

© CERN. Compact Muon Solenoid (detalle).

FIGURAS REVISTA ACADÉMICA DE INVESTIGACIÓN

ISSN 2683-2917

Vol. 1, núm. 3, julio-octubre 2020

<https://doi.org/10.22201/fesa.figuras.2020.1.3>



Esta obra está bajo una licencia  
Creative Commons Atribución-NoComercial-  
CompartirIgual 4.0 Internacional

## Uso de técnicas de *Machine Learning* en el experimento CMS

### *Use of Machine Learning techniques in the CMS experiment*

<https://doi.org/10.22201/fesa.figuras.2020.1.3.121>

 Cristina Oropeza-Barrera

Universidad Iberoamericana, Departamento de Física y Matemáticas, Ciudad de México

## Resumen

La física experimental de partículas se encuentra en una era dorada llena de retos tecnológicos. Para superarlos, las grandes colaboraciones del LHC (*Large Hadron Collider*) han implementado técnicas de *Ma-*

*chine Learning* en sus operaciones con resultados impresionantes. En esta reseña se resumen algunas de las aplicaciones principales del aprendizaje automatizado, en particular de las redes neuronales artificiales, en el experimento CMS (*Compact Muon Solenoid*). Además se resalta la importancia del trabajo colaborativo e interdisciplinario para la correcta implementación e interpretación de estas técnicas de análisis. El objetivo del presente trabajo consiste en despertar el interés por estos temas entre los miembros de las comunidades de ciencias de la computación y física de partículas, con el fin de ampliar las posibilidades de trabajos de investigación conjuntos.

## Introducción

### La física de partículas elementales

La física de partículas busca describir los componentes fundamentales de la materia, así como las interacciones entre éstos; la teoría que engloba el conocimiento que tenemos hasta ahora del mundo de las partículas elementales es el Modelo Estándar (ME).

En los 60 años que han transcurrido desde su nacimiento, constantemente se ha puesto a prueba el Modelo Estándar y sus predicciones han sido verificadas experimentalmente. El último gran descubrimiento sucedió en 2012, cuando el bosón de Higgs fue observado<sup>1</sup> por los experimentos ATLAS (*A Toroidal LHC ApparatuS*) y CMS (*Compact Muon Solenoid*) del Gran Colisionador de Hadrones (LHC por sus siglas en inglés); ante tal hallazgo, se otorgó el premio

<sup>1</sup> ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B* 716, (2012): 1.

CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B* 716, (2012): 30.

Nobel a Peter Higgs y François Englert, los dos físicos teóricos que propusieron su existencia.<sup>2</sup>

A pesar de sus numerosos éxitos, el ME no puede ser considerado como una teoría final o completa, ya que se han elaborado preguntas que todavía el modelo no logra responder. Esto ha llevado a que la comunidad científica busque indicios de física más allá del ME en los experimentos actuales.

Con la revolución tecnológica en materia de detectores y aceleradores de partículas, los experimentos son cada vez más grandes, más sensibles y más precisos.

#### Experimentando con partículas

Para detectar una partícula, ésta debe interactuar con el material del detector y depositar energía de alguna forma. En la primera mitad del siglo XX se utilizaban principalmente cámaras de niebla y cámaras de burbujas. Cuando las partículas con carga atravesaban el material, sus trayectorias eran captadas mediante fotografías que eran –posteriormente– reveladas y analizadas, una por una, por técnicos especializados. Si bien el arduo trabajo rendía frutos, la capacidad de procesamiento y análisis de datos era bastante limitada.

<sup>2</sup> F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.* 13, (1964): 321.

P. W. Higgs, “Broken symmetries, massless particles and gauge fields,” *Phys. Lett.* 12, (1964): 132.

P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.* 13, (1964): 508.

Esto cambiaría con la llegada de los amplificadores con transistores y la creación en 1968 de las cámaras proporcionales (*multiwire proportional chambers*), detectores de gas capaces de registrar millones de trazas por segundo.<sup>3</sup> Con la revolución tecnológica en materia de detectores y aceleradores de partículas, los experimentos son cada vez más grandes, más sensibles y más precisos.

