0)	(0,1)	(0,2)	(1,3)	(1,0)	(1,1)	(1,2)	(1,3)	(2,0)	(2,1)
ELM	100	100	100	90	100	100	90	100	70	90
MLP	5	10	5	7	5	7	12	5	5	12
ESN	30	15	45	15	15	15	15	15	15	15
	(2,2)	(2,3)	(3,0)	(3,1)	(3,2)	(3,3)	(4,0)	(4,1)	(4,2)	(4,3)
ELM	100	70	80	80	80	90	100	100	80	90
MLP	5	5	12	6	6	16	7	7	5	8
ESN	15	15	15	30	45	15	15	15	60	15

5.1.3.2 Índice SP

Nas Figuras 42 e 43 são apresentados os resultados para as melhores redes em cada uma das três técnicas. Na Figura 42 são as *boxplot* para os treinos utilizando reamostragem *Jackknife* e na Figura 43 são as *boxplots* para os treinos reamostrados via *K-fold*. Cada *boxplot* para a técnica de reamostragem *Jackknife* é composta de 10 sorteios, os quais contêm todas as possibilidades de combinação entre os subconjuntos possíveis. Enquanto que as *boxplots* com a técnica *K-fold* contém 50 sorteios das 210 combinações possíveis para treino ($C_{10,6}$) entre os subconjuntos de teste e treino.

Em preto, são os resultados para o MLP e apresentam a menor dispersão entre as técnicas, menos de 10 pontos entre o menor e o maior índice SP atingido. Na cor vermelha são os resultados para as redes ELM, nestes, os resultados obtidos se aproximaram dos índices alcançados pelo MLP, sugerindo equivalência nos resultados obtidos. Já em azul, os resultados para as redes ESN, nestes, apesar de alcançarem picos de índice SP superiores às duas anteriores, em todas as regiões da base, possuem maior dispersão entre os resultados alcançados.

Figura 42 – Boxplot ELM × ESN × MLP para cada região. Nesta, os resultados foram obtidos utilizando a técnica de reamostragem *Jackknife*. Sobre a indicação de cada região está a boxplot para a ELM, à esquerda a boxplot do MLP e à direita a boxplot da ESN.

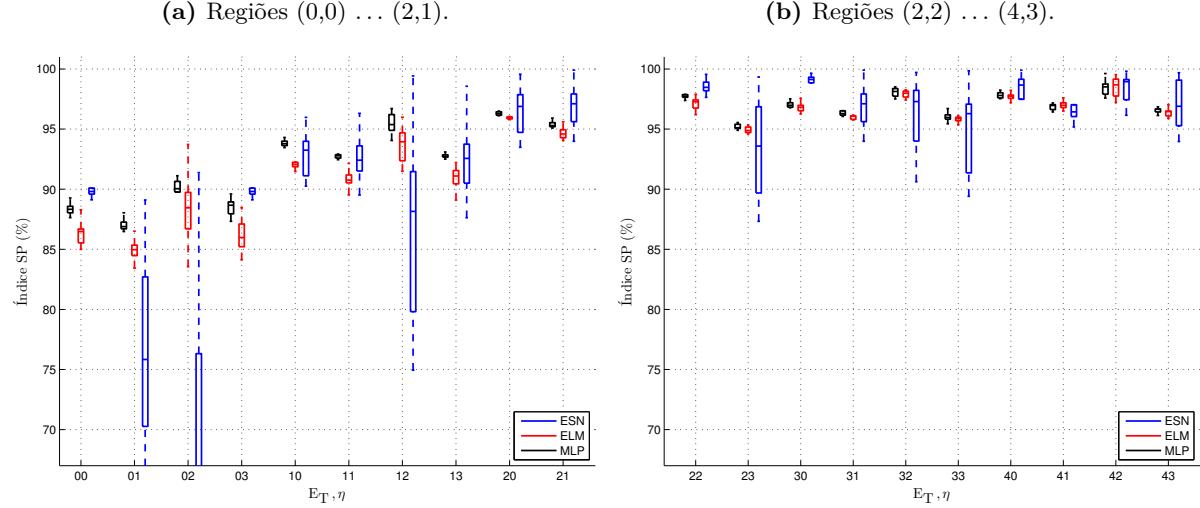
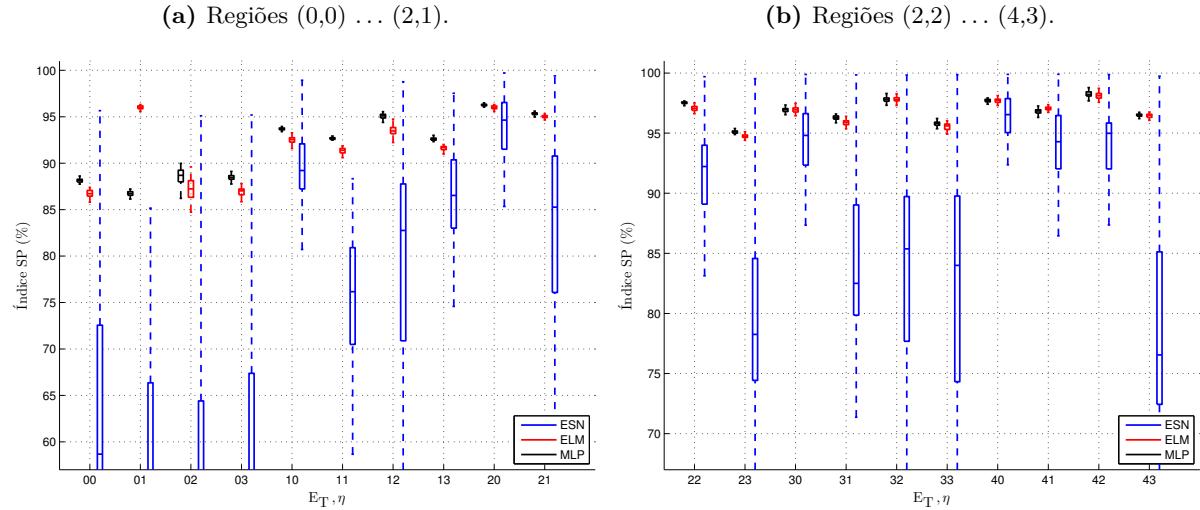


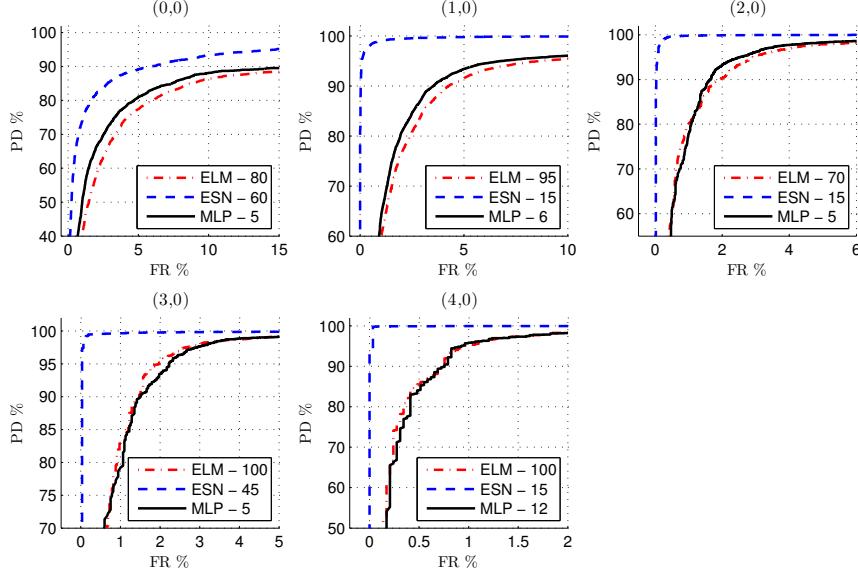
Figura 43 – Boxplot ELM × ESN × MLP para cada região. Nesta, os resultados foram obtidos utilizando a técnica de reamostragem *K-fold*. Sobre a indicação de cada região está a boxplot para a ELM, à esquerda a boxplot do MLP e à direita a boxplot da ESN.



A seguir, serão apresentadas as curvas ROC para as melhores boxplot dos classificadores nas técnicas: MLP, ELM e ESN. Os resultados referem-se aos treinamentos que utilizaram a técnica de reamostragem por *Jackknife*, uma vez que os resultados obtidos possuem baixa dispersão em comparação com os resultados obtidos com a técnica *K-fold*. Para auxiliar na análise, as curvas ROC foram organizadas em conjuntos de 5, referentes aos níveis de energia transversa (Δ_{ET}) simulados para a base em cada uma das regiões de η .

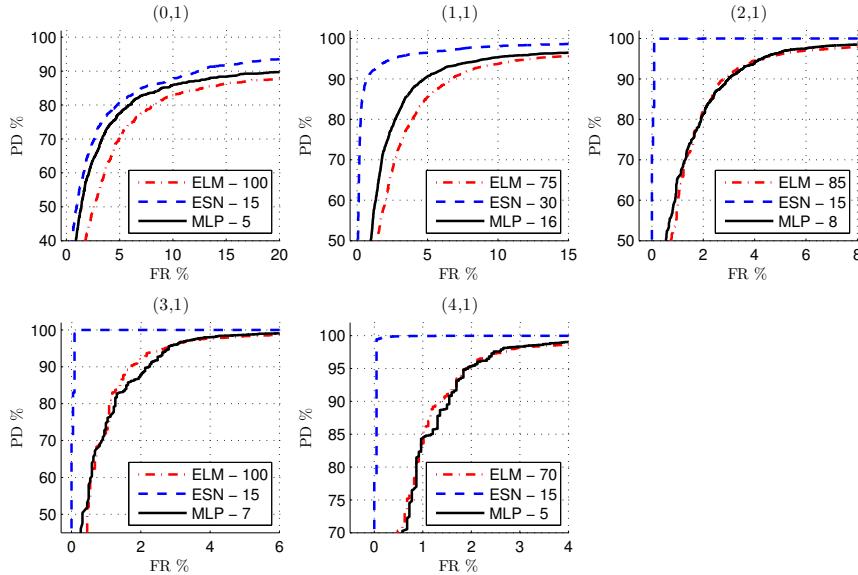
Na Figura 44 é exibido o conjunto das curvas ROC para as regiões: (0,0), (1,0), (2,0), (3,0) e (4,0). Nestas cinco regiões ($0 \leq \Delta_{|\eta|} \leq 0,8$) da base simulada, observa-se que o desempenho dos classificadores com redes ESN é superior em todas as regiões às redes ELM e MLP. Sobre o desempenho das redes ELM, percebe-se uma similaridade de desempenho na comparação com as redes MLP, nas regiões (1,0), (2,0), (3,0) e (4,0) nas quais as curvas quase se sobrepõem.

Figura 44 – ROCs para os melhores índices SP, regiões (0,0), (1,0), (2,0), (3,0) e (4,0).



Na Figura 45 é exibido o conjunto das curvas ROC para as regiões: (0,1), (1,1), (2,1), (3,1) e (4,1). Nestas cinco regiões ($0,8 \leq \Delta_{|\eta|} \leq 1,37$) da base simulada, nota-se que o desempenho dos classificadores com redes ESN é superior em todas as regiões às redes ELM e MLP. Observando o desempenho das redes ELM, nota-se similaridade quando comparado ao desempenho das redes MLP. Em especial, as regiões (2,1), (3,1) e (4,1) observa-se uma superposição das curvas.

