

Transferência de Conhecimento para Filtragem Online de Partículas Baseada em Calorimetria de Altas Energias

Rafael Rodrigues Vianna
Laboratório de Processamento de
Sinais (LPS) - COPPE/POLI
UFRJ
Rio de Janeiro, Brasil
rafael.rv@lps.ufrj.br

Juan Lieber Marin
Laboratório de Sistemas Digitais
PPGEE/UFBA
Salvador, Bahia
juan.lieber@ufba.br

José Manoel de Seixas
Laboratório de Processamento de
Sinais (LPS) - COPPE/POLI
UFRJ
Rio de Janeiro, Brasil
seixas@lps.ufrj.br

Resumo—No contexto do sistema de filtragem online de partículas de um dos principais experimentos do LHC, maior acelerador de partículas em operação no mundo, a demanda por um número maior de eventos raros e instáveis detectáveis exige um aumento da banda passante no sistema de seleção *online* de partículas, garantindo, assim, a seleção de bons candidatos para a análise *offline* posterior. Esse processamento *online*, ou seja, simultâneo às colisões, é fundamental para reduzir o enorme volume de dados gerados pelos sistemas de detecção (aproximadamente 70Tb/s), descartando os eventos de Física ordinária e filtrando aqueles de interesse das análises. Nesse sentido, este trabalho propõe uma solução para o experimento ATLAS baseada em conhecimento de máquina para a seleção *online* de fótons, partículas comumente formadas no processo de decaimento de muitos daqueles eventos de interesse. Os classificadores neurais são gerados a partir de modelos já existentes e pré-treinados com elétrons - cuja interação com a matéria se dá de maneira similar a dos fótons (eletromagnética) – e retreinados com dados de fótons, estratégia conhecida como transferência de aprendizagem. Os resultados obtidos mostram melhora do desempenho frente ao método tradicional de seleção empregado.

Palavras-chave—ATLAS, Trigger, Neural Ringer, Redes Neurais, Transferência de Aprendizagem.

I. INTRODUÇÃO

O Grande Colisor de Hádrons (LHC) é, hoje, o maior acelerador de partículas da Organização Europeia para Pesquisas Nucleares (CERN), em Genebra, Suíça [1]. Com seus quase 27 km de circunferência, sua concepção enquanto projeto representou uma grande conquista para a humanidade, posto que visa saciar o desejo ancestral de entender a composição e funcionamento do universo e da matéria que o compõe.

O ATLAS (do inglês, A Toroidal LHC ApparatuS) é um dos principais experimentos do acelerador [2][3], que realiza colisão de feixes prótonicos a cada 25. No LHC, busca-se recriar as condições da origem do universo, observando as transformações da matéria nuclear para melhor entender sua estrutura e, assim, validar modelos da Física de Altas Energias. Inicialmente, o LHC foi projetado para operar com 25 colisões por pacote, operando a 90 colisões no final da segunda temporada de coleta de dados (*Run 2*) e, para a terceira tomada de dados (*Run 3*), deve haver ainda mais. Operar nesse regime de altas taxas de colisão (luminosidade) é importante para acumular estatística suficiente para observação de eventos

considerados raros, como a primeira vez que fora observada a existência do Bóson de Higgs [4]. Há, no entanto, um contraponto: ao atuar com alta luminosidade, aumenta-se o número de interações próton-próton a cada cruzamento dos pacotes, isso implica necessariamente em um acréscimo do número de partículas interagindo com o detector e que podem gerar sobreposição do sinal caso incidam em regiões próximas, fenômeno conhecido como empilhamento de sinal (do inglês, *pile-up*) [3]. Esse fenômeno resulta em uma série de desafios e exige maior desempenho da reconstrução em diversos estágios da análise.

Dentre os sistemas de detectores de partículas que compõem o ATLAS, o detector de traços (ID), o sistema de calorimetria e o espectrômetro de múons são os principais. O primeiro possibilita a identificação da trajetória (traço) e momento linear de partículas carregadas, enquanto o terceiro é responsável exclusivamente pela detecção de múons. O sistema de calorimetria consiste em calorímetros segmentados de amostragem que envolvem praticamente todo o experimento e possibilitam medições de energia através da absorção das partículas que por eles atravessam, principalmente de elétrons, fótons e hádrons [2][3]. Esse sistema é dividido em duas seções: a primeira e mais interna, o calorímetro eletromagnético (ECAL), e a segunda e mais externa, o calorímetro hadrônico (HCAL). O ECAL possui quatro camadas, cada qual com granularidade, número de células e segmentação característicos. Já o HCAL, por sua vez, possui um total de três camadas. A Fig. 1 abaixo mostra uma visão do ATLAS em corte, exibindo seus sistemas de detecção.

