MILLENNIUM INSTITUTE FOR SUBATOMIC PHYSICS AT HIGH-ENERGY FRONTIER SAPHIR

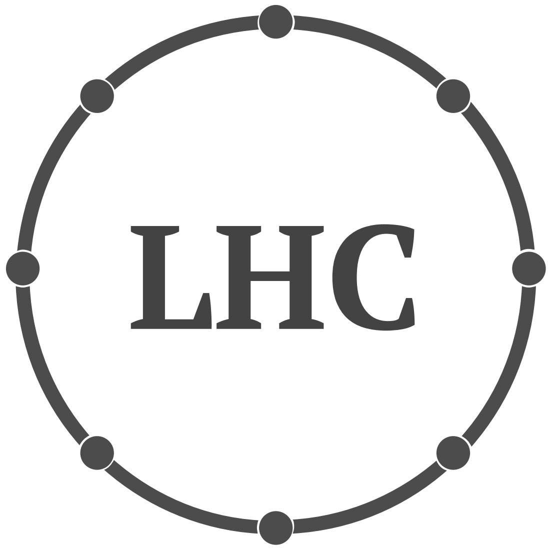
Universidad Andrés Bello

Machine Learning at SAPHIR

SAPHIR ARM 2023

Sebastian Olivares | January 2023

Machine Learning is a method to teach computers to learn from **data**, without being explicitly programmed



Data at a rate of about: 1 petabyte/second

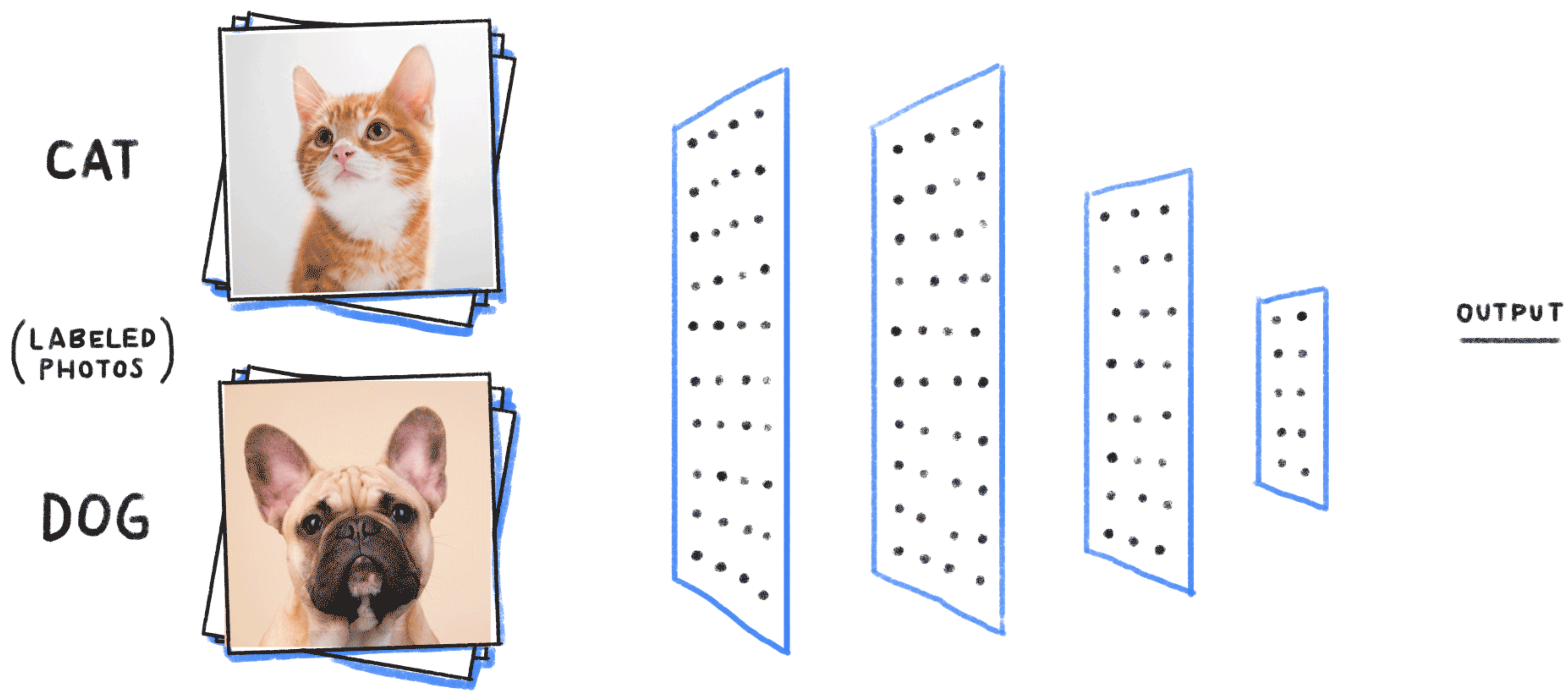
Higgs boson is produced only once every few billion proton-proton collisions

