

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

**Clasificación de colisiones de partículas
de alta energía utilizando redes de
convolución**

Daniela Alejandra Merizalde Aguirre

Física

Trabajo de titulación presentado como requisito
para la obtención del título de

Licenciado en Física

17 de mayo de 2023

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

**HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE
CARRERA**

Clasificación de colisiones de partículas
de alta energía utilizando redes de
convolución

Daniela Alejandra Merizalde Aguirre

Nombre del profesor, Título académico: Edgar Carrera, PhD

17 de mayo de 2023

© Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Daniela Alejandra Merizalde Aguirre

Código: 00210719

Cédula de Identidad: 1723679740

Lugar y fecha: 17 de mayo de 2023

ACLARACIÓN PARA LA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>

Dedicado a mi madre Sandra Aguirre y a mi padre Diego Merizalde. Gracias por escucharme y darme fuerzas para seguir adelante. Gracias por su apoyo incondicional y su confianza. Sin ustedes nunca habría podido culminar esta etapa de mi vida.

Agradecimientos

Agradezco a Edgar Carrera por su apoyo, esfuerzo y sabiduría. Sus consejos me guiaron a lo largo de todo este proyecto. Agradezco a cada uno de mis profesores. Todo lo que conozco sobre Física, se lo debo a ellos. Agradezco a mis compañeros y amigos porque siempre me apoyaron. Agradezco a mis hermanos, Gabriela Merizalde y Diego Bayardo Merizalde por su ilusión y entusiasmo en cada proyecto mío. Agradezco a Mónica Merizalde Bucheli quién fue una pieza fundamental para que estudie en esta universidad. Finalmente, agradezco a toda mi familia.

Resumen

En física de altas energías se requiere diferenciar distintos procesos considerando una gran cantidad de información. Por esta razón, se aplican algoritmos de aprendizaje automático para analizar los datos. El proyecto aplica una red neuronal de convolución para clasificar distintos decaimientos. Los procesos que se consideran son Higgs, W +Jets, J/ψ y Z cuyo estado final es un dimuon. Para ello, se requiere interpretar la colisión como una imagen que almacene la energía, el tipo de partículas y su posición dentro del detector. Se utilizaron simulaciones Monte Carlo a 13 TeV para el entrenamiento. Esta información pertenece a los datos abiertos del CERN liberados en el año 2015.

Palabras clave: *colisiones de partículas, física de partículas, redes neuronales*

Abstract

Collision analysis at high energy physics involves working with a lot of data. One way to manage that amount of data applies machine learning algorithms. This project uses a convolutional neural network to classify particles in high-energy physics by creating an image of the event. The events considered were Higgs, W+Jets, J/ψ y Z and all of the events had a final dimuon state. The training data are Monte Carlo simulations at 13 TeV taken from the CERN Open Data portal.

Keywords: *Convolutional neural network, classify particles.*

Índice general

1. Introducción	15
1.1. Modelo Estándar	15
1.2. Jets	20
1.3. Gran Colisionador de Hadrones	21
1.4. Sistema de referencia del Solenoide Compacto de Muones	23
1.5. Masa invariante	24
1.6. Técnicas convencionales de análisis de datos	26
2. Redes Neuronales	30
2.1. Arquitectura del perceptrón	32
2.2. Red Convolutiva	34

	9
2.2.1. Resnet 50	36
2.3. Conceptos generales	37
2.3.1. Tipos de aprendizaje	37
2.3.2. Clasificación de datos	37
3. Generación de imágenes	39
3.1. Extracción de datos	39
3.1.1. CERN Open Data Portal	39
3.1.2. Docker	41
3.1.3. CMSSW	42
3.1.4. Physics Object Extractor Tool	43
3.2. CERN SWAN	43
3.3. Tipos de decaimientos seleccionados	44
3.3.1. $W \rightarrow \mu\nu$	44
3.3.2. $J/\psi \rightarrow \mu^- \mu^+$	44
3.3.3. $Z \rightarrow \mu^- \mu^+$	45

	10
3.3.4. $H \rightarrow \mu^- \mu^+$	45
3.4. Conjunto de datos	46
3.5. Representación de las colisiones	46
3.6. Respositorio de GitHub	49
3.7. Bases de Datos	49
3.8. Detalles del entrenamiento	51
3.9. Matrices de confusión	51
4. Resultados	53
5. Conclusiones	59
Bibliografía	61
A. Acción invariante frente a transformaciones locales	65
B. Deducción Ecuación de Dirac	67

Índice de cuadros

1.1. Descripción sobre quarks	17
1.2. Descripción sobre leptones.	17
1.3. Descripción sobre bosones.	19
3.1. Rango Masas Invariantes	45
3.2. División conjunto de imágenes sin jets	50
3.3. División conjunto de imágenes con un jet	50
3.4. División conjunto de imágenes con dos jets	50
4.1. Resumen de resultados	58

Índice de figuras

1.1. Gran Colisionador de Hadrones[1]	21
1.2. Sistema de referencia del Solenoide Compacto de Muones	23
1.3. Análisis multivariado. a,b)Distribución de dos variables hipotéticas observables x_1 y x_2 c)Densidades bivariadas. d,e) Densidades marginadas unidimensionales. f)Discriminante lineal.[2]	28
2.1. Partes de una neurona biológica	31
2.2. Idea detrás del proceso de una neurona artificial. Los pesos w_i clasifican la información de entrada.	31
2.3. Entrada y proceso de datos dentro de un perceptrón	33
2.4. Pooling	36
2.5. Esquema de red convolucional	36

	13
3.1. Descripción del evento transformado a una imagen	47
3.2. Descripción del ajuste de las imágenes	48
3.3. Dimuones con dos jets. a) Decaimiento a un Higgs. b) Decaimiento W+jets. c) Decaimiento a un Bosón Z.	49
4.1. Simulaciones Monte Carlo. J/ψ con masa invariante $m_{J/\psi} = 3.1(\text{GeV}/c^2)$	53
4.2. Simulaciones Monte Carlo. Bosón Z con masa invariante $m_Z =$ $91(\text{GeV}/c^2)$	54
4.3. Simulaciones Monte Carlo. Bosón Higgs con masa invariante $m_H =$ $125(\text{GeV}/c^2)$	54
4.4. Matriz de confusión. El conjunto de imágenes no considera trazar jets solo muones	55
4.5. Matriz de confusión normalizada.El conjunto de imágenes no con- sidera trazar jets solo muones	55
4.6. Matriz de confusión. El conjunto de imágenes considera trazar un jet y muones.	56
4.7. Matriz de confusión normalizada.El conjunto de imágenes considera trazar un jet y muones.	56

- 4.8. Matriz de confusión. El conjunto de imágenes considera trazar dos jets y muones. 57
- 4.9. Matriz de confusión normalizada.El conjunto de imágenes considera trazar un jet y muones. 57

Capítulo 1

Introducción

1.1. Modelo Estándar

El comportamiento de las partículas fundamentales se describe por cuatro tipos de interacciones: gravedad, electromagnética, débil y fuerte. La fuerza que mantiene el núcleo de los átomos estable es la fuerza fuerte. De otra forma, los protones deberían repelerse.

Así como las partículas cargadas están asociadas a un campo electromagnético, las partículas asociadas a la fuerza fuerte se caracterizan por un color. La teoría que describe estas interacciones se llama cromodinámica cuántica. Esta teoría predice la existencia de gluones como mediadores de la fuerza fuerte. Por otro lado, la fuerza débil explica ciertos tipos de decaimientos radio activos [3].

Existe un régimen de energía donde no se perciben diferencias entre la fuerza débil y la fuerza electromagnética. Las interacciones en este caso se describen por la fuerza electrodébil y contiene cuatro mediadores [4]. El mediador en la fuerza electromagnética es el fotón neutral. Mientras que el mediador de la fuerza débil, son los bosones W^+ , W^- y Z^0 [3].

El modelo estándar abarca las interacciones entre la fuerza electromagnética, la fuerza débil y la fuerte. Además, describe la materia en términos de las partículas fundamentales. Esta teoría no logra asociar aún la gravedad y su relación con otras fuerzas. Existen dos clases fundamentales de partículas: fermiones y bosones. Los fermiones son partículas con espín semi-entero mientras que los bosones tienen un espín entero[5].

Los fermiones fundamentales se subdividen en dos grupos, quarks y leptones. Para diferenciar entre distintos tipos de fermiones, se observa si interactúan por medio de la fuerza fuerte. Los quarks interactúan entre sí por medio de la fuerza fuerte.

Los quarks son partículas con espín semi-entero cuya carga puede ser $-1/3$ o $+2/3$ [5]. Experimentalmente se han encontrado seis sabores de quarks descritos en Cuadro 1.1. Además, cada quark es una tripleta con tres números cuánticos asociados a un color como se mencionó. Los colores pueden ser rojo, azul y verde[1].

La cromodinámica cuántica establece que los quarks se agrupan para formar partículas compuestas que carecen de color. Otra característica es que los quarks siempre se encuentran en la naturaleza como partículas compuestas. Este fenómeno

se denomina confinamiento.[4] El confinamiento describe el hecho de que los objetos cargados con color no se observan solos, se encuentran como combinaciones que mantienen un color neutral. Los objetos sin color que consisten de un quark y otro anti-quark se llaman mesones[1]. Los objetos sin color que consisten de tres quarks se llaman bariones.

Quarks			
Partícula	masa [MeV]	carga [e]	espín [\hbar]
u	2,4	2/3	1/2
d	4,8	-1/3	1/2
c	$1,27 \times 10^3$	2/3	1/2
s	104	-1/3	1/2
t	$171,2 \times 10^3$	2/3	1/2
b	$4,2 \times 10^3$	-1/3	1/2

Cuadro 1.1: Descripción sobre quarks

Los leptones interactúan por medio de la fuerza electromagnética, débil o gravitacional. Hay tres leptones cargados. El leptón más liviano es el electrón. Además, cada leptón cargado tiene su leptón neutral llamado neutrino [4]. Todos los posibles leptones y sus características se encuentran en el Cuadro 1.2.

Leptones			
Partícula	masa [MeV]	carga [e]	espín [\hbar]
e	0,511	1	1/2
μ	105,7	1	1/2
τ	$1,777 \times 10^3$	1	1/2
ν_e	$< 2,2 \times 10^{-6}$	0	1/2
ν_μ	$< 0,17$	0	1/2
ν_τ	$< 15,5$	0	1/2

Cuadro 1.2: Descripción sobre leptones.

El concepto matemático que describe el comportamiento de las partículas fundamentales es la teoría cuántica de campos. El primer acercamiento significativo fue realizado por Paul Dirac, quien sugirió una extensión invariante de Lorentz para la ecuación de Schrödinger [1]. La densidad Lagrangiana para una partícula se expresa como:

$$\mathcal{L} = i\bar{\Psi}\gamma_{\mu}\partial^{\mu}\Psi - m\bar{\Psi}\Psi, \quad (1.1)$$

donde Ψ es un objeto cuadri-dimensional, llamado espinor, γ^{μ} son las matrices de Dirac y m es la masa en reposo del fermión [1]. La aplicación del formalismo de Euler-Lagrange a la ecuación anterior da como resultado la ecuación de Dirac para partículas que no interactúan[6]:

$$(\gamma^{\mu}p_{\mu} - m)\Psi(x) = 0 \quad (1.2)$$

donde p_{μ} es operador de momento $i\partial_{\mu}$. La ecuación de Dirac describe las partículas relativistas de espín-semi entero y predice la existencia de la anti-materia. Las teorías de gauge son esenciales para describir la interacción entre las partículas[4].