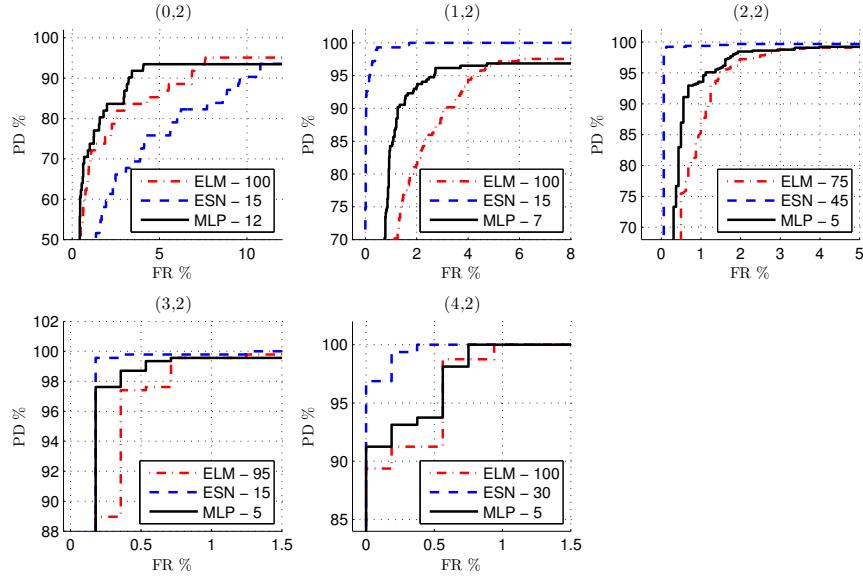
Figura 45 – ROCs para os melhores índices SP, regiões (0,1), (1,1), (2,1), (3,1) e (4,1).



Na Figura 46 é exibido o conjunto das curvas ROC para as regiões: (0,2), (1,2), (2,2), (3,2) e (4,2). Nestas cinco regiões ($1,37 \leq \Delta_{|\eta|} \leq 1,54$) da base simulada, observa-se que o desempenho dos classificadores com redes ESN é superior às redes ELM e MLP em quatro das regiões, perdendo para as redes MLP e com desempenho equivalente à ELM quando $15 \leq E_T \leq 20$.

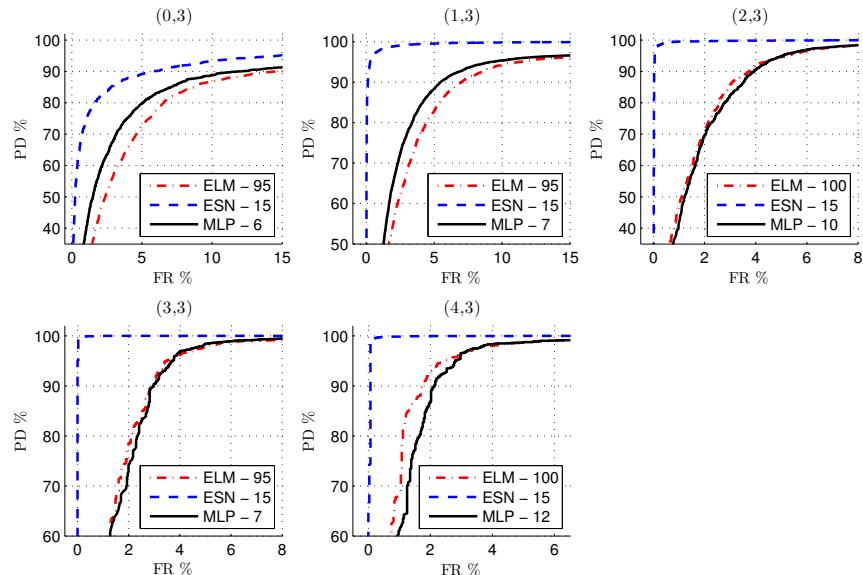
GeV (0,2). Quanto ao desempenho das redes ELM, nota-se similaridade de desempenho quando comparado ao desempenho das redes MLP, somente nas regiões (0,2), (2,2) e (4,2).

Figura 46 – ROCs para os melhores índices SP, regiões (0,2), (1,2), (2,2), (3,2) e (4,2).



Na Figura 47 é exibido o conjunto das curvas ROC para as regiões: (0,3), (1,3), (2,3), (3,3) e (4,3). Nestas cinco primeiras regiões ($1,54 \leq \Delta_{|\eta|} \leq 2,5$) da base simulada, observa-se que o desempenho dos classificadores com redes ESN é superior em todas as regiões às redes ELM e MLP. Quanto às redes ELM, nota-se similaridade de desempenho quando comparado às redes MLP nas regiões (2,3), (3,3) e (4,3), nas quais, as curvas, quase se sobrepõem, como ocorrido na Figura 44.

Figura 47 – ROCs para os melhores índices SP, regiões (0,3), (1,3), (2,3), (3,3) e (4,3).



A seguir na [Tabela 10](#), os máximos valores de índice SP alcançados para as melhores redes de cada técnica. Na tabela, é possível confirmar o desempenho observado nas curvas ROC, porém, análise será feita quanto à qualidade do resultado. Ou seja, avaliar o nível de incerteza associada ao máximo valor para o índice SP alcançado.

Para se ter uma referência, os valores de incerteza alcançados pelas redes MLP estão destacados (em negrito) afim de auxiliar na identificação. Iniciando pelas redes ELM, em três regiões, (3,2), (3,3) e (4,1) a incerteza foi menor do que a incerteza para as redes MLP. Nas demais regiões, a incerteza no valor do máximo índice SP alcançado foi maior. Entretanto, os valores não foram elevados. No melhor caso, a incerteza para mais, foi de 0,024% sendo registrada na região (4,3), e o caso onde se registrou a maior incerteza foi na região região (1,3), com diferença de 0,725%.

No que se refere à incerteza para os resultados alcançados pelas redes ESN, quatro regiões obtiveram valores de incerteza próximo aos alcançados pelas redes MLP: (0,0), (0,3), (2,2) e (4,3). Sendo a região (0,3) a de menor valor, 0,885% que é 1,32 vezes. Já a maior incerteza foi registrada na região (2,3), 4,137% que é superior ao valor alcançado pela rede MLP 19,88 vezes.

A avaliação do nível de incerteza das técnicas é importante, pois mesmo que uma técnica alcance melhores índices SP, a incerteza pode descredibilizar o resultado, uma vez que abrirá uma margem larga para o valor verdadeiro, além do fato de ultrapassar o limite inferior da técnica de referência (MLP).

Tabela 10 – Índice SP para os melhores resultados da ELM e ESN comparados com os valores obtidos com o MLP, em cada região. Cada coluna representa uma posição no interior do detector e em cada linha pode-se observar a melhora nos índices com a elevação da energia (Δ_{E_T} , $\Delta_{|\eta|}$).

	(0,0)	(0,1)	(0,2)	(0,3)
ELM	$88,288 \pm 0,993$	$86,509 \pm 0,837$	$93,720 \pm 2,828$	$88,471 \pm 1,299$
MLP	$89,270 \pm \mathbf{0,441}$	$88,035 \pm \mathbf{0,586}$	$94,673 \pm \mathbf{2,483}$	$89,600 \pm \mathbf{0,667}$
ESN	$92,146 \pm 0,885$	$89,106 \pm 8,641$	$91,378 \pm 15,417$	$92,146 \pm 0,885$
	(1,0)	(1,1)	(1,2)	(1,3)
ELM	$93,596 \pm 0,642$	$92,145 \pm 0,740$	$95,993 \pm 1,340$	$92,215 \pm 0,883$
MLP	$94,302 \pm \mathbf{0,247}$	$93,340 \pm \mathbf{0,241}$	$96,711 \pm \mathbf{0,841}$	$93,094 \pm \mathbf{0,158}$
ESN	$99,006 \pm 3,280$	$96,310 \pm 1,990$	$99,421 \pm 7,847$	$98,563 \pm 2,958$
	(2,0)	(2,1)	(2,2)	(2,3)
ELM	$96,545 \pm 0,361$	$95,611 \pm 0,510$	$97,890 \pm 0,498$	$95,347 \pm 0,260$
MLP	$96,909 \pm \mathbf{0,239}$	$95,921 \pm \mathbf{0,265}$	$98,268 \pm \mathbf{0,249}$	$95,543 \pm \mathbf{0,208}$
ESN	$99,566 \pm 2,840$	$99,905 \pm 1,843$	$99,554 \pm 0,618$	$99,339 \pm 4,137$
	(3,0)	(3,1)	(3,2)	(3,3)
ELM	$97,567 \pm 0,486$	$96,889 \pm 0,448$	$99,427 \pm 0,559$	$96,485 \pm 0,318$
MLP	$97,515 \pm \mathbf{0,250}$	$97,056 \pm \mathbf{0,292}$	$99,427 \pm \mathbf{0,563}$	$96,712 \pm \mathbf{0,347}$
ESN	$99,644 \pm 1,151$	$99,905 \pm 1,843$	$99,714 \pm 3,995$	$99,847 \pm 3,594$
	(4,0)	(4,1)	(4,2)	(4,3)
ELM	$98,228 \pm 0,295$	$97,604 \pm 0,316$	$99,530 \pm 0,777$	$97,036 \pm 0,325$
MLP	$98,238 \pm \mathbf{0,239}$	$97,732 \pm \mathbf{0,369}$	$99,624 \pm \mathbf{0,614}$	$97,221 \pm \mathbf{0,301}$
ESN	$99,896 \pm 2,532$	$99,775 \pm 2,780$	$99,813 \pm 1,206$	$99,686 \pm 2,078$

A seguir na [Tabela 11](#), os valores de probabilidade de detecção (PD) associados aos

índices SP ([Tabela 10](#)) da melhor *boxplot* para cada uma das técnicas, MLP, ELM e ESN. A análise feita tomando esse parâmetro como referência é importante, pois refere-se ao nível de acerto na detecção de elétrons. Indica a qualidade de classificação, correta de elétrons e jatos. Dessa forma, elevados índices PD são desejados pois reduz-se a chance de perda ou descarte de um evento que contenha um elétron que possa estar relacionado com a física de interesse.

Como ocorreu com os índices SP, os valores de PD para a ELM são bem próximas, tanto nos valores alcançados, quanto na incerteza registrada. Na região (0,2), foi onde a ELM obteve o melhor índice PD, sendo superior à probabilidade de detecção atingida tanto pelo MLP quanto pela ESN.

Em relação à ESN, os valores de PD registrados são melhores que o MLP e a ELM em 18 das vinte regiões, com exceção das regiões (0,2), na qual a ELM foi superior, e na região (4,2), que possui o menor número de assinaturas, região onde houve desempenho equivalente para as três técnicas. Porém, a incerteza é elevada em relação MLP.

Tabela 11 – Probabilidade de detecção (PD) para os melhores resultados, ELM e ESN comparados com os valores obtidos com o MLP, em cada região. Cada coluna representa uma posição no interior do detector e em cada linha pode-se observar a melhora nos índices com a elevação da energia (Δ_{E_T} , $\Delta_{|\eta|}$).