100 million detection elements

After data-reduction executed in real-time: 50 terabytes/second

Store as much data every hour as Facebook collect globally in a year

Particle physics is governed by quantum mechanics



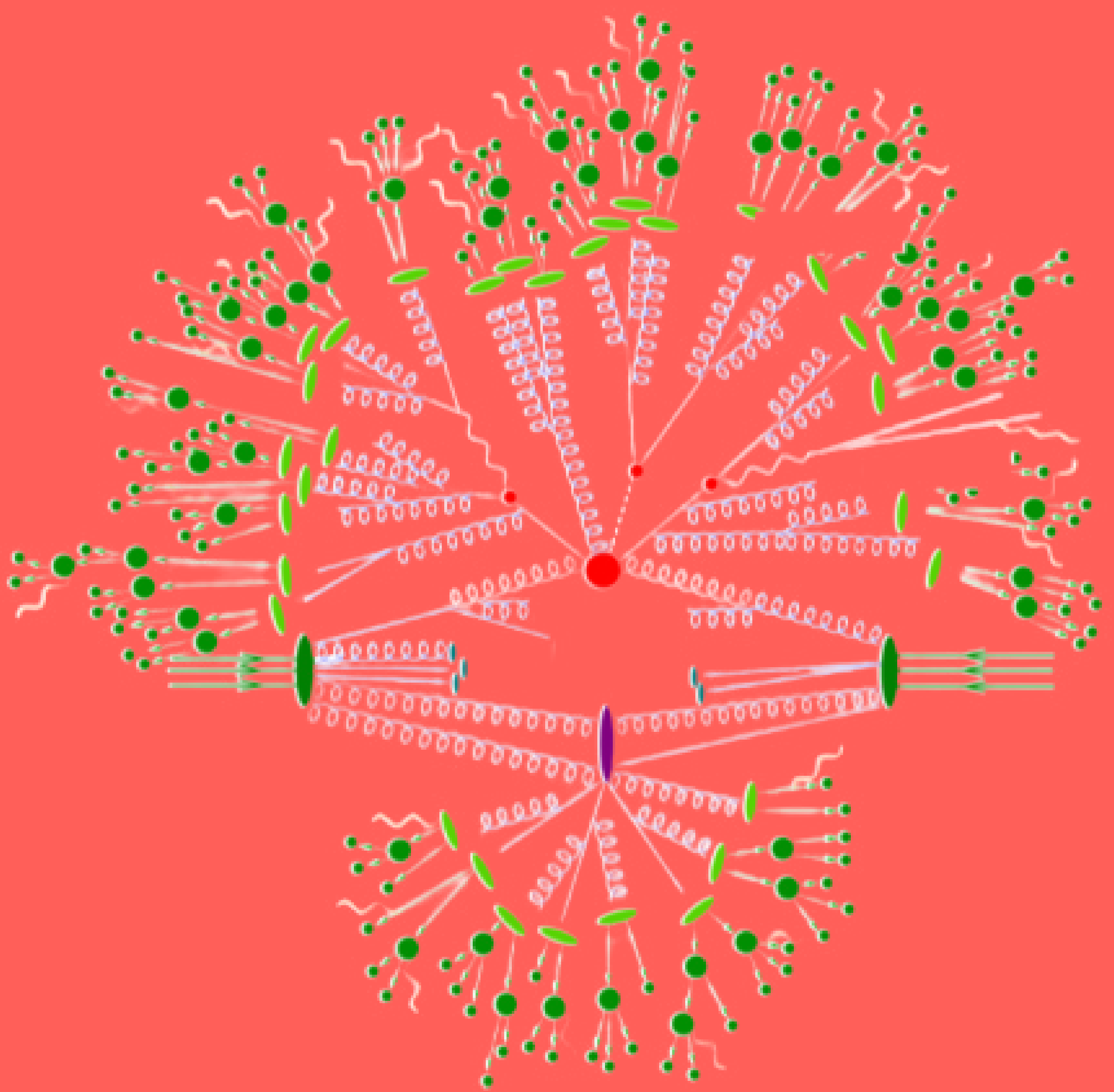
Particle physics is governed by quantum mechanics