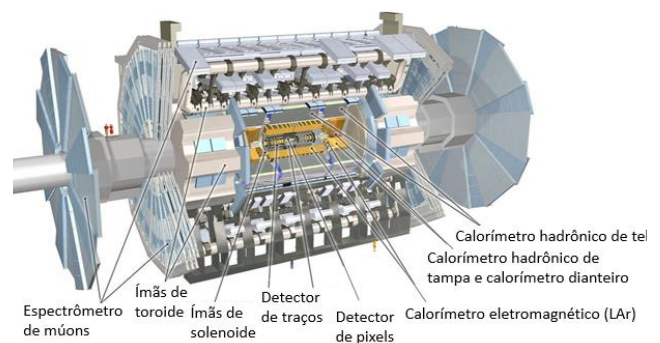


Fig. 1: ATLAS em corte. Destacam-se o sistema de calorimetria formado pelo calorímetro eletromagnético, em verde, e pelo hadrônico, em amarelo, além da formação dos chuveiros. (Fonte: adaptado do banco de figuras públicas do CERN/ATLAS).

No contexto das análises das colisões, é interessante adotar um sistema de coordenadas que acompanha a simetria do experimento e a direção de propagação das partículas oriundas da colisão [5][3]. Dessa forma, utilizam-se os eixos ϕ e z do sistema de coordenadas cilíndricas com origem no ponto de colisão, com cálculo da variável η , chamada de pseudorapidez, a qual representa a direção da projeção das partículas após a colisão. Esta última é importante porque contribui na formação das regiões do espaço de fase, que dividirá a operação do sistema no treinamento e operacionalização dos modelos classificadores.

Sistema de Coordenadas do ATLAS

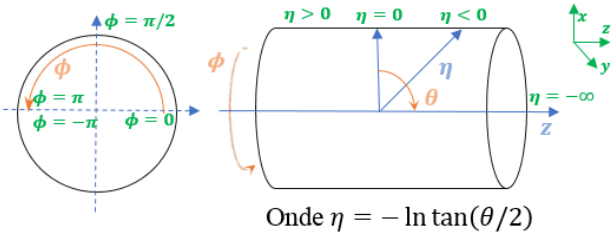


Fig. 2: Representação do sistema de coordenadas adotado pelo ATLAS.

Elétrons, fótons e hádrons de altas energias provenientes das colisões geradas pelo LHC interagem com o meio material que forma o calorímetro, depositando energia sobre as células que o compõem, o que resulta tanto em um espalhamento longitudinal de energia, quanto transversal, uma ação progressiva a qual prolonga-se até a dissipação total da partícula, processo conhecido como chuva de partículas [6][7]. Tal chuva pode se desenvolver de duas formas: eletromagnética (EM) e hadrônica. No caso da partícula incidente ser um fóton ou elétron, seu desenvolvimento fica, praticamente, contido no ECAL. Nesse sentido, elétrons e fótons têm padrões de deposição de energia muito similares, diferindo, em termos de assinatura, apenas pelo fato de elétrons também serem percebidos pelo Detector de Traços, ao contrário dos fótons [8].

Por outro lado, feixes hadrônicos são formados por partículas que interagem com os núcleos dos átomos, produzindo chuvas mais largas e profundas [6][9]. Cerca de 67% da cascata hadrônica é composta por outras partículas também hadrônicas (HAD), e acabam atingindo também o HCAL, que é posterior ao eletromagnético. Entretanto, os 33% restantes interagem eletromagneticamente [6][9], podendo, portanto, produzir chuvas mais curtas e estreitas. Justamente essa última característica da cascata hadrônica faz com que esses feixes sejam a principal forma de ruído na detecção de fótons e elétrons: uma vez que interagem com o calorímetro na mesma região onde elétrons e fótons também interagem, falseando suas assinaturas.

Quando se faz o estudo da interação de partículas com o calorímetro [7], vê-se que duas informações do chuva de partículas são mais utilizadas para a seleção de elétrons e fótons: as suas larguras lateral e longitudinal. Além disso, também se observa que os chuvas se propagam por uma região lateralmente mais ampla do calorímetro conforme o seu desenvolvimento. Logo, uma boa descrição do chuva se dá por um formato de cone (*pencil-like*). Uma alternativa às grandezas físicas tradicionalmente empregadas para a

detecção de elétrons utiliza esse conceito, formando anéis concêntricos de deposição de energia. O conceito é somar a energia das células dentro da região delimitada pelo anel dentro do cone para compor a informação anelada. Ao realizar isso para uma janela do calorímetro contendo a energia do chuva, detém-se uma representação que mantém sua largura lateral. Adicionalmente, se tal processo for repetido para todas as amostragens longitudinais obtidas dentro da janela, isso permite caracterizar a largura longitudinal, bem como as nuances da evolução lateral do chuva. Assim, obtém-se uma base representativa que mantém as características físicas do processo de interação do chuva, mas com menor dimensão que a base constituída pelas células de leitura.

Em [6], está descrito o algoritmo para se determinar os anéis. Divide-se o calorímetro em um total de 100 anéis, e essa informação anelada do padrão de deposição de energia é usada como entrada para os modelos neurais classificadores. A Tabela 1 abaixo descreve as amostragens longitudinais dos calorímetros compondo as camadas e seções no algoritmo de anelamento.

