



UNIVERSIDAD
NACIONAL
DE LA PLATA

Universidad Nacional de La Plata
Facultad de Ciencias Exactas
Departamento de Física

Identificación de fotones provenientes del decaimiento radiativo del bosón Z mediante Redes Neuronales Artificiales Convolucionales

Trabajo de Diploma

Santiago J. Noacco Rosende

Director

Dr. Fernando Monticelli

Marzo 2020

Índice general

Agradecimientos	1
Introducción	3
1. El Gran Colisionador de Hadrones y ATLAS	5
1.1. El detector de ATLAS	7
1.1.1. Sistema de coordenadas	8
1.1.2. Detector Interno de trazas (ID)	8
1.1.3. Calorímetros	11
1.1.4. Espectrómetro de Muones	13
2. Sistema de <i>Trigger</i>	15
2.1. Configuración del trigger	17
2.2. <i>L1</i>	17
2.3. <i>HLT</i>	18
3. Fotones en ATLAS	21
3.1. Reconstrucción	23
3.2. Identificación	26
3.2.1. Aislamiento de fotones	29

4. Redes Neuronales	31
4.1. Características del entrenamiento	33
4.1.1. Función de pérdida	34
4.2. Redes Neuronales Convolucionales	35
5. Identificación de fotones con CNNs	39
5.1. Almacenamiento de eventos	39
5.2. Definición y obtención de señal	41
5.2.1. Decaimiento radiativo del bosón Z	41
5.3. Definición y obtención de fondo	44
5.4. Creando las imágenes de los <i>topo-clusters</i>	46
5.5. Entrenando la <i>CNN</i>	48
6. Resultados	51
7. Conclusiones	55
Bibliografía	60

Agradecimientos

Aquellos que nos rodean nos moldean, y a ellos les debo el bello paisaje del camino recorrido. Todo el esfuerzo plasmado en este trabajo, síntesis de años de estudio, han sido posibles gracias al acompañamiento y sosten de mi familia; de quienes estoy inmensamente agradecido: mis abuelos Edith y Yeye, mis hermanos Merce y Manu, y en especial a mi mamá y mi papá. Junto a ellos quiero resaltar y agradecer la hermosa compañía de mi novia Pau, y a mis amigos que supieron darme aire y fuerzas en incontables ocasiones. Finalmente agradecer al grupo que me dio la oportunidad de realizar el trabajo, a todo el equipo de ATLAS de la Plata, en especial a Fer quién supo ser una guía incondicional, siempre amable, positivo y que hizo de un trabajo final una experiencia única.

Introducción

En el experimento ATLAS del CERN se realizan cientos de proyectos que buscan empujar el límite de nuestro conocimiento físico, y es allí donde se pone a prueba la teoría más completa y precisa que la comunidad científica ha logrado gestar, el Modelo Estándar (*Standard Model*). Sí bien posee un gran poder predictivo, es un modelo perfectible y acotado que no explica numerosos fenómenos físicos.

Uno de los métodos muy utilizados por ATLAS para explorar la posible existencia de nueva física consiste en estudiar las interacciones entre partículas conocidas del SM con el fin de descubrir, en medidas de precisión evidencia de nuevas partículas. En el programa de física de ATLAS, que contempla entre otros estudios, mediciones de las propiedades del bosón de Higgs, búsqueda de materia oscura, búsqueda de evidencia de Super Simetría (SUSY) y búsqueda de física más allá del SM en otros modelos, requieren contar generalmente con fotones en el estado final. Estos conforman un elemento clave, por ejemplo, para el estudio de QCD perturbativo en el contexto del SM , o para realizar búsquedas más allá del SM mediante resonancias de alta masa en estados finales con jets y fotones. Es por ello que resulta fundamental contar con un método eficiente en la reconstrucción e identificación de fotones y separación del fondo a partir de los datos de ATLAS.

En el marco de este trabajo evaluaremos la eficiencia en la identificación de fotones derivados del decaimiento radiativo del bosón Z entre el método vigente que emplea cortes rectangulares en las variables de la forma de las lluvias electromagnéticas en el calorímetro electromagnético, y un segundo método basado en datos y redes neuronales artificiales. El objetivo del trabajo será evaluar la posibilidad y viabilidad de implementar este método basado en Redes Neuronales en la etapa de reconstrucción rápida del sistema de selección “online” del trigger de alto nivel (*High Level Trigger*).

En los últimos años los avances en técnicas de aprendizaje automático y la mayor dis-

ponibilidad de recursos computacionales ha vuelto estas técnicas muy atractivas, en particular por su poder de clasificación, velocidad, escalabilidad e incluso la posibilidad de ser implementadas directamente en hardware. En este contexto utilizamos dos modelos de Redes Neuronales Artificiales Convolucionales, que entrenamos con muestras enriquecidas de fotones provenientes del decaimiento radiativo del bosón Z como señal. Y candidatos fotones que cumplen un criterio de aislación inversa como muestra enriquecida de fondo; obtenidas a partir de datos colectados por ATLAS. Con los modelos entrenados calculamos la eficiencia en la identificación de fotones y de rastro falso; este último también se puede entender como la eficiencia en el rechazo de fondo. Finalmente lo comparamos con los resultados del cálculo de las eficiencias mediante el método que emplea los cortes rectangulares en las variables de las lluvias electromagnéticas.

Capítulo 1

El Gran Colisionador de Hadrones y ATLAS

En el marco de los proyectos de la Organización Europea para la Investigación Nuclear (**CERN**) se encuentra el Gran Colisionador de Hadrones (*Large **H**adron **C**ollider*), una imponente obra de ingeniería y tecnología que permite llevar a cabo los más ambiciosos y vanguardistas experimentos de partículas en altas energías. El proyecto que comenzó su construcción en 1998 a partir del preexistente LEP vio su primera colisión entre protones a fines del 2009. El colisionador consiste en un doble anillo de 26,7 *km* de diámetro soterrados entre 40 y 170 *m* en una superficie que abarca una porción de Suiza (Ginebra) y Francia. En el interior de los tubos circulan dos haces de protones en direcciones opuestas que colisionan en cuatro lugares distintos, dónde se encuentran los detectores ALICE, ATLAS, CMS y LHCb. [1] [2] [3]

Tras la finalización del período conocido como Run2 (2015 - 2018), y previo al comienzo del Run3, se ha optimizado el LHC para colisionar protones (e iones pesados) a una energía de centro de masa de $\sqrt{s} = 13 \text{ TeV}$ (5,2 *TeV*). El proceso de aceleración de las partículas consta de varias etapas que involucran distintos aceleradores, cuyo objetivo es aumentar paulatinamente la energía de las partículas (ver figura 1.1), culminando en el LHC.

Las cavidades o tubos por dónde se propagan los haces de partículas presentan condiciones de extremo vacío, logrando una presión de 10^{-10} torr . Además, en total se utilizan alrededor de 10000 imanes superconductores mantenidos a una temperatura de 1,9 *K*. Dentro de este conjunto, 1232 dipolos magnéticos se emplean para mantener los haces en la órbita circular del acelerador; junto con otros 392 cuadrupolos magnéticos que man-

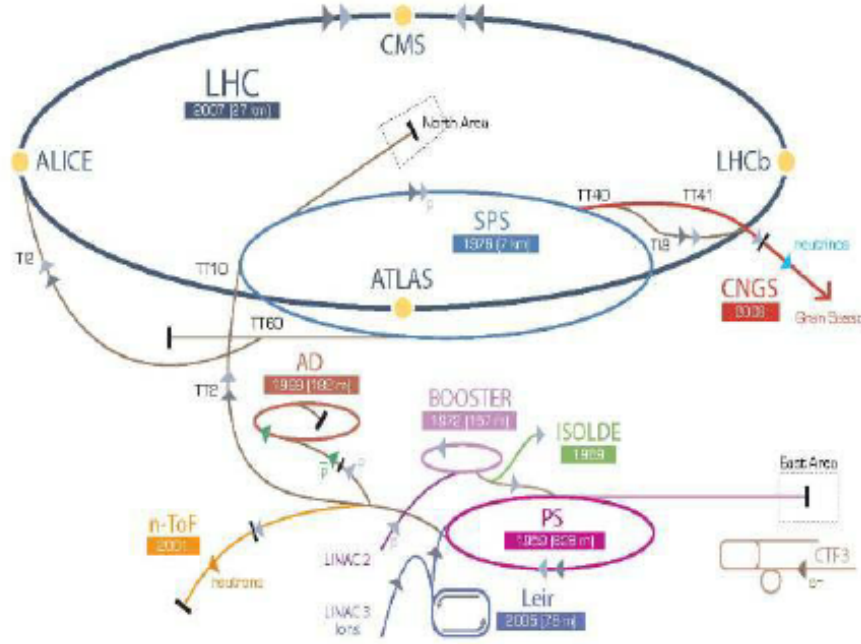


Figura 1.1: Esquema del LHC, los pre aceleradores y los sitios de detección.

tienen el haz enfocado. El resto de los imanes son de orden superior y su rol es corregir imperfecciones del campo magnético en las extremidades de los dipolos.

Los haces que circulan dentro del LHC no son continuos, por el contrario se conforman de paquetes (*bunches*) de $\sim 10^{11}$ protones cada uno. Estos paquetes viajan en trenes de hasta 2808 paquetes por tren, permitiendo colisiones cada 25 ns en promedio; dando lugar a una tasa de colisiones por paquete de 40 MHz.

Un importante parámetro del acelerador, que permite caracterizar su funcionamiento es la luminosidad instantánea (*instantaneous luminosity*) \mathcal{L} . La misma se define como el número de partículas por unidad de tiempo y unidad de área:

$$\mathcal{L} = f_{rev} n_b \frac{N_1 N_2}{A}, \quad (1.1)$$

donde f_{rev} es la frecuencia de revolución (~ 11 kHz), n_b es el número de paquetes por haz, N_i el número de partículas en cada paquete y A es la sección eficaz del haz.

$$A = \frac{4\pi\epsilon_n\beta^*}{\gamma F}, \quad (1.2)$$

donde ϵ_n es la emitancia transversal normalizada (la dispersión transversal media de las partículas del haz en el espacio de coordenadas e impulsos), β^* es la función de amplitud

en el punto de interacción, relacionada al poder de focalización de los cuadrupolos), γ es el factor relativista de Lorentz y F es un factor de reducción geométrico, debido al ángulo de cruce de los haces en el punto de interacción.

Cercanamente relacionada con estas variables, se suele utilizar la medida de la luminosidad integrada temporalmente (*time-integrated luminosity*), expresada en femtobarns inversos (fb^{-1}), para medir el número eventos por colisión. Durante Run2, el LHC produjo colisiones protón-protón equivalentes a $\sim 140 fb^{-1}$ a una energía de centro de masas de $13 TeV$.

1.1. El detector de ATLAS

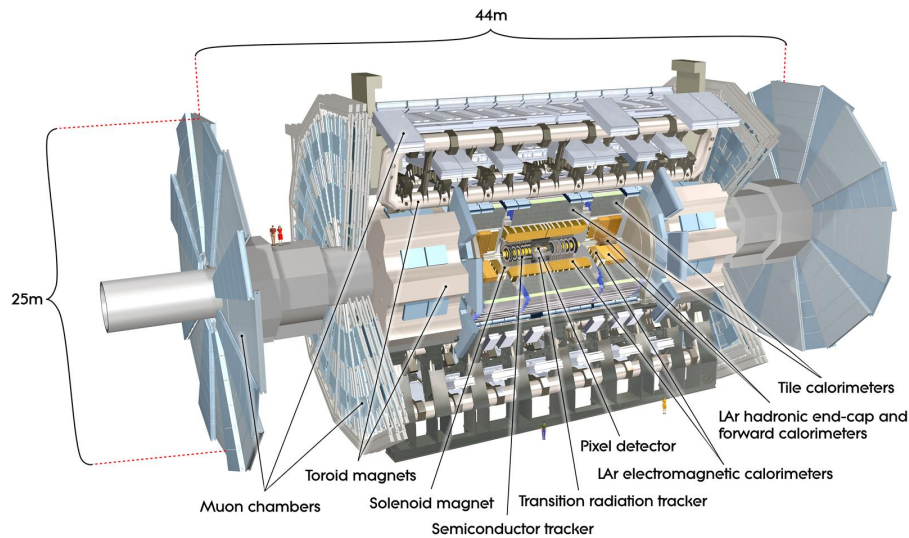


Figura 1.2: Esquema del detector ATLAS donde se pueden observar los distintos subdetectores y regiones que lo conforman.

El detector de ATLAS es un detector multipropósito diseñado para observar partículas producidas en colisiones de protón-protón (pp) e iones-pesados (HI) a altas energías. Posee una geometría cilíndrica aproximadamente simétrica, capaz de proveer una cobertura de ángulo sólido de 4π . Está compuesto por un detector de trazas (*Inner tracking Detector*) en la región interna alrededor del punto de interacción, rodeado de un fino solenoide superconductor (produciendo un campo magnético axial de $2 T$), calorímetros electromagnéticos y hadrónicos; además de un espectrómetro de muones. Este último

incorpora un arreglo de 3 grandes toroides magnéticos. [4] [5] [6]

1.1.1. Sistema de coordenadas

ATLAS utiliza un sistema de coordenadas acorde a la regla de la mano derecha, cuyo origen es el punto de interacción nominal (*nominal Interaction Point*), que se encuentra en el centro del detector; y el eje z a lo largo del tubo del haz. [4] [3] El eje x apunta desde el IP al centro del anillo del LHC, mientras el eje y apunta hacia arriba. En el plano transverso se utilizan coordenadas cilíndricas (r, ϕ) , siendo ϕ el ángulo acimutal alrededor del tubo del haz. Sin embargo, resulta útil definir la **psudorapidez** en términos del ángulo polar (θ) :

$$\eta = -\ln(\tan(\theta/2)), \quad (1.3)$$

de tal forma que se define la distancia angular ΔR :

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (1.4)$$

1.1.2. Detector Interno de trazas (ID)

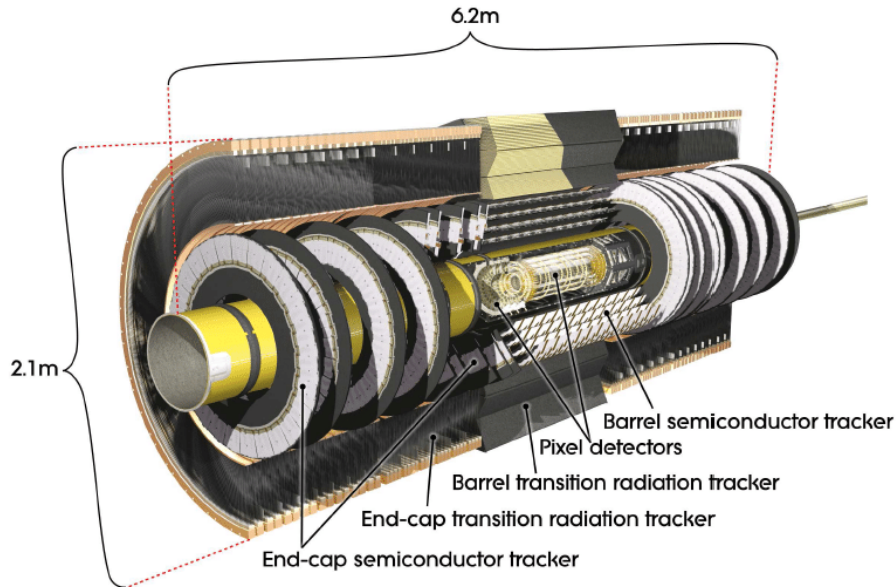


Figura 1.3: Esquema global del Detector Interno de ATLAS, donde se pueden observar sus componentes y dimensiones.