Was there a Higgs boson in this event?

Remarkably accurate simulated data

We have MC!

Detectors accurate from 10^-6 to 100 m



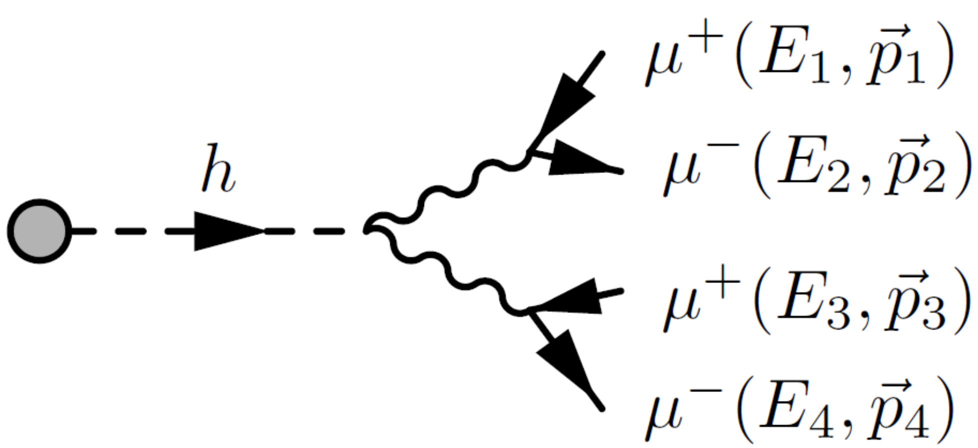
How-to do an analysis

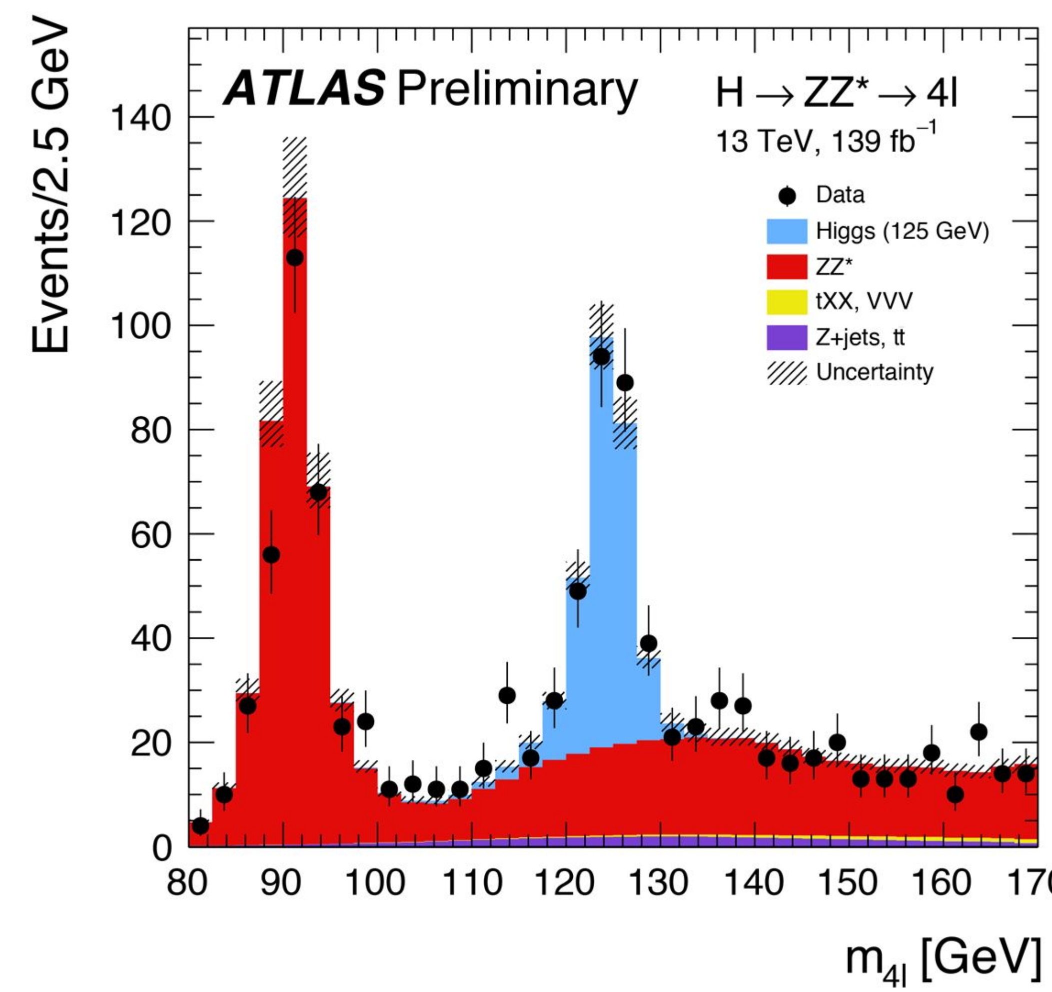
1. Process raw sensor data into low-level objects (e.g. calorimeter clusters and tracks)
2. Estimate the energy, momentum and identity of individual particles
3. Event selection algorithms selects subsets of the collision data for further analysis on the basis of the information associated to individual events
4. Cumulative product of these steps reduce the dimensionality of the problem to a number small enough to allow the missing statistical model to be estimated using simulated samples