TABELA 1: Descrição do número de anéis formados no algoritmo de anelamento por seção e camada do sistema de calorimetria

Seção	Camada	Nº de anéis
Calorímetro Eletromagnético (EM)	PreSampler (PS)	8
	EM1	64
	EM2	8
Calorímetro Hadrônico (HAD)	EM3	8
	HAD1	4
	HAD2	4
	HAD3	4

Adaptada de [6]

Atualmente, já existe um método de seleção rápida de elétrons baseado no emprego de rede neural de perceptrons multicamadas (MLP) que interpreta o perfil de deposição energética nos anéis, chamado de NeuralRinger, o qual possui resultados satisfatórios nesse sentido [6][7]. O modelo que atua na seleção de elétrons é um *ensemble* de classificadores em pseudorapidez e energia transversa ($E_T \times \eta$). Assim, para certo intervalo (espaço de fase) de pseudorapidez e energia transversa - esta representa a fração de energia projetada perpendicularmente ao feixe de partículas - há um classificador especializado.

O objetivo do presente trabalho é propor uma solução em *ensemble* de classificadores neurais de informação anelar capaz de identificar padrões de deposição de energia de fótons [10], para substituir o atual método de seleção *online* empregado no ATLAS, baseado em cortes lineares das variáveis de calorimetria, aqui referido como *Cutbased*. Desta maneira, procura-se atender à necessidade de manter-se elevada eficiência de detecção mesmo em condições de elevado empilhamento de sinais [8].

O método que fundamenta o trabalho é baseado no paradigma de transferência de aprendizagem de redes neurais (do inglês, *transfer learning*) [11]. Dada a similaridade dos padrões dos perfis dos anéis de fótons e elétrons, modelos classificadores pré-treinados com elétrons funcionarão como base para a construção de modelos

classificadores de fótons. O que serão transferidos em termos de aprendizagem são justamente os padrões anelares comuns a fótons e elétrons que já foram aprendidos no treinamento anterior dos modelos.

As vantagens de se aplicar transferência de aprendizagem são muitas: vão desde diminuir o tempo de treinamento, aumentando a taxa de aprendizagem, e melhora no modelo, até aumento da generalização e do desempenho final [12][11].

O documento está dividido da seguinte forma: a Seção II apresenta a divisão do banco de dados em espaços de fase para a composição do *ensemble* classificador, e a Seção III descreve a topologia das redes neurais pré-treinadas com elétrons. Prosseguindo com a Seção IV, nela está descrito o método empregado para transferência da aprendizagem daqueles modelos, bem como as configurações de treinamento. Na Seção V, são apresentados os resultados iniciais do trabalho e sua análise. E, por fim, a Seção VI busca concluir e expor as perspectivas do desenvolvimento futuro da pesquisa.

II. BANCO DE DADOS DO TREINAMENTO

A estratégia de empregar um *ensemble* foi benéfica por promover uma resposta mais suave dos classificadores em função da região do espaço de fase que ocupam [6]. Os espaços de fase foram definidos baseados na própria geometria do ATLAS e avaliados pela estratégia de *Cutbased*, a qual atua tanto em E_T , quanto em η . Optou-se por dividir o espaço de fase em cinco regiões em energia transversa (região cinemática) a partir de 15 GeV, e quatro regiões de pseudorapidez.

A composição da estatística para cada espaço de fase deu-se por simulações de Monte Carlo (MC) para a coleta de fótons e jatos com mais de 15 GeV. Utilizou-se a verdade da simulação de Monte Carlo a seleção dos fótons e de eventos de ruído físico (jatos). O volume de estatística disponível para o treinamento dos modelos para cada região é exibido na Fig. 3 abaixo.

η		Estatística disponível de Sinal/Ruído									
		s: 289034		s: 345330		s: 395594		s: 382824		s: 1821715	
3: 1.54 → 2.37		b: 2257068		b: 2367100		b: 1109731		b: 518850		b: 505414	
2: 1.37 → 1.54		s: 52159		s: 64609		s: 85994		s: 90120		s: 322647	
1: 0.80 → 1.37		b: 410499		b: 453386		b: 219728		b: 102531		b: 102198	
0: 0.00 → 0.80		s: 204503		s: 264559		s: 276582		s: 278324		s: 1301010	
		b: 1725472		b: 1767962		b: 778466		b: 351064		b: 347076	
		s: 314190		s: 415769		s: 415769		s: 388976		s: 1845070	
		b: 2776105		b: 2827188		b: 1255151		b: 563416		b: 551373	
		0: 15 → 20		1: 20 → 30		2: 30 → 40		3: 40 → 50		4: 50 → ∞	
		$E_T [GeV]$									

Fig. 3: Número de amostras disponíveis em cada região do espaço de fase.