La densidad lagrangiana permanece invariante frente a transformaciones locales. Un ejemplo de una transformación local es:

$$\Psi'(x) \rightarrow e^{i\alpha(x)}\Psi(x) \quad (1.3)$$

Como todos los observables son reales, la teoría no debería depender de fases complejas como $e^{i\alpha(x)}$. Por tanto, la densidad lagrangiana de la ecuación 1.1 es

invariante frente a transformaciones descritas por la ecuación 1.3. Esta suposición lleva a una densidad lagrangiana descrita por la ecuación 1.4.

$$\mathcal{L} = \bar{\Psi} (i\gamma_\mu \partial^\mu - m) \Psi + e\bar{\Psi}\gamma_\mu A^\mu \Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.4)$$

con $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Un vector potencial A_μ fue introducido para cumplir con la invarianza frente a transformaciones locales. El término $e\bar{\Psi}\gamma_\mu A^\mu \Psi$ representa la interacción del campo entre el fermión Ψ y el campo del vector potencial A_μ [6]. El término $F_{\mu\nu}F^{\mu\nu}$ es el tensor electromagnético y contiene las ecuaciones de Maxwell [5].

Al considerar las masas de los bosones gauge aparece un problema teórico porque al añadir este término deja de ser invariante frente a transformaciones de gauge [3]. La solución planteada originó la formulación de un campo de Higgs donde se formula la generación de la masa de los bosones Z y W [1]. Así, el campo de Higgs asociado al bosón de Higgs da masa a los bosones gauge y también otorga la cualidad de masa a los fermiones. En el Cuadro 1.3 se describen las características de los bosones.

Bosones			
Partícula	masa [MeV]	carga [e]	espín [\hbar]
γ	0	0	1
g	0	0	1
Z^0	$91,2 \times 10^3$	0	1
W^\pm	$80,4 \times 10^3$	± 1	1
H	$125,3 \times 10^3$	0	0

Cuadro 1.3: Descripción sobre bosones.

El modelo estándar ha sido extensamente probado en varios experimentos. Aún así, todavía se tienen preguntas relacionadas con la física partículas. El primer problema es la inclusión de la fuerza gravitacional[4]. Además, las observaciones cosmológicas sugieren la presencia de energía y materia oscura que aún no se encuentra dentro del modelo estándar.

1.2. Jets

Como se mencionó, los quarks no se encuentran solos en la naturaleza. Esto sucedía por la naturaleza del color asociada a la fuerza fuerte. Por tanto, los quarks siempre están asociados a un número arbitrario de gluones. Así, cada experimento que involucre la creación de quarks, como el evento $e^+e^- \rightarrow q\bar{q}$, ese quark o anti-quark siempre estará asociado con un número arbitrario de partículas. La nube de partículas colimadas se llaman jets[5].

1.3. Gran Colisionador de Hadrones

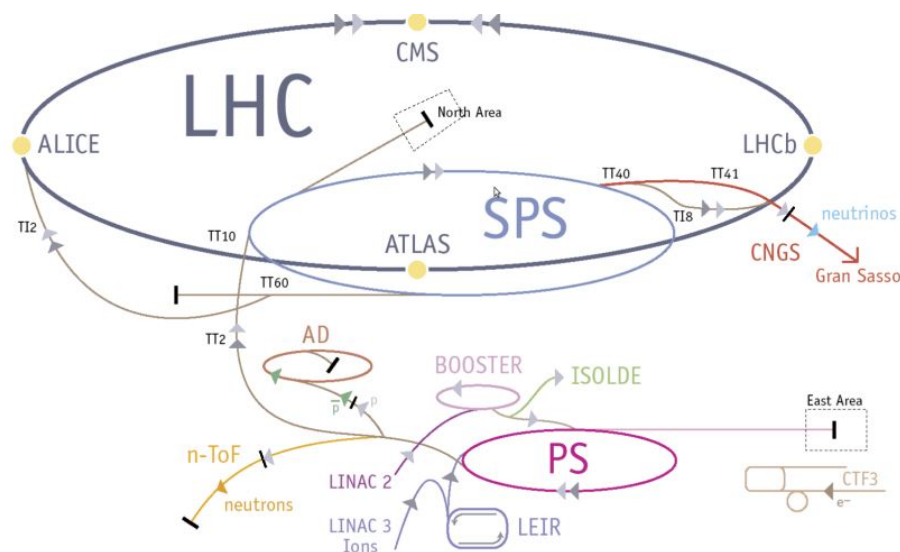


Figura 1.1: Gran Colisionador de Hadrones[1]

El Gran Colisionador de Hadrones es un acelerador de partículas hadrónicas. Los datos recolectados en el Gran Colisionador de Hadrones incluyen un gran número de colisiones y una gran complejidad en cada colisión. Los haces en este acelerador se agrupan como conjuntos de protones. Cada conjunto de protones, se cruza con una frecuencia aproximada de 40 MHz [1]. Además, cada colisión tiene el potencial de producir un gran número de partículas. Por tanto, el orden del número de detectores es de 10^8 y es necesario para capturar la información de cada una de las partículas. La cantidad de datos producida es necesaria porque los productos interesantes son raros.

Se obtienen protones al arrancar los electrones del átomo de hidrógeno. Luego, ingresan a un amplificador con una energía de 50 MeV donde se aceleran hasta adquirir una energía de 1.4 GeV. Luego, el haz continúa hacia el sincrotrón de

protones donde es acelerado hasta que adquieran una energía de 25 GeV. Más tarde, los protones se envían al súper protón sincrotrón donde terminan acelerándose hasta alcanzar los 450 GeV. [1]. Finalmente, se transfieren los protones en ambas direcciones por 20 minutos hasta que su energía nominal sea la requerida para los experimentos. Luego, los haces que colisionarán son enviados a los experimentos [1].

Como se mencionó anteriormente los haces consisten de agrupaciones de protones. El número nominal de agrupaciones por haz es 2808, con 10^{11} protones por agrupación. El espacio entre cada agrupación es de 25ns que corresponde a una frecuencia de 40 MHz[1]. Una gran densidad de agrupaciones en un haz y una gran cantidad de partículas en cada agrupación da como resultado una gran luminosidad, pero, también puede causar el efecto del amontonamiento, cuando muchas colisiones ocurren en un solo evento[1].

Existen 6 experimentos instalados en el Gran Colisionador de Hadrones: el experimento del Gran Colisionador de Iones (ALICE), el aparato toroidal del Gran Colisionador de Hadrones (ATLAS), Solenoide Compacto de Muones belleza (LHCb), el Gran Colisionador Hacia Adelante (LHCf) y el TOTAL Eléctrico y medición de sección transversal difractiva (TOTEM). ALICE, ATLAS, CMS y LHCb están instalados en cuatro puntos de interacción del LHC[1].

1.4. Sistema de referencia del Solenoide Compacto de Muones

El proyecto utiliza la información liberada por el Solenoide Compacto de Muones. Por esta razón, se describe su sistema de referencia a continuación. El acelerador tiene forma de cilindro. Por tanto, es conveniente utilizar una variable asociada al ángulo polar θ y una variable para el ángulo azimutal ϕ . La variable asociada al ángulo azimutal se denomina pseudo-rapidez. Tiene la siguiente definición:

$$\eta \equiv -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]$$

Es conveniente utilizar esta definición porque la mayoría de productos se detectan en el cuerpo del colisionador y no en sus tapas.

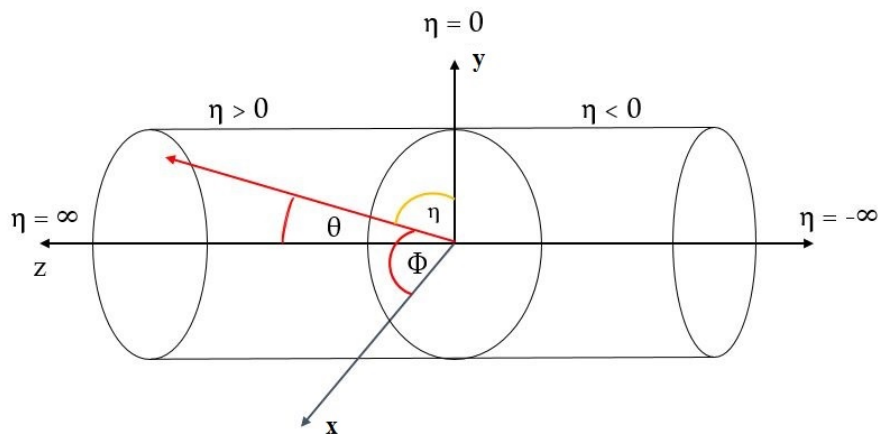


Figura 1.2: Sistema de referencia del Solenoide Compacto de Muones

1.5. Masa invariante

La masa invariante es la energía asociada a la partícula considerando un sistema de referencia en reposo. Esto implica que el sistema se mueve comóvil con la partícula. Además, esta cantidad debe ser invariante frente a transformaciones de Lorentz. Por tanto, se utiliza la cuadrivelocidad U^α y la expresión del cuadrimomento P^α para describir la masa invariante. La definición de la cuadrivelocidad es:

$$U^\mu = (\gamma c, \gamma \vec{v}) \quad (1.5)$$

donde γ es el factor de Lorentz y \vec{v} es la velocidad de la partícula en el sistema de referencia comóvil. El cuadrimomento tiene la siguiente definición:

$$P^\mu = mU^\mu \quad (1.6)$$

$$P^\mu = \left(\frac{E}{c}, \vec{P} \right) \quad (1.7)$$

Si se considera la métrica de Minkowski (eso es $\eta^{00}=1, \eta^{11}=-1, \eta^{22}=-1, \eta^{33}=-1$), el sistema comóvil de la partícula y unidades naturales, $c = 1$, la contracción del cuadrimomento se deduce como :

$$\begin{aligned} P_\mu P^\mu &= \eta^{\mu\alpha} P_\mu P_\alpha \\ &= \eta^{00} P_0 P_0 + \eta^{11} P_1 P_1 + \eta^{22} P_2 P_2 + \eta^{33} P_3 P_3 \\ &= E^2 - |\vec{P}|^2 \end{aligned}$$

Así, la expresión del cuadrimomento para en sistema co-móvil con la partícula es:

$$m^2 = E^2 - |\vec{P}|^2 \quad (1.8)$$

Al aplicar la ecuación 1.8 para un sistema con dos partículas, considerar la conservación del momento y la energía en reposo como $m_1^2 = E_1^2$ y $m_2^2 = E_2^2$, se obtiene la ecuación 1.9.

$$m^2 = m_1^2 + m_2^2 + 2(E_1 E_2 - P_1 \cdot P_2) \quad (1.9)$$

donde m es la masa invariante del sistema compuesto, m_1 y m_2 las partículas producidas después de la colisión. Tomando en consideración el sistema de referencia del CMS, la definición del momento transversal P_T , despreciando la masa de los productos y haciendo una aproximación donde $|P_T| \approx E$, se obtiene ecuación:

$$m^2 = 2P_{T1}P_{T2} (\cosh(\eta_1 - \eta_2) - \cos(\Phi_1 - \Phi_2)) [7] \quad (1.10)$$

Donde la expresión para P_T considerando el sistema de referencia del CMS es la ecuación ??.

$$|\vec{P}_T| = P_T \cosh(\eta) \quad (1.11)$$

1.6. Técnicas convencionales de análisis de datos

El modelo estándar describe el comportamiento de las partículas fundamentales y sus características. Para comprobar experimentalmente la teoría, se utilizan aceleradores de partículas. Entre estos aceleradores se encuentra el Gran Colisionador de Hadrones. La información recolectada del acelerador pretende comprobar, entre otras cosas, las teorías relacionadas con la naturaleza de la masa y la naturaleza de la materia oscura.