El ID está inmerso en un campo magnético de 2 T , proveyendo la reconstrucción precisa de las trazas de las partículas cargadas, en el rango de pseudorapidez $|\eta| < 2,5$, gracias a la información combinada de tres subdetectores. [4] [6] Donde por **trazas** entendemos a la secuencia de puntos coordinados que describen el rastro dejado por partículas cargadas tras interactuar con el detector. Este detector consiste en una región central (*barrel*) dispuesta alrededor del tubo del haz, que aporta cobertura total para $|\eta| < 1,5$; y dos tapas (*endcaps*). En la región de tapas se encuentran discos ubicados perpendicularmente a la dirección del haz, con el fin de cubrir la región $1,5 < |\eta| < 2,5$. Los subdetectores que conforman al ID (ver figuras 1.3 y 1.4) son: el detector de pixeles de silicio de alta granularidad, un detector de radiación de transición (**T**ransition **R**adiation **T**racker), y la capa insertable B (*I*nstable *B* *L*ayer).

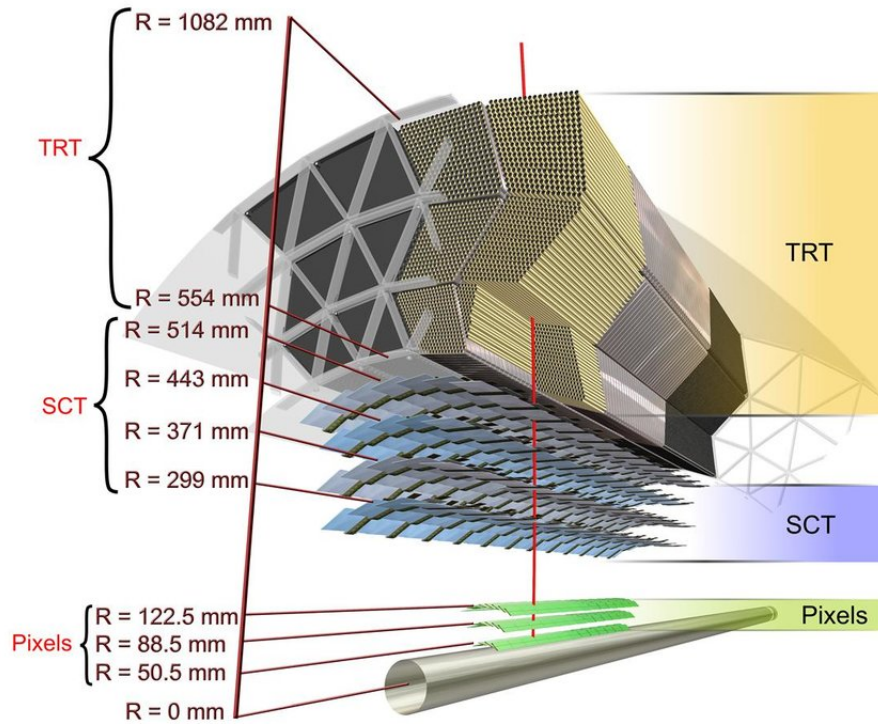


Figura 1.4: Esquema del corte lateral del ID, donde se observan los subdetectores que lo componen y su respectiva distancia al *IP*.

IBL

Siendo la parte del ID más cercana al *IP*, el *IBL* posee una distancia media al punto de interacción (dado por el radio medio) de 33 *mm*. [4] [3] Este detector está compuesto por 8 millones de canales de rápida lectura y con sensores de silicio, que detectan el paso de partículas cargadas mediante la deposición de carga inducida. El tamaño de los píxeles es de $50 \times 250 \mu m^2$, con una resolución de 8 μm ($R - \phi$) y 40 μm (z). El objetivo del mismo es mejorar la eficiencia en la identificación de trazas, vértices, y en la identificación de bottom quarks, que típicamente decaen fuera del radio del *IBL*. Los **vértices** son los puntos dónde se identifica la convergencia de dos o más trazas.

Detector de pixeles

Partiendo del *IP* se encuentra el detector de pixeles cubriendo la región del vértice, dividida en 3 capas en la región *barrel* y 3 discos a cada lado del punto de interacción (*IP*) reforzando la región *endcap*. [3] Cada capa de la región *barrel* se ubica a una distancia media del *IP* de 50,5 (*B-Layer*), 88,5 y 122,5 *mm* respectivamente (ver figura 1.4).

Su objetivo es medir la posición de las trazas de las partículas cargadas con alta precisión. Así mismo, aporta información esencial en la reconstrucción de vértices primarios y secundarios. Este provee típicamente cuatro medidas tridimensionales por traza.

El principio de detección para partículas cargadas es la medida de la deposición de la carga inducida en una capa de silicio por ionización. El sistema contiene un total de 80 millones de sensores, cada uno con una resolución de 10 μm ($R - \phi$) y 115 μm (z).

Detector Semiconductor de Trazas

Cubriendo el detector de pixeles, se encuentra el detector semiconductor de trazas (*SemiConductor Tracker*) compuesto por módulos de sensores. [4] [3] En la región *barrel* los módulos de SCT están dispuestos en 4 capas concéntricas, mientras que en la región *endcap* consiste en 9 discos transversales al eje del haz.

Cada módulo consta de dos sensores de silicio consecutivos segmentados en microbandas, dada la baja multiplicidad de partículas, aportando lectura estéreo de ángulo pequeño. Su función es medir las trazas con alta precisión en la zona intermedia del de-

tector. Típicamente provee ocho impactos 2D (*hits*) que se traducen en cuatro medidas 3D, por traza a una distancia media del *IP* de $(275 < r < 560)$ *mm*. Logrando una resolución de $17 \mu\text{m}$ ($R - \phi$) y $580 \mu\text{m}$ (z).

TRT

La región más externa del ID ($563 < r < 1066$) *mm* y en el rango $\eta < 2,0$, está cubierta por un Rastreador de Transiciones Radiativas (***Transition Radiation Tracker***). [4] [6] El *TRT* consiste en tubo finos apilados (*straw drift tubes*) llenos de una mezcla gaseosa compuesta por: 70 % *Xe*, 27 % *CO*₂ and 3 % *O*₂, intercalado con radiadores de transición de polipropileno/polietileno.

Permite identificar trazas de vértices secundarios, dando lugar una eficiente reconstrucción de conversión de fotones, hasta una distancia radial de 80 *cm* de la línea del haz. Para partículas cargadas con momento transverso mayor a 0,5 *GeV* el *TRT* provee típicamente 35 impactos por traza.

1.1.3. Calorímetros

El sistema del calorímetro consta tanto de componentes electromagnéticas (***ElectroMagnetic Calorimeter***) como hadrónicas (***Hadronic CALorimeter***). Está diseñado para aportar total cobertura en ϕ y cobertura de pseudorapidez en el rango $|\eta| < 4,9$, con granularidad más fina en la región asociada al detector interno. [6]

EMC

Este detector es un calorímetro de muestreo de plomo/argón-líquido (LAr) con geometría de acordeón. Está dividido en dos medio-tubos o *half-barrels* cubriendo la región de pseudorapidez ($-1,475 < \eta < 0$ y $0 < \eta < 1,475$), y dos tapas *endcap* cubriendo la región ($1,375 < |\eta| < 3,2$). La región de transición entre *barrel* y *endcap* ($1,37 < |\eta| < 1,52$) contiene una cantidad significativa de material inactivo. Sobre la región dedicada a medidas de precisión ($\eta < 2,5$ excluyendo la región de transición), el calorímetro EM está segmentado en tres capas longitudinales con respecto a la profundidad de las lluvias.

La primer capa posee un ancho de alrededor de 4,4 longitudes de radiación (X_0) en $\eta =$

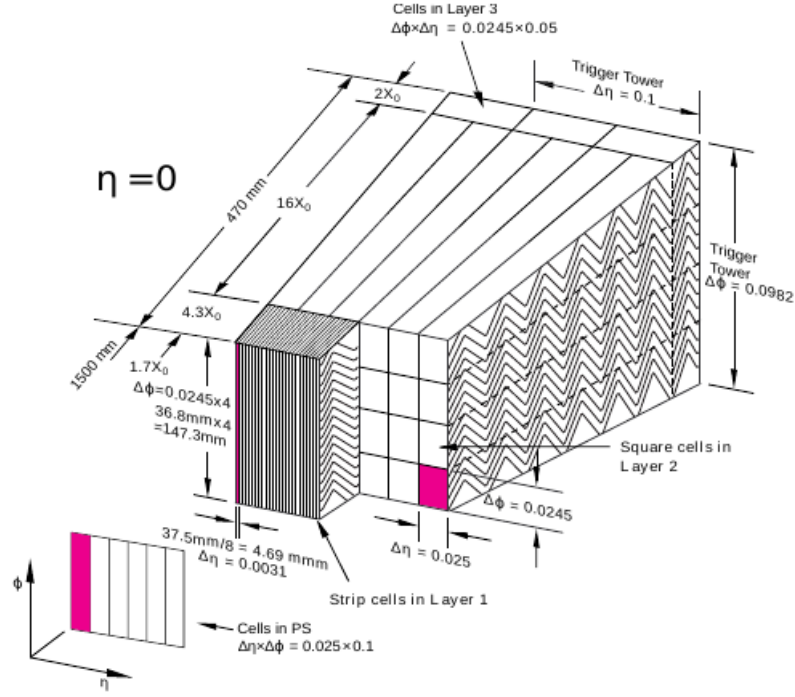


Figura 1.5: Esquema de la segmentación lateral y longitudinal del calorímetro electromagnético de ATLAS alrededor de $\eta = 0$. [4]

0. En el rango $|\eta| < 1,4$ y $1,5 < |\eta| < 2,4$, esta capa se divide en tiras de fina granularidad en la dirección de η , con un tamaño típico de celda para la región *barrel* de $0,003 \times 0,0982$ en $\Delta\eta \times \Delta\phi$. Para $1,4 < |\eta| < 1,5$ y $2,4 < |\eta| < 2,5$ la segmentación en η de la primera capa es más granular, con celdas de tamaño $0,025 \times 0,0982$ en $\Delta\eta \times \Delta\phi$. La fina granularidad en η de las tiras es suficiente como para ofrecer una excelente discriminación entre fotones aislados y pares de fotones cercanamente espaciados, provenientes del decaimiento $\pi^0 \rightarrow \gamma\gamma$ para un momento transversal de hasta $O(100) \text{ GeV}$.

Para electrones y fotones con alta energía transversal, la mayoría de la energía es recolectada en la segunda capa. Esta posee un ancho de aproximadamente $16 X_0$, contando con una granularidad de $0,025 \times 0,025$ en el espacio $(\Delta\eta, \Delta\phi)$, tal como se puede observar en la figura 1.5.

La tercera capa provee medidas de energía depositada en la cola de las lluvias. Posee una granularidad de $0,05 \times 0,0245$ en $(\Delta\eta, \Delta\phi)$, y una profundidad de alrededor de $2 X_0$ en $\eta = 0$. Se utiliza para corregir las pérdidas de lluvias de alta energía en materiales por fuera del *EMC*.

Frente al calorímetro acordeón, una fina capa de pre muestreo (***PreSampler layer***), cubriendo el intervalo de pseudorapidez $|\eta| < 1,8$ es utilizada para corregir la pérdida de energía (*upstream*) antes de llegar al calorímetro. La capa *PS* consiste en capa de *LAr* activa de $1,1(0,5)$ cm de ancho en la región *barrel* (*endcap*), y posee una granularidad de $0,025 \times 0,0982$ en $(\Delta\eta, \Delta\phi)$. El material de subida (*upstream*) de *PS* tiene un ancho de $2 X_0$ para $|\eta| < 0,6$. En la región $0,6 < |\eta| < 0,8$ el ancho incrementa linealmente de $2 X_0$ a $3 X_0$. Para $0,8 < |\eta| < 1,8$ el ancho es ligeramente superior a $3 X_0$, con la excepción de la región de transición entre *barrel* y *endcap* y la región cercana a $|\eta| = 1,7$, donde alcanza $5 - 6 X_0$. [4]

HCAL

Tres capas de calorímetro hadrónico rodean los *EMC*. Consiste en un calorímetro de losas de acero/centellador en la región central ($|\eta| < 1,7$), y un calorímetro de muestreo *LAr* con absorbedores de cobre y tungsteno en las regiones *endcap* ($1,5 < |\eta| < 3,2$) y *forward* ($3,1 < |\eta| < 4,9$). Para electrones y fotones, estos proveen discriminación extra del fondo mediante medidas de energía hadrónica. [4]

1.1.4. Espectrómetro de Muones

Las capas más externas del detector ATLAS, rodeando los calorímetros, consisten en un espectrómetro de muones (***Muon Spectrometer***) en el rango de pseudorapidez $|\eta| < 2,7$. Incorpora tres grandes ensambles de imanes toroidales con ocho bobinas cada uno, cámaras rastreadoras de precisión (*precision tracking chambers*) y detectores rápidos para usar el trigger en $|\eta| < 2,4$. La integral de los campos de los toroides abarca entre $2,0$ y $6,0$ Tm para mayor aceptación. [4]

Capítulo 2

Sistema de *Trigger*

Los eventos de interés son seleccionados mediante un sistema de trigger y adquisición de datos (*Trigger and Data Acquisition system*) de dos niveles, el cuál se esquematiza en la figura 2.1. El Primer Nivel (*Level 1*) basado en hardware y el Trigger de Alto Nivel (*High-Level Trigger*) basado en software. [4] [5] [7]

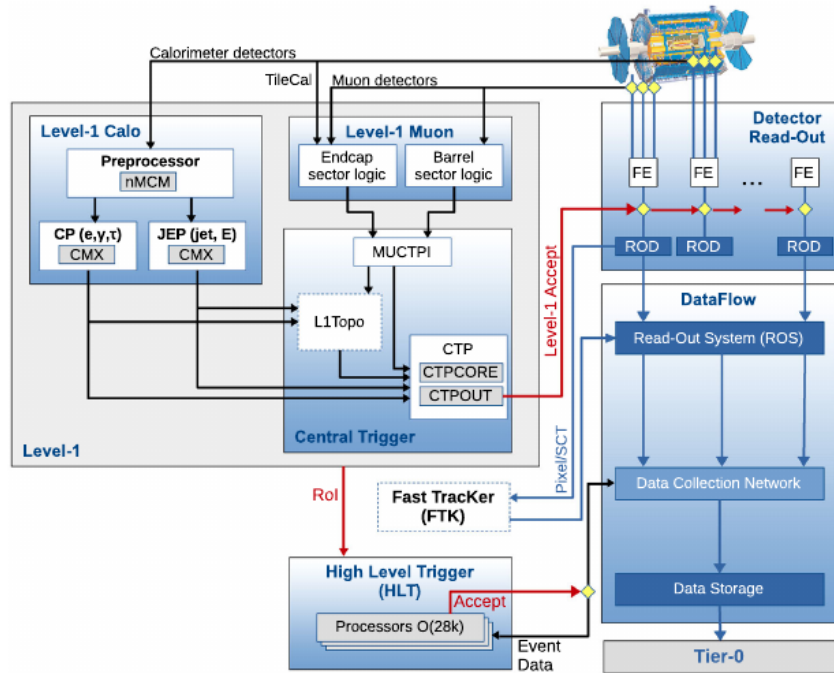


Figura 2.1: Esquema del sistema TDAQ de ATLAS para Run2. Las flechas indican la dirección de los datos partiendo del detector hacia las distintas secciones del sistema. [7]

Una secuencia de los algoritmos de *L1* y *HLT* se denomina “trigger” y su fin es identificar una o más partículas dado un tipo y un umbral de energía o momento transverso. Por ejemplo, los triggers de fotones y electrones están pensados para seleccionar eventos con uno o más electrones o fotones en el detector.

Los datos crudos (*raw data*) medidos por el detector son transmitidos al *L1* en formato de *bit-streams*, conteniendo un identificador de la parte del detector y los datos medidos en la misma. El *L1* utiliza señales de gruesa granularidad provenientes de los calorímetros y del espectrómetro de muones para reducir la tasa de eventos generados en las colisiones del LHC (*bunch crossing rate*) de 40 *MHz* a menos de 100 *kHz*. También definen regiones de interés (*RoIs*) que contienen clusters del calorímetro con alta energía transversa E_T , o trazas de muones provenientes de las cámaras de muones. La energía transversa se define como

$$E_T = \frac{E}{\cosh \eta}, \quad (2.1)$$

con E la energía del candidato. Las decisiones del *L1* son tomadas por el Procesador Central del Trigger (***C*entral *T*rigger *P*rocessor**), el cuál recibe datos del calorímetro *L1* (*L1Calo*) y del trigger *L1* de muones (*L1Muon*) como así también de otros subsistemas cómo: Minimum Bias Trigger Scintillators (MBTS), el contador Cherenkov LUCID y el Zero-Degree Calorimeter (ZDC). El *CTP* es además responsable de aplicar un tiempo muerto preventivo, para evitar superposición de lectura y la saturación de los *Buffers* de los *front-end*.