Dependendo da necessidade da análise física, é interessante ajustar a qualidade (critério) de seleção [13]. Para elétrons, existem quatro critérios de seleção que tentam balancear volume e qualidade estatística: os critérios *vloose* (*very loose*) e *loose* priorizam grande volume de estatística, sendo o primeiro minimamente restritivo. O critério *medium* busca um equilíbrio entre a probabilidade de detecção (P_D) e a contaminação por ruído (probabilidade de falso alarme – P_F). Por outro lado, o critério *tight* prioriza a

pureza estatística, ou seja, o volume de eventos aprovados pode ser menor que dos critérios anteriores, mas a probabilidade de falso alarme é inferior – é um critério mais restritivo. Para fótons, não há o critério *vloose*, apenas os outros três. Sendo assim, deve, no final, haver um *ensemble* com um total de sessenta classificadores neurais de fótons, três para cada espaço de fase.

III. MODELOS PRÉ-TREINADOS COM ELÉTRONS

Os modelos neurais que compõem o *ensemble* de elétrons são modelos treinados em 2017 com dados de simulação Monte Carlo contendo decaimentos $Z \rightarrow ee$ (um bóson Z em um par elétron-pósitron) para coleta de elétrons e jatos hadrônicos, sobre os quais aplicou-se o método *Tag and Probe* (T&P) [6] para selecionar os elétrons.

Esses modelos neurais utilizam perceptrons multicamadas totalmente conectados com uma camada de entrada de tamanho 100 – aqueles mesmos cem anéis gerados na janela do calorímetro – ; uma camada densa escondida com função de ativação $\tanh(\cdot)$ e número de neurônios variando de 5 a 6, dependendo do ponto de operação ou do espaço de fase; e uma camada de saída com 1 neurônio e função de ativação linear. O fato de esta última se ajustada após o treinamento com tal função de ativação é porque os modelos possuem patamar de decisão ajustado linearmente com correção de *pile-up*, de forma que a eficiência dos modelos independa do empilhamento. Construir um classificador que seja resiliente a este fenômeno confere vantagem em relação ao método *Cutbased*, já que este não possui tal correção [6].

Uma maneira de avaliar o poder discriminante dos classificadores pré-treinados em relação a fótons, isto é, verificar se tais modelos preservam atributos úteis para identificar essas partículas, é avaliar o histograma de separação de sinal e ruído para todos os modelos de cada região do espaço de fase – propagaram-se os dados de fótons e de jatos através das redes neurais e se verifica o comportamento das saídas.

A Fig. 4 apresenta o histograma de separação do modelo de elétron na primeira região de E_T (15 – 20 GeV) e η (0.00 – 0.80) operando sob critério *tight*, quando são propagados fótons e jatos que atingiram aquela região do espaço de fase:

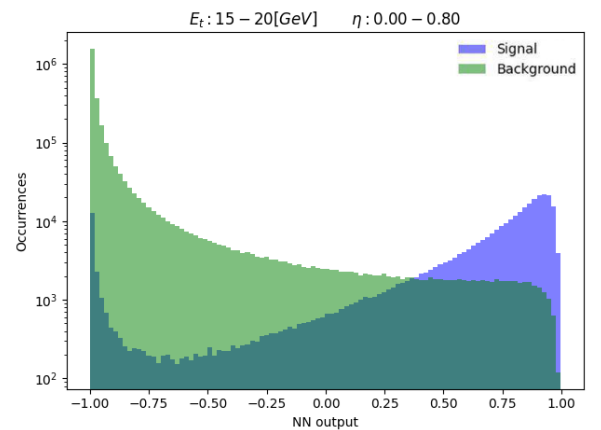


Fig. 4: Histograma de separação de sinal (azul) e ruído físico (verde) para a primeira região cinemática e de pseudorapidez.

Como é de interesse avaliar o comportamento do modelo antes da correção de empilhamento, reinsere-se a função $\tanh(\cdot)$ à saída da rede.

É evidente o caráter assimétrico de ambas as distribuições de sinal e ruído. O fato de haver um pico aparente de sinal (fótons) em +1 e de ruído de fundo (jatos) em -1, seguido de uma diminuição do número ocorrências no sentido contrário, permite a constatação de que a rede possui atributos que são úteis na discriminação de fótons. Não fosse a presença desses atributos aprendidos, haveria um padrão de distribuição normal, sinal de que a rede possuiria pouco, ou nenhum, poder discriminante. Padrões de distribuição semelhantes ao da Fig. 4 foram encontrados no histograma de separação dos modelos das outras regiões do espaço de fase. Essa avaliação qualitativa foi suficiente para decidir investir no método de transferência de aprendizado, com o objetivo de elevar o desempenho do *ensemble*.

IV. MÉTODO DE TRANSFERÊNCIA DE APRENDIZAGEM APLICADO

A fim de construir um *ensemble* de classificadores de fótons, dadas similaridades com elétrons no que tange o perfil de deposição de energia nos anéis, é proposta a ideia de aplicar transferência de aprendizado em classificadores de elétrons, retreinando-os com dados de fótons.

Devido ao caráter da aplicação de *transfer learning*, o algoritmo de transferência aqui usado deve ser, segundo [11], baseado em atributos (*feature-based algorithm*), uma vez que a aprendizagem a ser transferida corresponde ao subespaço gerado pelos atributos comuns ao domínio de origem (*source domain*) e de alvo (*target domain*). Essa abordagem mira minimizar a divergência de domínio e diminuir as taxas de erro ao buscar por atributos bem representativos do domínio fonte [12].