Sin embargo, existe un problema con el análisis de datos. Se calcula que suceden aproximadamente cuarenta millones de colisiones protón-protón cada segundo dentro de los detectores CMS y ATLAS[8]. La información recolectada por los experimentos en altas energías es compleja y se compone por un gran número de variables[9]. Asimismo, nuevos descubrimientos de partículas están determinados por señales raras que deben diferenciarse entre una gran cantidad de ruido. Por esta razón, se han realizado esfuerzos para implementar algoritmos que faciliten analizar una gran cantidad de datos. De esta manera, el objetivo se vuelve diferenciar estos procesos del modelo estándar, procesos exóticos que revelen física nueva o fenómenos de baja sección eficaz en el mismo modelo estándar.

En particular, es complejo identificar los decaimientos considerados en este proyecto. Se seleccionó Z y W en asociación con jets porque la mayor fuente de ruido en la búsqueda de física relacionada con el modelo estándar[10]. Mientras que la fracción de decaimiento del bosón de Higgs en dimuones es del 0.02%[10].

Para resolver este problema se puede recurrir a la forma tradicional de análisis donde se resuelve este problema utiliza secuencias de decisiones booleanas y métodos estadísticos para seleccionar los datos de interés[9]. Sin embargo, la dificultad del análisis incrementa entre mayor sea el número de variables. Una forma de resolver este problema sugiere utilizar algoritmos de análisis multivariado. El análisis multivariado considera que cada colisión tiene información sobre caudrivectores, sus cantidades cinemáticas deducidas y todas las variables tienen correlación[2].

El primer paso en el análisis multivariado es el preprocesamiento de los datos donde se extrae o selecciona el conjunto de características. Sobre este conjunto se aplica una transformación que exhibe las propiedades físicas de interés. El preprocesamiento puede implicar escalar las variables o una transformación más sofisticada. En algunos casos, el preprocesamiento es el único tratamiento de análisis necesario. Pero, en otros solo es el punto de partida para un análisis más refinado[2].

Un ejemplo del análisis multivariable se ilustra en la Figura 1.3. La Figura 1.3a y la Figura 1.3b muestran las distribuciones de dos variables observables x_1 y x_2 . Aunque, la Figura 1.3d y la Figura 1.3e no establecen una clara división entre ambos conjuntos de datos. La proyección en dos dimensiones del conjuntos evidencia una división en la Figura 1.3c . Por tanto, se puede aplicar un corte entre ambos datos utilizando una función lineal $\tilde{y} = ax_1 + bx_2$ conocido como discriminante lineal[2]. La separación de datos mejora su clasificación, como se puede observar en la Figura 1.3f. La función lineal que separa los dos conjuntos de datos describe un ejemplo simple de condición borde. La discriminación más óptima, en términos simplificados, es un proceso que minimiza la probabilidad de equivocarse en la

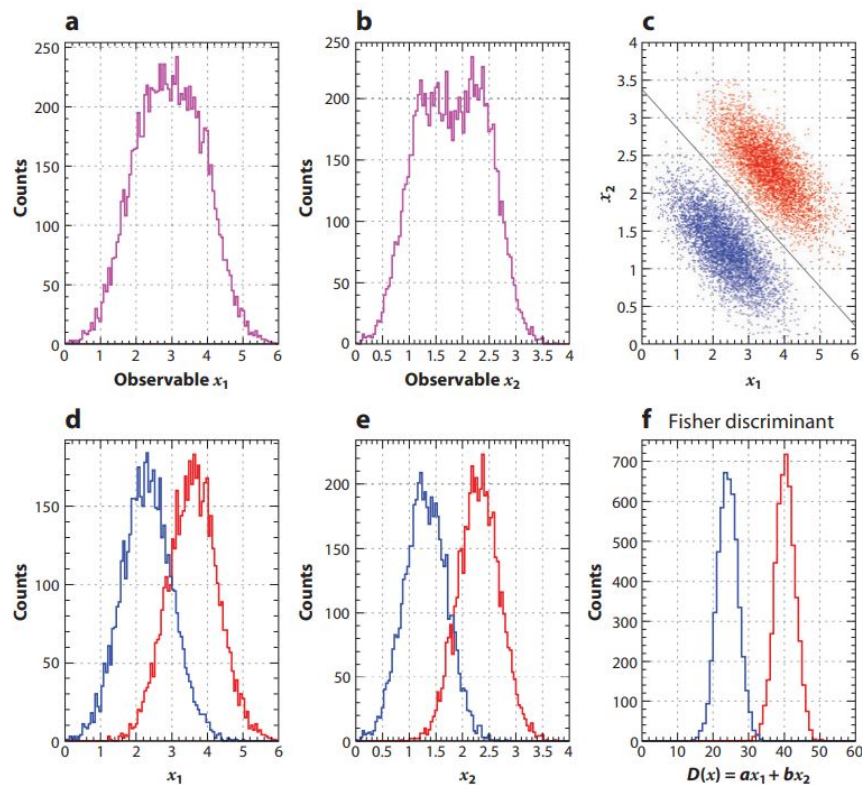


Figura 1.3: Análisis multivariado. a,b) Distribución de dos variables hipotéticas observables x_1 y x_2 c) Densidades bivariadas. d,e) Densidades marginadas unidimensionales. f) Discriminante lineal.[2]

clasificación.

El análisis multivariado utiliza técnicas de inteligencia artificial. Aunque, la inteligencia artificial se utiliza principalmente dentro de los árboles de decisión anidados. Estas técnicas logran facilitar la clasificación de datos y evidencian la necesidad de utilizar inteligencia artificial para procesar una gran cantidad de datos.

El análisis multivariado mejora el análisis de los datos, se podrían utilizar otras técnicas de inteligencia artificial. Entre estas técnicas se encuentran las redes neu-

ronales. El resurgimiento de las redes neuronales comenzó alrededor del 2012 con la aparición de algoritmos que mejoraron el desempeño de redes neuronales más complejas. Los nuevos algoritmos pueden aplicarse sobre problemas multidimensionales y alta complejidad[11]. Por esta razón, esta investigación aplica una red de neuronal de convolución utilizando datos abiertos del CERN. La estructura del proyecto se describe a continuación. El **Capítulo 2** introduce conceptos básicos de redes neuronales y describe el proceso de entrenamiento de una ResNet 50. El **Capítulo 3** describe cómo se generan las imágenes y el proceso de entrenamiento. En el **Capítulo 4** se presentan los resultados. Finalmente el **Capítulo 5** se presentan las conclusiones.

Capítulo 2

Redes Neuronales

En la búsqueda por automatizar el aprendizaje se desarrollan las redes neuronales cuyo origen comienza con la creación del perceptrón alrededor de 1960[12]. El concepto detrás de las redes neuronales consiste en simular el proceso biológico del aprendizaje. De esta manera, el objetivo del aprendizaje automático consiste en clasificar datos recibidos. Para realizar la clasificación, la red utiliza datos de entrenamiento donde se ajustan parámetros que determinan cómo clasificar la información.

El sistema nervioso contiene células, referidas como neuronas. Las neuronas están conectadas unas a otras por medio de dendritas y axones. En la Figura 2.1, se muestra la conexión entre neuronas. La región de conexión entre neuronas se llama sinapsis. Dos neuronas se comunican porque existe una diferencia de potencial en la sinapsis. La fuerza entre conexiones sinápticas cambia si existe algún estímulo externo. Este mecanismo biológico se simula en una red neuronal artificial donde

el estímulo externo proviene de la información recibida y la fuerza entre conexiones se establece por pesos[13].

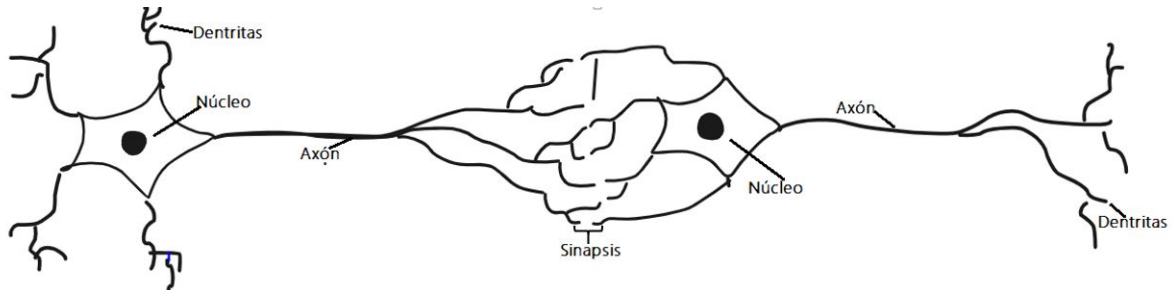


Figura 2.1: Partes de una neurona biológica

Las neuronas artificiales se conectan a través de pesos que simulan la fuerza entre conexiones sinápticas. Cada entrada está escalada por un peso, este peso afecta el cálculo final de salida donde se determinará una clasificación para la información. El concepto se representa en la Figura 2.2. Donde w_i representan los pesos de la neurona.

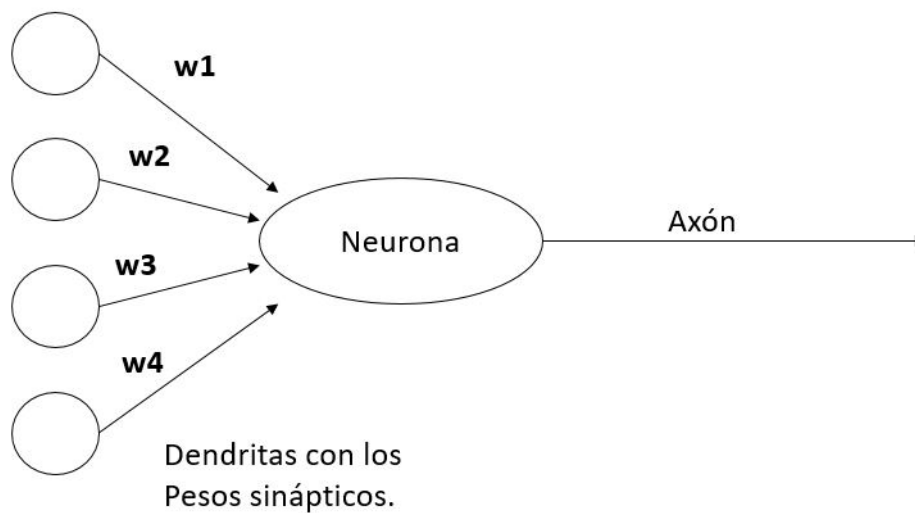


Figura 2.2: Idea detrás del proceso de una neurona artificial. Los pesos w_i clasifican la información de entrada.