	(0,0)	(0,1)	(0,2)	(0,3)
ELM	$86,412 \pm 0,967$	$82,929 \pm 1,748$	$95,082 \pm 5,446$	$87,004 \pm 1,299$
MLP	$87,622 \pm \mathbf{1,311}$	$85,125 \pm \mathbf{1,459}$	$93,443 \pm \mathbf{4,179}$	$87,555 \pm \mathbf{1,055}$
ESN	$90,042 \pm 1,633$	$86,084 \pm 6,391$	$93,548 \pm 16,041$	$90,042 \pm 1,633$
	(1,0)	(1,1)	(1,2)	(1,3)
ELM	$93,088 \pm 0,463$	$92,707 \pm 0,781$	$97,193 \pm 2,118$	$93,923 \pm 1,208$
MLP	$94,295 \pm \mathbf{0,412}$	$93,089 \pm \mathbf{0,427}$	$96,154 \pm \mathbf{1,224}$	$94,195 \pm \mathbf{0,589}$
ESN	$98,744 \pm 2,160$	$95,250 \pm 1,303$	$99,301 \pm 6,385$	$98,706 \pm 2,336$
	(2,0)	(2,1)	(2,2)	(2,3)
ELM	$97,470 \pm 0,581$	$96,764 \pm 0,865$	$98,773 \pm 1,024$	$97,484 \pm 0,428$
MLP	$97,473 \pm \mathbf{0,270}$	$97,261 \pm \mathbf{0,453}$	$98,466 \pm \mathbf{0,510}$	$97,328 \pm \mathbf{0,396}$
ESN	$99,550 \pm 1,767$	$99,946 \pm 1,108$	$99,233 \pm 0,593$	$99,362 \pm 3,068$
	(3,0)	(3,1)	(3,2)	(3,3)
ELM	$98,529 \pm 0,398$	$97,493 \pm 0,485$	$99,567 \pm 0,596$	$98,610 \pm 0,537$
MLP	$98,710 \pm \mathbf{0,358}$	$98,377 \pm \mathbf{0,448}$	$99,567 \pm \mathbf{0,385}$	$98,432 \pm \mathbf{0,385}$
ESN	$99,477 \pm 0,701$	$99,946 \pm 1,108$	$99,784 \pm 4,158$	$99,782 \pm 2,734$
	(4,0)	(4,1)	(4,2)	(4,3)
ELM	$98,278 \pm 0,284$	$98,254 \pm 0,525$	$100,000 \pm 1,290$	$97,958 \pm 0,513$
MLP	$98,887 \pm \mathbf{0,373}$	$98,123 \pm \mathbf{0,444}$	$100,000 \pm \mathbf{1,043}$	$98,296 \pm \mathbf{0,301}$
ESN	$99,895 \pm 1,456$	$99,695 \pm 1,931$	$100,000 \pm 1,155$	$99,676 \pm 1,468$

A seguir na [Tabela 12](#), os valores de falso alarme (FR) associados ao índices SP da melhor *boxplot* para cada uma das técnicas, MLP, ELM e ESN apresentados na [Tabela 10](#). Esse parâmetro está associado aos jatos identificados erroneamente como elétrons. Idealmente, deseja-se que a taxa de falso alarme seja zero. Entretanto, elevados índices, acarretam em processamento futuro de eventos classificados de maneira errada, demando tempo de processamento e armazenamento desnecessários. Não há perda de eventos raros, como ocorre na situação de baixos índices de PD,

os quais sugerem que informação relevante, de rara ocorrência, pode ser descartada.

Nos resultados, somente na região (0,2) que o MLP produziu o menor índice de FR, nas demais, obteve valores acima da ELM e/ou ESN. A ELM produziu o seu menor índice na região (4,0). Já a ESN, obteve o menor resultado nas 18 demais. Também há de notar que a elevada incerteza já percebida com o índice SP esteve presente tanto no índice PD ([Tabela 11](#)) como nos valores do índice FR da [Tabela 12](#).

Tabela 12 – Taxa de falso alarme (FR) para os melhores resultados, ELM e ESN comparados com os valores obtidos com o MLP, em cada região. Cada coluna representa uma posição no interior do detector e em cada linha pode-se observar a melhora nos índices com a elevação da energia (Δ_{E_T} , $\Delta_{|\eta|}$).

	(0,0)	(0,1)	(0,2)	(0,3)
ELM	$9,817 \pm 1,497$	$9,835 \pm 1,134$	$7,632 \pm 2,168$	$10,050 \pm 2,026$
MLP	$9,067 \pm \mathbf{0,891}$	$9,007 \pm \mathbf{1,318}$	$4,088 \pm \mathbf{2,379}$	$8,332 \pm \mathbf{0,901}$
ESN	$5,725 \pm 1,325$	$7,820 \pm 12,683$	$10,767 \pm 18,364$	$5,725 \pm 1,325$
	(1,0)	(1,1)	(1,2)	(1,3)
ELM	$5,894 \pm 1,087$	$8,416 \pm 1,147$	$5,200 \pm 1,847$	$9,477 \pm 1,087$
MLP	$5,691 \pm \mathbf{0,413}$	$6,409 \pm \mathbf{0,351}$	$2,731 \pm \mathbf{1,168}$	$8,000 \pm \mathbf{0,556}$
ESN	$0,731 \pm 4,743$	$2,625 \pm 2,873$	$0,458 \pm 10,188$	$1,580 \pm 3,751$
	(2,0)	(2,1)	(2,2)	(2,3)
ELM	$4,376 \pm 0,421$	$5,535 \pm 0,789$	$2,989 \pm 0,697$	$6,767 \pm 0,348$
MLP	$3,654 \pm \mathbf{0,562}$	$5,409 \pm \mathbf{0,443}$	$1,930 \pm \mathbf{0,410}$	$6,225 \pm \mathbf{0,527}$
ESN	$0,417 \pm 4,006$	$0,136 \pm 2,702$	$0,125 \pm 0,901$	$0,685 \pm 5,260$
	(3,0)	(3,1)	(3,2)	(3,3)
ELM	$3,390 \pm 0,761$	$3,712 \pm 0,646$	$0,713 \pm 0,867$	$5,617 \pm 0,616$
MLP	$3,672 \pm \mathbf{0,454}$	$4,255 \pm \mathbf{0,502}$	$0,713 \pm \mathbf{0,931}$	$4,993 \pm \mathbf{0,584}$
ESN	$0,188 \pm 1,608$	$0,136 \pm 2,702$	$0,357 \pm 4,510$	$0,089 \pm 5,273$
	(4,0)	(4,1)	(4,2)	(4,3)
ELM	$1,823 \pm 0,454$	$3,045 \pm 0,721$	$0,938 \pm 0,772$	$3,883 \pm 0,547$
MLP	$2,408 \pm \mathbf{0,342}$	$2,658 \pm \mathbf{0,544}$	$0,751 \pm \mathbf{0,488}$	$3,848 \pm \mathbf{0,591}$
ESN	$0,103 \pm 3,676$	$0,145 \pm 3,668$	$0,375 \pm 1,490$	$0,304 \pm 3,025$

5.1.3.3 Tempos de Treinamento

A seguir, na [Tabela 13](#), os tempos de treinamento utilizados para cada uma das três técnicas. Dentre as três técnicas, as redes MLP foram as que mais precisaram de tempo para o seu treinamento. No melhor caso, 2,48 s para a região (4,2), contra 0,15 s para a ELM e 0,33 s para a ESN. Para o pior caso, 284,92 s para a região (0,3), contra 31,14 s para ELM e 58,69 s para a ESN.

As redes ELM foram as que necessitaram de menos tempo para o seu treino, em todas as regiões. Em comparação com a técnica de referência (MLP) a diferença mínima foi da ordem de 5,58 vezes, na região (0,2), a qual possui o sexto menor número de assinaturas, 98.232, já a maior diferença foi registrada na região (4,3) com tempo de treino 16,53 vezes menor, a qual possui o sétimo menor número de assinaturas, 100.640.

Tabela 13 – Tempo de treinamento em segundos, para os melhores resultados, ELM × ESN × MLP, em cada região.

Tempos, em segundos, por Região na base				
	(0,0)	(0,1)	(0,2)	(0,3)
t_{ELM}	$21,400 \pm 2,885$	$21,590 \pm 3,152$	$5,420 \pm 1,029$	$31,140 \pm 4,161$
t_{MLP}	$242,080 \pm 33,410$	$166,430 \pm 23,709$	$30,250 \pm 5,217$	$284,920 \pm 47,683$
t_{ESN}	$58,690 \pm 11,812$	$13,720 \pm 6,155$	$7,280 \pm 1,626$	$58,690 \pm 11,812$
	(1,0)	(1,1)	(1,2)	(1,3)
t_{ELM}	$23,420 \pm 2,842$	$12,380 \pm 1,695$	$3,340 \pm 0,673$	$25,610 \pm 3,446$
t_{MLP}	$216,030 \pm 31,058$	$109,000 \pm 23,037$	$32,900 \pm 4,827$	$252,900 \pm 35,670$
t_{ESN}	$35,180 \pm 7,550$	$13,170 \pm 5,478$	$4,950 \pm 1,097$	$35,910 \pm 7,831$
	(2,0)	(2,1)	(2,2)	(2,3)
t_{ELM}	$16,530 \pm 2,548$	$10,340 \pm 1,448$	$0,760 \pm 0,137$	$16,030 \pm 1,204$
t_{MLP}	$158,580 \pm 22,600$	$96,540 \pm 14,778$	$10,480 \pm 1,601$	$100,680 \pm 21,517$
t_{ESN}	$27,650 \pm 6,120$	$13,720 \pm 2,741$	$2,200 \pm 0,458$	$19,210 \pm 4,252$
	(3,0)	(3,1)	(3,2)	(3,3)
t_{ELM}	$19,660 \pm 2,674$	$12,050 \pm 1,408$	$0,390 \pm 0,071$	$12,420 \pm 1,782$
t_{MLP}	$138,990 \pm 18,764$	$109,530 \pm 17,318$	$3,670 \pm 0,924$	$125,180 \pm 18,856$
t_{ESN}	$30,920 \pm 6,250$	$13,720 \pm 2,741$	$0,730 \pm 0,151$	$8,340 \pm 3,392$
	(4,0)	(4,1)	(4,2)	(4,3)
t_{ELM}	$7,840 \pm 1,167$	$1,980 \pm 0,578$	$0,150 \pm 0,049$	$5,770 \pm 0,975$
t_{MLP}	$55,820 \pm 7,997$	$24,830 \pm 5,666$	$2,480 \pm 0,569$	$40,980 \pm 6,117$
t_{ESN}	$9,770 \pm 2,237$	$6,570 \pm 1,518$	$0,330 \pm 0,113$	$7,400 \pm 1,693$

Para as redes ESN o tempo de treinamento necessário para 18 das 20 regiões, estiveram entre o tempo utilizado pelas redes ELM e o tempo necessário para as redes MLP, exceto nas regiões (0,1) e (3,3). Nas regiões (0,0) e (0,3) (dois maiores números de assinaturas) foram registrados os maiores tempos de treinamento para as redes ESN. No entanto, as redes foram treinadas 4,12 e 4,85 vezes mais rápido do que o treino nas redes MLP. E, na região (4,1) (quinto menor número de assinaturas, 89.425), foi registrado o menor tempo de treinamento, sendo 7,55 vezes mais rápido do que o tempo gasto pelas redes MLP.

Cabe salientar, que as redes ESN possuem em sua estrutura uma etapa recorrente, mas, que, apesar dessa estrutura, ainda pôde ser treinada 4,85 vezes mais rápido do que as redes MLP na região (0,3), a qual possui o maior número de assinaturas da base, 618.352.

Ao se comparar os tempos de treinamento para as redes ELM em relação às redes ESN, observa-se que na região (4,1) (quinto menor número de assinaturas, 84.425), foi registrado um tempo 3,31 vezes menor que o tempo para treino das redes ESN, e, na região (0,1) (quarto maior número de assinaturas, 384.619), foi registrado que o tempo de treinamento foi 0,635 vezes menor.

5.1.4 Análise Estatística

A seguir os resultados obtidos com o teste de Student para os melhores resultados obtidos com a base de dados simulados, utilizando as expressões das Equações 4.7, 4.8 e 4.9. Os dados de

entrada das equações foram os índices SP das melhores inicializações para cada uma das técnicas. Da mesma maneira que foi feita na base de dados experimentais, três análises serão apresentadas: ELM × MLP, ESN × MLP e ELM × ESN. A hipótese nula, H_0 , é de que os classificadores possuem o mesmo desempenho de separação.