Nesse sentido, propõe-se o uso da técnica de ajuste fino (do inglês, *fine tuning* [12]). É um método bastante usado no contexto das técnicas de transferência de aprendizagem de modelos de *deep learning*. Consiste em uma espécie de “cirurgia” feita nas redes neurais: retira-se a camada final da rede e a substitui por uma nova, com novas inicializações e, possivelmente, com uma nova arquitetura. Em seguida, a rede passa por um novo treinamento, com dados contidos no domínio alvo [12]. Uma forma de visualizar a aprendizagem de uma rede neural é pensar que as camadas iniciais aprendem atributos genéricos, enquanto as camadas finais são responsáveis pela identificação de padrões mais sutis, isto é, são mais restritivas, mais discriminantes [14]. Por isso, é interessante congelar os pesos de certas camadas antes do novo treinamento, isto é, fixá-los. Isso possibilita, em geral, alcançar um melhor desempenho com menos tempo de treino [12].

Devido ao grande volume de dados e à quantidade de modelos gerados, embora a arquitetura e o tipo das redes neurais sejam relativamente simples, há necessidade de certo poder computacional. Para tanto, foi usada a infraestrutura de *software* Ringer Atlas, desenvolvida por pesquisadores do grupo, a qual, além de muitas outras ferramentas, conta com um *framework* de ajuste de treinamento que usa *tensorflow*. Ademais, o treinamento é realizado em uma infraestrutura computacional com GPUs dedicadas, importante para acelerar o processo.

Assim, foi feito o ajuste fino daqueles modelos de elétrons que operavam em regime *tight*, ver Fig. 5, – a escolha foi arbitrária, posto que o ajuste do limiar de decisão dos modelos de fótons é feito a posteriori. A camada de

saída foi extraída e substituída por uma camada escondida com ativação por $\tanh(\cdot)$, com número de neurônios variável de 2 a 10, e ligada à camada escondida da rede original e ao neurônio da nova camada de saída, também com função de ativação $\tanh(\cdot)$, para fins de treinamento, com alvo +1/-1 para fótons e jatos, respectivamente. Em seguida, foram congelados os pesos da primeira camada escondida, original do modelo. Assim, um modelo de elétrons gera nove modelos após o ajuste fino, diferenciando-se pelo número de neurônios na segunda camada escondida.

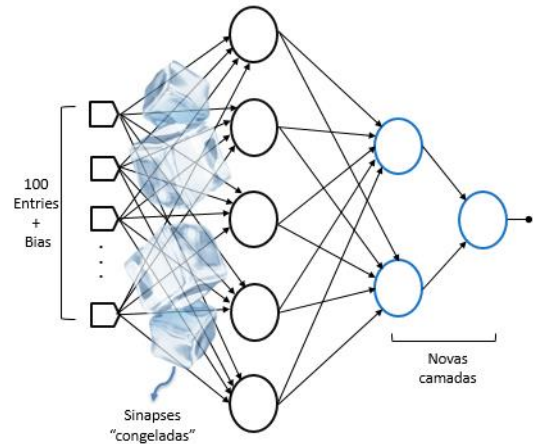


Fig. 5: Representação do processo de Fine Tuning descrito, com as sinapses da primeira camada escondida congeladas, ou fixados, e substituição da camada de saída por uma nova camada escondida e reconstrução da camada de saída.

A ideia de variar o número de neurônios na segunda camada escondida para gerar nove arquiteturas diferentes surge da dificuldade de determinar inicialmente a melhor arquitetura para a rede. Assim, construir modelos com alterações dessa quantidade de parâmetros permite obter variações no poder discriminante dos classificadores, o que é interessante para a escolha da melhor topologia.

Para cada uma das nove arquiteturas de rede ajustadas, cada uma com um número distinto e definido de neurônios na segunda camada escondida, é aplicado o método de validação cruzada *k-fold* com 10 subconjuntos. A validação cruzada permite avaliar os modelos com um conjunto de dados diferente daquele usado para treinamento, identificando, assim, classificadores mais generalistas e que não estejam ajustando excessivamente o conjunto de treino (*overfit*) ou não aprendendo com o treinamento (*underfit*). Além disso, para cada um subconjuntos de validação cruzada, são feitas 10 inicializações distintas dos pesos sinápticos não congelados. Posto que uma boa escolha de valores iniciais dos pesos pode ser muito útil no desenvolvimento de uma rede neural, propor diversas inicializações distintas ajuda a evitar os chamados pontos de mínimo local [14]. No final, totalizam-se de 900 redes por região de E_T e η para serem treinadas. Como são 20 espaços de fase, são treinados 18000 modelos de fótons, para futuramente serem analisados e, assim, resultar em um *ensemble* de 20 redes, uma para cada um dos 20 espaços de fase, sendo que cada uma tem seu corte de saída ajustado de três maneiras distintas para operar naqueles critérios de seleção *loose*, *medium* e *tight*.