Los datos de entrada simulan el rol de estímulos externos en las redes neuronales artificiales. Ingresan como entradas en la red neuronal para que realice predicciones sobre los valores de salida. Luego, estos valores se utilizan como retroalimentación para ajustar los pesos[14]. Si los pesos están mal ajustados el valor de salida predicho estará muy lejos del valor real etiquetado[13]. Entonces, se procede a corregir los pesos de tal manera que al introducir los datos otra vez la distancia entre la predicción y los datos reales sea mínima.

2.1. Arquitectura del perceptrón

La red neuronal más simple es el perceptrón porque contiene una sola capa de entrada y un solo nodo de salida. Para comprender el funcionamiento del perceptrón se analiza cómo logra su clasificación. Sea una situación donde cada instancia de entrenamiento tiene la forma (\bar{X}, y) donde $\bar{X} = [x_1, \dots, x_d]$ tiene d variables características y $y \in \{-1, +1\}$ contiene los valores observados de una clase binaria. La capa de entrada contiene d nodos asociados a d características $\bar{X} = [x_1, \dots, x_d]$ con pesos $\bar{W} = [w_1 \dots w_d]$ asociados a cada nodo. Luego, una función lineal $\bar{W} \cdot \bar{X} = \sum_{i=1}^d w_i x_i$ calcula cada nodo de salida. La predicción se calcula aplicando la función signo para obtener valor de +1 o -1, esto es apropiado para la clasificación binaria[13]. El resultado de evaluar la función signo se relaciona con la variable \hat{y} .

$$\hat{y} = \text{sgn}\{\bar{W} \cdot \bar{X}\} = \text{sgn}\left\{\sum_{j=1}^d w_j x_j\right\}$$

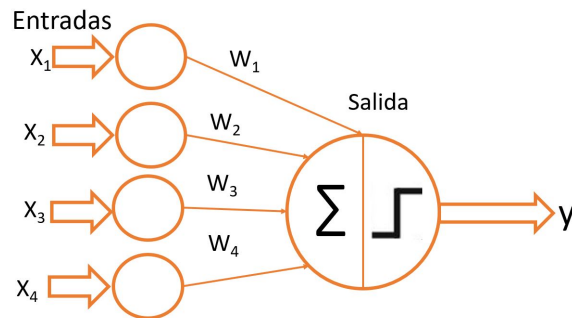


Figura 2.3: Entrada y proceso de datos dentro de un perceptrón

La función sgn se denomina función de activación. Distintas redes neuronales aplicarán otras funciones de activación que permitan clasificar los datos de salida. El error se calcula utilizando la siguiente definición [13]:

$$E(\bar{X}) = y - \hat{y}$$

donde y son las etiquetas reales y \hat{y} son los valores predichos por la red. A partir del error, se ajustan los parámetros. Para ajustar los parámetros se utiliza el gradiente descendiente. Este proceso es una optimización del gradiente, así se determina la variación y ajuste de los parámetros de tal manera que se minimice el error[13]. Por tanto se puede escribir,

$$L = \sum_{(\bar{X}, y) \in \mathcal{D}} (y - \hat{y})^2 = \sum_{(\bar{X}, y) \in \mathcal{D}} (y - \text{sign}\{\bar{W} \cdot \bar{X}\})^2$$

$$\nabla L = \sum_{(\bar{X}, y) \in \mathcal{D}} (y - \hat{y}) \bar{X}$$

donde \mathcal{D} es el conjunto que contiene las características que ejecuta la red.

Cada subgrupo del conjunto de entrenamiento se le llama batch. Es importante notar que cada batch obtiene una predicción. Después cada parámetro se ajusta de la siguiente manera:

$$\bar{W} \leftarrow \bar{W} + \alpha(y - \hat{y})\bar{X}$$

donde α es un parámetro que regula el entrenamiento de la red. El algoritmo del perceptrón se repite, en forma aleatoria, varias veces por ciclo a través de todos los ejemplos de entrenamiento e iterativamente ajusta los pesos hasta que la convergencia se alcance [13]. Como se había mencionado, cada recorrido sobre todos los datos de entrenamiento se llama época.

El perceptrón muestra la idea general detrás de una red neuronal. El objetivo principal es predecir correctamente datos no observados, en contraste con los datos estadísticos convencionales donde se prueban varios modelos matemáticos para que se ajusten al comportamiento de los datos. En este método, se infiere una automáticamente la función.

2.2. Red Convolutiva

La diferencia entre una red neuronal y una red convolutiva es la forma que tienen los datos de entrada. Los valores de entrada son matrices tridimensionales que generalmente representan imágenes[13]. Tienen altura, ancho y profundidad. La palabra profundidad en este contexto es una característica de la imagen y no determina el número de capas que tiene la red.

Una imagen a color se puede descomponer en tres representaciones distintas que contienen un solo color. Este color puede ser rojo, azul o verde. Al seleccionar alguna de estas representaciones a color, se obtiene una matriz cuyas entradas representan la intensidad de luz de ese color. El siguiente paso es realizar una convolución de esta representación con una ventana o kernel. El objetivo al aplicar una determinada ventana es resaltar las características de la imagen. La ecuación que describe la convolución es:

$$z_i = \sum_{i=1}^k w_i \cdot a_i$$

donde z_i es la i -ésima entrada de la matriz convolución, w_i es la i -ésima entrada en la ventana y a_i i -ésima entrada de la imagen, k es la dimensión de la ventana[13]. Al resultado de la convolución se aplica la función ReLU, cuya definición es:

$$f(z_i) = \text{máx}(0, z_i) = \begin{cases} z_i & \text{si } x > 0 \\ 0 & \text{otra forma} \end{cases}$$

$$m_i = \text{ReLU}(z_i),$$

donde m_i es la i -ésima entrada de la matriz generada por la convolución. El número de matrices generadas es igual al número de kernels aplicados. Este procedimiento reduce las dimensiones de la imagen original. Luego, se aplica un proceso denominado pooling[13], donde a cada matriz resultante se le aplica una ventana 2×2 y se extrae el mayor valor entre los números de la ventana. Este proceso se ilustra en la Figura 2.4

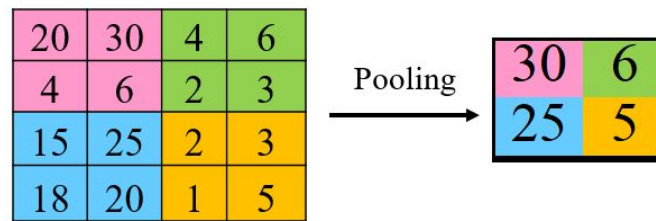


Figura 2.4: Pooling

Luego, toda la matriz se vuelve un solo arreglo y se transforma en una sola red multiconectada. Esta red contiene pesos y se aplica el mismo concepto de entrenamiento descrito para el perceptrón pero compuesto por varias neuronas. En la Figura 2.5 se representa gráficamente el proceso de una red convolucional.

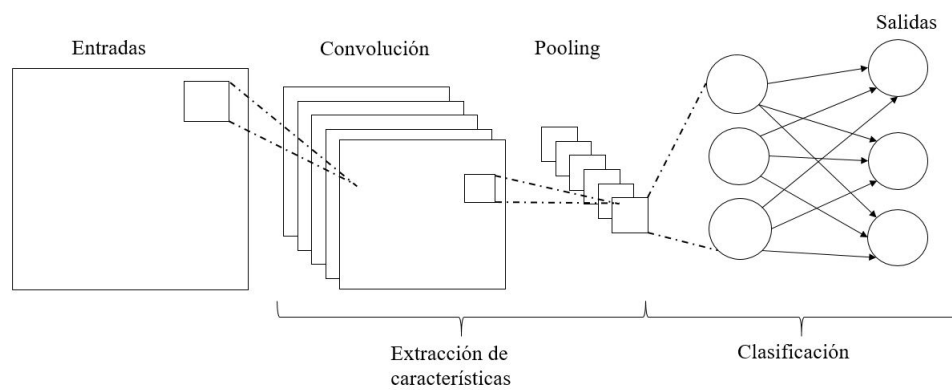


Figura 2.5: Esquema de red convolucional

2.2.1. Resnet 50

Uno de los desafíos que enfrenta la red neuronal por retropropagación es la desaparición del gradiente. Después de cada entrenamiento, el cambio sobre los parámetros producido por el cálculo del gradiente se vuelve muy pequeño. [15]

Esto provoca que la red termine el entrenamiento sin optimizar lo mejor posible la función pérdida. La idea detrás de una red residual consiste en añadir atajos de dos o tres capas para que el proceso sea más rápido. Este atajo cambia la forma de calcular el gradiente y se salta a las capas útiles. En particular, ResNet 50 tiene 50 capas de profundidad. Además, puede cargar una versión preentrenada de la red en más de un millón de imágenes desde la base de datos ImageNet [7].

2.3. Conceptos generales

2.3.1. Tipos de aprendizaje

Se puede clasificar en aprendizaje supervisado, no supervisado y aprendizaje por recompensa. El aprendizaje supervisado asocia los datos de entrenamiento con etiquetas de clasificación, mientras que los datos no supervisados no se encuentran asociados a etiquetas. Por otra parte, el aprendizaje por refuerzo implica un algoritmo que aprende al interactuar y cambiar según su entorno. La robótica y los video juegos son ejemplos de aprendizaje por refuerzo. El tipo que de aprendizaje que se emplea en este proyecto es el aprendizaje supervisado[11].

2.3.2. Clasificación de datos

El conjunto de datos se dividen por la función que cumplen dentro de la red. Ciertos datos se utilizan para el entrenamiento pues la red ajusta sus parámetros

dependiendo de qué tan lejos se encuentran estos datos de la predicción real. La duración del entrenamiento depende del número de que se repita este proceso sobre todo este conjunto de datos. Cada iteración completada sobre todo el conjunto de datos se le llama época.

El conjunto de validación, se utiliza para verificar como la red se desempeña con datos no vistos, pero, no se utiliza para realizar correcciones sobre los pesos[11]. Sirve como herramienta para comprobar que la red no esté sobre ajustada y logre predecir sobre datos con la misma clasificación pero distinto contexto. También, se utiliza para modificar otros hiper parámetros, por ejemplo, el número de épocas.

Finalmente, el último conjunto sobre el que se aplica la red es el conjunto de prueba. Su función es evaluar el desempeño de la red después de todas las modificaciones descritas anteriormente[11].

Capítulo 3

Generación de imágenes

3.1. Extracción de datos

Desde el 2014, la colaboración CMS, con el objetivo de incentivar la investigación, libera la información de las colisiones[8].

3.1.1. CERN Open Data Portal

Para descargar el conjunto de datos apropiado se debe ingresar a CERN Open Data Portal[16]. La clasificación de datos se determina por el experimento que extrajo los datos y el tipo de datos. Como se mencionó existen varios experimentos que pertenecen al Gran Colisionador de Hadrones. En particular, los datos se extrajeron del experimento CMS[8]. Además, la información se clasifica entre si-

mulaciones y colisiones reales. Los datos relacionados con las colisiones provienen directamente de los detectores, mientras que los datos descritos como simulaciones es información recreada por métodos Monte Carlo. El nombre de cada conjunto de datos describe el tipo de datos, el año de extracción o corrida e información sobre la naturaleza de los objetos físicos [8].