O valor de referência tabelado para uma distribuição de Student com nove graus ($k-1 = 9$, Equação 4.7) de liberdade é $t_{tab} = 2,262$ para um nível de significância de 95%. Os dados utilizados são os dados que produziram as melhores *boxplot* da Figura 42

- ELM × MLP: Na Tabela 14 são apresentados os resultados do teste de comparação entre os classificadores. Observa-se que os resultados para as regiões de $E_T = 0$ e $E_T = 1$ (O que corresponde a faixa de energia [15;30] GeV), os classificadores não possuem mesmo desempenho de classificação, destacados em negrito. Nas demais o teste indicou semelhança, ou seja, $t_{calc} < t_{tab}$, sendo $t_{tab} = 2,262$.

Tabela 14 – Resultados do teste de Student, ELM × MLP, para cada uma das regiões ($E_T, |\eta|$) da base.

		$ \eta $			
		0	1	2	3
E_T	0	3,797	3,575	2,937	4,331
	1	2,872	3,526	3,153	3,302
	2	0,767	1,241	1,109	0,510
	3	0,513	0,651	0,182	0,253
	4	0,224	0,117	0,005	0,387

- ESN × MLP: Na Tabela 15 são apresentados os resultados do teste de comparação entre os classificadores. Observa-se que somente em três regiões (em negrito), (0,1), (0,2) e (1,2), o valor de $t_{calc} > t_{tab}$, sendo $t_{tab} = 2,262$, nas demais os classificadores possuem mesmo desempenho de classificação.

Tabela 15 – Resultados do teste de Student, ESN × MLP, para cada uma das regiões ($E_T, |\eta|$) da base.

		$ \eta $			
		0	1	2	3
E_T	0	0,726	4,836	11,344	0,647
	1	0,352	0,056	3,557	0,166
	2	0,141	0,612	0,329	0,752
	3	0,760	0,187	1,006	0,563
	4	0,070	0,298	0,036	0,163

- ELM × ESN: Na Tabela 16 é apresentado o resultado da avaliação entre os classificadores. Neste teste os resultados indicam que nas regiões (0,1), (0,2) e (1,2), os classificadores não apresentam mesmo desempenho, pois o $t_{calc} > t_{tab}$, sendo $t_{tab} = 2,262$.

Tabela 16 – Resultados do teste de Student, ELM × ESN, para cada uma das regiões (E_T , $|\eta|$) da base.

		$ \eta $			
		0	1	2	3
E_T	0	1,647	4,064	10,836	1,695
	1	0,329	0,787	2,864	0,621
	2	0,040	0,920	0,600	0,643
	3	0,896	0,346	0,980	0,512
		4	0,018	0,331	0,038
					0,258

5.2 Análise dos Resultados

Nas duas bases utilizadas para avaliação das técnicas propostas, o desempenho obtido foi quantitativamente semelhante e satisfatório. Para as redes ELM, o desempenho alcançado para o índice SP foi muito próximo do desempenho alcançado pelo algoritmo de referência em utilização no *Neural Ringer*. A diferença entre o índice alcançado com a ELM esteve abaixo de 2%, sugerindo semelhança entre as técnicas, tanto para a base experimental quanto para a base de dados simulados.

Como critério qualitativo, as curvas ROC também indicaram que a ELM pode ser utilizada como alternativa ao MLP, pois as curvas estiveram próximas de uma superposição, para a base experimental. Na base de dados simulados, o comportamento de similaridade também foi observado, e em 10 das 20 regiões estiveram próximas de se sobrepor, sugerindo equivalência de desempenho. Nas demais regiões, não ficaram próximas de uma sobreposição, porém a diferença não foi significativa, o que pode ser verificado com o auxílio da Tabela 10 que registra a proximidade dos índices alcançados.

Em relação às redes ESN, seu desempenho quantitativo é expressivo e superior as demais técnicas, na base experimental e em quase todas as regiões da base de dados simulados, exceto em uma região (0,2), quando o seu desempenho foi inferior ao desempenho das redes MLP e foi muito semelhante ao desempenho das redes ELM, com as curvas quase sobrepostas. A análise qualitativa é feita observando o grau de incerteza alcançado para os melhores treinos, os quais foram elevados quando comparados com o MLP, sendo de até 12,93%.

Outro ponto importante a se comentar é que o desempenho dos classificadores é elevado ao passo que a energia aumenta. Esse fato pode ser observado em todas as quatro regiões de $\eta = [0, 1, 2, 3]$, (E_T, η), nas Figuras 44, 45, 46 e 47.

Quanto ao tempo total de treinamento necessário, as duas técnicas propostas (ELM e ESN) são mais rápidas que o MLP. E isso considerando todos os resultados obtidos, tanto com a base experimental quanto a base simulada. No pior caso, a diferença foi de pelo menos 3,78 vezes mais rápido que o MLP, o que representa uma redução de 73,54%. E no melhor, a diferença foi de 16,53 vezes, representando uma redução de 93,95%. Nesse quesito, as técnicas propostas demonstram que seu tempo de treinamento podem permitir um número maior de ensaios de configurações, objetivando a rede classificadora de máxima eficiência dentro do mesmo tempo

gasto para treino de uma rede de máxima eficiência do tipo MLP.

Por fim, foi realizada a análise estatística da qualidade dos resultados obtidos com os três classificadores. Três comparações foram realizadas: ELM×MLP, ESN×MLP e ESN×ELM. Para a base de dados experimentais, o resultado demonstrou que há diferenças significativas entre os classificadores ELM×MLP. O resultado do teste foi de $t_{calc} = 5,191$, enquanto que o valor tabulado é de $t_{tab} = 2,262$, neste caso $t_{calc} > t_{tab}$, logo, a hipótese nula é descartada. Pois os classificadores possuem diferenças significativas.

Na comparação ESN×MLP, o resultado indicou que os classificadores não possuem diferenças significativas. Pois o resultado foi de $t_{calc} = 1,146$, inferior ao valor tabulado. Dessa forma os classificadores possuem resultados similares. E na última comparação, ESN×ELM, o resultado também indicou similaridade entre os classificadores, pois $t_{calc} = 0,500$.

Na base de dados segmentada, com dados simulados, a comparação entre ELM e MLP resultou na separação da base em dois grupos. O primeiro, são as regiões onde E_T está na faixa de até 30 GeV, (0,0) até (1,0), oito regiões. Nestas regiões, o teste indicou que há semelhanças significativas entre os classificadores, com o mínimo valor calculado foi de 2,872 na região (1,0), e o máximo de 3,797, região (0,0). Os valores de t_{calc} foram superiores ao de $t_{tab} = 2,262$. No segundo grupo, E_T está na faixa acima de até 30 GeV, (2,0) até (4,3), 12 regiões. Nestas, o indicou indicou que os classificadores são semelhantes.

No trabalho de **SIMAS FILHO, SEIXAS e CALÔBA (2010, p 52-53)**, uma análise indicou, que o nível de energia envolvida no evento registrado, possui influência na separação entre elétrons e jatos. A [Figura 13](#) exibiu tal fato. Observou-se, que a separação entre jatos e elétrons fica mais saliente para níveis de energia acima de 20 GeV. Tal fato, pode auxiliar na compreensão dos resultados indicarem semelhança significativa. Pois, nessa faixa de energia, o classificador baseado em ELM apresentou mais dificuldade em separar os elétrons dos jatos,

Na comparação entre ESN e MLP, houve um maior número de semelhanças entre os classificadores. Somente as regiões (0,1), (0,2) e (1,2), que o teste indicou diferenças significativas entre os classificadores. Os valores para o t_{calc} foram de 4,836, 11,344 e 3,557, respectivamente. Nas demais 17, todos valores ficaram abaixo do $t_{tab} = 2,262$, indicando que os classificadores ESN e MLP para esses regiões são semelhantes.

A última comparação realizada, ESN e ELM, produziu resultado semelhante à comparação entre ESN e MLP. Somente as regiões (0,1), (0,2) e (1,2), que o teste indicou diferenças significativas entre os classificadores. Os valores para o t_{calc} foram de 4,064, 10,836 e 2,864, respectivamente. As demais 17 regiões, os valores calculados ficaram abaixo do $t_{tab} = 2,262$, indicando semelhança entre os classificadores ESN e ELM.

Na base experimental, o teste estatístico rejeitou a hipótese de semelhança entre os classificadores ELM e MLP. Enquanto que na comparação ESN × MLP e ELM × ESN, indicou semelhança entre os classificadores. É possível que ajustes na forma de gerar os pesos da camada oculta, e o ajuste do número de neurônios utilizados na camada oculta, produzam melhores resultados, aproximando o desempenho da ELM do MLP.

Um detalhe a ser observado, é o fato de as bases de dados possuírem características

significativamente distintas. Visto que, uma não possui segmentação (base experimental), ou seja, todas as variações, em níveis de energia, registradas nas colisões foram fornecidas à rede para que efetuasse a classificação, sem nenhum tipo de tratamento. Enquanto que na base segmentada, os classificadores foram projetados por faixa de energia e região onde o evento foi registrado. Também, há de se observar que a base não segmentada (experimental) possui um total 416.011 assinaturas, para jatos e elétrons, enquanto a base segmentada, possui regiões com número assinaturas superior a toda base experimental.

5.2.1 Resumo

A seguir na Tabela 17 é exibido o resumo dos resultados obtidos nessa esta pesquisa de mestrado com as técnicas propostas MLP, ELM e ESN.

Tabela 17 – Resumo dos melhores resultados obtidos com os classificadores, MLP, ELM e ESN, para as duas bases utilizadas, tendo como técnica de reamostragem o *Jackknife*.

Base Experimental	
Tempo de Treinamento:	A ELM e ESN foram treinadas num tempo significativamente menor, comparado com o MLP, 12,65 e 11,18 vezes mais rápido, uma redução de 92,09% e 91,06%, respectivamente.
Índice SP:	A ELM obteve desempenho de $92,875\% \pm 0,460\%$, enquanto o MLP $93,411\% \pm 0,399\%$. Já a ESN alcançou o melhor índice $98,564\% \pm 3,924\%$.
Teste Estatístico:	Utilizando os dados da melhor inicialização para cada uma das técnicas, o resultado indicado, foi de similaridade na comparação ESN×MLP e ESN×ELM. Porém, na comparação ELM×MLP o teste indicou que os classificadores projetados possuem desempenho de classificação com diferenças significativas.
Base Simulada	
Tempo de Treinamento:	As técnicas propostas, ELM e ESN, foram treinadas num tempo de pelo menos 4,12 vezes mais rápido do que o MLP. A ELM obteve o menor tempo na região (4,3), sendo 16,53 vezes menor que o MLP, no pior caso, região (0,2) a diferença foi de 5,58 vezes. A ESN teve o treino mais rápido na região (4,2), com tempo 7,55 vezes menor que o MLP, no pior caso, região (0,0) (segundo maior número de assinaturas), o tempo foi 4,12 vezes menor. O tempo de treinamento da ESN ficou entre o tempo para o MLP e a ELM em 18 regiões, exceto para as regiões (0,1) e (3,3), nas quais foi a mais rápida.
Índice SP:	A ELM obteve desempenho semelhante ao MLP, $92,875\% \pm 0,460\%$, enquanto o MLP $93,411\% \pm 0,399\%$. Já a ESN alcançou o melhor índice $98,564\% \pm 3,924\%$.