Actualmente, el centro de investigación sobre partículas más importante del mundo es la Organización Europea de Investigación Nuclear (CERN, por sus siglas en francés). En él se encuentra el LHC, un acelerador de partículas de 27 km de circunferencia que colisiona protones y núcleos de átomos pesados a altas energías. El LHC cuenta con cuatro experimentos principales: ALICE (*A Large Ion Collider Experiment*),<sup>4</sup> ATLAS,<sup>5</sup> CMS<sup>6</sup> y LHCb (*Large Hadron Collider beauty*).<sup>7</sup> Cada uno tiene tecnologías de detección y metas de investigación distintas. CMS es un experimento multipropósito. El detector está conformado por diferentes sub-detectores distribuidos radialmente en capas. Cada sub-detector tiene una función específica, y en conjunto proporcionan toda la información necesaria para la reconstrucción de las colisiones.

<sup>3</sup> G. Charpak, *et al.*, “The use of multiwire proportional counters to select and localize charged particles,” *Nucl. Instr. Methods* 62, (1968): 262.

<sup>4</sup> ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *JINST* 3, (2008): S08002.

<sup>5</sup> ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider,” *JINST* 3, (2008): S08003.

<sup>6</sup> CMS Collaboration, “The CMS experiment at the CERN LHC,” *JINST* 3, (2008): S08004.

<sup>7</sup> LHCb Collaboration, “The LHCb detector at the LHC,” *JINST* 3, (2008): S08005.

El LHC ha operado de manera extraordinaria en los últimos años. No sólo ha incrementado la energía de las colisiones, sino que también aumentó la luminosidad. Esto tiene repercusiones importantes para los experimentos. Los haces de protones que circulan por el LHC se cruzan en el centro de CMS 40 millones de veces por segundo. Adicionalmente, en cada cruce de protones pueden ocurrir múltiples colisiones, por lo que el detector debe ser capaz de identificar, seleccionar y reconstruir eventos en un ambiente denso. Esto lleva a la colaboración a buscar nuevas formas para enfrentarse a estos retos.

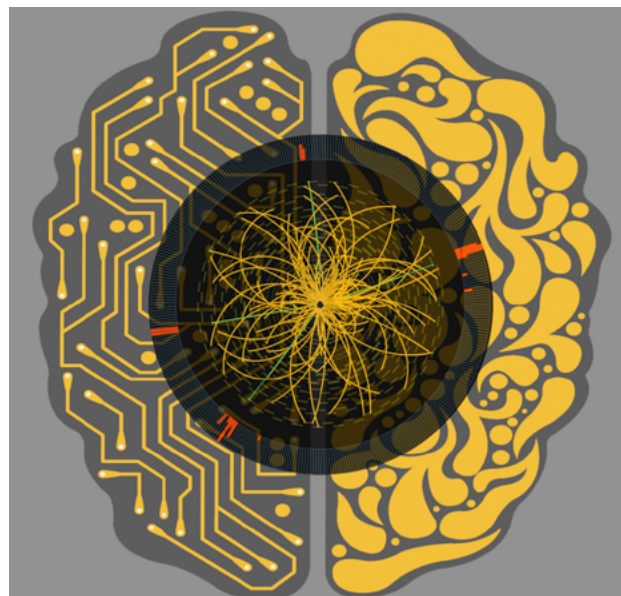
### ¿Qué es *Machine Learning*?

El aprendizaje automatizado, o *Machine Learning* (ML), es una rama de la inteligencia artificial que busca desarrollar algoritmos que puedan aprender de los datos, identificar patrones y tomar decisiones con mínima intervención humana.<sup>8</sup>

### Machine Learning en CMS

El volumen de datos recabados por los experimentos del LHC ha crecido rápidamente y los retos –de análisis y computacionales– son cada vez mayores. Desde hace tiempo se utilizan redes neuronales en estudios de CMS, pero en años recientes sus aplicaciones han ido en aumento. En las siguientes secciones se describirán brevemente algunas de ellas.