Se les llama objetos físicos a las partículas percibidas por los detectores. Existen partículas cuyo tiempo de decaimiento es muy pequeño, por tanto, el detector no los percibe directamente[8]. Asimismo, existen partículas que no interactúan con el detector. Es por ello que se infiere la presencia de estas partículas a partir de sus productos de decaimiento. Los objetos físicos pueden ser[8]:

- Muones
- Taus
- Electrones
- Fotones
- Jets
- Energía perdida

Para entrenar la red se utilizaron datos simulados en el año 2015 que corresponden a la segunda corrida del CMS[8]. Esta información viene en formato MiniAOD. El tamaño de este formato es más pequeño en comparación al formato AOD. El nombre de las simulaciones localizado dentro de CERN Open Data Portal consiste de

tres partes separadas por '/'. La primera parte corresponde al proceso simulado, la segunda parte corresponda al periodo de simulación y la tercera parte corresponde al formato. Por ejemplo:

- /GluGluHToTauTau-M125-13TeV-powheg-pythia8-PythiaFragment-Up
/RunIIFall15MiniAODv2-PU25nsData2015v1 76X mcRun2-asymptotic-v12-
v1/MINIAODSIM

La primera parte es GluGluHToTauTau-M125-13TeV-powheg-pythia8. Aquí se describe es el proceso de decaimiento [8]. En este caso es la fusión de dos gluones que decaen en Higgs y finalmente decae en dos leptones tau. El código M125-13TeV establece una masa invariante de $125 \frac{GeV}{c^2}$ cuyo centro de masa en el detector es 13 TeV[8]. Luego, se establece el programa que genera la simulación. En este ejemplo es Pythia 8. La segunda parte es RunIIFall15MiniAODv2-PU25nsData2015v1 76X mcRun2-asymptotic-v12-v1. RunIIFall comprende la segunda corrida realizada en otoño. MiniAODv2 es el formato MiniAOD. 25nsData2015v1 establece que la extracción de los datos se agrupó cada 25ns. La tercera parte es MINIAODSIM donde se recalca que tiene un formato MiniAOD.

3.1.2. Docker

Para descargar la información es necesario instalar el contenedor Docker. En pocas palabras, un contenedor crea un ambiente donde se ejecutan librerías y cualquier distribución, incluso sus versiones anteriores[8]. Esto es útil para realizar

el análisis de datos abiertos porque cada conjunto de datos se extrae utilizando las versiones de software establecidas para ese año[8].

Para instalar Docker se debe tomar en cuenta el sistema operativo de la computadora. En particular, si la computadora tiene Windows se debe instalar una aplicación que simule un kernel de Linux y genere una terminal. La aplicación que provee esta función se llama WSL2 o “Subsistema de Linux para Windows”. Además, entre las opciones de configuración permite trabajar con Ubuntu. Una vez completado este paso, se procede con la instalación del contenedor Docker. Es importante comprender la diferencia entre la imagen del contenedor y el contenedor para manipular el contenido su contenido. La imagen del contenedor son las instrucciones que se deben ejecutar para crear el contenedor, mientras que el contenedor es el resultado de aplicar esas instrucciones. En particular, la imagen genera un ambiente remoto necesario para extraer la información de los datos abiertos. Al copiar la imagen se carga el CMSSW o software CMS[8].

3.1.3. CMSSW

Para comunicarse con el contenedor se ejecutan los comandos en la terminal de Ubuntu asociada al WSL2. El software añadido en el contenedor es CMSSW o software CMS. Este software integra todas las librerías necesarias para extraer y analizar los datos. Los programas están escritos en C++ pero su configuración está escrita en Python. Para editar cualquiera de estos programas se utiliza un editor de texto y si se requiere compilar todos los archivos dentro de este contenedor se ejecuta el comando `scram b` y el comando `cmsRun` ejecuta el configurable que

descarga los datos en un archivo ROOT. La versión del software CMS es CMSSW 7_6_7[8].

3.1.4. Physics Object Extractor Tool

Para extraer la información necesaria de los datos se clona un repositorio de github llamado herramienta de extracción de objetos físicos o Physics Object Extractor Tool. Este repositorio cuenta con ejemplos que facilitan la extracción de datos. Todo el código C++ que procesa la información de los objetos físicos se llaman analizadores o Analyzers. El código C++ tiene la estructura de programación orientada a objetos. Es por ello que cada analizador se comporta como clase y generan objetos tipo TTrees. Finalmente, cada analizador asociado al objeto se integran en el configurador de Python. Se compila con el comando `scram b` y se ejecuta el configurador con `cmsRun`. El resultado de la configuración es un archivo en el formato `.root` con la información de cada objeto físico[8].

3.2. CERN SWAN

Toda la extracción de datos `.Root` se transforman a una sola tupla a través de una función dentro del POET. Esta tupla tipo `Root` se carga en una plataforma para análisis de datos. Esta plataforma es interactiva y permite ejecutar Jupyter Notebooks utilizando GPU[8]. La configuración que se utilizó para realizar el entrenamiento fue `102b Cuda 11.7.1 GPU` con una memoria de 16GB.

3.3. Tipos de decaimientos seleccionados

Los decaimientos seleccionados se enfocaron en seleccionar procesos cuyo producto final son dos muones. De esta manera, la principal característica que se evalúa en el proyecto es la masa invariante.

3.3.1. $W \rightarrow \mu\nu$

El decaimiento de W +jets se toma como eventos de fondo. Es importante identificarlos para separarlos de esta manera las investigaciones se dirigen hacia la búsqueda de partículas más allá del modelo estándar [17]. El conjunto de datos que se analizó considera la producción de un bosón W y jets para un centro de energía de 13 TeV en el canal de los muones. Como los muones son leptones, el bosón W puede decaer en un muon y un neutrino.

3.3.2. $J/\psi \rightarrow \mu^- \mu^+$

Otro conjunto de datos que se analizó fue un mesón decaimiento J/ψ en muon y un anti-muon. Este mesón consta de un quark y un anti quark encanto. Su masa invariante es $3,1 \frac{GeV}{c^2}$ [7]. El decaimiento de este mesón podría revelar la dinámica entre interacciones considerando varios rangos de energía. Sin embargo, se consideró para el proyecto porque su masa invariante tiene un valor muy distante a la masa invariante de otras partículas.

3.3.3. $Z \rightarrow \mu^- \mu^+$

Al igual que el bosón W este bosón se considera un evento de fondo. Conviene identificarlo para descartar su presencia y encontrar señales más exóticas que vayan más allá del modelo estándar. Las partículas asociadas a la fuerza débil son Z y W. La masa invariante de un bosón Z es $91 \frac{\text{GeV}}{c^2}$. [7] Además, este bosón es eléctricamente neutro y se acopla con cualquier partícula que lleve asociada una carga.

3.3.4. $H \rightarrow \mu^- \mu^+$

El modelo estándar predice que las masas de los fermiones se determinan por la fuerza de acoplamiento al campo de Higgs. Por tanto, existe una fracción de bosones de Higgs que decaen a un muon y un anti-muon [5]. El porcentaje de decaimiento es del 0.02%, mucho más pequeño en comparación a la fracción que decae en b-quark (58.1%) o un par tau (6.3%) [18]. Por otro lado, cuando un Higgs decae en un dimuon $H \rightarrow \mu^- \mu^+$ es el decaimiento con fermiones más limpio, permitiendo una mejor sensibilidad [18]. El rango de masas invariantes que se consideró se describe en el Cuadro 3.1 .

Clase	Objeto	Rango masa invariante (GeV/c^2)
1	J/Ψ	[2,94, 3,24]
2	Z	[83,69, 98,69]
3	H	[117,81, 132,79]

Cuadro 3.1: Rango Masas Invariantes

3.4. Conjunto de datos

Todos los datos que se utilizaron para el entrenamiento, validación y prueba son simulaciones Monte Carlo. Los datos, como se mencionó, se extrajeron del CERN Open Data Portal [16]. Todos los datos pertenecen a simulaciones cuyos decaimientos corresponden a W +jets, J/ψ , Z y Higgs. Se consideró cuatro tipos distintos de decaimientos.

Tipo de decaimiento	Nombre conjunto de datos
W +jets	WJetsToLNu-TuneCUETP8M1-13TeV-amcatnloFXFX-pythia8[12]
J/ψ	JpsiToMuMu-JpsiPt8-TuneCUEP8M1-13TeV-pythia8 [19]
Z	GluGluToContinToZZTo2e2mu-13TeV-MCFM701-pythia8 [20]
H	VBF-HToMuMu-M125-13TeV-powheg-pythia8 [21]

3.5. Representación de las colisiones

Para generar la imagen del evento se toma en consideración el tipo de partículas involucradas, el ángulo polar ϕ , pseudo-rapidez η y la energía. Toda la información proviene de los datos abiertos del CMS[16]. Cada objeto físico está representado por circunferencias cuyo centro coincide con los valores del ángulo polar ϕ , pseudo-rapidez η y el radio es proporcional a su energía.

El tamaño de la imagen es de 224×224 pixeles. Mientras que las dimensiones de la ventana están dentro de un rango $[-6,6]$ para la pseudo-rapidez η y $[-2\pi, 2\pi]$

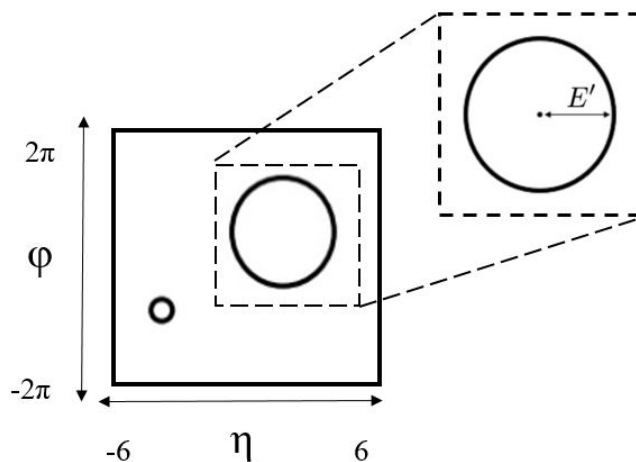


Figura 3.1: Descripción del evento transformado a una imagen

para el ángulo polar ϕ . Al ajustar el tamaño de la ventana se aplicaron factores de conversión. El factor de conversión para el eje η es $12/224$ y a lo largo del eje ϕ es $4\pi/224$.

Es útil utilizar circunferencias porque existen partículas cuyos centros coinciden lo que produce un solapamiento. Asimismo, la red interpreta mejor los resultados, si cada objeto físico traza una circunferencia completa dentro de la imagen. Por esta razón, la energía que reproduce el radio de cada partícula se escala de la siguiente manera:

$$E' = \log_{1.1}(E),$$

donde E es la energía. Se escoge un logaritmo con base 1.1 porque la mayoría de las partículas tiene una energía 5-200 GeV. En la gráfica que genera la función, se observa que estos valores son re-escalados entre 10-60 GeV.