Higgs boson discovery with the four-lepton channel

$$m = \sqrt{(E_1 + E_2 + E_3 + E_4)^2 - (\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_4)^2}$$

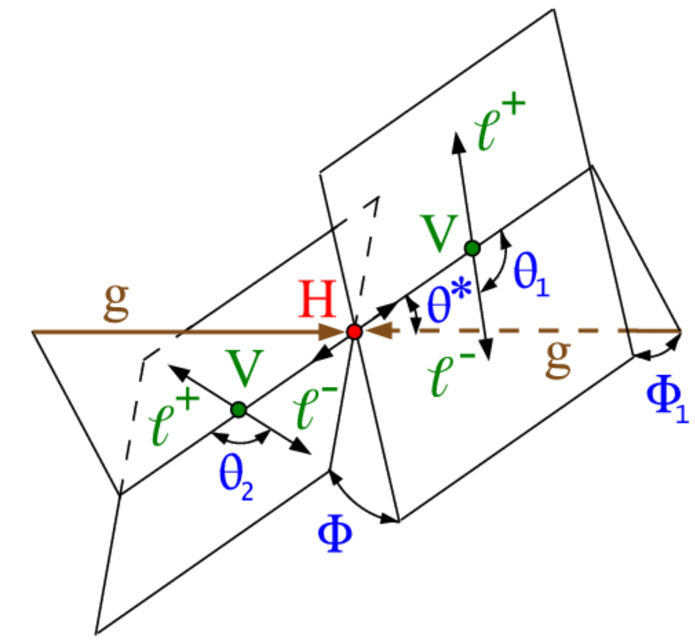
Occurs roughly once every few billion proton-proton collisions