Los eventos aceptados por el *L1* son almacenados en el sistema de lectura (***R*ead *O*ut *S*ystem**) y procesados por el *HLT*. Este debe reducir el número de eventos almacenados en disco a una tasa promedio de alrededor de 1 *kHz*, en unos pocos segundos. A cada paso el *HLT* realiza la selección de candidatos a partículas, de tal manera que si se produce un error en algún paso, los subsecuentes no se ejecutan. Este procedimiento es esencial para reducir el tiempo que le toma al *HLT* reconstruir el evento y tomar una decisión. Los eventos aceptados por el *HLT* son trasferidos a sitios de almacenamiento local, y exportados a centros de cómputo *Tier-0* del CERN, para la reconstrucción *offline*.

Además de los triggers descriptos arriba, hay *rerun* triggers los cuales jamás aceptan un evento por sí mismos, sino que están configurados para correr sólo sobre eventos aceptados por otros triggers. Su decisión es gravada para uso *offline* y la información recopilada se utiliza para estudios en la eficiencia del trigger, la cual es calculada separadamente para cada objeto (leg) de los triggers multiobjetos.

2.1. Configuración del trigger

La configuración del trigger es controlada por el “trigger menu”, el cuál define una lista completa de los triggers del $L1$ y del HLT . Para asegurar un óptimo menú de trigger contenido en las restricciones de las tasas impuestas por la luminosidad del LHC, es posible aplicar factores de preescala tanto a los triggers del $L1$ como a los del HLT independientemente y configurarlos durante la recopilación de datos. Esto le permite al experimento deshabilitar los triggers completamente o establecer una fracción de eventos que pueda ser aceptada por ellos. El *trigger menu* consiste en:

- *primary triggers* utilizados para análisis físico, típicamente no están preescalados.
- *support triggers* utilizados para medidas de eficiencia y rendimiento o para monitoreo, típicamente pre escalados.
- *alternative triggers* utilizando algoritmos de reconstrucción alternativos (nuevos o experimentales), comparados con los dos triggers mencionados anteriormente. Usualmente solapando los *primary triggers*.
- *backup triggers* utilizados como opción de respaldo si algún *primary trigger* dispara con un rate muy alto.
- *calibration triggers* utilizado para seleccionar eventos que son de interés para hacer calibraciones con datos del detector.

Los *primary triggers* cubren todas las señales relevantes para el programa físico de ATLAS, incluyendo electrones, fotones, muones, leptones tau, (b-) jets y energía transversa perdida E_T^{miss} ; los cuales son usados para medidas de precisión del Modelo Estándar (SM) incluyendo decaimientos del Higgs, bosones W y Z, y búsquedas de física más allá del SM como partículas pesadas, supersimetría o partículas exóticas.

2.2. $L1$

Los algoritmos de $L1Calo$ realizan sus selecciones utilizando objetos “torre de cluster” (*cluster towers*), ilustrados en la figura 2.2. Tanto los triggers de electrones, fotones y taus identifican RoI como 2×2 torres de clusters de los EMC , para los cuales la suma de

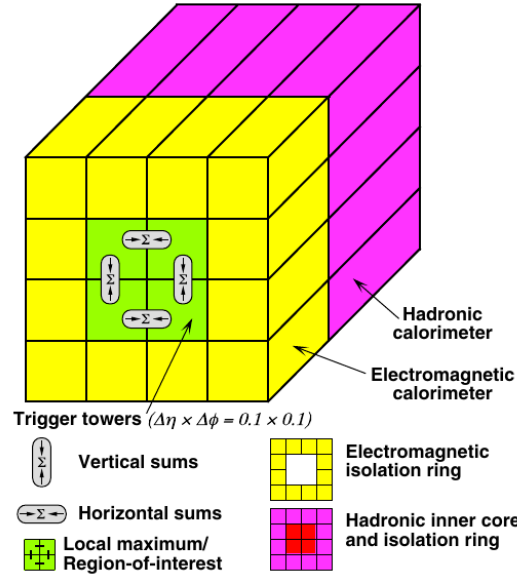


Figura 2.2: Vista esquemática de las torres de clusters, utilizadas por los algoritmos de *L1Calo* como *input.i*

la energía transversa de al menos uno de los cuatro posibles pares de torres vecinas más cercanas (1×2 o 2×1) excedan un límite predefinido. Es decir que los *RoIs* contienen información de coordenadas y tipo de detector. El límite en E_T puede ser fijado para diferentes regiones de η con una granularidad de $0,1 \eta$, con el fin de corregir la respuesta en energía del detector.

Los resultados del preprocesador *preprocessor* son transmitidos tanto a los subsistemas Procesador de Cluster (*Cluster Processor*) y al Procesador de Jet/Energía-sumada (*Jet/Energy-sum Processor*) en paralelo. El subsistema *CP* identifica candidatos electrones/fotones y leptones tau con E_T por encima de un umbral programable, y de ser necesario, satisfaciendo cierto criterio de aislamiento. La información como la suma de E_T , las coordenadas $\eta - \phi$ y los umbral de aislamiento (de ser necesarios), son almacenados en objetos trigger (*trigger objects*). [5] [8]

2.3. HLT

Tras haber sido aceptados por el trigger *L1*, los eventos son procesados por el *HLT* utilizando información calorimétrica de fina granularidad, medidas de precisión de los espectrómetros de muones, e información de trazas proveniente de ID; los cuales no están

disponibles en $L1$. La reconstrucción en HLT puede ser ejecutada tanto en el marco de los RoI identificados en $L1$ como en el detector completo (*full-scan*). Para reducir el tiempo de procesamiento, la mayoría de los triggers del HLT siguen un esquema de dos etapas, una primera de rápida reconstrucción para rechazar la mayoría de los eventos, y una más lenta de reconstrucción precisa de los eventos restantes. Como resultado se cuenta con candidatos a una cierta partícula, es decir clusters y trazas a las que se les asigna una categoría de partícula.

Capítulo 3

Fotones en ATLAS

El programa de física del experimento ATLAS contempla un amplio conjunto de mediciones de precisión del SM, de propiedades del bosón de Higgs descubierto en 2012 y búsquedas de nueva física sin precedentes en el marco del LHC proveyendo colisiones de protones a 13 TeV. Gran parte de estas mediciones y búsquedas se realizan con eventos con fotones en el estado final.

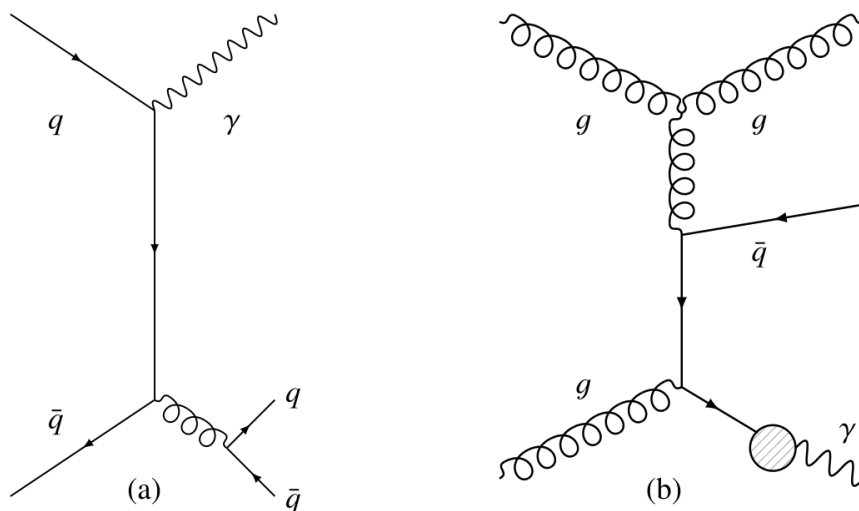


Figura 3.1: Ejemplo de diagramas de producción de eventos $\gamma + j + j$ vía (a) producción directa y (b) proceso de fragmentación de fotones.

Por ejemplo, la producción de fotones directos asociados con al menos un jet de partículas [9] o en asociación con dos jets [10] producidos como en los diagramas de la figura 3.1 en colisiones protón-protón (pp) proveen excelentes escenarios para estudiar y probar QCD

perturbativo dentro del SM . En colisiones pp , fotones no originados en forma secundaria por decaimiento de hadrones son considerados fotones directos.

Búsquedas más allá del MS en ATLAS se realizan mediante una vasta serie de análisis como ser búsqueda de resonancias de alta masa en estados finales con jets y fotones [11], o búsquedas de resonancias de partículas neutras masivas decayendo a un fotón y un bosón Z , W o Higgs, éstos a su vez decayendo hadrónicamente [12] o búsquedas como las de la referencia [13] en la que la identificación de fotones altamente colimados (o “jet de fotones”) en el estado final pueden ser indicio de nuevos bosones de gauge escalares, pseudo-escalares o vectoriales que sólo pueden decaer en fotones.

Fotones en el estado final están presentes también en muchísimos procesos físicos de interacciones en escenarios de Súper Simetría (SUSY) como por ejemplo en procesos representados en los diagramas de la figura 3.2 [14]. En estos escenarios, el estado final está compuesto por fotones, jets proveniente de la hadronización quarks y las partículas SUSY mas livianas que al no ser detectadas dejan en el detector un desbalance en la energía transversa del evento.

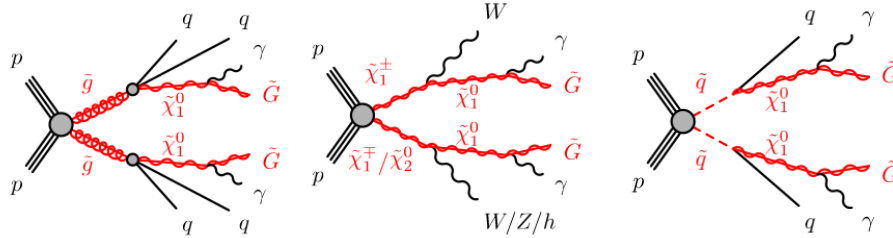


Figura 3.2: Procesos típicos de producción y decaimiento de (izquierda) gluinos (centro) neutrilinos y (derecha) squarks.

El bosón de Higgs fue observado en 2012 por ambas colaboraciones ATLAS [15] y CMS [16] en su canal de decaimiento a dos fotones (ver figura 3.3) y dos bosones Z decayendo a cuatro leptones. Desde ese histórico descubrimiento gran parte del programa de física de ATLAS está abocado a las mediciones de sus propiedades como ser su masa [17] o mecanismos de producción y acoplamientos [18, 19] en su canal de decaimiento a dos fotones.

De modo que la reconstrucción e identificación de fotones en el experimento ATLAS del LHC, tanto en el sistema de Trigger **online** como en el análisis **offline**, cumple con un rol fundamental en el programa de física del experimento. Es en este marco en el que se desarrolla este trabajo de diploma.

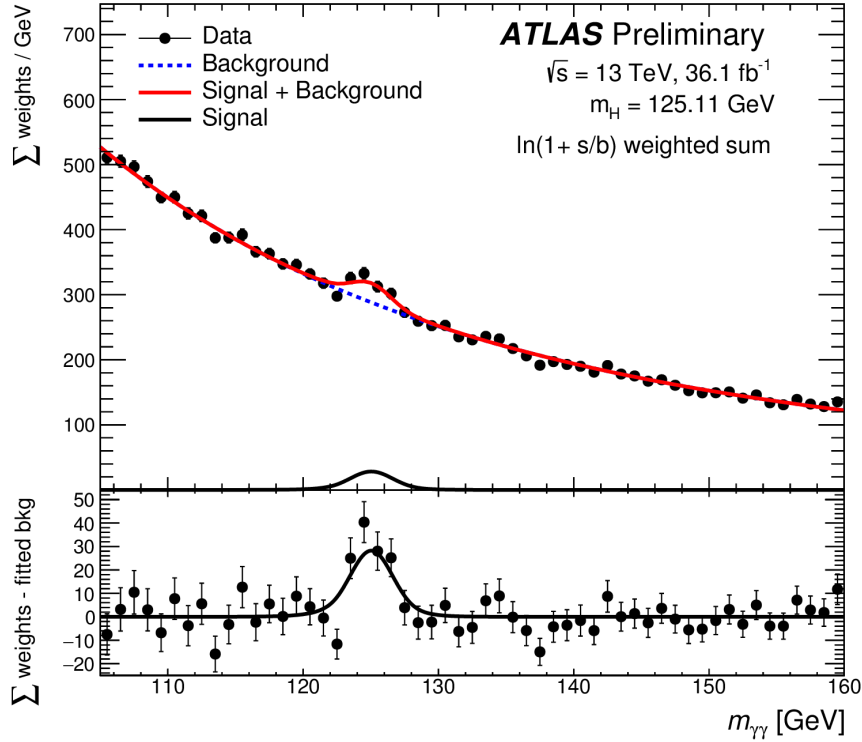


Figura 3.3: Espectro de masa invariante de cuatro leptones seleccionados en medidas de $H \rightarrow \gamma\gamma$. La línea roja es el ajuste a los datos donde el modelo ajustado incluye la forma esperada para el bosón de Higgs como también el fondo.

Los fotones en ATLAS son reconstruidos e identificados tanto *online* en el sistema de Trigger como *offline* una vez que el evento es aceptado por el Trigger y almacenado para posterior análisis. La diferencia entre ambas instancias radica en que la primera se realiza en tiempo real y está optimizada y simplificada para que pueda ser ejecutada dentro de las limitaciones de tiempo y de recursos del HLT, mientras que la segunda está optimizada para aumentar la *performance* y separación entre señal y fondo y prácticamente no tiene que cumplir con limitaciones temporales o de recursos computacionales. A pesar de la diferencia en los tiempos empleados para la reconstrucción e identificación, los algoritmos que realizan dicha tarea son similares tanto *online* como *offline*.

3.1. Reconstrucción

La reconstrucción *offline* de electrones y fotones utiliza clusters dinámicos de tamaño variable de los depósitos de energía medidos en celdas del *EMC* y *HCAL* topológicamente

conectadas, llamados *topo-clusters*; para recopilar energía de fotones de bremsstrahlung o de electrones originados en conversión de fotones. [4] [5] [8] Luego de aplicar correcciones a la posición inicial y calibración de energía a los topo-clusters, estos son asignados a trazas de ID reajustadas para tener en cuenta el bremsstrahlung y reconstruir los candidatos electrones. [20] [21]

Aquellos topo-clusters que no fueron asignados a ninguna traza o vértice de conversión son reconstruidos como candidatos-fotones. Posteriormente se recalibra la energía de los candidatos electrones y fotones que se utilizarán para análisis. En algunos casos, un objeto puede ser reconstruido como electrón y fotón a la vez. [22]

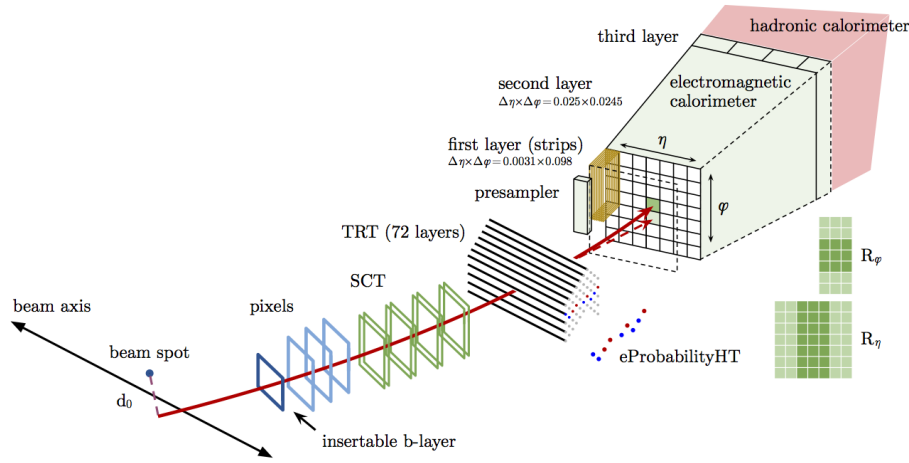


Figura 3.4: Esquema del proceso de reconstrucción de los objetos a través del detector.