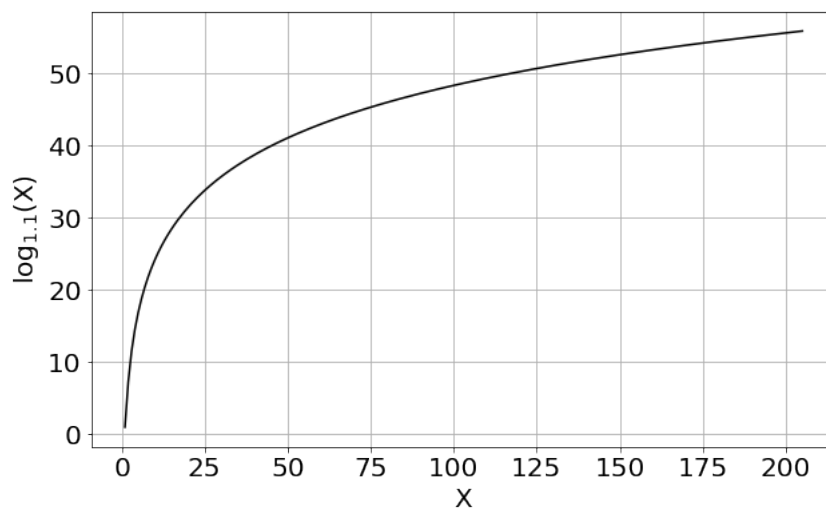


Figura 3.2: Descripción del ajuste de las imágenes

Además, las imágenes identifican dos muones pero cada conjunto de entrenamiento toma en consideración un número distinto de jets. El primer conjunto de entrenamiento no considera la presencia de un jet. El segundo conjunto toma en consideración la presencia un jet y el tercer conjunto toma en considera la presencia de dos jets. Cada circunferencia asociada a un jet tiene color naranja, su radio es proporcional a la energía y el centro coincide con los valores de pseudo rapidez η y ángulo azimutal ϕ . Como se observa, estas son las mismas condiciones con las que se recrean las circunferencias asociadas a los muones.

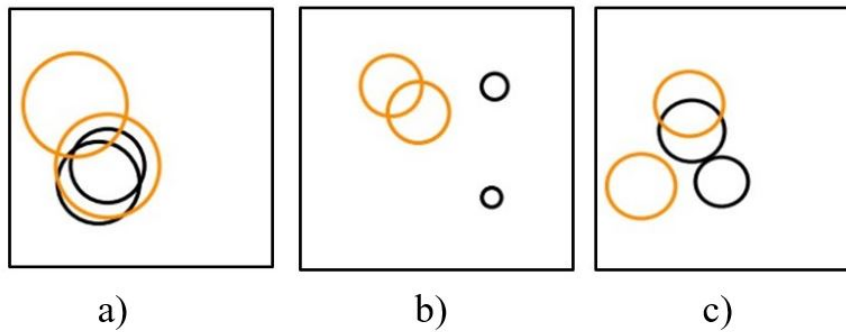


Figura 3.3: Dimuones con dos jets. a) Decaimiento a un Higgs. b) Decaimiento W +jets. c) Decaimiento a un Bosón Z .

3.6. Repositorio de GitHub

El código que generó las imágenes y el proceso entrenamiento se encuentra en el siguiente repositorio: <https://github.com/Danymery/Tesis.git>. El código se realizó sobre Jupyter Notebook; además, requiere importar dos librerías `split folders` y `uproot`. La primera librería sirve para separar los datos de entrenamiento. La segunda librería permite extraer la información de los archivos Root generados por el Physics Object Extractor Tool.

3.7. Bases de Datos

Se entrenó tres conjuntos de imágenes. Cada entrenamiento separa todas las imágenes generadas en tres carpetas: entrenamiento, validación y prueba. En el código las carpetas aparecen con el nombre de: `val`, `train` y `test`. Los datos fueron separados en una proporción de 60|20|20. Esto implica que 60% de todas las imágenes generadas se utilizaron para el entrenamiento de la red y 20% para el

conjunto de validación y prueba. El número total de imágenes generadas para el conjunto de datos sin jets fue 48000. La distribución de imágenes se describe en el Cuadro 3.2.

Decaimiento	Entrenamiento	Validación	Prueba
J/ψ	7200	2400	2400
Higgs	7200	2400	2400
W + jets	7200	2400	2400
Z	7200	2400	2400
Total	28800	9600	9600

Cuadro 3.2: División conjunto de imágenes sin jets

El número total de imágenes generados donde se traza un jet fue 3900. La distribución de imágenes se describe en el Cuadro 3.3.

Decaimiento	Entrenamiento	Validación	Prueba
Higgs	900	200	200
W + jets	900	200	200
Z	900	200	200
Total	2700	600	600

Cuadro 3.3: División conjunto de imágenes con un jet

El número total de imágenes generados donde se traza un jet fue 15602. La distribución de imágenes se describe en el Cuadro 3.3.

Decaimiento	Entrenamiento	Validación	Prueba
Higgs	3600	800	800
W + jets	3600	800	800
Z	3600	800	800
Total	10802	2400	2400

Cuadro 3.4: División conjunto de imágenes con dos jets

3.8. Detalles del entrenamiento

Para entrenar las imágenes se utilizó la librería de TensorFlow. Esta librería proporciona las herramientas necesarias para entrenar datos utilizando redes neuronales. Además, es una librería de código abierto[22]. Otra librería necesaria para entrenar los datos fue Keras. De la misma manera, Keras es una librería de código abierto y puede ejecutarse sobre TensorFlow[23]. Ambas librerías tienen asociadas estructuras de redes neuronales para el entrenamiento de datos. En particular, se utilizaron estas librerías porque el código está escrito en Jupyter Notebooks y se puede ejecutar ambas librerías con Python[23][22].

El optimizador que se utilizó para el entrenamiento es Adam y la función pérdida “categorical crossentropy”. La métrica del entrenamiento fue “accuracy”. La función de activación fue “soft-max”. La función de activación utilizada es “ReLU”. El número de épocas es 10 y cada grupo dentro de la época tienen 10 imágenes. El modelo de red neuronal que se utilizó es ResNet 50. Como se mencionó el número de capas que considera esta red es 50.

3.9. Matrices de confusión

Se llama modelo a la distribución de pesos de la red después del entrenamiento. Así, al aplicar este modelo sobre una imagen, el algoritmo logra obtener una clasificación. Este modelo se obtiene después del entrenamiento de la red.

Para evaluar el modelo entrando, se utilizan las imágenes de validación y prueba. Estas imágenes generan las matrices de confusión. Las matrices de confusión representan el error generado al utilizar el modelo. Esto es utiliza el modelo entrenado para predecir la clasificación de una imagen. Como el algoritmo puede recuperar la etiqueta con la verdadera clasificación, evalúa si el modelo entrenado se equivocó o no en la predicción.

Capítulo 4

Resultados

Para verificar que los datos pertenecen a los procesos descritos anteriormente se obtuvo gráficas de la masa invariante. En la Figura 4.1 se muestra la masa invariante de J/ψ producida por simulaciones Monte Carlo.

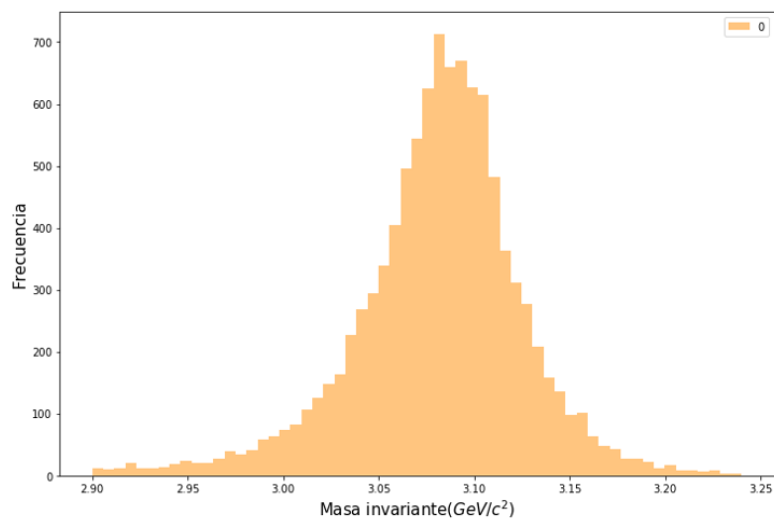


Figura 4.1: Simulaciones Monte Carlo. J/ψ con masa invariante $m_{J/\psi} = 3.1(\text{GeV}/c^2)$.

En la Figura 4.2 se muestra la masa invariante del bosón Z producida por simulaciones Monte Carlo.

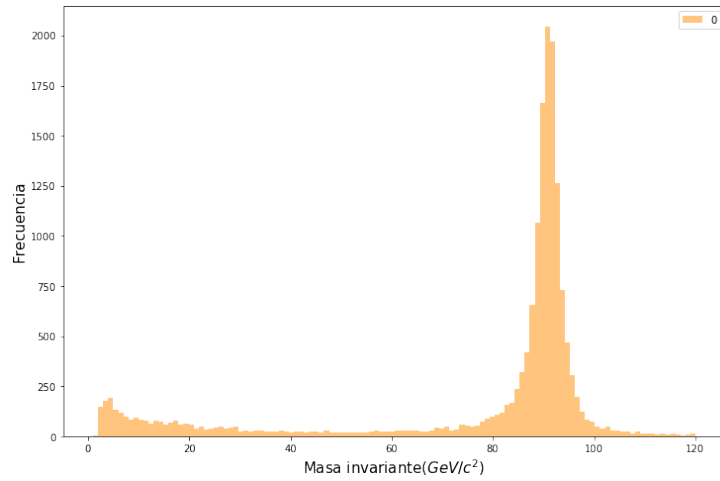


Figura 4.2: Simulaciones Monte Carlo. Bosón Z con masa invariante $m_Z = 91(\text{GeV}/c^2)$.

En la Figura 4.3 se muestra la masa invariante del bosón Higgs producida por simulaciones Monte Carlo.

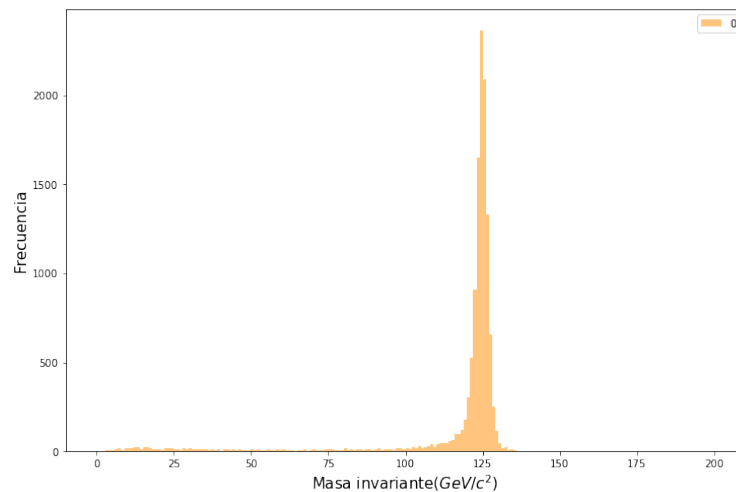


Figura 4.3: Simulaciones Monte Carlo. Bosón Higgs con masa invariante $m_H = 125(\text{GeV}/c^2)$.

Las imágenes generadas para el primer conjunto de entrenamiento no consideran jets, solo la generación un muon y un antimuon. La matriz de confusión se puede notar en la Figura 4.4. La matriz de confusión normalizada se puede notar en la Figura 4.5.