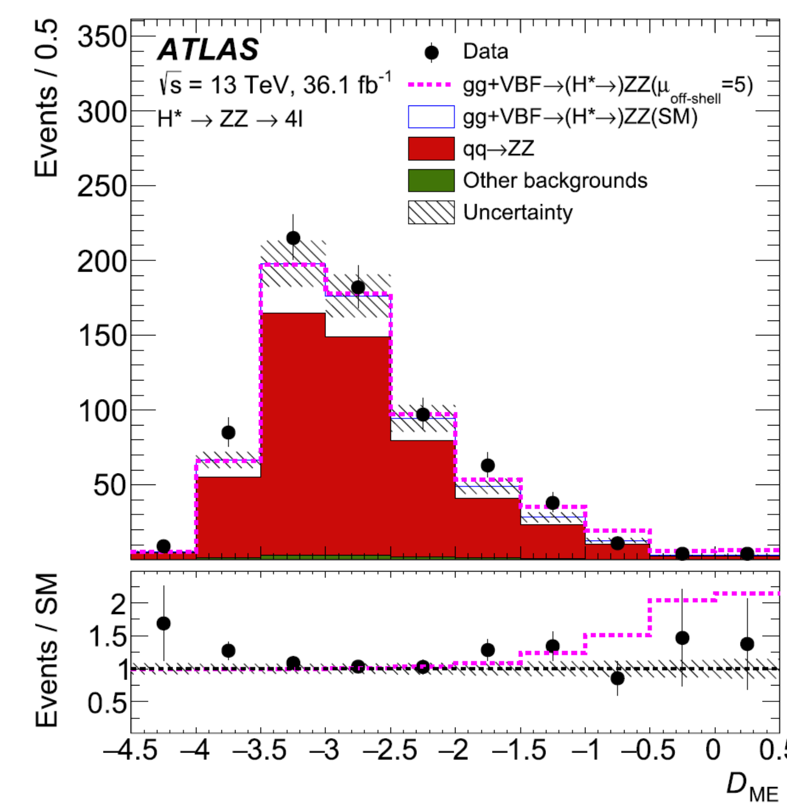


What feature is the optimal way of statistically discriminating a signal from its background?

Higgs boson discovery with the four-lepton channel



Particle physicists have sought to improve the power of their analysis by employing algorithms that utilize **mutiple variables** simultaneously