La reconstrucción de fotones convertidos y no-convertidos (ver figura 3.4 y 3.5) se puede resumir de la siguiente manera: [4]

- Una ventana de tamaño 3×5 en unidades de $\Delta\eta \times \Delta\phi = 0,025 \times 0,0245$ (ver sección 1.1.3), correspondiente a la granularidad de la segunda capa del *EMC*, es utilizada para buscar “semillas” de clusters electromagnéticos como torres longitudinales cuya energía transversa total sea superior a $2,5 \text{ GeV}$. Las variables cinemáticas del cluster son reconstruidos utilizando una ventana extendida dependiendo de la posición del cluster en el calorímetro.
- Las trazas reconstruidas en el ID son emparejadas a las semillas de cluster. Estos sirven de entrada para la reconstrucción por vértices de conversión.
- Se realiza un arbitraje dependiente de las propiedades de las trazas y de los vértices de conversión asociados a un dado cluster electromagnético, para determinar si el objeto

es reconstruido como un electrón, como un fotón convertido o no-convertido, ó ambos como un objeto electrón y fotón en los siguientes casos ambiguos:

- * candidatos fotones no-convertidos: a partir de aquellos clusters a los cuales durante la reconstrucción de electrones no se les asignó ni un vértice de conversión ni una traza.
- * candidatos fotones convertidos: a partir de aquellos clusters asignados a un vértice de conversión.



Figura 3.5: Diagrama de flujo del proceso de reconstrucción offline de objetos en ATLAS.

Se define el verdadero-fotón-convertido como un fotón que atraviesa una conversión a un par electrón-positrón, dentro de una distancia $r < 80 \text{ cm}$ del IP . La eficiencia en la reconstrucción de dichos objetos es superior al 70 % para fotones simulados con verdadero $E_T > 20 \text{ GeV}$. La misma eficiencia aumenta para valores menores de μ ; siendo μ el número medio de interacciones por cruce de haces (“beam bunch crossing”), llegando a ser superior al 75 % para $\mu \sim 0$ y disminuyendo a 65 % cuando $\mu \sim 60$. La fracción de

verdaderos-fotones-convertidos en simulaciones que son erróneamente reconstruidos como fotones convertidos, es menor a 9 % para $\mu = 60$, y disminuye en función de μ hasta 1 % para $\mu < 24$.

A diferencia de la reconstrucción *offline*, en el HLT no se utiliza la información del detector de trazas para identificar conversiones.

Esto significa que en el trigger de ATLAS la reconstrucción de fotones se hace solamente con la información de los calorímetros bajo la hipótesis de fotones no convertidos. Además el proceso de reconstrucción cuenta con dos pasos, uno de reconstrucción rápida (*online*) en el *HLT* y un segundo paso de reconstrucción de precisión (*offline*). En el primero sólo se aplican selecciones que tengan buen poder de discriminación entre fondo y señal, como ser la E_T del cluster, y variables de forma de las lluvias (*shower shapes*) como R_{had} , E_{ratio} o R_η .

La otra gran diferencia con la reconstrucción *offline* es que en el trigger, los clusters de celdas del EMC que dan lugar a los candidatos fotones no son *topo-clusters* si no clusters rectangulares producidos por un algoritmo de *sliding-window*.

El algoritmo ***sliding-window*** opera sobre una grilla dónde las celdas están divididas en torres proyectivas. El algoritmo escanea la grilla y posiciona la ventana de tal forma que la energía transversa contenida dentro de la ventana sea un máximo local. Si dicho máximo se encuentra por encima de un dado umbral, se forma un cluster mediante la suma de celdas dentro de una dada ventana de aglomeración (*clustering*) rectangular. Para cada capa se le determina el baricentro a cada celda, y luego todas las celdas contenidas en una ventana fija alrededor del baricentro son incluidas en el cluster. Si bien el tamaño de la ventana de aglomeración está fijo, la posición del centro de la venta puede variar ligeramente en cada capa del calorímetro, dependiendo de cómo está distribuida la energía de las celdas.

3.2. Identificación

La identificación de candidatos-fotones en ATLAS se realiza mediante requisitos de selección rectangulares tanto *online* en el trigger como *offline*. Este requisito de selección está basado en variables que describen la forma y las propiedades asociadas a lluvias electromagnéticas (*shower shapes*) depositadas en los calorímetros las cuales proveen buena

separación entre fotones *prompt* y fotones falsos. Los fotones falsos pueden ser el resultado de fotones *non-prompt* originados por el decaimiento de hadrones neutros en jets ($\pi^0 \rightarrow \gamma\gamma$), o por QCD jets depositando una gran fracción de energía en el *EMC*. A estos requerimientos se les suma el aislamiento de los candidatos a otras partículas del evento. Por último dichas selecciones se optimizan por separado para aquellos candidatos fotones que se convirtieron en un par electrón-positrón antes de alcanzar el *EMC* (*fotones convertidos*), y aquellos candidatos que no son asociados a una conversión (*fotones no convertidos*).

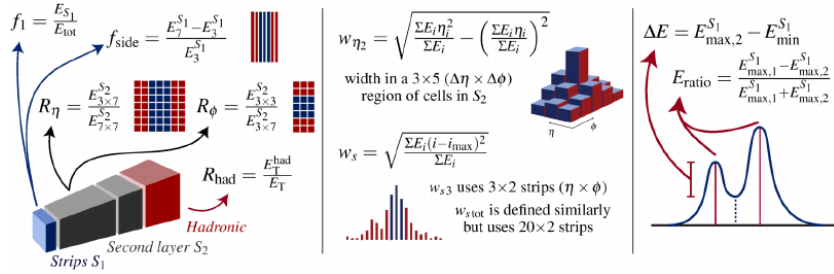


Figura 3.6: Representación esquemática de las variables de discriminación utilizadas para la identificación de fotones. $E_C^{S_N}$ identifica la energía recolectada en la N-ésima capa longitudinal del *EMC* en un cluster. Donde C denota las propiedades del cluster, identificando el número y/o propiedades de las celdas seleccionadas. E_i es la energía en la i-ésima celda, η_i la pseudorapidez del centro de la celda.

Para fotones se definen dos puntos de trabajo de identificación (*identification Working Points*), “loose” y “tight”. La selección *loose* se basa en las formas de las lluvias *EM* de la segunda capa del *EMC* y en la energía depositada en el *HCAL*, ver figura 3.6. Estas cantidades caracterizan la evolución de las distribuciones laterales y longitudinales en el *EMC*, como también la fracción de lluvia perdida en *HCAL*. Además del criterio de selección *loose*, la selección *tight* utiliza información de la primera capa finamente segmentada del *EMC*; esta última optimizada por separado para fotones convertidos y no convertidos, con tal de incorporar el perfil más ancho del rastro lateral dejado por estos últimos. Para una colección de fotones irradiados de electrones y muones del decaimiento del Z y $E_T > 25 \text{ GeV}$, la eficiencia integrada sobre los datos de la selección *loose* (*tight*) para fotones no asignados a ninguna traza es 98,9 % (87,5 %), y para fotones asignados a vértices de conversión es 96,3 % (87,6 %). [23]

Los electrones “*prompt*” que ingresan a la región central del detector ($|\eta| < 2,47$) son seleccionados utilizando una identificación basada en la verosimilitud (likelihood-based),

Table 1: Discriminating variables used for *loose* and *tight* photon identification.

Category	Description	Name	<i>loose</i>	<i>tight</i>
Acceptance	$ \eta < 2.37$, with $1.37 \leq \eta < 1.52$ excluded	—	✓	✓
Hadronic leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $ \eta < 0.8$ or $ \eta > 1.52$)	R_{had_1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \eta < 1.37$)	R_{had}	✓	✓
EM middle layer	Ratio of the energy in $3 \times 7 \eta \times \phi$ cells over the energy in 7×7 cells centered around the photon cluster position	R_η	✓	✓
	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	w_{η_2}	✓	✓
EM strip layer	Ratio of the energy in $3 \times 3 \eta \times \phi$ cells over the energy of 3×7 cells centered around the photon cluster position	R_ϕ		✓
	Lateral shower width, $\sqrt{(\sum E_i (i - i_{max})^2)/(\sum E_i)}$, where i runs over all strips in a window of $3 \times 2 \eta \times \phi$ strips, and i_{max} is the index of the highest-energy strip calculated from three strips around the strip with maximum energy deposit	$w_{s\ 3}$		✓
	Total lateral shower width $\sqrt{(\sum E_i (i - i_{max})^2)/(\sum E_i)}$, where i runs over all strips in a window of $20 \times 2 \eta \times \phi$ strips, and i_{max} is the index of the highest-energy strip measured in the strip layer	$w_{s\ tot}$		✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	f_{side}		✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE_s		✓
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in the secondary maximum in the cluster to the sum of these energies	E_{ratio}		✓
	Ratio of the energy in the first layer to the to the total energy of the EM cluster	f_1		✓

Figura 3.7: Tabla con la definición de algunas de las variables que describen la distribución de las lluvias.

la cual explota rasgos característicos de los depósitos de energía en el *EMC* (distribuciones de la forma de lluvia longitudinales y laterales), calidad de trazas, asociación trazas a clusters, e identificación de partículas mediante el *TRT*. Las funciones densidad de probabilidad (pdfs) en el rango de 4,5 a 15 *GeV* son derivadas de eventos $J/\Phi \rightarrow ee$, y para $E_T > 15$ *GeV* de $Z \rightarrow ee$. El umbral de discriminación es ajustado linealmente como función del número de vértices reconstruidos para preservar un rechazo estable de electrones de fondo. Para los candidatos electrones se definen tres puntos de operación, correspondientes a un mayor umbral para el discriminante (*LikeliHood*), ellos son *LHloose*, *LHmedium* y *LHtight*.

Para calcular la eficiencia en la identificación ϵ_{ID} de fotones; se utilizan tres técnicas

basadas en datos. *Radiative Z decays*, la cual selecciona una muestra de fotones de baja energía provenientes del decaimiento radiativo del bosón Z ($Z \rightarrow \ell\ell\gamma$). Permite medidas de eficiencia en identificación ϵ_{ID} desde $E_T = 10 \text{ GeV}$, por debajo de las cuales los fotones no son reconstruidos, hasta $E_T \sim 100 \text{ GeV}$, a partir de los cuales no se cuenta con suficientes eventos. Las otras dos técnicas son *Electron extrapolation* e *Inclusive photons*.

3.2.1. Aislamiento de fotones

Para reducir el fondo de jets mal identificados y de decaimiento de hadrones livianos y pesados dentro de los jets, a los candidatos fotones y leptones se les requiere que estén aislados. Si por el contrario, se requiere que los candidatos fotones no estén aislados se consigue una muestra de falsos fotones a partir de datos. En ATLAS se define el aislamiento de fotones (*photon isolation*) basado en la energía transversa en un cono cuyo tamaño angular ΔR alrededor de la dirección del candidato-fotón. Esta energía transversa está caracterizada por dos cantidades, la aislación calorimétrica E_T^{iso} (*calorimeter isolation*) y la aislación de trazas p_T^{iso} (*track isolation*).

La aislación E_T^{iso} se calcula como la suma de energías transversas de los topo-clusters en los calorímetros en un cono alrededor del candidato; luego de sustraer evento por evento la energía depositada por los candidatos-fotones. Y la contribución de los eventos subyacentes y el *pile-up*.

La aislación p_T^{iso} se obtiene sumando el momento transverso de todas las trazas, excluyendo las trazas asociadas a fotones-convertidos, cumpliendo: $p_T > 1 \text{ GeV}$ y teniendo una distancia de mínimo acercamiento al vértice primario a lo largo del eje del haz tal que $|z_0 \sin \theta| < 3 \text{ mm}$.

Capítulo 4

Redes Neuronales

El término Red Neuronal Artificial (*Artificial Neural Network*) hace referencia a un algoritmo computacional basado en neuronas artificiales (en adelante simplemente neuronas), el nodo elemental del algoritmo, (ver figura 4.1). [24] [25] Una neurona es una función matemática definida:

$$y_j = \Phi\left(\sum_{i=1}^n w_{ij}x_i + b_j\right), \quad (4.1)$$

$$\mathbf{y} = \Phi(\mathbf{w} \cdot \mathbf{x} + \mathbf{b}), \quad (4.2)$$

dónde \mathbf{x} es un vector de valores reales de entrada o “señal” (*input*), \mathbf{y} un valor de salida, “respuesta” (*output*) o activación. \mathbf{w} es un vector de pesos y b una desviación; conocida como sesgo (*bias*). El rol de b es correr la umbral de decisión independientemente del valor de entrada, y en ocasiones se expresa contenido en \mathbf{w} imponiendo $b_j = w_{j0}$ con $x_0 = 1$. Finalmente Φ es una función, denominada función de transferencia (o función de activación); su rol es establecer un límite que caracterice la respuesta de la neurona. Un ejemplo típico de este objeto es el perceptrón implementado para realizar clasificación binaria; caracterizado por la función $\Phi = f$ como la función *Heaviside*:

$$f(\mathbf{x}) = \begin{cases} 1 & \text{si, } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ 0 & \text{si no,} \end{cases} \quad (4.3)$$

Por lo tanto, podemos representar al perceptrón de la siguiente manera:

Las ANN son sistemas de procesamiento computacional inspirados por la manera en que operan los sistemas nerviosos biológicos. Se conforman mediante la interconexión de

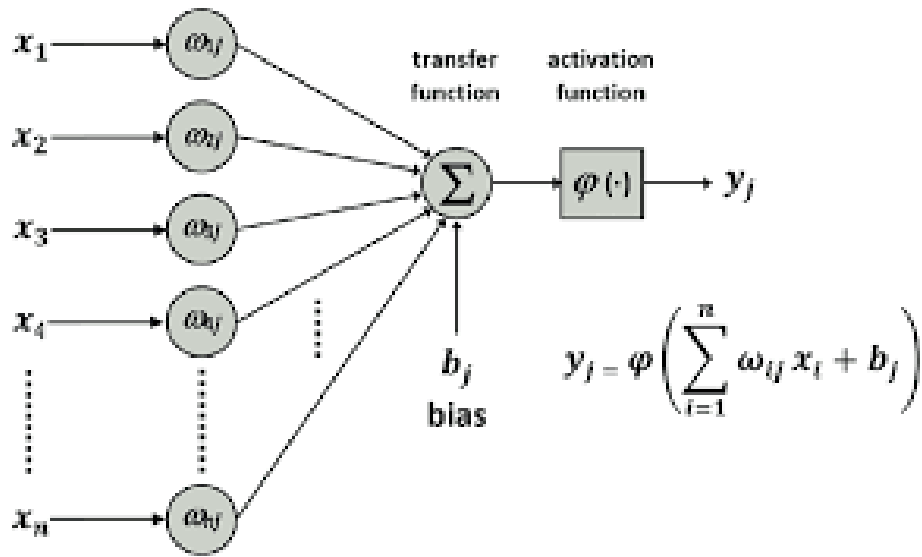


Figura 4.1: Esquema de una neurona artificial. Cada componente del vector de entrada posee un peso asociado, el nodo suma cada contribución y aplica la función de activación para generar una respuesta.

múltiples nodos computacionales, en analogía a la sinapsis, donde dichos nodos trabajan entrelazados de manera distribuida para aprender colectivamente a partir de los datos de entrada con el objetivo de optimizar el resultado de salida. Un algoritmo con uno o más neuronas conectadas o independientes se denomina modelo.