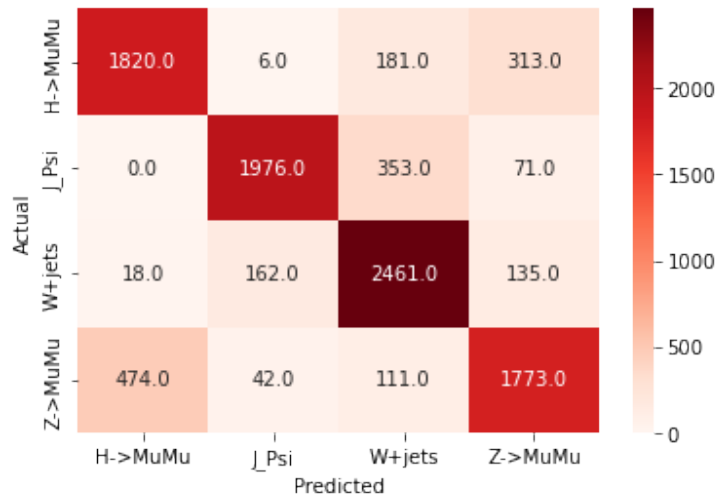


Figura 4.4: Matriz de confusión. El conjunto de imágenes no considera trazar jets solo muones

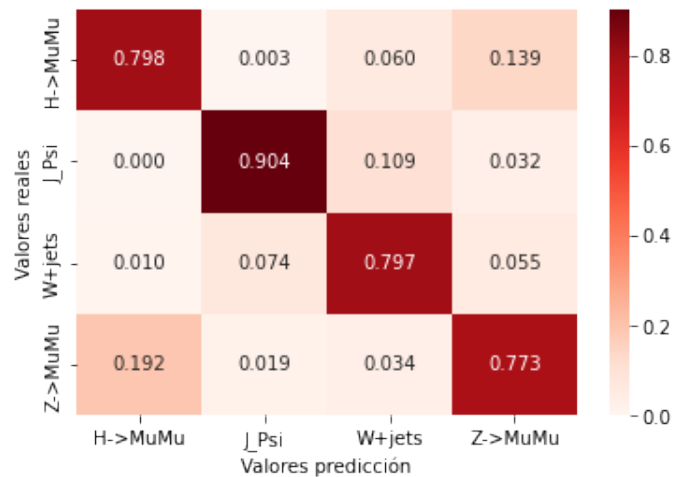


Figura 4.5: Matriz de confusión normalizada. El conjunto de imágenes no considera trazar jets solo muones

A continuación, se presentan las imágenes generadas para el segundo conjunto de entrenamiento donde se consideran un jet. La matriz de confusión se puede notar en la Figura 4.6 La matriz de confusión normalizada se puede notar en la Figura 4.7.

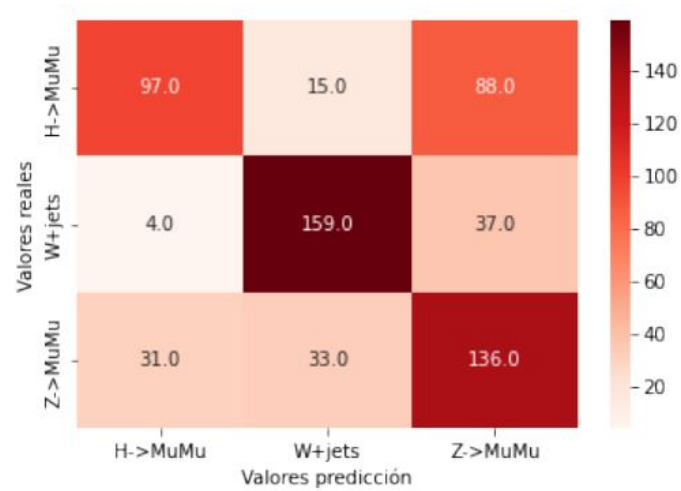


Figura 4.6: Matriz de confusión. El conjunto de imágenes considera trazar un jet y muones.

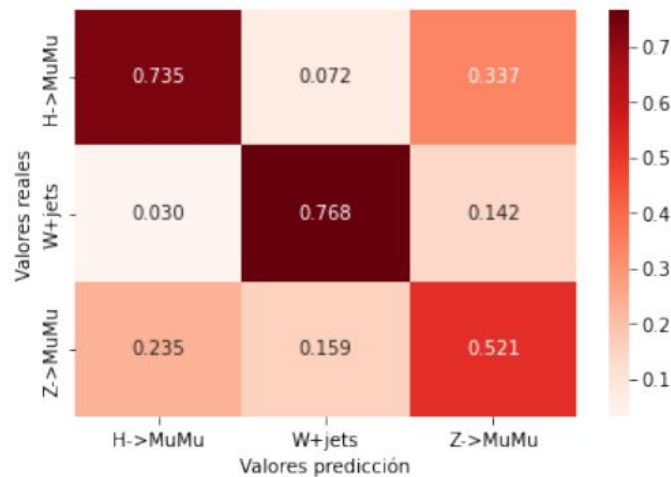


Figura 4.7: Matriz de confusión normalizada. El conjunto de imágenes considera trazar un jet y muones.

A continuación, se presentan las imágenes generadas para el tercer conjunto de entrenamiento consideran dos jets. La matriz de confusión se puede notar en la Figura 4.8. La matriz de confusión normalizada se puede notar en la Figura 4.9.

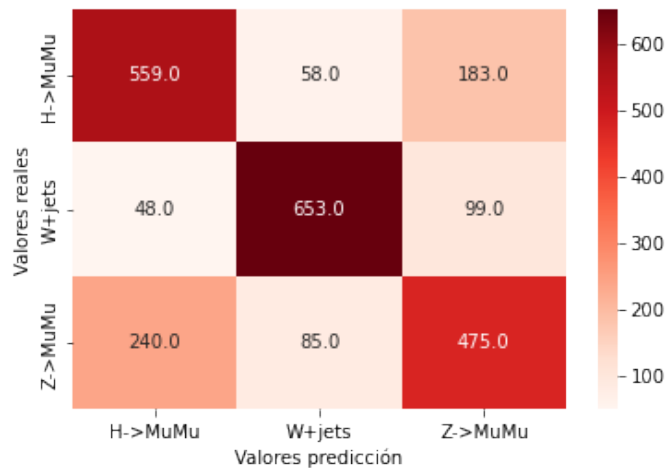


Figura 4.8: Matriz de confusión. El conjunto de imágenes considera trazar dos jets y muones.

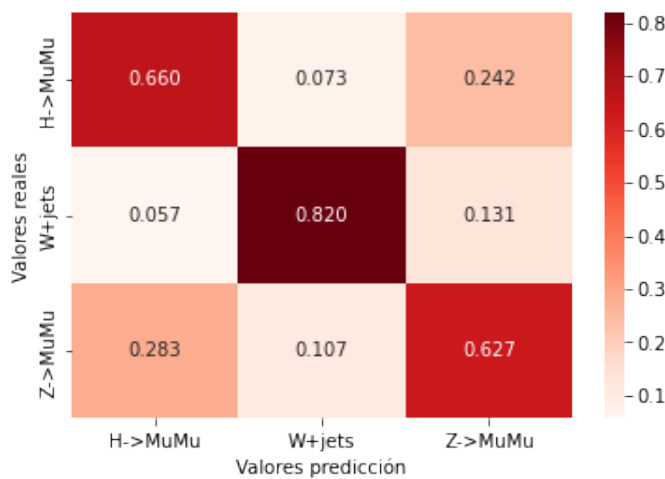


Figura 4.9: Matriz de confusión normalizada. El conjunto de imágenes considera trazar un jet y muones.

Dado que el régimen de energía no coincide con la generación de jets y el mesón J/ψ , no se consideró dentro de las imágenes con jets. Los resultados de la precisión y la función pérdida se muestran en el Cuadro 4.1.

Jets	Error	Precisión
0	47.70 %	80.67 %
1	74.56 %	65.33 %
2	70.29 %.	69.35 %

Cuadro 4.1: Resumen de resultados

Capítulo 5

Conclusiones

El proyecto verifica el desempeño de una red neuronal para la clasificación de colisiones. Para cumplir con este objetivo, se transforma la información del evento en imágenes. Las imágenes son circunferencias que representan la energía y posición del objeto físico en el evento. En particular el tipo de red neuronal que se utilizó es la ResNet 50. Además, se consideran los decaimientos del bosón Higgs, W +jets, bosón Z y mesón J/ψ con un estado final de un muon y un antimuon. Todos los datos considerados son simulaciones Monte Carlo obtenidas del Open Data Portal. Estos datos pertenecen a la colaboración CMS y el año de liberación de los datos es el 2015.

Los resultados muestran que la red tiene un mejor desempeño cuando no se consideran jets. Por tanto, se puede concluir que el tamaño de los círculos de los muones y su posición es información suficiente para la clasificación. En otras palabras, la característica principal para la clasificación del evento es la masa

invariante. Como se puede observar en la matriz de confusión, se genera un error de clasificación entre la clasificación de un bosón Z y bosón de Higgs. Ambos decaimientos tienen masas invariantes más cercanas. Mientras que, la red logra clasificar con un error menor entre el mesón J/ψ y el resto de decaimientos. Esto se debe a que la masa invariante del este mesón tiene una magnitud muy diferente en comparación a la masa invariante de los otros decaimientos.

Otra característica que se debe tener en cuenta es la información que describe la imagen. La imagen no debe contener información que presente una ventaja al entrenamiento. Por ejemplo, podría consistir en cambiar en el color de los muones dependiendo del tipo de decaimiento. Si se entrena imágenes con esta información la clasificación será perfecta, pero, se aleja de las condiciones bajo las cuales se analizan los datos reales. En los datos reales, no se obtiene información sobre el tipo de decaimiento. Por tanto, no se podría utilizar el modelo de clasificación.

Asimismo, se debe preservar la información que aporta la imagen. Las imágenes utilizan circunferencias y no círculos porque existen eventos donde el objeto físico coincide con el ángulo polar y la pseudorapidez. Si las imágenes generadas son círculos se pierde información sobre estos objetos. Además, se debe tomar los factores que determinan el tamaño de la ventana donde se encuentran las circunferencias. Ninguna circunferencia debe quedar incompleta o desaparecer dentro de la imagen porque provoca que la red confunda los datos de entrenamiento. Por tanto, se debe modificar la ventana de tal forma que todos los objetos físicos trazados tengan circunferencias completas. Todos estos detalles preservan la información que describe la imagen y aportan la información considerada para el entrenamiento.

Por otro lado, otra característica importante del proyecto es el uso de GPU para el entrenamiento. Emplear GPU aumenta la velocidad del entrenamiento al optimizar el tiempo permite realizar distintas pruebas. Se puede concluir que las redes neuronales son una solución para el análisis de datos en física de altas energías. Esto implica que se pueden emplear para la clasificación sobre de distintos tipos de decaimientos y facilitar el análisis considerando un gran número de datos. En particular, se puede transformar un evento a una imagen que preserve la información sobre la colisión. Así se podrá aplicar el tipo de red neuronal ResNet 50 para el entrenamiento del modelo. Otros proyectos podrían considerar el modelo generado al considerar solo dimuones y aplicar este modelo sobre datos reales.