Continua na próxima página

Tabela 17 – Continuação da página anterior

Base Simulada

Teste Estatístico:	No teste ELM×MLP, as regiões (0,0) … (1,3), com E_T na faixa [15, 30] GeV, os classificadores projetados com ELM não são semelhantes, pois o $t_{calc} > 2,262$ para uma distribuição de probabilidade bicaudal. Nas demais regiões, com E_T acima de 30 GeV, os classificadores projetados possuem semelhança. Nos testes, ESN×MLP e ELM×ESN, três regiões tiveram $t_{calc} > 2,262$: (0,1), (0,2) e (1,2). Ou seja, os classificadores projetados com ELM e ESN para essas regiões não são similares. Nas demais regiões os testes, ESN×MLP e ELM×ESN, indicaram que os classificadores possuem desempenho semelhante.
--------------------	--

6 Conclusões

As pesquisas na área de físicas de partículas ou física de altas energias (HEP - *High Energy Physics*) desenvolvidas no CERN, tem como objetivo a compreensão da natureza constituinte da matéria, descobertas na área de físicas de partículas e validação de modelos físicos teóricos. Para o alcance desse objetivos foi necessário o desenvolvimento do maior acelerador de partículas em operação, o LHC (*Large Hadron Collider*).

Neste trabalho foi proposto a utilização de Máquinas de Aprendizado Extremo (ELM) e de Redes com Estados de Eco (ESN) como alternativa ao classificador neural utilizado no detector ATLAS. Atualmente o detector utiliza como classificador o *Neural Ringer* (NR), algoritmo proposto por [Anjos et al. \(2006\)](#) para realizar a separação da informação de interesse e o ruído de fundo produzido durante as colisões no interior do detector.

Desde o início do funcionamento o níveis de energia do LHC passaram de 450 GeV por feixe para 7 TeV. Isso implica no incremento no volume de informação, e como consequência, produz desafios à operação dos algoritmos de seleção *online* de eventos. Neste contexto o NR precisa estar atualizado para operar em diferentes condições. Entretanto, o tempo disponível para o processamento vem sendo reduzido, à medida que as atualizações ocorrerem, demandando capacidade de rápido treinamento dos algoritmos de seleção dos eventos.

Nos testes realizados em duas bases de dados, uma experimental e uma com dados simulados, os classificadores baseados em ELM e ESN, produziram resultado de classificação equivalente aos resultados das redes com MLP utilizado como referência.

Os resultados demonstraram que os classificadores propostos produzem resultado equivalente ao resultado alcançado pelo classificador com redes MLP. Os resultados obtidos com as redes ELM foram muito próximos aos resultados para as redes MLP, havendo similaridade tanto para os máximos valores de índice SP alcançados quanto na incerteza.

Os resultados obtidos com as redes ESN indicaram superioridade de desempenho de classificação, alcançando índices SP superiores, tanto ao MLP quanto à ELM. Entretanto, as incertezas associadas foram superiores ao valores de incerteza tanto do MLP quanto da ELM.

No que se refere ao tempo de processamento necessário para o treinamento dos classificadores propostos, os resultados foram satisfatórios. Em todas as redes ELM e ESN de melhor desempenho, o tempo de treinamento foi inferior ao tempo gasto pelas redes MLP em pelo menos 3,78 vezes para o pior caso, e 16,53 no melhor caso. Considerando o tempo de treinamento para a ELM, no melhor caso, o tempo foi 16,53 vezes menor e no pior caso foi 5,58 menor do que o tempo gasto para o treino da melhor rede MLP na mesma região. Na ESN, o tempo de treinamento foi 15,01 vezes menor no melhor caso, e no pior caso foi 3,78 vezes menor do que o tempo para treino da rede MLP na mesma região.

No teste estatístico, os resultados obtidos foram semelhantes, no que se refere à comparação ELM×MLP. Na base experimental, o teste rejeitou a hipótese nula de semelhança entre os

classificadores, com o valor calculado sendo mais que o dobro do valor tabulado de referência. Já na base de dados simulados, houve semelhanças de desempenho entre os classificadores, de forma parcial. A comparação ELM×MLP, indicou que os classificadores ELM não apresentam mesmo desempenho de classificação para a faixa de E_T até 30 GeV. Nas regiões de maior energia, o desempenho é semelhante.

Nos testes estatístico ESN×MLP e ESN×ELM, para os resultados obtidos com a base experimental, houve semelhança de desempenho. Na base de dados simulados, somente três regiões apresentaram diferenças significativas entre os classificadores, nos dois testes: (0,1), (0,2) e (1,2). Nas demais regiões os classificadores possuem desempenho semelhante.

Com base nos testes e treinos realizados é possível utilizar as redes ELM e ESN como alternativas para o classificador utilizado no NR, mantendo o desempenho de classificação, e essencialmente, num tempo de treino significativamente menor. Tal vantagem pode permitir o projeto de um número maior de estruturas a ser utilizada, ou ainda, pode permitir obter resultados em tempo reduzido ao tempo necessário com a técnica em uso atualmente.

6.1 Trabalhos Futuros

A pesquisa desenvolvida apontou que técnicas de treinamento rápido, com estrutura diferente da utilizada pelo MLP podem alcançar níveis satisfatórios de classificação. Nesse sentido, há espaço para aprofundar a pesquisa otimizando os parâmetros das redes especialistas, com o objetivo de obter melhoria nos desempenhos alcançados, especialmente no que se refere à incerteza.

Trabalhos utilizando técnicas de estimadores M associados à redes ELM sugerem melhorias para a técnica elevando sua robustez a *outliers*, o que melhora seu desempenho de classificação (**BARRETO; BARROS, 2016**). Levando esse fato em consideração e os resultados alcançados, existem pontos de melhoria a serem trabalhados e otimizados.

Uma perspectiva futura, é a aplicação das técnicas no ambiente *online* do detector ATLAS. O objetivo é validá-las com alternativas ao classificador neural em uso. Para tanto, é necessário aprimorar os resultados alcançados.

Por se tratar de um problema de classificação com elevado volume de informação, e uma elevada taxa de eventos, problemas com características semelhantes podem ser tratados por meio das técnicas propostas.

Referências

- AAD, G. et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, v. 114, p. 191803, 2015. Citado na página 6.
- ABDI, H.; WILLIAMS, L. J. Jackknife. *Neil Salkind (Ed.), Encyclopedia of Research Design*, 2010. Disponível em: <<http://www.utd.edu/~Herve/abdi-Jackknife2010-pretty.pdf>>. Citado na página 48.
- ACHENBACH, R. et al. The ATLAS level-1 calorimeter trigger. *Journal of Instrumentation*, v. 3, n. 03, p. P03001, 2008. Disponível em: <<http://stacks.iop.org/1748-0221/3/i=03/a=P03001>>. Citado na página 25.
- AIUBE, F. A. L. *Modelagem dos preços futuros de commodities: abordagem pelo filtro de partículas*. 183 p. Tese de Doutorado — PUC-Rio, 2005. Disponível em: <https://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=7604@1>. Citado na página 44.
- ALVES, E. J. *Métodos de bootstrap e aplicações em problemas biológicos*. Dissertação (Mestrado) — UNESP, Rio Claro, 2013. Disponível em: <https://alsafi.ead.unesp.br/bitstream/handle/11449/94336/alves_ej_me_rcla.pdf?sequence=1&isAllowed=y>. Citado na página 48.
- ANJOS, A. dos et al. Neural triggering system operating on high resolution calorimetry information. *Nuclear Instruments and Methods In Physics Research A*, v. 559, p. 134–138, Abril 2006. Citado 3 vezes nas páginas 5, 25 e 75.
- ANTONELO, E.; SCHRAUWEN, B.; STROOBANDT, D. Event detection and localization for small mobile robots using reservoir computing. *Neural Networks*, v. 21, n. 6, p. 862 – 871, 2008. ISSN 0893-6080. Computational and Biological Inspired Neural Networks, selected papers from ICANN 2007. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0893608008001317>>. Citado 2 vezes nas páginas 42 e 43.
- ARAÚJO, M. P. de. Trabalho de Conclusão de Curso, *Aplicação de Redes com Estados de Eco ao Problema de Previsão de Vazão em Usina Hidrelétrica*. 2017. Citado na página 45.
- ATLAS Collaborartion. An computer generated image representing how ATLAS detects particles. 2013. Disponível em: <<https://cds.cern.ch/record/1505342>>. Citado na página 19.
- ATLAS Collaboration. ATLAS high level trigger, data acquisition and controls technical design report. *Tech. Rep., CERN, Genbra, Suíça*, 2003. Citado na página 25.
- ATLAS Collaboration. ATLAS experiment at CERN Large Hadron Collider. *Journal of Instrumentation*, v. 3, n. S08003, p. 1–407, 2008. Disponível em: <<http://stacks.iop.org/1748-0221/3/i=08/a=S08003>>. Citado na página 4.
- ATLAS Collaboration. Computer Generated image of the ATLAS calorimeter. 2008. Disponível em: <<https://cds.cern.ch/record/1095927>>. Citado na página 21.
- ATLAS Collaboration. ATLAS Fact Sheet : To raise awareness of the ATLAS detector and collaboration on the LHC. 2010. Disponível em: <<https://cds.cern.ch/record/1457044>>. Citado na página 4.
- ATLAS Collaboration. *Observation of an excess of events in the search for the Standard Model Higgs boson in the gamma-gamma channel with the ATLAS detector*. Geneva, 2012. Disponível em: <<http://cds.cern.ch/record/1460410>>. Citado 2 vezes nas páginas 5 e 6.
- ATLAS Collaboration. A particle consistent with the higgs boson observed with the ATLAS detector at the large hadron collider. *Science*, American Association for the Advancement of Science, v. 338, n. 6114, p. 1576–1582, 2012. ISSN 0036-8075. Disponível em: <<http://science.sciencemag.org/content/338/6114/1576>>. Citado na página 11.

ATLAS Collaboration. Measurement of the photon identification efficiencies with the ATLAS detector using lhc run-1 data. measurement of the photon identification efficiencies with the ATLAS detector using lhc run-1 data. *Eur. Phys. J. C*, v. 76, n. CERN-EP-2016-110. 12, p. 666. 57 p, Jun 2016. 44 pages plus author list (61 pages total), 19 figures, 2 tables, submitted to Eur. Phys. J. C, All figures including auxiliary figures are available at <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2013-04/>. Disponível em: <<https://cds.cern.ch/record/2158117>>. Citado 2 vezes nas páginas 18 e 23.

ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in lhc run 1. *Eur. Phys. J.*, C77, p. 490, 2017. Disponível em: <<https://link.springer.com/content/pdf/10.1140%2Fepjc%2Fs10052-017-5004-5.pdf>>. Citado 3 vezes nas páginas 19, 21 e 22.

ATLAS Collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. 2018. Citado na página 11.

BARRA, F. *Redes Neurais Artificiais*. 2013. Disponível em: <<https://petcivilufjf.wordpress.com/2013/07/05/redes-neurais-artificiais/>>. Acesso em: 21 set. 2017. Citado na página 30.

BARRETO, G. A.; BARROS, A. L. B. A robust extreme learning machine for pattern classification with outliers. *Neurocomputing*, v. 176, p. 3 – 13, 2016. ISSN 0925-2312. Recent Advancements in Hybrid Artificial Intelligence Systems and its Application to Real-World ProblemsSelected papers from the HAIS 2013 conference. Citado 2 vezes nas páginas 40 e 76.

BELLE, J.-L. V. *Reading Feynman*. 2017. Disponível em: <<https://readingfeynman.org/essentials/s>>. Acesso em: 08 mai. 2018. Citado na página 14.

BOCCATO, L. *Novas Propostas e Aplicações de Redes Neurais com Estados de Eco*. 210 p. Tese de Doutorado — Unicamp, 2013. Citado 4 vezes nas páginas 7, 30, 31 e 43.

BRAGA, A. de P.; CARVALHO, A. C. P. de Leon de; LUDEMRIR, T. B. *Redes Neurais Artificiais: teoria e aplicações*. 1. ed. Rio de Janeiro: LTC, 2007. Citado 4 vezes nas páginas 29, 33, 34 e 35.

BRAIBANT, S.; GIACOMELLI, G.; SPURIO, M. *Particles and Fundamental Interactions: An Introduction to Particle Physics*. 2. ed. [S.l.]: Springer Netherlands, 2012. (Undergraduate Lecture Notes in Physics). ISBN 9400724632,9789400724631. Citado 3 vezes nas páginas 12, 14 e 37.