La red se forma a partir de la conexión de sucesivos nodos, dónde un nodo es una neurona en particular, con su respectiva función de activación, matriz de pesos y sesgo. Por lo tanto, las conexiones expresan el flujo de los datos, es decir, la dirección y los nodos por los que pasa la señal. De esta manera, para establecer la conexión entre dos nodos basta con asociar la respuesta del primero a la señal del segundo. Cada instancia sucesiva de la red da lugar a una “capa” caracterizada por contar con n nodos independientes entre sí; dónde entendemos por nodos independientes aquellos que no comparten conexiones. Como consecuencia, cada nodo de una capa está conectado únicamente a uno o más nodos anteriores y posteriores. La primera capa, aquella que consiste en el vector de entrada, se conoce como capa de entrada (*input layer*), la capa de salida (*output layer*) contiene los valores de respuesta final de la red, y las capas intermedias (si las hay), se denominan capas ocultas (*hidden layers*). A las redes que consisten en una única capa interna se las conoce como redes neuronales simples, si por el contrario contamos con más de una capa interna hablamos de redes neuronales profundas (*Deep Neural Networks*). Incorporar más capas complejiza considerablemente el modelo pero a su vez permite aumentar la eficiencia

de la red y abre el abanico a arquitecturas más diversas y específicas. Un primer ejemplo de una ANN profunda es la red de perceptrones múltiples ilustrada en la figura 4.2. [26]

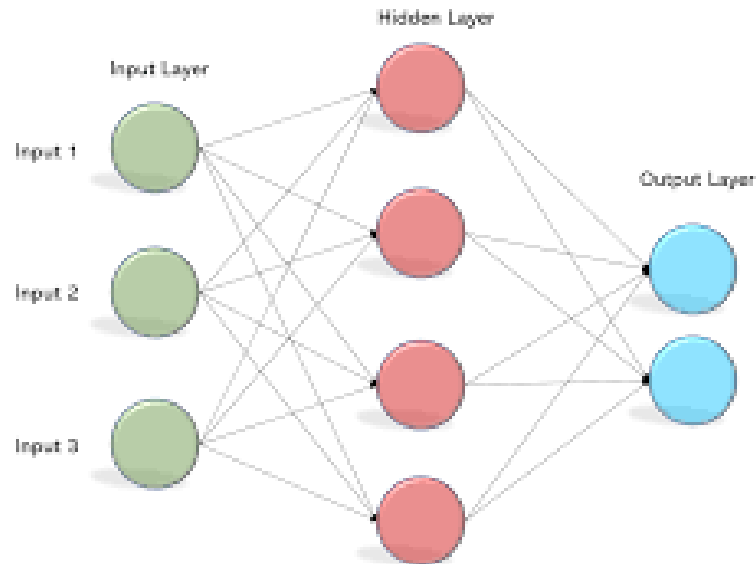


Figura 4.2: Esquema de una red neuronal artificial de múltiples perceptrones. Los nodos verdes representan cada uno un vector de entrada diferente, los rojos un perceptrón y los nodos celestes un resultados o salida. Las flechas indican las conexiones entre los nodos de la red, resulta explícito que hay 3 capas distintas: una de entrada, una de salida, y una “oculta” compuesta por perceptrones independientes entre sí.

La elección de la función de activación está sujeta a varios parámetros como: el tipo de señal a analizar, el tipo de respuesta buscada y el rol de la capa dentro de la arquitectura de red. La función *ReLU*, inspirada en el comportamiento biológico de neuronas, es utilizada para la toma de decisiones y por lo tanto resulta favorita en los problemas de clasificación como el estudiado en el marco de este trabajo. La misma se define:

$$f(\mathbf{x}) = \begin{cases} x & \text{para } x > 0 \\ 0 & \text{para } x \leq 0 . \end{cases} \quad (4.4)$$

4.1. Características del entrenamiento

El concepto de aprendizaje hace referencia a la labor que realizan las capas internas de la red, las cuales toman decisiones de capas anteriores y pesan cómo un cambio estocástico

mejora o empeora el resultado final. [26] En la actualidad se agrupan a las distintas arquitecturas dentro de alguno de los siguientes dos paradigmas de aprendizaje: Aprendizaje Supervisado (*Supervised learning*) y Aprendizaje No-supervisado (*Unsupervised learning*).

El aprendizaje supervisado se caracteriza por requerir de datos de entrada etiquetados que actúen como “objetivos” para el entrenamiento de la red. Para cada muestra de entrenamiento habrá un conjunto de valores (probablemente vectores) de entrada asociados a un designado valor de salida. El objetivo de esta forma de entrenamiento es reducir el error global de clasificación del modelo, mediante el calculo correcto del valor de salida. Por otro lado el aprendizaje no supervisado carece de etiquetas, y se basa en la capacidad de la red de reducir o no la función de coste/pérdida asociada. A continuación detallaremos sólo el paradigma de aprendizaje supervisado, puesto que contiene la arquitectura de *CNN*.

Ya sea una red simple o profunda, el proceso de entrenamiento es esencialmente idéntico, consiste en un algoritmo recursivo dividido en los siguientes pasos. El primer paso del algoritmo consiste en evaluar la red en sentido positivo o propagación delantera (*feed-forward*), es decir inicializar la matriz de pesos aleatoriamente, y procesar la señal desde la capa de entrada hasta obtener un resultado final en la capa de respuesta. El segundo paso es evaluar la función de pérdida entre los resultados del 1^{er} paso frente al resultado esperado. Es precisamente aquí dónde entra el paradigma del aprendizaje supervisado; para calcular la función de pérdida resulta indispensable contar con la información del resultado deseado para un dado vector de entrada. A continuación, en el tercer paso, se aplica un reajuste de la matriz de pesos en el sentido negativo de la propagación de la señal, es decir que primero se corrigen los pesos de la capa de respuesta, luego la anterior a esta y así sucesivamente hasta llegar a la primer capa. Este proceso se conoce como propagación inversa (*Backpropagation*), y ocupa un lugar fundamental en el proceso de aprendizaje de la red. La ejecución de estos primeros tres pasos es lo que se conoce como una **época** de entrenamiento. Finalmente, el cuarto paso consiste en repetir la propagación delantera con la nueva matriz de pesos hasta lograr minimizar la función de pérdida por debajo de la cota deseada, para ello se suelen implementar entrenamiento de varias épocas. [27]

4.1.1. Función de pérdida

En el paradigma del aprendizaje supervisado, la función de pérdida calcula qué tan apartado del valor esperado es la predicción realizada por la red (es decir el resultado), para cada época del entrenamiento. Por lo tanto, la elección de la función de pérdida está sujeta a varias condiciones, como por ejemplo el tipo de datos de entrada, la magnitud que se emplea para comparar los resultados con valores esperados (la métrica), o el tipo de problema que se está atacando, entre otros. Dos funciones de pérdida populares son la suma cuadrática de errores (*sum-of-squares*), dada por:

$$\epsilon(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N (\mathbf{y}(\mathbf{x}_n, \mathbf{w}) - t_n)^2 \quad (4.5)$$

y la entropía cruzada (*cross-entropy*)

$$\epsilon(\mathbf{w}) = - \sum_{n=1}^N t_n \ln(y_n) + (1 - t_n) \ln(1 - y_n) \quad (4.6)$$

donde t_n el n-ésimo valor esperado para la n-ésima señal.

4.2. Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales (*Convolutional Neural Network*) conforman un subgrupo dentro de las ya mencionadas ANN de aprendizaje supervisado, en tanto que constan de neuronas asociadas y distribuidas que se auto-optimizan mediante entrenamiento. [26] Se utilizan principalmente para resolver problemas de reconocimiento de patrones orientados a imágenes. Cada neurona recibe una entrada, realiza una operación y devuelve como salida una medida de probabilidad de pertenecer a una clase. Donde por clase entendemos a la categoría que caracteriza los posibles resultados de la red, si por ejemplo deseamos discriminar imágenes que contengan perros de gatos las clases serán “perro” y “gato”. Es posible obtener una idea del tamaño del modelo, estimando la cantidad de elementos que conforman la matriz de pesos, para ello basta con multiplicar:

$$Size = N^{in} \times Im_h \times Im_w \times Im_{ch}, \quad (4.7)$$

dónde N^{in} es el número e imágenes de entrada, Im_h la altura de una imagen, Im_w el ancho de esta, e Im_{ch} la cantidad de canales que conforman la imagen (1 si es monocromática, 3 para RGB, etc).

Cuándo los resultados en la métrica de entrenamiento son superiores a los resultados de la misma métrica en la evaluación, hablamos de sobre-ajuste (*Overfitting*). En este contexto se entiende que la NN desarrolló la capacidad de memorizar parte del conjunto de entrenamiento en lugar de aprender a generalizar a partir de las características globales del conjunto. Como consecuencia disminuye la capacidad de predicción de la red.

La arquitectura de *CNN* comprende generalmente tres tipos de capas:

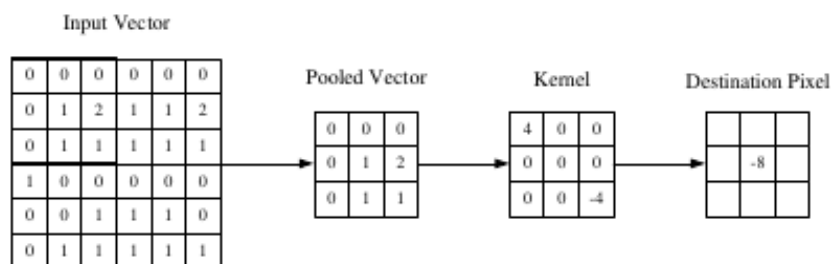


Figura 4.3: Ilustración del proceso de convolución. El elemento central del núcleo se posiciona sobre el vector de entrada, dónde luego se aplica la convolución y se remplaza por una suma pesada de sí mismo y de los pixels vecinos.

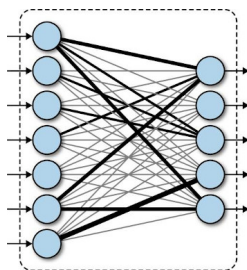


Figura 4.4: Esquema de una capa densa. El grosor y la tonalidad de las líneas hace referencia al peso de cada conexión.

- **capas convolucionales:** el elemento central de esta capa lo conforman los núcleos o filtros (*kernels*); lo cuales usualmente tienen dimensiones espaciales pequeñas, pero se distribuyen a lo largo de toda la profundidad de la entrada. Segmentan el volumen de entrada en regiones locales caracterizadas por un “paso”, y a cada región le computan el producto escalar con cada filtro para generar un mapa de activación 2D (*activation map*) por filtro. Luego al volumen de mapas de activación creado se aplica una función de transferencia, con el fin de que la red aprenda a activarse cuando detecta una característica específica en una dada posición espacial del volumen de entrada. La

figura 4.3 muestra un diagrama esquemático del proceso de convolución en una capa convolucional.

- **capas de agrupamiento** (*pooling layer*): el objetivo de esta capa es reducir gradualmente la dimensionalidad de la representación (de la entrada), y por ende reducir el número de parámetros y la complejidad computacional del modelo. Opera sobre cada mapa de activación de la entrada reescalando su dimensión mediante alguna función de agrupamiento, comúnmente se utilizan las funciones máximo o promedio. Su dinámica está sujeta al “paso” (*step/ stride*) asignado, el cuál determina las dimensiones del área donde aplicar la función de agrupamiento.
- **capas completamente conectadas o densas** (*Fully Connected layer*): Se caracteriza por contener neuronas que están directamente conectadas a las neuronas de las dos capas adyacentes (anterior y posterior), sin estar mutuamente conectadas, como se ilustra en la figura 4.4. Resultan particularmente útiles para realizar tareas de clasificación, y generalmente se utiliza la función de activación ReLu entre capas para mejorar el rendimiento.

La superposición secuencial de estas capas da lugar a una arquitectura *CNN*.

Capítulo 5

Identificación de fotones con CNNs

5.1. Almacenamiento de eventos

Como se mencionó previamente en la sección 2 los datos de los eventos que no fueron descartados por el *trigger* son almacenados en un formato llamado ***Raw Objects Data***, tanto los *trigger objects* como los candidatos partícula. Después de la reconstrucción e identificación offline, mencionada en la sección 3, los objetos físicos son almacenados en contenedores (como objetos de clases **C++**) según el tipo de partícula que contienen. Es decir que si se identifica un fotón, los datos como la energía y las coordenadas (η, ϕ) y los criterios de selección serán almacenados en el contenedor de fotones. Dichos contenedores son almacenados en archivos, cuyo formato puede ser utilizado por la plataforma **ROOT**. El formato de los archivos mencionados se conoce como ***Event Summary Data***, del cual se obtiene una versión de tamaño reducido (100 kB/evento) denominado ***Analysis Objects Data***.

Con el objetivo de entrenar la red neuronal con representaciones precisas de los *topo-clusters*, fue necesario escribir un software de análisis implementado como un paquete de **ATHENA**, llamado **CellDumpingPackage**. [28]. Este paquete extrae de los eventos los candidatos partícula, las cantidades asociadas relevantes a estos, aplica la selección del decaimiento radiativo del bosón Z (detallada a continuación en la sección 5.2.1), y obtiene las celdas que conforman los *topo-clusters*. Esta información, necesaria para reconstruir los *topo clusters* dejados por las partículas en el calorímetro, es almacenada en un archivo **ROOT**. [3] [29] [30] [31] [32] [33]

Para cumplir con el objetivo de comparar la eficiencia en identificación de fotones entre métodos basados en datos, se emplearon técnicas *data driven* utilizadas por *ATLAS* para definir muestras de datos enriquecidas tanto con señal y como con fondo. [4] Una vez seleccionada la muestra es posible entrenar la *CNN*, clasificar los eventos y calcular la eficiencia en la identificación de una u otra categoría comparando los resultados obtenidos mediante el método actual, basado en cortes rectangulares en las *shower shapes* (descrito en la sección 3.1) y otro basado en *CNNs* desarrollado en el contexto de esta tesis. En este análisis se utilizaron dos tipos de archivos: AODs y DAODs. Los AODs almacenan los contenedores de un período de toma de datos y un rango de energías; mientras que DAODs son una submuestra (*Derivation*) de los AODs a la cual se le aplicó una pre selección de eventos. Para el análisis se utilizaron eventos recolectados en todo el 2018 por *ATLAS* para obtener eventos de señal, mientras que para obtener eventos de fondo utilizamos los datos de sólo un *Run* de *ATLAS*. Donde *Run* corresponde típicamente a la adquisición de datos durante un Fill del LHC, que dura aproximadamente 10 hrs.

Si bien es posible realizar un análisis utilizando todo el *EMC*, en este estudio nos restringimos a utilizar datos medidos en la región *barrel* del *EMC* puesto que las dimensiones en el arreglo de celdas son distintas entre *barrel* y *endcap*. Esto implica que la clasificación entre señal y fondo necesariamente tiene que definirse y optimizarse separadamente para *barrel* y para *endcap*. En este análisis, sólo utilizamos la señal depositada en la segunda capa (2^{nd} layer) del *EMC*, debido a que como se mencionó previamente, es en esta capa dónde los electrones y fotones con elevado E_T depositan la mayor cantidad de energía. Así mismo, uno de los objetivos finales de este estudio es probar la técnica en el *FastReco* del *HLT*, utilizando únicamente de los datos de la 2^{nd} layer.

De los archivos de datos AODs y DAODs obtenemos los datos de los eventos de la 2^{da} capa del *EMC*, haciendo uso del paquete [CellDumpingPackage](#). Dado un evento, obtenemos la cantidad de objetos fotones y/o electrones que contiene, y todas sus propiedades medidas en la reconstrucción de *ATLAS*, incluyendo la distribución de energía de las celdas del cluster originados por el fotón y el resultado de la identificación de fotones *tight*.