<sup>8</sup> SAS Institute Inc., “Machine Learning. What it is and why it matters,” *SAS-THE POWER TO KNOW*. (2018) [https://www.sas.com/en\\_us/insights/analytics/machine-learning.html](https://www.sas.com/en_us/insights/analytics/machine-learning.html) Revisado el 20 de marzo, 2020.



Fermilab/CERN.

### Reconstrucción

La reconstrucción de trazas es un elemento vital para prácticamente todo análisis que se realice en CMS. Inferir las trayectorias de las partículas cargadas a partir de las mediciones de cada sensor es una tarea mayúscula, que se ve severamente afectada por el número de trazas presentes en el evento (número de colisiones por cada cruce de los protones). El uso de técnicas de ML (GNNs: *graph neural networks*) resulta útil para mejorar la resolución de los detectores en zonas muy densas.<sup>9</sup>

### Trigger

La decisión de almacenar o no la información de un evento en CMS se lleva a cabo utilizando un *trigger*, el cual toma información de diferentes sub-detectores y ejecuta un algoritmo primitivo de reconstrucción. Las redes neuronales convolucionales se utilizan en la reconstrucción de trazas en el detector interno para

<sup>9</sup> S. Farrell *et al.*, “The HEP.TrkX Project: deep neural networks for HL-LHC online and offline tracking,” *EPJ Web Conf.* 150, (2017): 00003.

seleccionar las combinaciones de señales de pares de sensores que están asociadas al paso de una misma partícula, de aquellas que son falsas. Estos algoritmos deben tomar la decisión en tiempos muy cortos y han demostrado ser capaces de obtener una eficiencia del 99%.<sup>10</sup>

## Análisis

Discriminar jets provenientes de un quark tipo b, de otros jets, es una tarea crucial para el programa de física de CMS. Estos jets se caracterizan por tener un vértice desplazado. Redes neuronales son usadas con frecuencia para realizar esta labor; sin embargo, recientemente se han empezado a utilizar redes neuronales convolucionales para incrementar la cantidad de información usada para la toma de decisión, con resultados prometedores.<sup>11</sup> Estos mismos algoritmos han sido adaptados para identificar otros objetos pesados, como quarks tipo t, mostrando una alta eficiencia.<sup>12</sup>

Técnicas de análisis multivariado son empleadas con frecuencia en los estudios de las propiedades del boson de Higgs. Un ejemplo claro es el caso del canal de decaimiento  $H \rightarrow \gamma\gamma$  en el que, entre otras aplicaciones, se utilizan árboles de decisión (BDTs, *boosted decision trees*) para clasificar los eventos en los diferentes modos de producción.<sup>13</sup>

---

<sup>10</sup> A. Di Florio *et al.*, “Convolutional Neural Network for Track Seed Filtering at the CMS HighLevel Trigger,” *J. Phys.: Conf. Ser.* 1085, (2018): 042040.

<sup>11</sup> CMS Collaboration, “Heavy flavor identification at CMS with deep neural networks,” *CMS-DP-2017-005*, (2017).

<sup>12</sup> CMS Collaboration, “Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment,” *CMS PAS JME-18-002*, (2019).

<sup>13</sup> CMS Collaboration, “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *CMS-HIG-16-040*, (2018).

## Calidad de datos

El aprendizaje supervisado se utiliza frecuentemente en la certificación semi-automática de los datos.<sup>14</sup> Antes, un experto deberá revisar cada una de las corridas (divididas en pequeñas secciones de luminosidad) y evaluar si los datos eran buenos para análisis o no. Ahora, ML es utilizado para predecir la decisión que tomará el experto. Puede etiquetar los datos como *good* (blanco) or *bad* (negro) a partir de criterios previamente definidos. Si la decisión cae en una zona gris, la corrida debe ser revisada por un experto.

En el monitoreo de los detectores de muones DT también se utilizan redes neuronales convolucionales para clasificar imágenes del número de *hits* detectados en los módulos.<sup>15</sup> Si hay mucho ruido, los datos asociados a dicho periodo de tiempo se etiquetan como problemáticos.