A função de custo adotada para o ajuste do treinamento foi o erro quadrático médio (*MSE*) e o método de

retropropagação, ADAM [12]. O tamanho das bateladas (*batch size*) foi definido pela amostragem da menor classe, e utilizou-se de um limite de 5000 épocas de treinamento (*Epochs*), isso foi possível porque foi usado o critério de parada que avalia a evolução no subconjunto de validação por meio do critério SP, onde o índice soma-produto (SP), definido em [15] e cuja expressão é descrita em (1), busca medir o quanto a probabilidade de falso alarme (P_F) influencia na probabilidade de detecção (P_D). Durante o treinamento, busca-se, em intervalos de 25 épocas, o valor máximo de SP. Caso não haja, o modelo retrocede ao último ponto de máximo e o treinamento é interrompido.

$$SP = \sqrt{P_D(1 - P_F)} \left(\frac{P_D + (1 - P_F)}{2} \right) \quad (1)$$

É esperado, assim, obter melhorias nos modelos quando operam com a mesma probabilidade de detecção que a referência. A referência adotada foi o *Cutbased* atuando no HLTCalo, que corresponde ao segundo nível de filtragem *online* do ATLAS que emprega informação do Sistema de Calorimetria [6].

V. RESULTADOS

Optou-se por treinar, inicialmente, somente os modelos da primeira região de E_T e η , e avaliar os resultados, a fim de detectar possíveis erros de configuração, como de inserção correta dos caminhos para diretórios, ou de outra natureza, como possíveis *bugs*, e ajustá-los antes do treinamento dos inúmeros outros modelos.

O ajuste das redes, como já adiantado, deu-se com o algoritmo ADAM e o critério de parada é o máximo valor do índice soma-produto ($SP_{\text{máx}}$). Todas os modelos ajustados, bem como os históricos de evolução, foram salvos com o objetivo de possibilitar a posterior análise do treinamento.

Para o ajuste dos limiares de decisão, fixou-se que todas as redes operem com a mesma probabilidade de detecção que no HLTCalo (referência), e tomou-se a probabilidade de falso alarme resultante dessa fixação. Isso é feito através de uma curva de operação do receptor (do inglês, *Receiver Operating Characteristic (ROC)* [7][16]).

Das 10 inicializações de cada subconjunto de validação cruzada das 9 diferentes topologias de rede, filtra-se aquelas de maior índice SP. A Tabela 2 abaixo agrupa os resultados de eficiência (P_D , P_F e SP) para a referência, operando no ponto mais próximo ao P_D , e apresenta o valor médio e 1σ das melhores inicializações dos subconjuntos de validação cruzada de cada topologia para a primeira região cinemática, E_T , e de pseudorapidez, η .

Observa-se acentuada redução da probabilidade de falso alarme de uma maneira geral, para todas as topologias, ao considerar a rede operando com a mesma probabilidade de detecção que a referência: de 14.88% para, em média, 5.95%. Note que não há nenhum valor discrepante, de forma que os valores médios estão dentro das incertezas. Haja vista que no HLTCalo, opera-se com a mesma probabilidade de detecção nos regimes *loose*, *medium* e *tight*, as tabelas para os três pontos de operação são idênticas.

TABELA 2: Valores de eficiência (P_D , SP e P_F) para a referência (HLTCalo) medida em toda estatística disponível na primeira região do espaço de.

0.00 < η < 0.80 15 < E_T [GeV] < 20			
	P_D [%]	SP [%]	P_F [%]
Ref.	89.54	87.32	14.88
mlp2	89.54±0.99	91.64±0.99	6.24±0.99
mlp3	89.54±0.75	91.99±0.75	5.52±0.75
mlp4	89.54±0.83	91.83±0.83	5.85±0.83
mlp5	89.54±0.70	91.72±0.70	6.08±0.70
mlp6	89.54±0.91	91.84±0.91	5.84±0.91
mlp7	89.54±0.75	91.68±0.75	6.16±0.75
mlp8	89.54±1.00	91.96±1.00	5.60±1.00
mlp9	89.54±0.70	91.77±0.70	5.98±0.70
mlp10	89.54±0.71	91.64±0.71	6.24±0.71

Outra forma de avaliar o poder discriminante dos classificadores é construir novos histogramas de separação, assim como feito na Fig. 3. Um formato de histograma que se repete para a maioria dos histogramas de separação dessas modelos é o da Fig. 6.

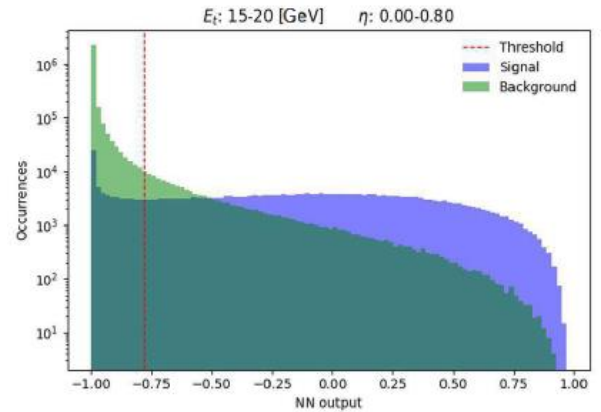


Fig. 6: Histograma de separação da terceira inicialização modelo de arquitetura 100-5-3-1 treinado com o sétimo subconjunto de validação cruzada.