5.2. Definición y obtención de señal

Para llevar a cabo el análisis necesario para entrenar y clasificar eventos de señal, se utilizaron fotones reconstruidos del decaimiento radiativo del bosón Z (en adelante nos referiremos a él por el acrónimo *Zrad* de ***Z radiative decay***); en particular considerando el canal $Z \rightarrow e^+e^-$. Esto se debe a que es posible obtener una muestra de eventos con poco fondo, aplicando tan sólo unos pocos criterios de selección, lo cual implica contar con una muestra de señal poco contaminada. En el LHC, a partir de las colisiones de protones, puede producirse un bosón Z que a su vez decaiga en un par electrón positrón. Si bien la tasa de decaimiento a un par electrón positrón es del 3,3% [34] la producción de Z en el LHC es suficientemente copiosa para reconstruir en *ATLAS* miles de candidatos $Z \rightarrow e^+e^-$ por cada Run. En el caso en el que uno de los leptones producto del decaimiento irradie un fotón, este puede identificarse como una señal aislada en los datos de *ATLAS*. De modo que el decaimiento radiativo del Z provee una fuente de bajo fondo y alta pureza de fotones en los datos de *ATLAS*. Los procesos del Modelo Estándar que proveen un estado final con dos leptones y un fotón a partir de la producción del Z se muestran en la figura 5.1. Es por ello que en *ATLAS* se utiliza el proceso *Zrad* para la estimación de rendimiento (*performance*), mencionado en la sección 3.

5.2.1. Decaimiento radiativo del bosón Z

Para obtener los eventos de señal utilizamos únicamente los archivos DAOD-EGAM3, un subtipo de DAOD caracterizado por tener preselecciones de eventos que permiten contar con un conjunto de datos rico en candidatos *Zrad*. En concreto, la preselección aplicada a DAOD-EGAM3 para eficiencia en identificación de fotones y aislamiento es:

- electrones: par electrón positrón, identificación *LH medium*, $p_T > 9,5 \text{ GeV}$, masa invariante del par mayor a 40 GeV .
- fotones: $E_T > 9,5 \text{ GeV}$.

Por todo lo anterior, los eventos de señal serán aquellos que cumplan la selección *Zrad* que se detalla a continuación. [35] [36]

Se utilizaron eventos disparados por triggers de electrones, con el objetivo de contar con fotones no sesgados por el trigger. Estos triggers requieren o bien un candidato electrón

(e^\pm) aislado que pase la selección *LHtight* y con al menos 26 GeV de momento transversal, o dos candidatos electrones que pasen la selección *LHloose* con al menos 17 GeV de momento transversal cada uno. En la reconstrucción *offline* del evento del decaimiento radiativo del bosón Z se requiere la presencia de un candidato-fotón (γ) y un par leptónico de carga opuesta ($\ell\bar{\ell}$) de candidatos electrones (e^\pm). Por lo tanto, los requisitos que deben satisfacer

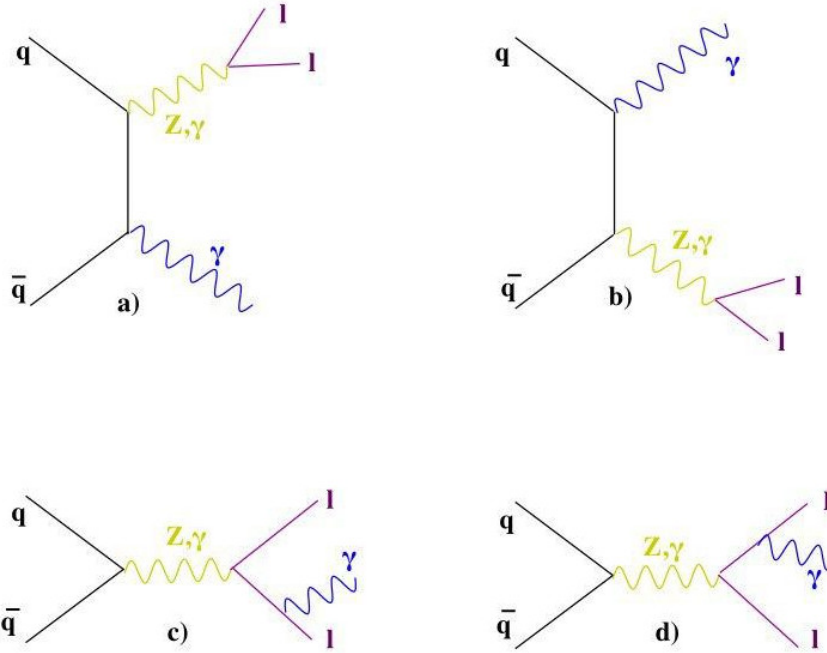


Figura 5.1: Diagramas de Feynman en el *SM* del decaimiento radiativo del Z . Los diagramas a) y b) son **Initial State Radiation**. Los diagramas c) y d) son **Final State Radiation**.

los candidatos fotones son:

1. tener un momento transversal $E_T^\gamma > 10 \text{ GeV}$,
2. y una pseudorapidez en el rango $|\eta^\gamma| < 1,37$ ó $1,52 \leq |\eta^\gamma| < 2,37$.
3. Al resultado nominal se le aplica el criterio de aislamiento *loose*.

Se evita aplicar cualquier otro criterio de selección al candidato-fotón para evitar introducir sesgos en la definición de la categoría de señal y en la medida de eficiencia. En cuanto a los requisitos que deben cumplir los candidatos-electrones se encuentran:

1. deben contar con $E_T^{e^\pm} > 10 \text{ GeV}$,

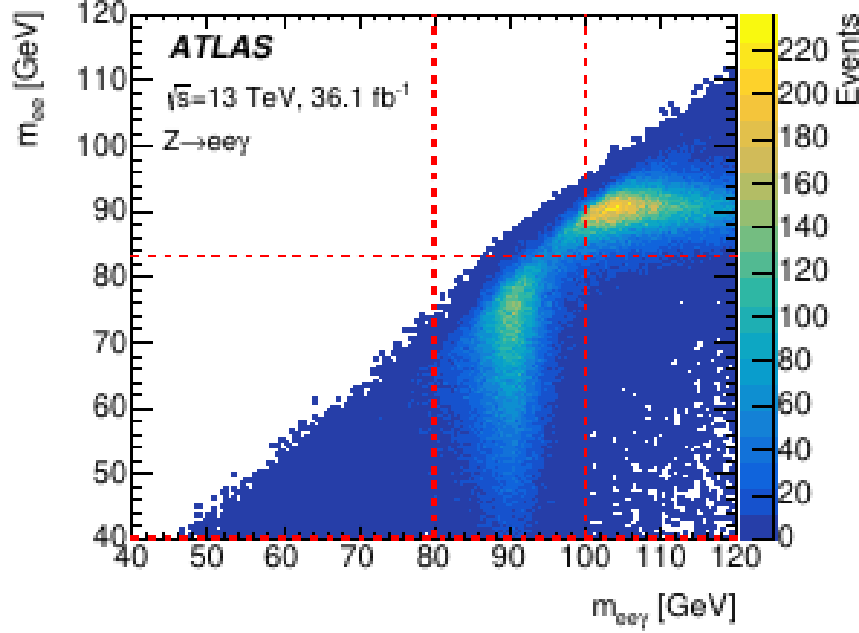


Figura 5.2: Distribución de $m_{ee\gamma}$ vs m_{ee} para eventos que satisfacen todos los criterios de selección $Z \rightarrow ee\gamma$ excepto los de $m_{ee\gamma}$ y m_{ee} . Las líneas punteadas horizontales y verticales muestran la selección utilizada para aislar los decaimientos radiativos, $40 < m_{ee} < 83 \text{ GeV}$ y $80 < m_{ee\gamma} < 100 \text{ GeV}$. En la región $m_{ee\gamma} > 100 \text{ GeV}$ y $m_{ee} \sim 90 \text{ GeV}$ se observan eventos provenientes del proceso $Z \rightarrow ee + \gamma$ con γ un fotón en *Initial State Radiation*; [5] los cuales son fácilmente descartados sabiendo que la masa del par leptónico es igual a la masa del Z .

2. $|\eta^{e^\pm}| < 1,37$ ó $1,52 \leq |\eta^{e^\pm}| < 2,37$.
3. Deben satisfacer el criterio de identificación *LHmedium*, el cuál se basa principalmente en información sobre la forma del cluster asociado en el EMC y de las emisiones radiativas de transición en el *TRT*.
4. Así mismo, los electrones debe cumplir con el criterio de aislación de trazas *loose*.

Sumados a los requisitos anteriores los eventos asociados al decaimiento radiativo Z deben satisfacer, $40 \text{ GeV} < m_{\ell\ell} < 83 \text{ GeV}$ y $80 \text{ GeV} < m_{\ell\ell\gamma} < 100 \text{ GeV}$, donde hemos llamado $m_{\ell\ell}$ a la masa invariante del sistema dileptónico y $m_{\ell\ell\gamma}$ a la masa invariante del par leptónico más el fotón. Estas selecciones tienen por objetivo aislar los decaimientos radiativos de eventos dónde los fotones se originaron a partir de radiación de los estados iniciales, ver figura 5.1. Las distribuciones de las cantidades $m_{\ell\ell}$ y $m_{\ell\ell\gamma}$ para el canal de

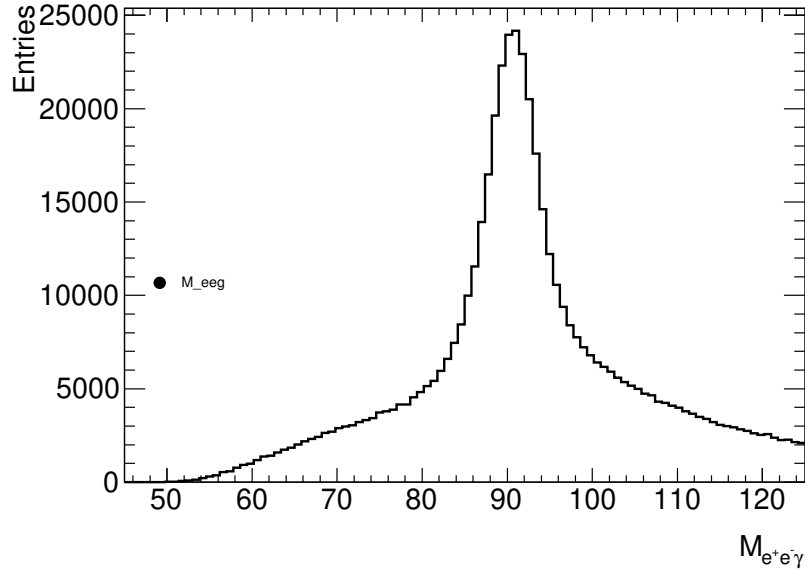


Figura 5.3: Figura masa invariante del fotón y el par electrón positrón, cumpliendo la selección $Zrad$ salvo por los cortes en masa para $M_{ee\gamma}$.

electrones se muestran en la figura 5.2. Finalmente resulta necesario que entre el fotón y el electrón más cercano exista una separación de $\Delta R > 0,4$. Este último criterio se aplica para evitar sesgos en las variables de las distribuciones de las lluvias *shower shapes* del fotón y en las variables de aislamiento.

Existe un pequeño fondo de contaminación debido a que el decaimiento $Z \rightarrow \ell\ell$ es acompañado por un jet que se puede ser identificado incorrectamente como un fotón, en particular cuando el fotón posee un momento transverso chico.

En la figura 5.3 se puede ver la distribución de masa invariante de la muestra de señal, calculada con los tres objetos electrón, positrón y fotón, cumpliendo los criterios de selección del evento $Zrad$ menos los cortes en la masa del sistema de tres objetos.

5.3. Definición y obtención de fondo

Los eventos de fondo, a diferencia de los de señal, se obtienen a partir de los archivos *AOD* dado que no poseen ninguna preselección de eventos. En la reconstrucción e identificación de fotones en ATLAS, el fondo más importante está dado por jets de partículas en los que un π^0 contiene la mayoría de la energía del jet. El decaimiento del π^0 da un par de

fotones muy cercanos que pueden dejar en el detector una señal muy parecida a la de un fotón real. Si bien todos aquellos eventos que no cumplen la selección $ZRad$ pertenecen al **fondo real** del evento, para evitar sesgar la muestra de fondo se aplica un criterio de selección inversa en el criterio de aislamiento de los candidatos fotones (muestra no sesgada) del evento para separar el fondo de la señal. De esta manera el **fondo** estará dado por los candidatos fotones que no estén aislados. Por lo tanto, para obtener una muestra enriquecida en **fondo** se utilizan criterios de aislamiento calorimétrico invertido, (recordar su definición en la sección 3.2.1) con el propósito de que repliquen el fondo real con exactitud. En el análisis de fondo invertimos E_T^{cone20} para los fotones, pidiendo

$$E_T^{cone20,\gamma} > p_T^\gamma \times 0,07, \quad (5.1)$$

donde *cone20* hace referencia a un cono alrededor del candidato fotón de tamaño $\Delta R = 0,2$

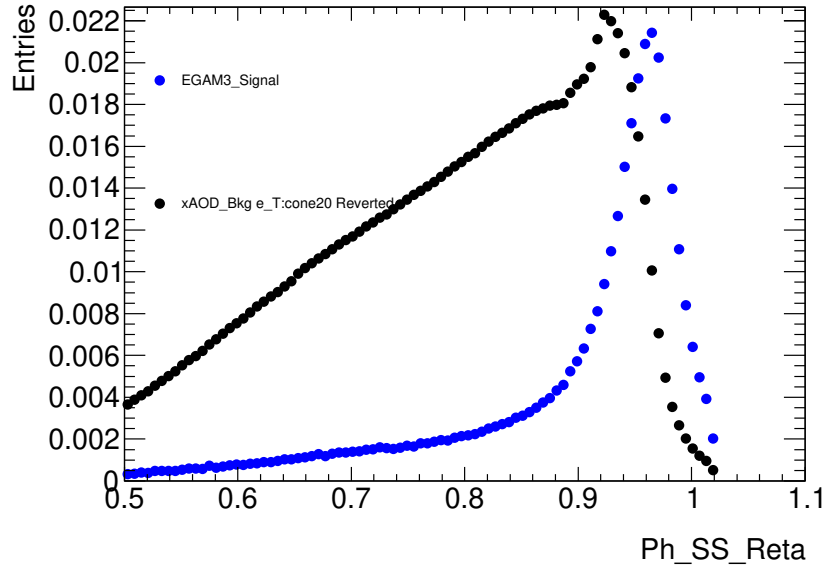


Figura 5.4: Histogramas normalizados de la *shower shape* R_η para eventos clasificados como señal (azul) y fondo (negro), mediante aislamiento invertido E_T^{cone20} .

A la hora de comparar los criterios resulta útil utilizar la distribución de R_η , dado que se espera que los fotones aislados reales depositen concentradamente su energía y por ende que $R_\eta \simeq 1$. La variable R_η es una de las *shower shape* mencionadas en 3.2, definida como

$$R_\eta = \frac{E_{3 \times 7}^{S2}}{E_{7 \times 7}^{S2}}, \quad (5.2)$$

y entendida como el perfil lateral de energía en η para la 2^{da} capa del *EMC*. Siendo $E_{i \times j}^{S2}$ la suma de las celdas en la segunda capa del calorímetro electromagnético contenidas en una ventana $i \times j$ (en $\Delta\eta \times \Delta\phi$). De hecho R_η es una de las variables utilizadas en la identificación de fotones *tight* que más poder de discriminación tiene entre fotones reales y fondo.

En la figura 5.4 se observa la distribución de R_η , comparando candidatos fotones en la muestra enriquecida de señal via *Zrad* (obtenida de un archivo *EGAM3*) y de fondo a partir del corte invertido de aislamiento E_T^{cone20} (obtenida de un archivo *AOD*). Se puede ver que los eventos de señal se distribuyen cerca de $R_\eta \rightarrow 1$, mientras que los eventos de fondo cubren las colas de la distribución. Tal como se observa en la (figura 5.4) ambas selecciones replican exitosamente el fondo para valores de $R_\eta < 0,93$, y la diferencia en la cantidad de eventos entre ambas es despreciable. Por estos motivos, y teniendo en cuenta que p_T^{cone20} utiliza información de trazas y no calorimétrica, optamos por utilizar este criterio para definir los eventos de **fondo**.

5.4. Creando las imágenes de los *topo-clusters*

Para realizar el análisis necesitamos recrear las imágenes de los *topo-clusters* de los fotones a partir de la distribución de energía depositada en la 2^{da} capa del *EMC*. Con la información almacenada por el *CellDumpingPackage*, para cada fotón de cada evento de nuestra muestra podemos acceder a la información del cluster que se utilizó en la reconstrucción que generó dicho fotón. Un cluster está formado por las celdas del calorímetro dónde la partícula depositó su energía, es decir que a cada celda la podemos entender como un vector almacenando:

$$celda = (\eta, \phi, E). \quad (5.3)$$

De esta manera, un cluster resulta un volumen discreto en el espacio (η, ϕ, E) .