Bibliografía

- [1] Mikhail Karnevskiy. *Measurement of the Z boson production with the ATLAS experiment at the LHC*. PhD thesis, Hamburg U., 2012.
- [2] Pushpalatha C. Bhat. Multivariate analysis methods in particle physics. *Annual Review of Nuclear and Particle Science*, 61(1):281–309, 2011.
- [3] D. H. Perkins. *Introduction to high energy physics*. 1982.
- [4] M. Robinson, K. Bland, G. Cleaver, and J. Dittmann. A simple introduction to particle physics, 2008.
- [5] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008.
- [6] Andrew J. Larkoski. *Elementary Particle Physics: An Intuitive Introduction*. Cambridge University Press, 2019.
- [7] Ignacio Heredia. Large-scale plant classification with deep neural networks. *CoRR*, abs/1706.03736, 2017.
- [8] E. Carrera et al. Cms open data workshop, 2022. Accessed on Mayo 8, 2023.

- [9] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep learning and its application to LHC physics. *Annual Review of Nuclear and Particle Science*, 68(1):161–181, oct 2018.
- [10] V. Khachatryan and et al A.M. Measurement of the zz production cross section and z to $4l$ branching fraction in pp collisions at $s=13$ tev. *Physics Letters B*, 763:280–303, 2016.
- [11] Celia Fernández Madrazo, Ignacio Heredia Cacha, Lara Lloret Iglesias, and Jesús Marco de Lucas. Application of a convolutional neural network for image classification to the analysis of collisions in high energy physics, 2017.
- [12] CMS Collaboration (2021). Simulated dataset wjetstolnu-tunecuetsp8m1-13tev-amcatnlofxfx-pythia8 in miniaodsim format for 2015 collision data. DOI:10.7483/opendata.cms.zyy4.gyid.
- [13] Charu C. Aggarwal. *Neural networks and Deep Learning: A textbook*. 2018.
- [14] Dimitri Bourilkov. Machine and deep learning applications in particle physics. *International Journal of Modern Physics A*, 34(35):1930019, dec 2019.
- [15] Minghao Yin, Xiu Li, Yongbing Zhang, and Shiqi Wang. On the mathematical understanding of resnet with feynman path integral. *CoRR*, abs/1904.07568, 2019.
- [16] Cern. Open data cern, 2023.
- [17] G. Aad, B. Abbott, and J. Abdallah et al. Study of jets produced in association with a w boson. *Physical Review D*, 85(9), may 2012.

- [18] Xunwu Zuo. A Search for the Standard Model Higgs Boson Decaying into Two Muons in the Vector Boson Associated Production Mode at the CMS Experiment, 2021.
- [19] CMS Collaboration(2021). Simulated dataset jpsitomumu-jpsipt8-tunecuep8m1-13tev-pythia8 in miniaodsim format for 2015 collision data. cern open data portal. **DOI:**10.7483/opendata.cms.2tx8.nv9c.
- [20] CMS Collaboration(2021). Simulated dataset vbf-htomumu-m125-13tepowheg-pythia8 in miniaodsim format for 2015 collision data. cern open data portal. **DOI:**10.7483/opendata.cms.qiq9.zh9w.
- [21] CMS Collaboration(2021). Simulated dataset gluglu-contin-zzt-2e2mu in miniaodsim format for 2015 collision data. cern open data portal. **DOI:**10.7483/opendata.cms.dx8k.3i2a.
- [22] Tensor Flow. Crea modelos de aprendizaje automático de nivel de producción con tensorflow, 2023. <https://www.tensorflow.org/?hl=es-419> [Accessed: (14-5-2023)].
- [23] Keras. Keras: Deep learning for humans, 2023. <https://keras.io/> [Accessed: (14-5-2023)].

Apéndice A

Acción invariante frente a transformaciones locales

La acción viene dada por:

$$S = \int dx^4 \bar{\psi} (i\gamma_\mu \partial^\mu - m) \psi + e \bar{\psi} \gamma_\mu A^\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (\text{A.1})$$

el tensor electromagnético tiene la siguiente definición:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

Consideremos las siguientes transformaciones locales:

$$\begin{aligned} \psi' &= e^{i\alpha(x)} \psi \\ A'_\mu &= A_\mu + \frac{1}{e} \partial_\mu \alpha(x) \end{aligned}$$

donde e es una constante. Se aplica la transformación:

$$S = \int d^4x \left[\bar{\psi}' (i\gamma_\mu \partial^\mu - m) \psi' + e \bar{\psi}' \gamma_\mu A^\mu \psi' - \frac{1}{4} F'_{\mu\nu} F^{\mu\nu'} \right]$$

$$S = \int d^4x \left[e^{-i\alpha(x)} \bar{\psi} i\gamma_\mu e^{i\alpha(x)} (i\partial^\mu \alpha(x) \psi + \partial^\mu \psi) - \bar{\psi} m \psi + e \bar{\psi} \gamma_\mu A^\mu \psi + \bar{\psi} \gamma_\mu \partial^\mu \alpha(x) \psi - \frac{1}{4} F'_{\mu\nu} F^{\mu\nu'} \right]$$

$$S = \int d^4x \left[-\bar{\psi} \gamma_\mu \partial^\mu \alpha(x) \psi + \bar{\psi} i\gamma_\mu \partial^\mu \psi - \bar{\psi} m \psi + e \bar{\psi} \gamma_\mu A^\mu \psi + \bar{\psi} \gamma_\mu \partial^\mu \alpha(x) \psi - \frac{1}{4} F'_{\mu\nu} F^{\mu\nu'} \right]$$

$$S = \int d^4x \bar{\psi} (i\gamma_\mu \partial^\mu - m) \psi + e \bar{\psi} \gamma_\mu A^\mu \psi - \frac{1}{4} F'_{\mu\nu} F^{\mu\nu'}$$

Se demuestra que $F'_{\mu\nu} F^{\mu\nu'}$ es invariante:

$$= \eta^{\mu\sigma} \eta^{\beta\nu} \left(\partial_\mu A_\nu + \frac{1}{e} \partial_\mu \partial_\nu \alpha(x) - \partial_\nu A_\mu - \frac{1}{e} \partial_\nu \partial_\mu \alpha(x) \right) \left(\partial_\sigma A_\beta + \frac{1}{e} \partial_\sigma \partial_\beta \alpha(x) - \partial_\beta A_\sigma - \frac{1}{e} \partial_\beta \partial_\sigma \alpha(x) \right)$$

$$= \eta^{\mu\sigma} \eta^{\beta\gamma} (\partial_\mu A_\gamma - \partial_\gamma A_\mu) (\partial_\sigma A_\beta - \partial_\beta A_\sigma)$$

$$= (\partial_\mu A_\nu - \partial_\nu A_\mu) (\partial^\mu A^\nu - \partial^\nu A^\mu)$$

$$= F_{\mu\nu} F^{\mu\nu}$$

Por tanto, la acción es invariante frente a transformaciones locales.

Apéndice B

Deducción Ecuación de Dirac

Para deducir la ecuación de Dirac notamos que la densidad lagrangiana que se considera depende de lo siguiente:

$$\mathcal{L} = \mathcal{L}(\bar{\Psi}, \Psi, \partial^\mu \Psi)$$

Se realizan variaciones sobre esta acción y la expresión queda:

$$\delta S = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial(\bar{\Psi})} \delta \bar{\Psi} + \frac{\partial \mathcal{L}}{\partial(\Psi)} \delta \Psi + \frac{\partial \mathcal{L}}{\partial(\partial^\mu \Psi)} \delta(\partial^\mu \Psi) \right]$$

Para encontrar la variación se encuentra otra expresión para el último término de la acción. Por la regla del producto, se puede expandir de la siguiente manera:

$$\int d^4x \left[\partial^\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial^\mu \Psi)} \delta(\Psi) \right) \right] = \int d^4x \left[\partial^\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial^\mu \Psi)} \right) \delta \Psi + \frac{\partial \mathcal{L}}{\partial(\partial^\mu \Psi)} \delta(\partial^\mu \Psi) \right]$$

Utilizando el teorema de la divergencia y anulando los términos en la frontera:

$$\int d^4x \left[\partial^\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial^\mu \psi)} \delta(\psi) \right) \right] = \int d\sigma \left[\frac{\partial \mathcal{L}}{\partial (\partial^\mu \psi)} \delta(\psi) \right] = 0$$

Entonces, el último término se puede reescribir así:

$$\int d^4x \left[\partial^\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial^\mu \psi)} \right) \delta\psi \right] = - \int d^4x \left[\frac{\partial \mathcal{L}}{\partial (\partial^\mu \psi)} \delta(\partial^\mu \psi) \right]$$

La ecuación que queda después del cambio es:

$$\delta S = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial (\bar{\Psi})} \delta(\bar{\Psi}) + \frac{\partial \mathcal{L}}{\partial (\Psi)} \delta(\Psi) - \partial^\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial^\mu \Psi)} \right) \delta(\Psi) \right]$$

Agrupando las variaciones e imponiendo que sean cero:

$$\delta S = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial (\bar{\Psi})} \delta\bar{\Psi} + \left(\frac{\partial \mathcal{L}}{\partial (\Psi)} - \partial^\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial^\mu \Psi)} \right) \right) \delta\Psi \right]$$

Se obtiene las siguientes ecuaciones:

$$\frac{\partial \mathcal{L}}{\partial (\bar{\Psi})} = 0 \tag{B.1}$$

$$\frac{\partial \mathcal{L}}{\partial (\Psi)} - \partial^\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial^\mu \Psi)} \right) = 0 \tag{B.2}$$

Ahora, consideramos la siguiente densidad lagrangiana y aplicamos la ecuación

B.2:

$$\begin{aligned} \mathcal{L} &= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - m\bar{\Psi}\Psi \\ \frac{\partial \mathcal{L}}{\partial \Psi} &= \frac{\partial (i\bar{\Psi}\gamma_\mu\partial^\mu\Psi)}{\partial \Psi} - \frac{\partial (m\bar{\Psi}\Psi)}{\partial \Psi} \end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \Psi} &= -m\bar{\Psi} \\ \frac{\partial \mathcal{L}}{\partial (\partial^\mu \Psi)} &= \frac{\partial (\bar{\Psi} \gamma_\mu \partial^\mu \Psi)}{\partial (\partial^\mu \Psi)} = i\bar{\Psi} \gamma_\mu \\ -m\bar{\Psi} - i\partial^\mu \bar{\Psi} \gamma_\mu &= 0 \\ i\gamma_\mu \partial^\mu \bar{\Psi} + m\bar{\Psi} &= 0\end{aligned}$$

Resulta en la ecuación conjugada de Dirac:

$$(i\gamma_\mu \partial^\mu + m)\bar{\Psi} = 0$$

Si se utiliza la ecuación B.1:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \bar{\Psi}} &= \frac{\partial (i\bar{\Psi} \gamma_\mu \partial^\mu \Psi)}{\partial \bar{\Psi}} - \frac{\partial (m\bar{\Psi} \Psi)}{\partial \bar{\Psi}} \\ &= i\gamma_\mu \partial^\mu \Psi - m\Psi\end{aligned}$$

La última expresión es la ecuación de Dirac:

$$(i\gamma_\mu \partial^\mu - m)\Psi = 0$$