## Cómputo

Una de las posibles áreas de oportunidad para la implementación de técnicas de ML es el análisis de los metadatos de los procesos computacionales de CMS, así como encontrar patrones sobre el uso de los sitios del *Grid*,<sup>16</sup> el flujo de trabajo, la transferencia de datos, servicios, accesos, etc. En lugar de simplemente monitorear el uso reciente para resolver problemas urgentes, se puede hacer un estudio profundo para analizar todos esos metadatos disponibles y diseñar un modelo computacional efectivo.

---

<sup>14</sup> K. Deja, “Using machine learning techniques for Data Quality Monitoring in CMS and ALICE,” *Proceedings of Science (LHCP 2019)*: 236.

<sup>15</sup> C. Calabria, “Monitoring tools for the CMS muon detector: present workows and future automation,” *EPJ Web of Conferences (CHEP 2018)* 214, (2019): 06001.

<sup>16</sup> LCG Collaboration, “LHC Computing Grid - Technical Design Report,” *LCG-TDR-001* (2005).

## El boom de la interdisciplina

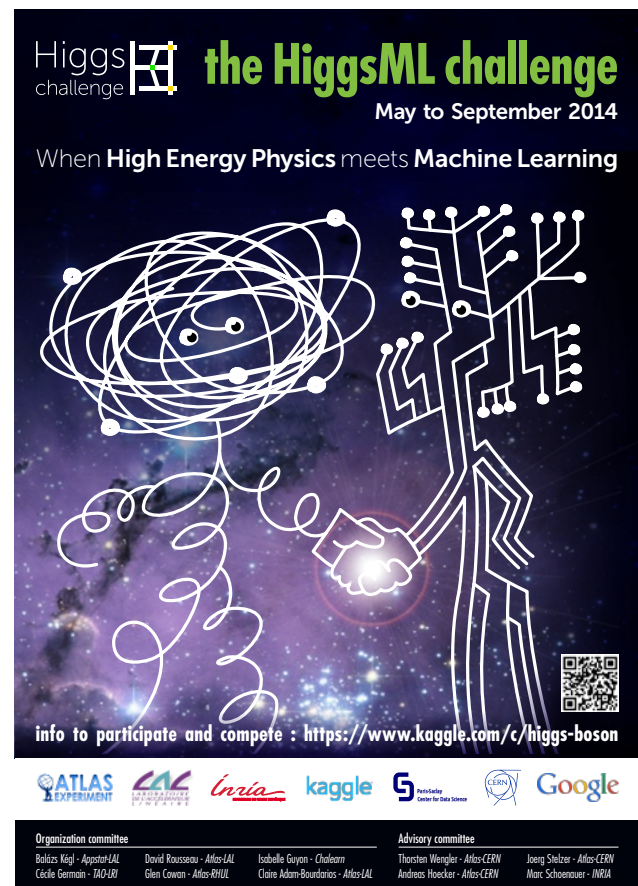
Dentro de CMS se ha conformado un grupo de investigación dedicado a la implementación de técnicas de ML. En 2016 se formó el grupo de trabajo IML (*Inter-Experimental LHC Machine Learning Working Group*) con representantes de los cuatro experimentos del LHC.<sup>17</sup> Sus objetivos son facilitar la comunicación entre los experimentos y la comunidad de ML, brindar soporte para la integración de técnicas de ML en el software del LHC, así como organizar eventos de difusión y entrenamiento para la comunidad experimental.

Las colaboraciones del LHC han extendido sus horizontes al lanzar retos de ML al público en general para buscar ideas. Uno de ellos fue el *Higgs Boson Machine Learning Challenge* lanzado por el experimento ATLAS en 2014.

Además del trabajo interno, las colaboraciones del LHC han extendido sus horizontes al lanzar retos de ML al público en general para buscar ideas. Uno de ellos fue el *Higgs Boson Machine Learning Challenge* lanzado por el experimento ATLAS en 2014, en el que

<sup>17</sup> IML, “Who we are. Ginebra, Suiza,” *IML: Inter-Experimental LHC Machine Learning Working Group* (2018). <https://iml.web.cern.ch/> Revisado el 20 de marzo, 2020.

se buscaba utilizar técnicas de ML para identificar al bosón de Higgs en un conjunto de datos.<sup>18</sup> Ninguno de los ganadores del concurso era físico de partículas. El primer lugar fue un joven matemático con maestría en inteligencia artificial, mientras que el segundo lugar contaba con un doctorado en econometría. Esto nos demuestra que el trabajo colaborativo e interdisciplinar es cada vez más relevante.