A continuación detallamos el procedimiento para realizar el mapeo de un objeto cluster a una matriz, vale notar que este proceso es independiente de cualquier identificación y sólo requiere de la información de las celdas del objeto, anteriormente mencionadas. La matriz que debemos obtener ha de ser la proyección de la distribución de energías en la 2^{da} capa del *EMC*. Dado que las coordenadas $(\eta, \phi) \in \mathbb{R}$ y que los índices de una matriz son \mathbb{N} , es necesario dar una regla de transformación. Para ello, tomamos como referencia los anchos de las celdas de la segunda capa del calorímetro (ver figura 1.5), dónde establece

que $(\Delta\eta, \Delta\phi) = (0,025, 0,0245)$; siendo $\Delta\eta$ el ancho en la pseudorapidez, $\Delta\phi$ el ancho en el ángulo acimutal. Estos últimos serán implementados como unidades discretas o “pasos”. Luego tomamos el origen de cada coordenada en una arista de la celda. Sin embargo, también es posible tomar el origen en el centro de la celda.

En dichas condiciones el primer paso consiste en ubicar el cluster en la región del calorímetro donde fue medido. Por lo tanto calculamos cuál es la distancia del cluster al origen de coordenadas $(\eta_0, \phi_0) = (0,0)$, mediante las siguientes ecuaciones:

$$step_\eta = round(\frac{(\eta + \Delta_\eta/2) * sig(\eta)}{\Delta_\eta} + \delta_\eta), \quad (5.4)$$

$$step_\phi = round(\frac{(\phi + \Delta_\phi/2) * sig(\phi)}{\Delta_\phi} + \delta_\phi). \quad (5.5)$$

Dónde hemos llamado $(step_\eta, step_\phi)$ al número de pasos de la celda hasta el origen de coordenadas. Así mismo, $(\delta_\eta, \delta_\phi)$ son las desviaciones necesarias para imponer que el origen de la celda se encuentre en la posición deseada, en nuestro caso centrado en ϕ y en el arista izquierda de η . De esta manera si imponemos que el origen de la celda

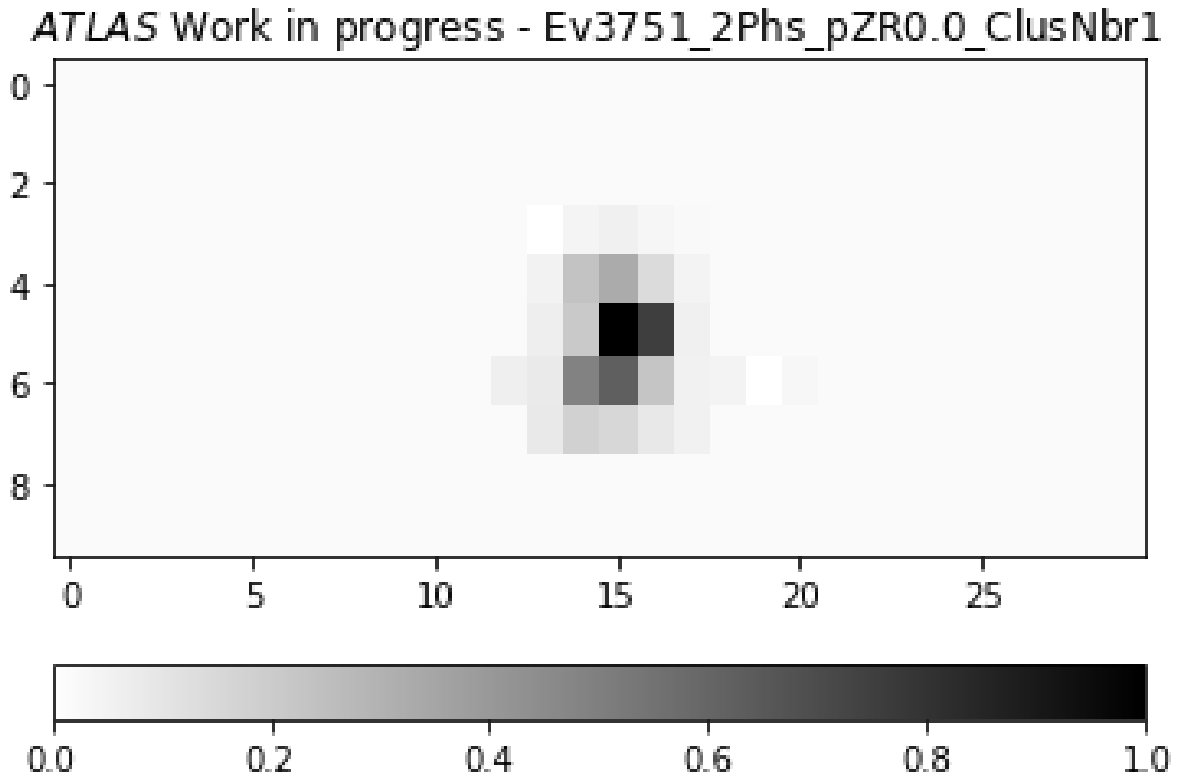


Figura 5.5: Imágen de un *topo-cluster* de **fondo**.

esté centrado basta con fijar $(\delta_\eta, \delta_\phi) = (0, 0)$, mientras que para se encuentre en la arista inferior izquierda $(\delta_\eta, \delta_\phi) = (-\Delta_\eta, -\Delta_\phi)$. Además la función *round* indica que se redondee su argumento al entero más cercano. Mediante este procedimiento nos construimos una matriz de 10×30 celdas centrada en el candidato fotón, que representa la distribución de energía de las celdas del topocluster que lo conforma. Finalmente se normaliza la distribución de energía de la matriz, con el fin de evitar introducir un sesgo en p_T a la hora de entrenar la red neuronal. Como resultado de este proceso obtenemos la imagen de un *topo-cluster*, ver figura 5.5.

5.5. Entrenando la CNN

El conjunto de datos de entrenamiento constó de treinta mil (30000) imágenes de *topo-clusters* de fotones de señal, y 30000 imágenes de fotones de fondo; siguiendo las características dadas anteriormente (ver sección 5.2). Dónde por fondo entendemos aquellos eventos que cumplen con la selección de aislamiento invertida de E_T^{cone20} , según la sección 5.3).

La implementación de la arquitectura se realizó utilizando **TensorFlow** y **Keras**, junto con librerías propias de esos paquetes. Definimos dos modelos de *CNN*, con las siguientes especificaciones:

- modelo 0: dos capas convolucionales, una capa de agrupamiento, dos capas densas y una de activación; dispuestas tal como se muestra en la figura 5.6.
- modelo 1: cuatro capas convolucionales, dos capas de agrupamiento, dos capas densas y una de activación, representada en la figura 5.7.

Para compilar los modelos se optó por medir la precisión, tomando como la función de error la entropía cruzada, y como optimizador el *RMSprop* con una tasa de entrenamiento de 0,0001; tal como evidencian las figuras 5.6 y 5.7. El optimizador *RMSprop* hace referencia al método empleado para minimizar la función $\epsilon(w)$. Es una variante del descenso por el gradiente con momento (*gradient decent with momentum*) cuya diferencia radica en la forma de calcular el gradiente, tornándolo más rápido. El poder de este optimizador radica en que varía la tasa de aprendizaje (*learning rate*), es decir la cantidad de pasos y la dirección en la que calcula el gradiente. Para todas las capas convolucionales y densas se asignó la función de activación *ReLU*.

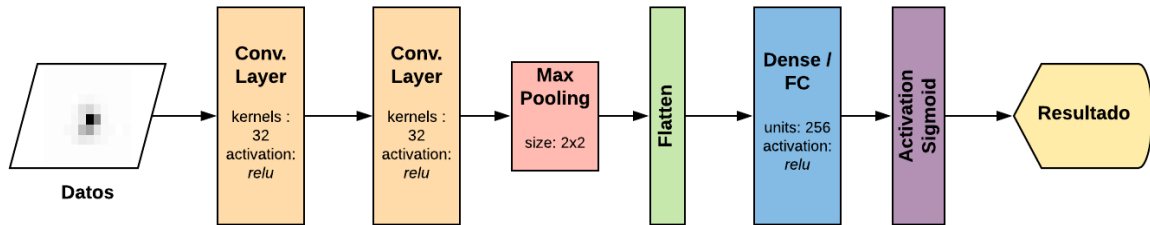


Figura 5.6: Esquema de la arquitectura del modelo 0.

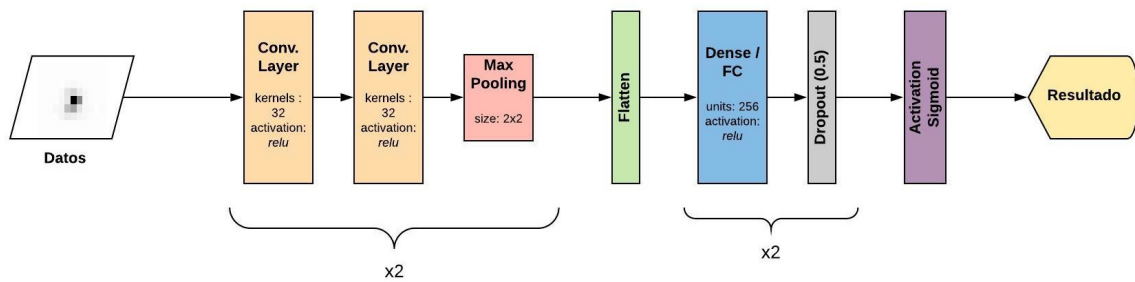
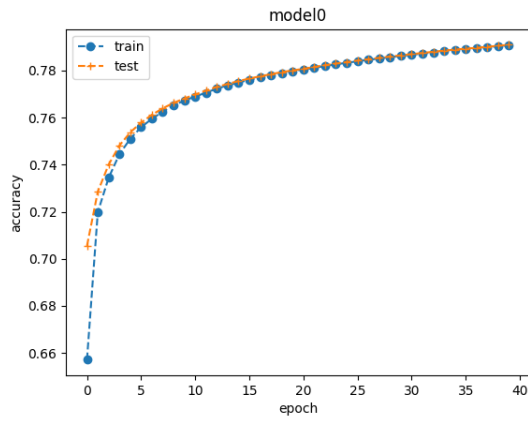
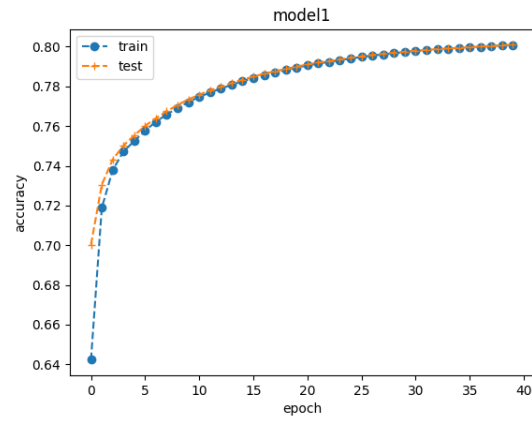


Figura 5.7: Esquema de la arquitectura del modelo 1, donde los corchetes indican que el conjunto de capas se repite en serie la cantidad de veces que figura.

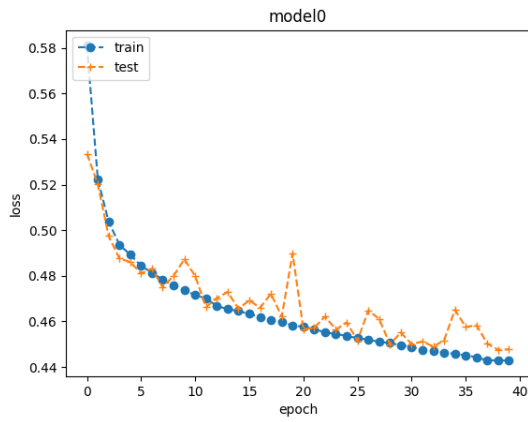
Cada modelo se entrenó durante 40 épocas, donde cada una utilizaba $N/40$ puntos de entrenamiento y de los cuales el 20 % se utilizaron para validar las predicciones en el cómputo de la función de pérdida. Los resultados del entrenamiento se resumen en las figuras 5.8a y 5.8c para el modelo 0, y en las figuras 5.8b y 5.8d para el modelo 1. La precisión (*accuracy*) se calcula tomando la fracción de etiquetas asignadas correctamente y dividiendo por el total de la muestra. [37] Es decir que dada una precisión del 70 % contaremos con un 30 % de probabilidad de incurrir en errores de Tipo *I* y/o Tipo *II*. La diferencia entre el cálculo de la precisión y de los errores (definido en la sección 4.1.1) radica en que la primera no tiene incidencia alguna en el entrenamiento, mientras que la última es utilizada para el ajuste de los pesos en la red. Dado que en ambos modelos se obtuvo una precisión predictiva marginalmente superior al 80 % esto impone una cota en el rendimiento de los modelos.



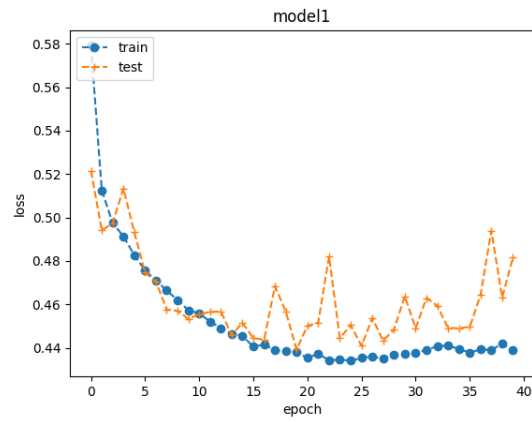
(a) modelo 0



(b) modelo 1



(c) modelo 0



(d) modelo 1

Figura 5.8: contraste de entrenamiento en función de las épocas de entrenamiento entre los dos modelos según la precisión (**arriba**) y según la función de pérdida *binary cross entropy* (**abajo**). En todos los casos mostramos para cada gráfico los cálculos tanto para el entrenamiento (*train* puntos azules) y la validación (*test* cruces naranjas).

Capítulo 6

Resultados

Una vez que contamos con los modelos entrenados y la herramienta de creación de imágenes a partir de archivos de ROOT, estamos en condiciones de comparar el rendimiento en identificación de fotones implementando las CNN.

Para evaluar el rendimiento de los métodos de identificación calculamos la eficiencia en identificación de señal y en el rechazo de fondo en función del momento transversal p_T y de la pseudorapidez η del candidato fotón, frente a la identificación *tight* por selección rectangular de cortes en las *shower shapes* expuesta en la sección 3. La eficiencia en identificación de señal se define

$$\epsilon_{analysis} = \frac{N(x)_{data}^{pass}}{N(x)_{data}^{total}}, \quad (6.1)$$

donde N_{data}^{total} y N_{data}^{pass} son el número total de eventos en la muestra completa, y en el subconjunto en el cual el fotón pasa los cortes del análisis, respectivamente. A su vez, x es el observable medido, en nuestro caso: η^γ y p_T^γ . Para calcular la eficiencia en la identificación de señal se tomó $N_{data}^{total} = zrad$, es decir todos los eventos de la muestra de que cumplen el criterio *zrad*. Mientras que para $N_{data}^{pass} = method$, donde *method* denota al método empleado para definir la selección que puede ser *tight* o alguno de los modelos de CNN. Por otro lado, para el cálculo de la eficiencia en rechazo de fondo se utilizó $N_{data}^{total} = reverse(E_T^{cone20})$, mientras que para el numerador se tomó nuevamente $N_{data}^{pass} = method$.