Note que há um pico evidente de ruído em -1 e uma redução crescente do número de ocorrências para saídas mais próximas de +1 – mais significativa, inclusive, que aquela redução observada no histograma de separação dos modelos treinados com elétrons, nos quais há um platô na ordem de milhares de ocorrências antes de cair abruptamente perto de +1.

Todavia, há um pico menos evidente de sinal à direita, tendendo, na verdade, a uma distribuição mais uniforme que a observada anteriormente. Como resultado, limiar de decisão (*threshold*) fixo da rede é definido para um valor bastante negativo (cerca de -0.783). A despeito da distribuição de sinal, o falso alarme é reduzido justamente pelo fato de ainda assim haver um volume de estatística de falsos fótons à esquerda do patamar de decisão que possui ordem de grandeza bastante elevada. Um possível motivo para isso pode ser o fato de existir bem mais estatística disponível para ruído que para sinal (vide Fig. 2).

A escolha do melhor modelo (histograma da Fig. 7) para cada região é feito filtrando os classificadores de cada topologia possuidores do maior valor do índice SP, significando um menor falso alarme. Se houver topologia com P_F inferior que o das demais, esta é escolhida. Se não houver discrepância significativa entre os falsos alarmes, o

classificador de menor custo computacional é escolhido, ou seja, o de menos neurônios na segunda camada escondida.

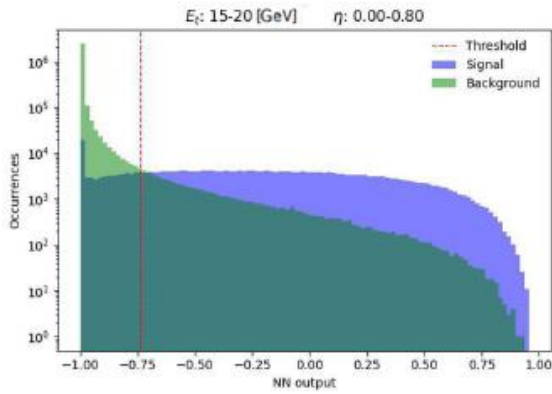


Fig. 7: Histograma de separação do modelo treinado de menor falso alarme na primeira região cinemática e de pseudorapidez.

Posto isso, o modelo selecionado para esta região foi o da primeira inicialização de arquitetura 100-5-8-1 treinado com o quinto subconjunto de validação cruzada. Operando sob mesma P_D que a referência (ver Fig. 4), apresenta probabilidade de falso alarme de, aproximadamente, 4.14%, o que representa uma redução de 72.18% em relação ao da referência. Note que o padrão de distribuição é semelhante ao da Fig. 5. O fato de não haver um pico positivo claro para sinal (fótons) no histograma implica em um corte bastante negativo. Isto indica que há aspectos de treinamento que ainda precisam ser aprimorados.

Para certificar que não houve sobreajuste ou subajuste [12][14], analisou-se as curvas de aprendizado do modelo - Fig. 8. Observando o padrão de decrescimento da curva do erro na etapa de treino, infere-se que a rede aprendeu, portanto não houve subajuste. Além disso, a partir da curva de erro na etapa de validação, vê-se que não há sobreajuste, dado que esta também tem um padrão de decrescimento e não há oscilação relevante. A evolução do índice soma-produto (SP) mostra o gradual decréscimo dos falsos alarmes a despeito do aumento dos eventos selecionados.

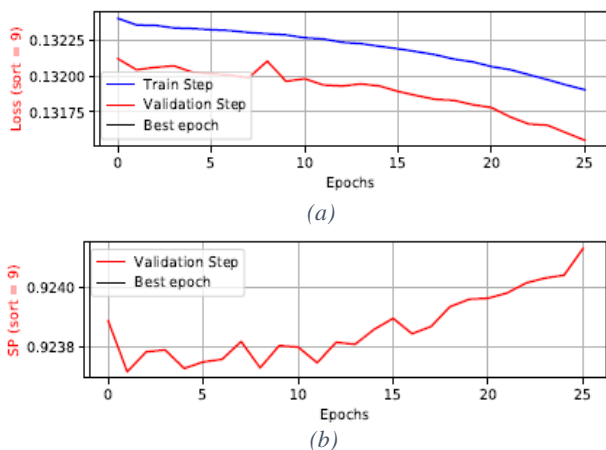


Fig. 8: curvas de aprendizado do modelo selecionado. Em (a), evolução do erro quadrático médio (MSE) por época, e, em (b), a evolução do índice soma-produto (SP).