BRUMFIEL, G. LHC prepares for data pile-up. *Nature Physics*, 2012. Disponível em: <<http://www.nature.com/news/lhc-prepares-for-data-pile-up-1.10596>>. Acesso em: 20 Set 2016. Citado na página 4.

CERN. *The ATLAS Computing Workbook*. 2015. Disponível em: <<https://twiki.cern.ch/twiki/bin/view/AtlasComputing/WorkBookAtlasExperiment>>. Acesso em: 12 nov. 2016. Citado na página 18.

CERN. *CERN Timelines All Events*. 2016. Disponível em: <<https://timeline.web.cern.ch/events>>. Acesso em: 30 ago. 2016. Citado na página 3.

CERN. LHC guide. 2017. Disponível em: <<https://cds.cern.ch/record/2255762>>. Citado 3 vezes nas páginas 14, 15 e 16.

CERN Collaborartion. The four main LHC experiments. 1999. Disponível em: <<http://cds.cern.ch/record/40525>>. Citado na página 1.

CERN-COLLABORATION. *CERN Homepage*. 2016. Disponível em: <[DisponívelEm:\\$<\\$http://www.cern.ch/\\$>\\$](DisponívelEm:$<$http://www.cern.ch/$>$)>. Acesso Em: 21 De Maio De 2016. Citado na página 1.

CIODARO, T. X. *Identificação Online de Sinais Baseada em Calorimetria de Altas Energias e com Fina Segmentação*. 147 p. Tese de Doutorado — UFRJ, 2012. Disponível em: <<http://pee.ufrj.br/teses/index.php?Resumo=2012122101>>. Citado na página 25.

COLIBORO, T. P. P. Trabalho de Conclusão de Curso, *Decomposição em valores singulares e aplicações*. 2008. Disponível em: <<http://repositorio.ufsc.br/xmlui/handle/123456789/96820>>. Citado 2 vezes nas páginas 38 e 39.

COSTA, G. G. de O. *Um procedimento inferencial para análise fatorial utilizando as técnicas Bootstrap e Jackknife: construção de intervalos de confiança e testes de hipóteses*. 196 p. Tese de Doutorado — PUC-Rio, 2006. Disponível em: <https://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=8741@1>. Citado 2 vezes nas páginas 47 e 48.

- COSTA, I. A. *Otimização dos algoritmos univariados e bivariados aplicados à identificação de elétrons no experimento ATLAS*. Dissertação (Mestrado) — UFJF, Juiz de Fora, 2016. Disponível em: <<https://repositorio.ufjf.br/jspui/handle/ufjf/4179>>. Citado na página 53.
- DVORKIN, E.; GOLDSCHMIT, M.; STORTI, M. Reconhecimento de acordes musicais: Uma abordagem via perceptron multicamadas. In: *Mecanica Computacional Vol. XXIX*. AMCA - Associação Argentina de Mecânica Computacional, 2010. p. 9169–9175. Disponível em: <<http://www.cimec.org.ar/ojs/index.php/mc/article/viewFile/3659/3572>>. Acesso em: 21 ago. 2014. Citado na página 36.
- ELLWANGER, U. *From the Universe to the Elementary Particles: A First Introduction to Cosmology and the Fundamental Interactions*. 1. ed. [S.l.]: Springer, 2012. (Undergraduate Lecture Notes in Physics). ISBN 3642243746,9783642243745. Citado 3 vezes nas páginas 12, 14 e 17.
- EVANS, L.; BRYANT, P. LHC machine. *Journal Of Instrumentation*, v. 3, n. Ss08001, p. 1–158, 2008. Citado 2 vezes nas páginas 1 e 15.
- FACELLI, K. et al. *Inteligência Artificial uma Abordagem de Aprendizado de Máquina*. Rio de Janeiro: LTC, 2011. v. 1. Citado na página 47.
- FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861 – 874, 2006. ISSN 0167-8655. ROC Analysis in Pattern Recognition. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016786550500303X>>. Citado na página 49.
- Fernandes Jr., A. C. L.; DAMIANI, F. Extração de tempo musical utilizando transformada wavelet e rede neural artificial. In: *12º Congresso de Engenharia de Audio da AES Brasil*. AES - Audio Engineering Society, 2014. p. 9–16. Disponível em: <http://aesbrasil.org/wp-content/uploads/2014/05/Anais_AESBR2014.pdf>. Acesso em: 21 ago. 2014. Citado na página 37.
- FERREIRA, J. E. V. et al. Graphical representation of chemical periodicity of main elements through boxplot. *Educación Química*, v. 27, n. 3, p. 209 – 216, 2016. ISSN 0187-893X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0187893X16300106>>. Citado 2 vezes nas páginas 50 e 51.
- FREUND, W. S. Trabalho de Conclusão de Curso, *Algoritmo Neural Para Detecção de Partículas em um Calorímetro de Altas Energias*. 2011. Disponível em: <<http://monografias.poli.ufrj.br/monografias/monopolis10003987.pdf>>. Citado 6 vezes nas páginas 2, 5, 15, 17, 24 e 36.
- FREUND, W. S. *Identificação de Elétrons Baseada em um Calorímetro de Altas Energias Finamente Segmentado*. 310 p. Tese de Doutorado — UFRJ, 2018. Disponível em: <<http://www.pee.ufrj.br/teses/textocompleto/2010121601.pdf>>. Citado 5 vezes nas páginas 17, 18, 21, 22 e 26.
- FREUND, W. S.; PINTO, J. V. da F.; SEIXAS, J. M. Redes neurais aplicadas na identificação de partículas em um detector finamente segmentado. In: *Anais do XXI Congresso Brasileiro de Automática*. [S.l.]: CBA - Congresso Brasileiro de Automática, 2016. p. 1072 – 1079. Acesso em: 3 a 7 out. 2016. Citado 4 vezes nas páginas 4, 5, 19 e 37.
- GALSTER, G. *ATLAS Trigger: Preparations for Run II*. 172 p. Dissertação (Dissertação de Mestrado) — University of Copenhagen, 2015. Citado na página 26.
- GANJEFAR, S.; GHASEMI, A. A. A novel-strategy controller design for maximum power extraction in stand-alone windmill systems. *Energy*, v. 76, p. 326 – 335, 2014. ISSN 0360-5442. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0360544214009608>>. Citado na página 44.
- GROSSMANN, C. *Modelo Padrão da Física: o que é e de onde veio?* 2013. Disponível em: <<https://hypescience.com/o-que-e-e-de-onde-veio-o-modelo-padroao-da-fisica/>>. Acesso em: 21 set. 2017. Citado na página 13.
- HAN, T.; HEWETT, J. L. Top-charm associated production in high-energy e^+e^- collisions. *Phys. Rev. D*, American Physical Society, v. 60, p. 074015, Sep 1999. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevD.60.074015>>. Citado na página 5.
- HAYKIN, S. *Sinais e Sistemas*. Porto Alegre: Bookman, 2001. Tradutor: José Carlos Baarbosa dos Santos. Citado na página 5.
- HAYKIN, S. *Neural Networks and Learning Machines*. 3rd ed. ed. Nova Jersey, Estados Unidos: Prentice Hall, 2008. Citado 6 vezes nas páginas 25, 29, 30, 31, 32 e 34.

- HEUER, R. D. General meeting january 2013. 2013. Disponível em: <<https://indico.cern.ch/event/219327/attachments/355208/494765/address-Jan2013.pdf>>. Acesso em: 21 fev. 2018. Citado na página 3.
- HORATA, P.; CHIEWCHANWATTANA, S.; SUNAT, K. Robust extreme learning machine. *Neurocomputing*, v. 102, p. 31 – 44, 2013. ISSN 0925-2312. Advances in Extreme Learning Machines (ELM 2011). Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925231212004171>>. Citado na página 40.
- HUANG, G. et al. Trends in extreme learning machines: A review. *Neural Networks*, v. 61, p. 32 – 48, 2015. ISSN 0893-6080. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0893608014002214>>. Citado 2 vezes nas páginas 6 e 40.
- HUANG, G.-B. What are extreme learning machines? filling the gap between frank rosenblatt's dream and john von neumann's puzzle. *Cognitive Computation*, v. 7, n. 3, p. 263–278, 2015. ISSN 1866-9964. Disponível em: <<http://dx.doi.org/10.1007/s12559-015-9333-0>>. Citado 5 vezes nas páginas 6, 7, 37, 38 e 89.
- HUANG, G.-B.; WANG, D. H.; LAN, Y. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, v. 2, n. 2, p. 107–122, 2011. ISSN 1868-808X. Disponível em: <<http://dx.doi.org/10.1007/s13042-011-0019-y>>. Citado 4 vezes nas páginas 6, 37, 38 e 89.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: a new learning scheme of feed-forward neural networks. In: *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*. [S.l.: s.n.], 2004. v. 2, p. 985–990 vol.2. ISSN 1098-7576. Citado 2 vezes nas páginas 37 e 89.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing*, v. 70, n. 1, p. 489 – 501, 2006. ISSN 0925-2312. Papers From The 7th Brazilian Symposium On Neural Networks (SBRN 04). Disponível em: <<http://www.sciencedirect.com/science/article/pii/s0925231206000385>>. Citado 5 vezes nas páginas 6, 7, 37, 38 e 89.
- INAMURA, T. T. Nagaoka's atomic model and hyperfine interactions.(report). The Japan Academy, v. 92B, n. NO. 4, April 2016. ISSN 0021-4280. Citado na página 9.
- JAEGER, H. The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, v. 148, n. 34, p. 13, 2001. Citado 4 vezes nas páginas 7, 41, 43 e 44.
- JAEGER, H. *A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the “echo state network” approach*. [S.l.]: GMD-Forschungszentrum Informationstechnik Bonn, 2002. v. 5. 46 p. Citado na página 7.
- JAEGER, H. Adaptive nonlinear system identification with echo state networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2003. p. 609–616. Citado na página 6.
- JAEGER, H. The “echo state” approach to analysing and training recurrent neural networks - with an Erratum note. [S.l.], 2010. Disponível em: <<https://pdfs.semanticscholar.org/8430/c0b9afa478ae660398704b11dca1221ccf22.pdf>>. Citado 2 vezes nas páginas 6 e 43.
- JAEGER, H.; HAAS, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, American Association for the Advancement of Science, v. 304, n. 5667, p. 78–80, 2004. ISSN 0036-8075. Disponível em: <<http://science.scienmag.org/content/304/5667/78>>. Citado na página 6.
- JONES, M. T. *Recurrent neural networks deep dive*. IBM, 2017. Disponível em: <<https://www.ibm.com/developerworks/library/cc-cognitive-recurrent-neural-networks/index.html>>. Acesso em: Acesso em 17/02/2018. Citado 2 vezes nas páginas 31 e 32.
- KANE, G. The dawn of physics beyond the standard model. *Scientific American*, v. 288, n. 6, p. 68, 2003. ISSN 00368733. Disponível em: <<http://search-ebscohost-com.ez10.periodicos.capes.gov.br/login.aspx?direct=true&db=aph&AN=9725646&lang=pt-br&zsite=ehost-live>>. Citado na página 12.
- KILBY, C.; ATLAS Collaboration. The design and performance of the ATLAS inner detector trigger for run 2 lhc collisions at 13 tev. *Journal of Physics: Conference Series*, v. 762, n. 1, p. 012029, 2016. Disponível em: <<http://stacks.iop.org/1742-6596/762/i=1/a=012029>>. Citado na página 26.