Para calcular la eficiencia de señal, utilizamos la muestra de fotones reconstruidos a partir de *zrad* descrito en la sección 5.2.1. El rechazo del fondo lo obtuvimos a partir del cálculo de la eficiencia utilizando la muestra de fotones falsos a partir de la inversión en el

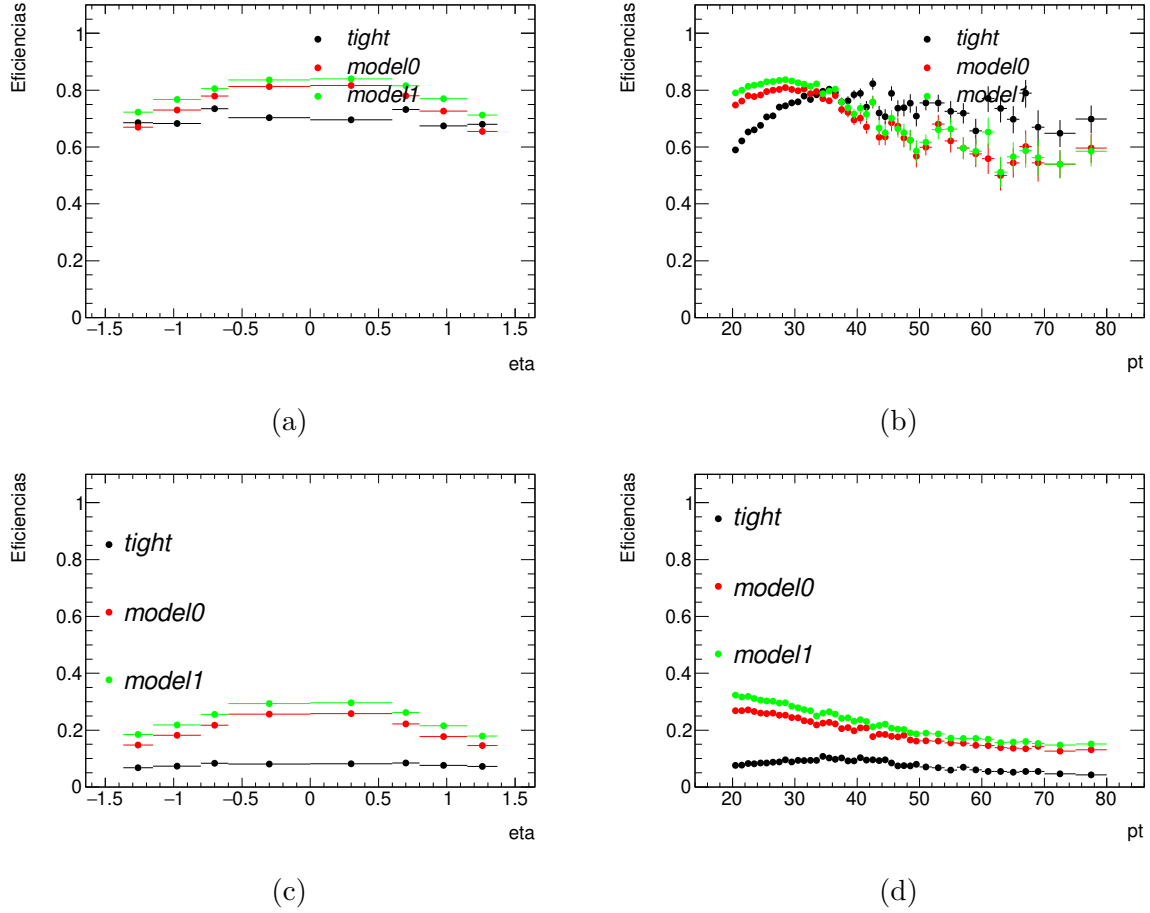


Figura 6.1: Comparación de eficiencias entre los métodos *tight*, *modelo 0* y *modelo 1* de CNN en eventos de señal(arriba) y de fondo (abajo). Las eficiencias se muestran en función de la pseudorapidez η (izquierda) y del momento transverso p_T (derecha) en GeV .

corte de aislamiento mencionado en la sección 3.2.1. Es por esto que utilizamos nuevamente un archivo *EGAM3* con alrededor de 3629000 eventos, y el análisis lo realizamos sobre los candidatos fotones únicamente. Una vez que contamos con una muestra óptima, el proceso de evaluación consiste en tomar el/los fotones del evento, obtener su valor de la selección *tight*, recrear la imagen de su *topo-cluster* asociado, y utilizar los modelos entrenados para predecir a partir de dicha imagen la categoría del candidato-fotón. Los resultados nos permiten calcular las eficiencias mediante la ecuación (6.1).

En la figura 6.1 se muestra la eficiencia en la identificación de fotones provenientes del decaimiento radiativo del Z (arriba) y en la identificación de fotones falsos (abajo), en función de η (izquierda) y p_T (derecha). En cada una de ellas se compara el resultado de la selección *tight*, la clasificación del *modelo 0* y del *modelo 1*. A partir de la figura 6.1a

se puede observar que la identificación de los tres métodos es comparable entre sí, aunque los modelos de *CNN* en el rango $|\eta| < 1$ tienen un mejor desempeño clasificando fotones reales como señal, llegando a marcar una diferencia del 10 %. Por el contrario, en la región $|\eta| > 1$ el *modelo 1* tiene mejor desempeño que los otros dos métodos, siendo este el más homogéneo de los tres alcanzando una eficiencia promedio del 75 %. Al observar la figura 6.1b el panorama está nuevamente separado en regiones. Para fotones con $p_T < 34 \text{ GeV}$ ambos modelos *CNN* ofrecen mejor identificación que *tight*. En el rango $34 \text{ GeV} \leq p_T < 40 \text{ GeV}$ los tres modelos presentan un comportamiento comparable, teniendo en cuenta las incertezas, mientras que a partir de 40 GeV el método *tight* muestra una mayor eficiencia de entre un 1 % y 3 %. Finalmente, se observa una tendencia global decreciente conforme aumenta p_T a partir de $p_T > 30 \text{ GeV}$.

En cuanto a la evaluación del rechazo del fondo, en general, tanto para η (ver figura 6.1c) como para p_T (ver figura 6.1d), los modelos *CNN* *modelo 0* y *modelo 1* muestran una mayor eficiencia. Esto significa que clasifican como fotones más eventos de fondo que el método *tight*, denotando un peor desempeño. En la figura 6.1c podemos apreciar que el *modelo 0* se desempeña mejor que el *modelo 1* en todo el rango de η , marcando una diferencia $\sim 5 \%$ mejor que el *modelo 1*. La identificación por *tight* de fotones falsos en función de η es notablemente superior en el rango $|\eta| < 1$ llegando a rendir un $\sim 15 \%$ mejor que ambos modelos *CNN*. Para el observable p_T (figura 6.1d) se observa una tendencia decreciente de la eficiencia conforme aumenta el p_T , en tanto que la identificación por *tight* supera a la de los modelos *CNN* por un $\sim 5 \%$ en todo el rango del observable. Además se puede ver que el *modelo 0* realiza un mejor trabajo en el rechazo del fondo que el *modelo 1*, por un margen de entre $\sim (0,5, 1) \%$.

Capítulo 7

Conclusiones

El programa de física del experimento ATLAS del LHC es vastamente amplio y contempla, entre otros, mediciones de precisión de procesos del Modelo Estándar, búsquedas genéricas de física más allá del Modelo Estándar, mediciones de las propiedades del bosón de Higgs, búsquedas de evidencia de Supersimetría, búsquedas de dimensiones extras y búsquedas de materia oscura. Todas estas mediciones y búsquedas se realizan, entre otros, con eventos colectados por ATLAS con fotones en el estado final. De modo que una buena identificación de fotones y separación del fondo a partir de los datos de ATLAS es fundamental para el programa de física del experimento.

En este trabajo evaluamos el uso de Redes Neuronales Convolucionales para la identificación de fotones a partir de datos colectados por ATLAS y lo comparamos con la identificación *tight* actualmente utilizada a partir de cortes rectangulares en las distribuciones de la forma de las lluvias *shower shapes* calculadas a partir de la señal depositada en el calorímetro electromagnético.

Para ello desarrollamos un paquete de software que construye una imagen a partir de la energía depositada en las celdas de la segunda capa del calorímetro electromagnético de los topo-clusters de los candidatos fotones en la zona central de pseudorapidez $|\eta|$ del detector (barril). A partir de los datos de ATLAS, utilizamos el proceso de decaimiento radiativo del bosón Z para obtener una muestra no sesgada de señal y un corte de inversión de aislamiento para obtener una muestra enriquecida de fondo.

Utilizamos estas muestras para alimentar y entrenar un conjunto de dos modelos de redes neuronales convolucionales con el objetivo de que aprenda a clasificar fotones reales

de falsos fotones (fondo) a partir de las imágenes construidas de la energía depositada en las celdas de los topo-clusters. Utilizando 60.000 eventos de ATLAS colectados en 2018, obtuvimos un 80 % de precisión en el entrenamiento de la red.

Con ambos modelos entrenados calculamos la eficiencia en la identificación de fotones y en el rechazo de fondo de éstos, y la contrastamos con las calculadas utilizando la identificación actual de cortes rectangulares.

Como resultado de esta comparación observamos que la eficiencia de identificación de fotones reales por parte de estas redes neuronales es más homogénea en función del momento transversal que la identificación *tight* con cortes rectangulares.

Por otro lado observamos que tanto en función de η como de p_T , la identificación *tight* provee un mejor rechazo de fondo, marcando una diferencia entre los métodos que ronda valores superiores al 15 %. Mientras que en la identificación de señal, ambos modelos de *CNN* presentan mayor eficiencia en la selección de fotones reales. En particular, comparado con la identificación *tight* la eficiencia de identificación de las *CNN* en función de η es superior por un margen de entre un 10 % y un 15 % dependiendo de la región. Finalmente en función de p_T la identificación de señal resultó más homogénea destacándose por un pequeño margen el *modelo 1* frente al resto para $p_T < 30 \text{ GeV}$, mientras que para $p_T > 45 \text{ GeV}$ se destaca levemente *tight*.

Como resultado de este trabajo hemos desarrollado un conjunto completo de herramientas de análisis que permite implementar una identificación de fotones sin utilizar cortes rectangulares, que dado que se realiza sólo a partir de la energía depositada en la segunda capa del calorímetro electromagnético es posible de utilizar en la etapa de reconstrucción rápida del HLT.

Bibliografía

- [1] L. Evans and P. Bryant, Journal of instrumentation **3** (2008) S08001.
- [2] *The Large Hadron Collider*,
<https://home.cern/science/accelerators/large-hadron-collider>. Accessed:
2020-02-03.
- [3] G. E. Orellana and H. P. Wahlberg, *Estudios de fondos de electrones reconstruidos como fotones en búsquedas de Supersimetría con el detector ATLAS*, 2017.
- [4] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, D. K. Abhayasinghe, S. H. Abidi, O. AbouZeid, N. Abraham, H. Abramowicz, et al., The European Physical Journal C **79** (2019) 205.
- [5] G. Aad, A. Kupco, S. Webb, T. Dreyer, Y. Wang, K. Jakobs, B. Le, M. Spousta, M. Cobal, P. Wang, et al., *Performance of electron and photon triggers in ATLAS during LHC Run 2*, tech. rep., ATLAS-TRIG-2018-05-003, 2019.
- [6] G. Aad, J. Butterworth, J. Thion, U. Bratzler, P. Ratoff, R. Nickerson, J. Seixas, I. Grabowska-Bold, F. Meisel, S. Lokwitz, et al., Jinst **3** (2008) S08003.
- [7] A. Collaboration, Eur. Phys. J. C **77** (2017) 317.
- [8] M. Aaboud, A. Kupco, S. Webb, T. Dreyer, Y. Wang, K. Jakobs, M. Spousta, M. Cobal, P. Wang, S. Schmitt, et al., JINST **14** (2018) P03017.
- [9] ATLAS Collaboration, M. Aaboud et al., *Phys. Lett.* **B780** (2018) 578–602,
[arXiv:1801.00112](https://arxiv.org/abs/1801.00112) [[hep-ex](#)].
- [10] ATLAS Collaboration, G. Aad et al., [arXiv:1912.09866](https://arxiv.org/abs/1912.09866) [[hep-ex](#)].
- [11] ATLAS Collaboration, M. Aaboud et al., *Eur. Phys. J.* **C78** (2018) 102,
[arXiv:1709.10440](https://arxiv.org/abs/1709.10440) [[hep-ex](#)].

-
- [12] ATLAS Collaboration, M. Aaboud et al., *Phys. Rev.* **D98** (2018) 032015, [arXiv:1805.01908 \[hep-ex\]](#).
- [13] ATLAS Collaboration, M. Aaboud et al., *Phys. Rev.* **D99** (2019) 012008, [arXiv:1808.10515 \[hep-ex\]](#).
- [14] ATLAS Collaboration, M. Aaboud et al., *Phys. Rev.* **D97** (2018) 092006, [arXiv:1802.03158 \[hep-ex\]](#).
- [15] ATLAS Collaboration, G. Aad et al., *Phys. Lett.* **B716** (2012) 1–29, [arXiv:1207.7214 \[hep-ex\]](#).
- [16] CMS Collaboration, S. Chatrchyan et al., *Phys. Lett.* **B716** (2012) 30–61, [arXiv:1207.7235 \[hep-ex\]](#).
- [17] ATLAS Collaboration, M. Aaboud et al., *Phys. Lett.* **B784** (2018) 345–366, [arXiv:1806.00242 \[hep-ex\]](#).
- [18] ATLAS Collaboration, M. Aaboud et al., *Phys. Lett.* **B784** (2018) 173–191, [arXiv:1806.00425 \[hep-ex\]](#).
- [19] ATLAS Collaboration Collaboration,, *Combined measurements of Higgs boson production and decay in the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels using $\sqrt{s} = 13$ TeV pp collision data collected with the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2017-047, CERN, Geneva, Jul, 2017. <http://cds.cern.ch/record/2273854>.
- [20] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, R. Aben, M. Abolins, O. AbouZeid, H. Abramowicz, H. Abreu, R. Abreu, et al., *European Physical Journal C* **77** (2017).
- [21] A. collaboration et al., *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047. 2012.
- [22] J. Saxon.
- [23] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, O. AbouZeid, N. Abraham, H. Abramowicz, H. Abreu, et al., *The European Physical Journal C* **77** (2017) 195.

-
- [24] I. Stephen, IEEE Transactions on neural networks **50** (1990) 179.
- [25] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [26] K. O'Shea and R. Nash, arXiv preprint arXiv:1511.08458 (2015).
- [27] F. Chollet, *Aprendizaje profundo con Python y Keras: la guía práctica del desarrollador de la biblioteca Keras*. 2018.
- [28] S. J. Noacco Rosende and M. F., *CellDumpingPackage*, <https://gitlab.cern.ch/snoaccor/CellDumpingPackage>, 2019.
- [29] ATLAS Collaboration Collaboration,, *ATLAS Computing: technical design report*. Technical Design Report ATLAS. CERN, Geneva, 2005. <https://cds.cern.ch/record/837738>.
- [30] for the ATLAS collaboration Collaboration, B. Lenzi, *The Physics Analysis Tools project for the ATLAS experiment*, Tech. Rep. ATL-SOFT-PROC-2009-006, CERN, Geneva, Oct, 2009. <https://cds.cern.ch/record/1214931>.
- [31] P. Calafiura, W. Lavrijsen, C. Leggett, M. Marino, and D. Quarrie, <https://cds.cern.ch/record/865624>.
- [32] R. Brun and F. Rademakers, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **389** (1997) 81–86.
- [33] CERN Collaboration,, *The Worldwide LHC Computing Grid*, <https://home.cern/about/computing/worldwide-lhc-computing-grid>, 2019. Accessed: 2020-02-03.
- [34] Particle Data Group Collaboration, M. Tanabashi et al., *Phys. Rev.* **D98** (2018) 030001.
- [35] CERN Collaboration, G. Marchiori, *egama xAOD derivations*, <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/EGammaxAODDerivations#Selection>, 2014.
- [36] CERN Collaboration, C. Anastopoulus, *xAOD Egamma Variables*, <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/EGammaD3PDtoxAOD>, 2020. Accessed:2020-02-03.

- [37] *Built in metrics*, 2019. <https://github.com/tensorflow/tensorflow/blob/r1.14/tensorflow/python/keras/metrics.py>.