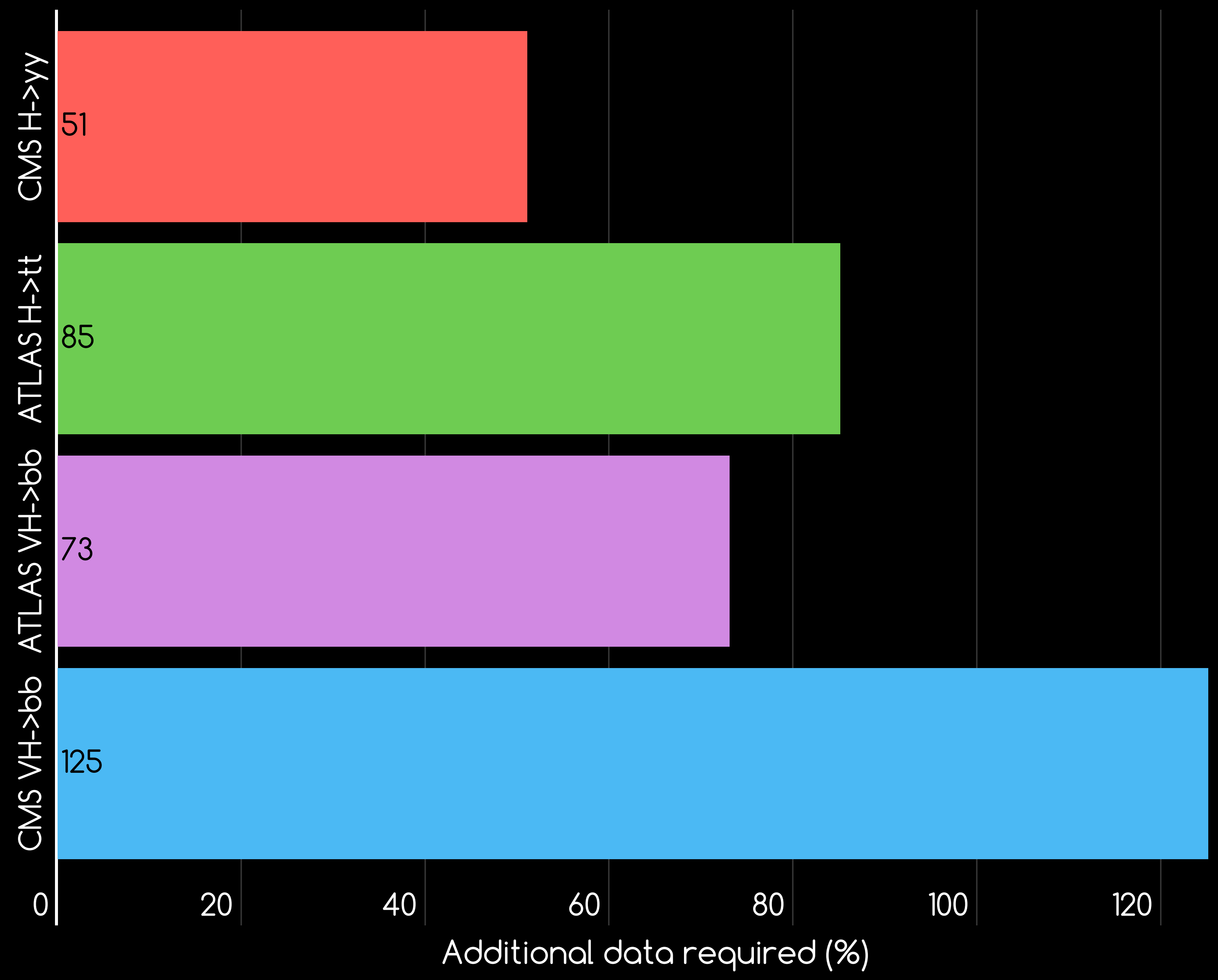


multi-variate analisis (MVA)