- KIM, T. K. T test as a parametric statistic. *Korean journal of anesthesiology*, v. 68, n. 6, p. 540–546, 2015. Citado na página 52.
- KUNCHEVA, L. I. *Combining Pattern Classifiers, Methods and Algorithms*. Burlington, MA: Wiley, 2004. ISBN 0-471-21078-1. Citado 2 vezes nas páginas 51 e 52.
- LENZI, T. *Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors*. Dissertação (Mestrado) — U. Brussels (main), 2013. Disponível em: <<http://inspirehep.net/record/1236817/files/arXiv:1306.0858.pdf>>. Citado na página 20.
- LUZ, I. A. dos S. et al. Sistema para redução da sobreposição de sinais em calorimetria de altas energias em operação com alta taxa de eventos. In: *Anais do XXI Congresso Brasileiro de Automática*. [S.l.]: CBA - Congresso Brasileiro de Automática, 2016. Acesso em: 3 a 7 out. 2016. Citado na página 4.
- MACAUSLAND, R. *Moore-Penrose Inver and Least Squares*. University of Puget Sound, 2014. Disponível em: <<http://buzzard.ups.edu/courses/2014spring/420projects/math420-UPS-spring-2014-macausland-pseudo-inverse.pdf>>. Acesso em: Acesso e 22/06/2018. Citado na página 39.
- MARSHALL, Z.; ATLAS Collaboration the. Simulation of pile-up in the ATLAS experiment. *Journal of Physics: Conference Series*, v. 513, n. 2, p. 022024, 2014. Disponível em: <<http://stacks.iop.org/1742-6596/513/i=2/a=022024>>. Citado na página 3.
- MARTÍNEZ, A. R.; ATLAS Collaboration. The run-2 ATLAS trigger system. *Journal of Physics: Conference Series*, v. 762, n. 1, p. 012003, 2016. Disponível em: <<http://stacks.iop.org/1742-6596/762/i=1/a=012003>>. Citado na página 26.
- MELZER, E. E. M.; AIRES, J. A. A história do desenvolvimento da teoria atômica: um percurso de dalton a bohr. *Amazônia: Revista de Educação em Ciências e Matemáticas*, v. 11, n. 22, p. 62–77, 2015. Citado na página 10.
- MOREIRA, M. A. O modelo padrão da física de partículas. *Revista Brasileira de Ensino de Física*, v. 31, n. 1, p. 1–158, 2009. Disponível em: <www.sbfisica.org.br>. Citado 4 vezes nas páginas 1, 13, 17 e 27.
- NATH, P. et al. The hunt for new physics at the Large Hadron Collider. *Nuclear Physics B - Proceedings Supplements*, v. 200, p. 185 – 417, 2010. ISSN 0920-5632. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0920563210001490>>. Citado na página 17.
- NÓBREGA, F. K.; MACKEDANZ, L. F. O LHC (Large Hadron Collider) e a nossa física de cada dia. *Revista Brasileira de Ensino de Física*, scielo, v. 35, p. 1 – 11, 03 2013. ISSN 1806-1117. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-11172013000100001&nrm=iso>. Citado na página 23.
- OLIVEIRA, A. G. de; BUTKEWITSCH, S. Redes neurais artificiais aplicadas à análise da variabilidade de processos produtivos. In: *CONEM - Congresso de Engenharia Mecânica - 2000*. ABCM - Associação Brasileira de Engenharia e Ciências Mecânicas, 2000. Disponível em: <<http://www.abcm.org.br/pt/wp-content/anais/conem/2000/OC8760.pdf>>. Acesso em: 22 dez. 2014. Citado na página 34.
- OLIVEIRA Ótom Anselmo de; FERNANDES, J. D. G. *Arquitetura atômica e molecular*. Natal: EDUFRN, 2006. 280 p. Organizador. ISBN 85-7273-278-0. Citado na página 10.
- PERALVA, B. S.-M.; SEIXAS, J. M. de. *The TileCal Energy Reconstruction for LHC Run2 and Future Perspectives*. [S.l.], 2015. Comments: 5 pages, 7 figures, LISHEP 2015, 2-9 August 2015, Manaus. Disponível em: <<https://cds.cern.ch/record/2057907>>. Citado na página 4.
- PIMENTA, J. J. M. et al. O bóson de higgs. *Revista Brasileira de Ensino de Física*, scielo, v. 35, p. 1 – 14, 06 2013. ISSN 1806-1117. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-11172013000200006&nrm=iso>. Citado 4 vezes nas páginas 1, 12, 14 e 27.
- PINHEIRO, L. A.; COSTA, S. S. C. da; MOREIRA, M. A. *Do átomo grego ao Modelo Padrão: os indivisíveis de hoje*. 2011. Textos de Apoio ao Professor de Física. Citado 2 vezes nas páginas 9 e 10.
- Portal Action. *Boxplot*. 2016. Disponível em: <<http://www.portalaction.com.br/estatistica-basica/31-boxplot>>. Acesso em: 19 set. 2016. Citado na página 50.

- PRATER, A. A. Comparison of echo state network output layer classification methods on noisy data. *ArXiv e-prints*, mar 2017. Citado na página 45.
- PRIBERAM. “átomo” - em dicionário Priberam da Língua Portuguesa. 2016. Disponível em: <[DisponívelEm:http://www.priberam.pt/DLPO/%C3%A1tomo](http://www.priberam.pt/DLPO/%C3%A1tomo)>. Acesso em: Acesso Em: 02 de Nov De 2016. Citado na página 9.
- QU, B. et al. Two-hidden-layer extreme learning machine for regression and classification. *Neurocomputing*, v. 175, Part A, p. 826 – 834, 2016. ISSN 0925-2312. Citado na página 40.
- REIS, R. L. dos. *Modelos autorregressivos periódicos para previsão e geração de séries de vazões médias mensais*. 108 p. Tese de Doutorado — UPS, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-13052013-111226/pt-br.php>>. Citado na página 43.
- REYES, J.; VELLASCO, M.; TANSCHEIT, R. Monitoramento e diagnóstico de múltiplos sensores por redes neurais auto-associativas. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, Scielo, v. 23, p. 121 – 137, 04 2012. ISSN 0103-1759. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-17592012000200001&nrm=iso>. Citado na página 34.
- ROCHA, J. F. M. *Origens e evolução das ideias da física*. Salvador: EDUFBA, 2002. 299 – 359 p. Organizador. Citado 2 vezes nas páginas 9 e 10.
- RUCKSTUHL, A. Robust fitting of parametric models based on M-Estimation. *Lecture notes.*, 2014. Disponível em: <<https://www.ethz.ch/content/dam/ethz/special-interest/math/statistics/sfs/Education/Advanced%20Studies%20in%20Applied%20Statistics/course-material/robust-nonlinear/robstat16E.pdf>>. Citado na página 40.
- SANTOS, M. S. dos. Trabalho de Conclusão de Curso, *Reconhecimento de Acordes Naturais de Guitarra com Redes Neurais Artificiais*. 2014. Citado na página 36.
- SCHAETTI, N.; SALOMON, M.; COUTURIER, R. Echo state networks-based reservoir computing for mnist handwritten digits recognition. In: *2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES)*. [S.l.: s.n.], 2016. p. 484–491. Citado na página 44.
- SEIXAS, J. et al. Neural second-level trigger system based on calorimetry. *Computer Physics Communications*, v. 95, n. 2, p. 143 – 157, 1996. ISSN 0010-4655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0010465596000124>>. Citado na página 19.
- SIMAS FILHO, E. F.; SEIXAS, J. M.; CALÔBA, L. P. Modified post-nonlinear ICA model for online neural discrimination. *Neurocomputing*, v. 73, n. 16–18, p. 2820 – 2828, 2010. ISSN 0925-2312. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925231210002511>>. Citado na página 72.
- SIMAS FILHO, E. F. de. *Análise Não-Linear de Componentes Independentes para uma Filtragem Online Baseada em Calorimetria de Alta Energia e com Fina Segmentação*. 310 p. Tese de Doutorado — UFRJ, 2010. Disponível em: <<http://www.pee.ufrj.br/teses/textocompleto/2010121601.pdf>>. Citado 3 vezes nas páginas 19, 20 e 24.
- SIMEÓN, E. J. A. *PROGNÓSTICO DE FALHAS BASEADO EM REDES NEURAIS COM ESTADOS DE ECO*. 125 p. Tese de Doutorado — UnB, 2015. Citado 5 vezes nas páginas 7, 41, 42, 43 e 44.
- SIQUEIRA, H. V. *Máquinas Desorganizadas para Previsão de Séries de Vazões*. 218 p. Tese de Doutorado — Unicamp, 2013. Citado na página 43.
- SOARES, F. A. A. et al. Recursive diameter prediction and volume calculation of eucalyptus trees using multilayer perceptron networks. *Computers and Electronics in Agriculture*, v. 78, n. 1, p. 19 – 27, 2011. ISSN 0168-1699. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0168169911001232>>. Citado na página 36.
- SOUTO, G. Trabalho de Conclusão de Curso, *Decomposição em Valores Singulares*. 2000. Citado na página 38.

- SOUZA, E. E. P. de. *Pré - Processamento Eficiente para um Sistema online de Classificação Baseado em Redes Neurais Artificiais*. Dissertação (Mestrado) — UFBA, Salvador, 2015. Disponível em: <<http://www.pgee.eng.ufba.br/teses/2563ead70510e2565a35c28e9b4377fd.pdf>>. Citado 4 vezes nas páginas 20, 48, 50 e 53.
- SOUZA, E. E. P. de et al. Sistema neural para seleção online de eventos com pré-processamento através de DWT. In: *Anais Do XX Congresso Brasileiro de Automática*. [S.l.]: CBA - Congresso Brasileiro De Automática, 2014. p. 1072 – 1079. Acesso em: 20 a 24 Set. 2014. Citado na página 37.
- SOUZA, E. E. P. de et al. Sistema neural segmentado de detecção online de elétrons utilizando pré-processamento estatístico. In: *Anais do XX Congresso Brasileiro de Automática*. CBA - Congresso Brasileiro de Automática, 2014. p. 2493 – 2499. Disponível em: <<http://www.swge.inf.br/cba2014/anais/PDF/1569935545.pdf>>. Acesso em: 20 A 24 Set. 2014. Citado na página 37.
- TERMENON, M. et al. Brain MRI morphological patterns extraction tool based on extreme learning machine and majority vote classification. *Neurocomputing*, v. 174, Part A, p. 344 – 351, 2016. ISSN 0925-2312. Disponível em: <<http://www.sciencedirect.com/science/article/pii/s0925231215011388>>. Citado 2 vezes nas páginas 7 e 40.
- TORRES, R. C. et al. Neural online filtering based on preprocessed calorimeter data. In: *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. [S.l.: s.n.], 2009. p. 530–536. ISSN 1082-3654. Citado na página 50.
- TRENTIN, E.; SCHERER, S.; SCHWENKER, F. Emotion recognition from speech signals via a probabilistic echo-state network. *Pattern Recognition Letters*, v. 66, p. 4 – 12, 2015. ISSN 0167-8655. Pattern Recognition in Human Computer Interaction. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865514003328>>. Citado na página 44.
- VAZQUEZ, W. P. The ATLAS data acquisition system: from run 1 to run 2. *Nuclear and Particle Physics Proceedings*, v. 273-275, p. 939 – 944, 2016. ISSN 2405-6014. 37th International Conference on High Energy Physics (ICHEP). Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2405601415006355>>. Citado na página 26.
- WANG, D.; HUANG, G.-B. Protein sequence classification using extreme learning machine. In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*. [S.l.: s.n.], 2005. v. 3, p. 1406–1411 vol. 3. ISSN 2161-4393. Citado 2 vezes nas páginas 38 e 40.
- WANG, L.; WANG, Z.; LIU, S. An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm. *Expert Systems with Applications*, v. 43, p. 237 – 249, 2016. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417415006120>>. Citado na página 45.
- WEN, G.; LI, H.; LI, D. An ensemble convolutional echo state networks for facial expression recognition. In: *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. [S.l.: s.n.], 2015. p. 873–878. Citado na página 44.
- WIGMANS, R. Calorimetry. In: . Texas: Scientifica Acta, 2008. v. 2, p. 18–55. Disponível em: <http://siba.unipv.it/fisica/ScientificaActa/Volume_2_1/Wigmans.pdf>. Citado na página 4.
- XU, L. et al. Stochastic cross validation. *Chemometrics and Intelligent Laboratory Systems*, v. 175, p. 74 – 81, 2018. ISSN 0169-7439. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169743917306482>>. Citado na página 47.
- YORIYAZ, H. Método de monte carlo: princípios e aplicações em física médica. *Revista Brasileira de Física Médica*, v. 3, n. 1, 2009. Disponível em: <<https://www.ipen.br/biblioteca/2008/14528.pdf>>. Citado 2 vezes nas páginas 49 e 53.
- ZHANG, K.; LUO, M. Outlier-robust extreme learning machine for regression problems. *Neurocomputing*, v. 151, Part 3, p. 1519 – 1527, 2015. ISSN 0925-2312. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0925231214012053>>. Citado na página 40.