Poster del Higgs Boson Machine Learning Challenge.

<sup>18</sup> Kaggle Inc., “Higgs challenge,” *Kaggle* (2018). <https://www.kaggle.com/c/higgs-boson> Revisado el 20 de marzo, 2020.

## Conclusiones

Las técnicas de *Machine Learning* juegan un papel cada vez más importante en las colaboraciones del LHC y en la física de partículas en general. La incorporación de más métodos –como *Deep Learning*– va en aumento con aplicaciones diversas. Sin embargo, es importante resaltar que se requiere de trabajo interdisciplinario con especialistas de diversas áreas para poder lidiar con la gran cantidad de información que se genera y, sobre todo, para poder interpretarla correctamente. El objetivo principal de esta reseña es iniciar una conversación entre integrantes de las distintas disciplinas para buscar formas de colaboración y beneficio mutuo. —

### Referencias

- ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *JINST* 3, (2008): S08002.
- ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B* 716, (2012): 1.
- ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider,” *JINST* 3, (2008): S08003.
- Calabria, C., “Monitoring tools for the CMS muon detector: present workows and future automation,” *EPJ Web of Conferences (CHEP 2018)* 214, (2019): 06001.
- Charpak, G. *et al.*, “The use of multiwire proportional counters to select and localize charged particles,” *Nucl. Instr. Methods* 62, (1968): 262.
- CMS Collaboration, “Heavy flavor identification at CMS with deep neural networks,” *CMS-DP-2017-005*, (2017).
- CMS Collaboration, “Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment,” *CMS PAS JME-18-002*, (2019).
- CMS Collaboration, “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *CMS-HIG-16-040*, (2018).
- CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B* 716, (2012): 30.
- CMS Collaboration, “The CMS experiment at the CERN LHC,” *JINST* 3, (2008): S08004.
- Deja, K., “Using machine learning techniques for Data Quality Monitoring in CMS and ALICE,” *Proceedings of Science (LHCP 2019)*: 236.
- Di Florio, A. *et al.*, “Convolutional Neural Network for Track Seed Filtering at the CMS HighLevel Trigger,” *J. Phys.: Conf. Ser.* 1085, (2018): 042040.
- Englert, F. and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.* 13, (1964): 321.
- Farrell, S. *et al.*, “The HEP.TrkX Project: deep neural networks for HL-LHC online and offline tracking,” *EPJ Web Conf.* 150, (2017): 00003.
- Higgs, P. W., “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.* 13, (1964): 508.
- Higgs, P. W., “Broken symmetries, massless particles and gauge fields,” *Phys. Lett.* 12, (1964): 132.
- IML, “Who we are. Ginebra, Suiza,” *IML: Inter-Experimental LHC Machine Learning Working Group* (2018). <https://iml.web.cern.ch/> Revisado el 20 de marzo, 2020.
- Kaggle Inc., “Higgs challenge,” *Kaggle* (2018). <https://www.kaggle.com/c/higgs-boson> Revisado el 20 de marzo, 2020.
- LCG Collaboration, “LHC Computing Grid-Technical Design Report,” *LCG-TDR-001* (2005).
- LHCb Collaboration, “The LHCb detector at the LHC,” *JINST* 3, (2008): S08005.
- SAS Institute Inc., “Machine Learning. What it is and why it matters,” *SAS-THE POWER TO KNOW.* (2018) [https://www.sas.com/en\\_us/insights/analytics/machine-learning.html](https://www.sas.com/en_us/insights/analytics/machine-learning.html) Revisado el 20 de marzo, 2020.

---

El texto pertenece al Seminario de Investigación de Ciencia, Tecnología, Ingeniería y Matemáticas. PAIDI/007/18 de la FES Acatlán, UNAM.