VI. CONCLUSÕES

O presente trabalho propôs o desenvolvimento de classificadores neurais de fótons através da transferência de

aprendizagem de modelos de elétrons, a fim de substituir o atual método de seleção *online* daquelas partículas. Foram apresentados os resultados de desempenho (P_D , P_F e SP) do treinamento realizado do modelo na primeira região de E_T e η , quando se verificou uma redução significativa (de mais de 70%) na probabilidade de falso alarme em relação à referência ao operar com a mesma probabilidade de detecção desta. Para além disso, após a conclusão do treinamento dos modelos das demais regiões, e efetuada a mesma análise feita neste trabalho, é feito o reajuste dos limiares de operação considerando os efeitos de empilhamento, os quais já são aplicados no estágio preciso do HLT e no método *offline*, realizando, assim, um ajuste linear da saída da rede em função do estimador de empilhamento. Nesse contexto, para evitar a propagação de não-linearidades no ajuste [6], retira-se a função de ativação $\tanh(\cdot)$ do neurônio da camada de saída da rede. Se tal ajuste for positivo para um classificador, isto é, se houver redução dos falsos alarmes, adota-se o limiar ajustado por empilhamento para aquela rede, caso contrário mantém-se o limiar fixo. Em seguida, é importante avaliar o impacto desses classificadores na coleta de fótons – isso é feito simulando o *Trigger* de fótons -, a fim de garantir que sejam entregues os eventos certos para a etapa posterior na cadeia de filtragem – a etapa precisa - fazer os cortes adicionais. Por fim, serão usados dados de física de colisões para treinar o *ensemble* gerado por transferência de aprendizagem, e serão comparadas as respostas das redes frente a um *ensemble* treinado somente com dados de fótons, a fim de determinar o método mais adequado para atuar no *Trigger* de fótons na próxima etapa de coleta de dados do ATLAS.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Os autores também agradecem ao CNPq, FAPERJ, FAPESB e RENAFAP (MCTI) e CERN pelo apoio. Além disso, agradecemos à colaboração TDAQ/ATLAS por prover os dados de simulação.

REFERÊNCIAS

- [1] T. Schömer-Sadenius (ed). “The Large Hadron Collider: Harvest of Run 1”. 1ª ed., Springer International Publishing, pp. 532, 2015.
- [2] ATLAS collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, JINST3(2008)S08003.
- [3] D. Green (ed). “At the leading edge: The ATLAS and CMS LHC Experiments”. 1ª ed., World Scientific Publishing, pp. 445, 2010.
- [4] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, Phys. Lett. B, v. 716, pp. 1-29, 2012.
- [5] Atlas collaboration “Performance of electron and photon triggers in ATLAS during LHC Run 2”. European Physical Journal C., v.80, pp. 1-41, 2020.
- [6] J. M. Seixas, L. P. Caloba, M. N. Souza, A. L. Braga, A. P. Rodrigues. “Neural second-level trigger system based on calorimetry”, Computer Physics Communication, Volume 95, Number 2-3, pp. 143-157, June, 1996.
- [7] ATLAS Collaboration. “An Ensemble of Neural Networks for Online Filtering Implemented in the ATLAS Trigger System”, Journal of Physics, Conf. Ser.1162 012039, pp. 8, 2018.
- [8] ATLAS Collaboration. “The Upgrade of the ATLAS Electron and Photon Triggers towards LHC Run 2 and their Performance”,

European Physical Society Conference on High Energy Physics 2015, pp.264., Vienna, Austria, 22 - 29 Jul 2015.

- [9] G. Barr, R. Devenish, R. Waleczak, T. Weidberg. "Particle Physics in the LHC era". 1^a ed., Oxford University Press, pp. 423, 2016.
- [10] ATLAS Collaborarion. "Photon identification with the ATLAS detector", 2017 European Physical Society Conference on High Energy Physics, , pp.760, Venice, Italy, 05 - 12 Jul 2017.
- [11] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, "Transfer Learning". Reino Unido, UK Cambridge University Press, p. 377, 2020.
- [12] D. Sarkar, R. Bali, and T. Ghosh, "Hands On Transfer Learning with Python Implement advanced deep learning and neural network models using TensorFlow and Keras". English Edition, ed. [S.] Packt Publishing Ltd, p. 554, 2018.
- [13] Gesualdi-Mello, A ; Anjos, A ; Armstrong, *et al.* "Overview of the High-Level Trigger Electronand Photon Selection for the ATLAS Experiment at the LHC". Nuclear Science, IEEE Transactions on, vol. 53, no. 5, pp. 2839 – 2843, 10 2006.
- [14] S. Haykin, "Redes Neurais Principios e Prática". 2^a ed Porto Alegre, BR Bookman, p. 902, 2001.
- [15] E. F. S. Filho, J. M. Seixas, L. P. Caloba. "Online neural filtering operating over segmented discriminating components". Proceedings of the 15th IEEE International Conference on Electronics, Circuits and Systems, 31 August–3 September 2008, St. Julien's, Malta. Washington DC, USA: IEEE, 2008: pp 530–533.
- [16] M. Gonen. "Analyzing Receiver Operating Characteristic Curves With SAS", 1^a ed., SAS Publishing, p. 153, 2007.