Apêndices

APÊNDICE A – Trabalhos Publicados

A.1 Artigos Publicados em Anais de Congressos e Simpósios

1. SANTOS, M. S.; SIMAS FILHO, E. F.; FARIAS, P. C. A. M; SEIXAS, J. M. Máquinas de aprendizado extremo para classificação online de eventos no detector ATLAS. In: *XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2017) (SBrT 2017)*. São Pedro, Brazil: [s.n.], 2017. p. 413–417

- **Resumo**

O ATLAS é um dos detectores do LHC (*Large Hadron Collider*), e está localizado no CERN (Organização Européia para a pesquisa Nuclear). Para adequada caracterização das partículas é preciso realizar uma precisa medição do perfil de deposição de energia à medida que ocorrem interações com o detector. No ATLAS os calorímetros são responsáveis por realizar a estimativa da energia das partículas e, neste sentido, utilizam mais de 100.000 sensores. Um dos discriminadores para a detecção *online* de elétrons utilizados no ATLAS é o *Neural Ringer*, no qual o perfil de deposição de energia é utilizado como entrada para um classificador neural tipo *perceptron* de múltiplas camadas. Este trabalho propõe o uso de Máquinas de Aprendizado Extremo (ELM) em substituição às redes do tipo *perceptron multilayer* no *Neural Ringer*. Os resultados obtidos de uma base de dados simulados apontam para uma significativa redução do tempo de treinamento, com desempenho de classificação semelhante.

2. SANTOS, M. S dos; SIMAS FLHO E. F de; FARIAS, P. C. A. M; SEIXAS, J. M. Máquinas de Aprendizado Extremo e Redes com Estados de Eco para Classificação *Online* de Eventos no detector ATLAS. In: *Anais Do XXII Congresso Brasileiro de Automática*. 09 a 12 de setembro. João Pessoa, Brasil[S.l.]: CBA - Congresso Brasileiro De Automática, 2018.

- **Resumo**

O ATLAS é um dos detectores do acelerador de partículas LHC e com sua estrutura cilíndrica que compreende diversas camadas de sensores é capaz de caracterizar os fenômenos de interesse que ocorrem após as colisões dos feixes de partículas. O sistema de medição de energia (calorímetro) do ATLAS é composto por mais de 100.000 sensores e fornece informações importantes para a seleção *online* dos eventos de interesse. Neste contexto, o *Neural Ringer* é um discriminador de partículas eletromagnéticas (elétrons e fótons) que opera no sistema *online* de filtragem (*trigger*) do ATLAS e utiliza uma rede neural tipo Perceptron de múltiplas camadas (MLP - *Multi-layer Perceptron*) para realizar a classificação das partículas a partir do perfil de deposição de energia medido nos calorímetros e formatado em anéis. Neste trabalho é proposta a substituição dos classificadores MLP do *Neural Ringer* por máquinas de aprendizado com reduzido custo computacional de

treinamento. São utilizadas máquinas de aprendizado extremo (ELM - *Extreme Learning Machines*) e redes com estados de eco (ESN - *Echo State Networks*) e resultados apontam que é possível obter eficiência de classificação semelhante ao sistema original com uma considerável redução do tempo de treinamento.

A.2 Resumo Publicado em Encontro

1. SANTOS, M. S.; SOUZA, E. E. P.; SIMAS FILHO, E. F.; FARIAS, P. C. A. M; SEIXAS, J. M; ANDRADE FILHO, L. M. Uso de Algoritmos de Treinamento Rápido para o Discriminador *Neural Ringer* no Detector ATLAS. In: *XXXIX Encontro Nacional de Física de Partículas e Campos (SBF - Sociedade Brasileira de Física 2018) (SBF 2018)*. 24 a 28 de setembro, Campos do Jordão, Brazil: [s.n.], 2018.

• Resumo

O Neural Ringer é um dos algoritmos atualmente utilizados para identificação de elétrons no segundo nível de filtragem online do detector ATLAS. Para prover a decisão de aceitação ou rejeição dos eventos, o Neural Ringer realiza um ordenamento topológico em forma de anéis concêntricos do perfil de deposição de energia medido nos calorímetros. Neste discriminador, uma rede neural tipo *perceptron* de múltiplas camadas é utilizada para classificação. Neste trabalho é proposta a utilização de outros modelos de rede neural de treinamento rápido para realizar a etapa de classificação no Neural Ringer. Foram testados a Máquina de Aprendizado Extremo (ELM - *Extreme Learning Machine*) e a Rede de Estado de Eco (ESN - *Echo State Network*). Utilizando dados simulados foi possível observar que os modelos de treinamento propostos obtiveram resultados de eficiência de classificação semelhantes à versão tradicional do Neural Ringer, porém num tempo de treinamento consideravelmente reduzido.

APÊNDICE B – Análise de Desempenho da ELM

B.1 Sensibilidade dos pesos à distribuição de probabilidade utilizada.

As redes ELM propostas inicialmente por [Huang, Zhu e Siew \(2004\)](#) são redes que não possuem em seu algoritmo de treino uma etapa de retropropagação do erro. Ou seja, não possui realimentação baseada no erro cometido pelo processo de treino. Sua base é a determinação da matriz \mathbf{H} , que representa os pesos da camada oculta da rede, expressa na [Equação B.1](#) em sua forma matricial.

$$\mathbf{y}_j = \sum_{i=1}^N \beta_i \Phi(\mathbf{w}_i \mathbf{x}_j + b_i), \quad j \in [1, M] \quad (\text{B.1})$$

A equação B.1 pode ser reescrita como $\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}$, sendo,

$$\mathbf{H} = \begin{pmatrix} \Phi(\mathbf{w}_1 \mathbf{x}_1 + b_1) & \dots & \Phi(\mathbf{w}_N \mathbf{x}_1 + b_N) \\ \vdots & \ddots & \vdots \\ \Phi(\mathbf{w}_1 \mathbf{x}_M + b_1) & \dots & \Phi(\mathbf{w}_N \mathbf{x}_M + b_N) \end{pmatrix}, \quad (\text{B.2})$$

e $\boldsymbol{\beta} = (\beta^T \dots \beta_N^T)^T$ e $\mathbf{Y} = (y_1^T \dots y_M^T)^T$.

Nos trabalhos de [Huang, Zhu e Siew \(2006\)](#), [Huang, Wang e Lan \(2011\)](#) e [Huang \(2015\)](#) são exibidos os teoremas que dão suporte e fundamentação matemática à técnica. Alguns teoremas apresentados e demonstrados são: sua capacidade de aproximador universal, capacidade de aprendizagem e convergência. E uma característica interessante é a forma que os pesos da camada oculta são gerados, os quais são gerados por uma função que produza números pseudo-aleatórios, e a função de ativação seja diferenciável continuamente, para que seja possível determinar os valores da matriz \mathbf{H} .

Neste trabalho foi feita uma avaliação da influência da característica dos número pseudo aleatórios utilizados na camada oculta. Pois, existem diferentes tipos de funções de distribuição de probabilidade utilizadas para produção de números pseudo-aleatórios. Três formas foram avaliadas: Número gerados com distribuição normal (N1), Números com distribuição uniforme (N2) e distribuição uniforme de números inteiros pseudo-aleatórios normalizados (N3) pelo maior módulo.

Os resultados foram obtidos utilizando uma das bases de dados disponíveis, a qual é segmentada em 16 regiões, ver [Tabela 18](#), (E_T , $|\eta|$) e são apresentados na [Tabela 19](#). Primeiro as redes foram treinadas e variando-se o número de neurônios na camada oculta até 100. em seguida, o número de neurônios que apresentou o maior índice SP dentro desse intervalo, foi utilizado para o teste de sensibilidade. Observa-se que os resultados obtidos com os número produzidos

com distribuição uniforme (N2) resultaram nos menores índices SP em todas as regiões da base de teste.

Tabela 18 – Segmentação base de dados utilizada.

Intervalos				
E_T [GeV]	[20; 30]	[30; 40]	[40; 50]	[50; 20.000]
$ \eta $	[0, 00; 0, 80]	[0, 80; 1, 37]	[1, 37; 1, 54]	[1, 54; 2, 5]

Tabela 19 – Índices SP para três métodos de produzir números pseudo-aleatórios para a ELM.

Índices SP para cada região da base de teste				
	(0,0)	(0,1)	(0,2)	(0,3)
N1	96,366 ± 0,590	94,911 ± 0,963	94,959 ± 2,262	91,918 ± 0,566
N2	94,003 ± 1,190	92,370 ± 0,989	93,376 ± 2,587	90,124 ± 1,111
N3	96,355 ± 0,528	95,169 ± 0,699	94,328 ± 2,378	92,020 ± 0,461
	(1,0)	(1,1)	(1,2)	(1,3)
N1	97,398 ± 1,024	95,278 ± 0,940	90,873 ± 3,201	92,197 ± 1,186
N2	95,708 ± 2,661	92,468 ± 1,814	89,789 ± 2,855	88,749 ± 2,605
N3	97,568 ± 0,899	95,681 ± 0,820	92,148 ± 3,455	91,546 ± 1,165
	(2,0)	(2,1)	(2,2)	(2,3)
N1	97,693 ± 1,058	96,734 ± 1,495	99,147 ± 4,761	96,348 ± 1,647
N2	95,938 ± 1,470	95,892 ± 1,877	96,655 ± 3,990	94,172 ± 3,678
N3	97,782 ± 0,869	96,230 ± 1,167	99,716 ± 3,805	95,151 ± 1,879
	(3,0)	(3,1)	(3,2)	(3,3)
N1	99,209 ± 0,256	98,541 ± 0,564	99,569 ± 1,993	98,081 ± 0,489
N2	98,757 ± 0,842	97,501 ± 0,775	98,270 ± 1,867	96,749 ± 1,674
N3	99,271 ± 0,204	98,652 ± 0,578	99,482 ± 2,025	98,140 ± 0,464

As menores diferenças alcançadas foram de 0,45%, enquanto que as maiores foram de 3,44% na comparação entre os métodos N2 e N1. Na comparação entre o método N3 e N1, a menor diferença foi de 0,34% e a maior diferença foi de 3,21%, em favor do método N3. Outro dado possível de observar, é a incerteza alcançada pelos métodos. Em todas as regiões o método N2 produziu resultados com incerteza superior aos métodos N1 e N3.

Os resultados obtidos indicam que apesar de a técnica ELM ser flexível quanto ao método de produção dos números pseudo-aleatórios para a camada de entrada, ela possui sensibilidade, quanto às características da distribuição utilizada. Em problemas com grande volume de dados a ser processado essa variação pode ser significativa e interferir nos resultados reduzindo o desempenho do classificador.