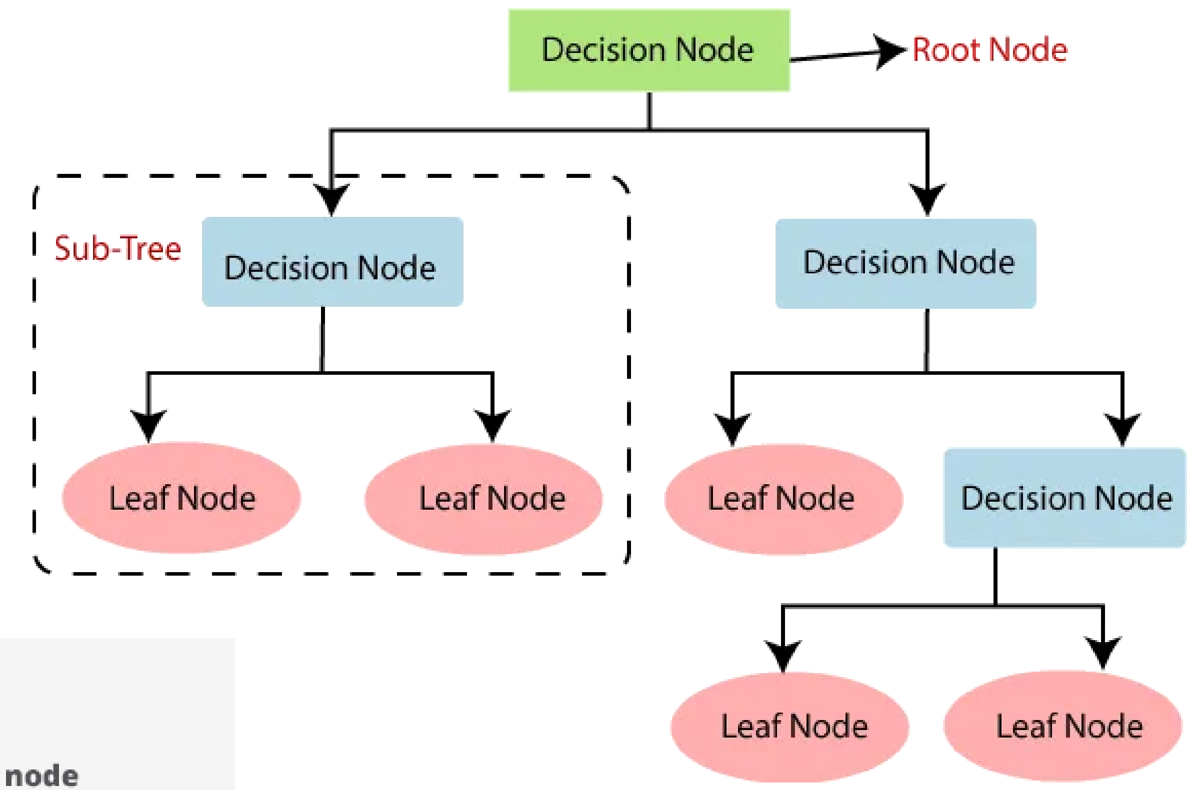
↕

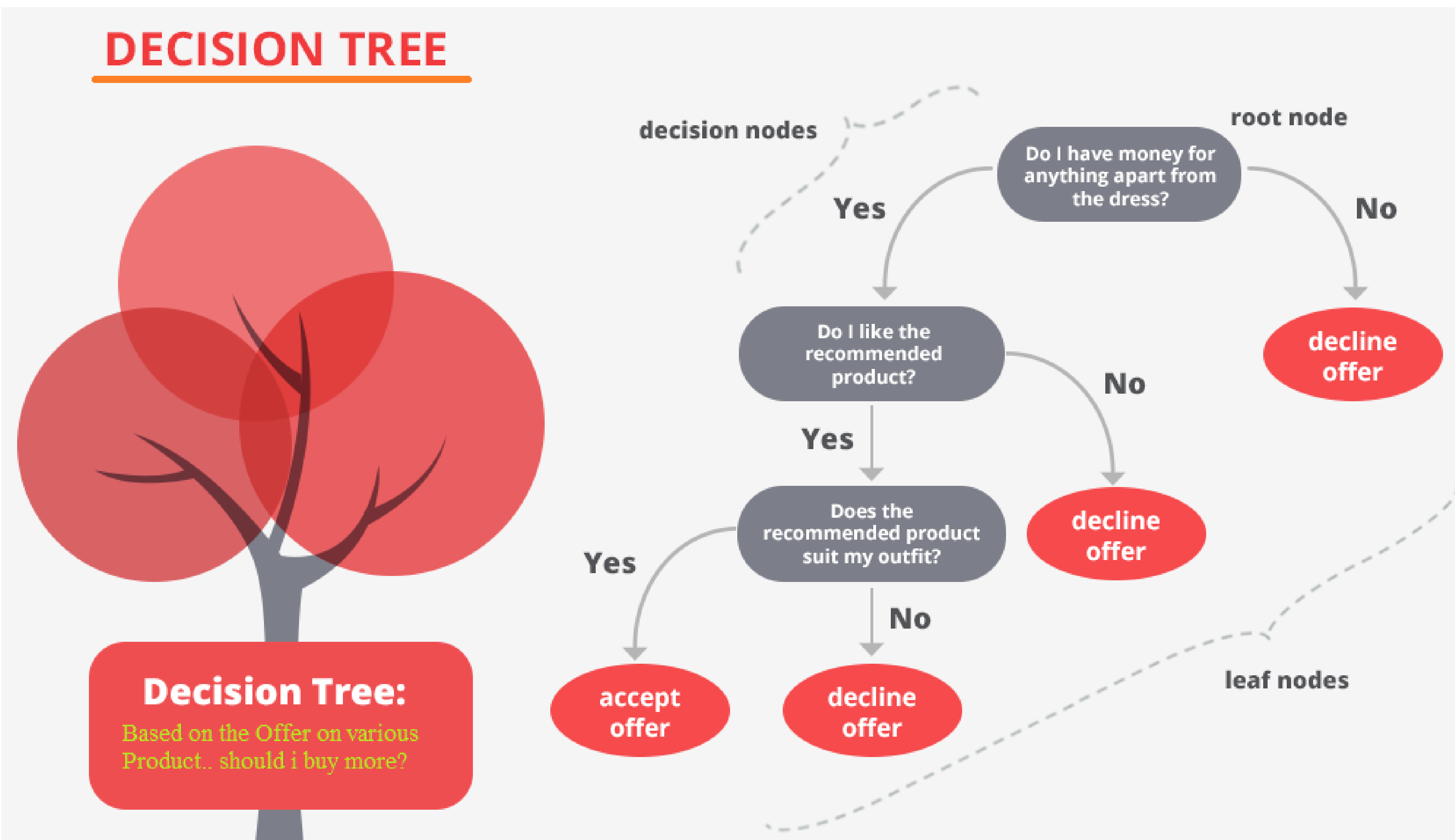
machine learning (ML)

What if we didn't use a MVA (2011-2012)?



For several years the status quo of machine learning in HEP was to use boosted decision trees (BDT) implemented in the software package TMVA





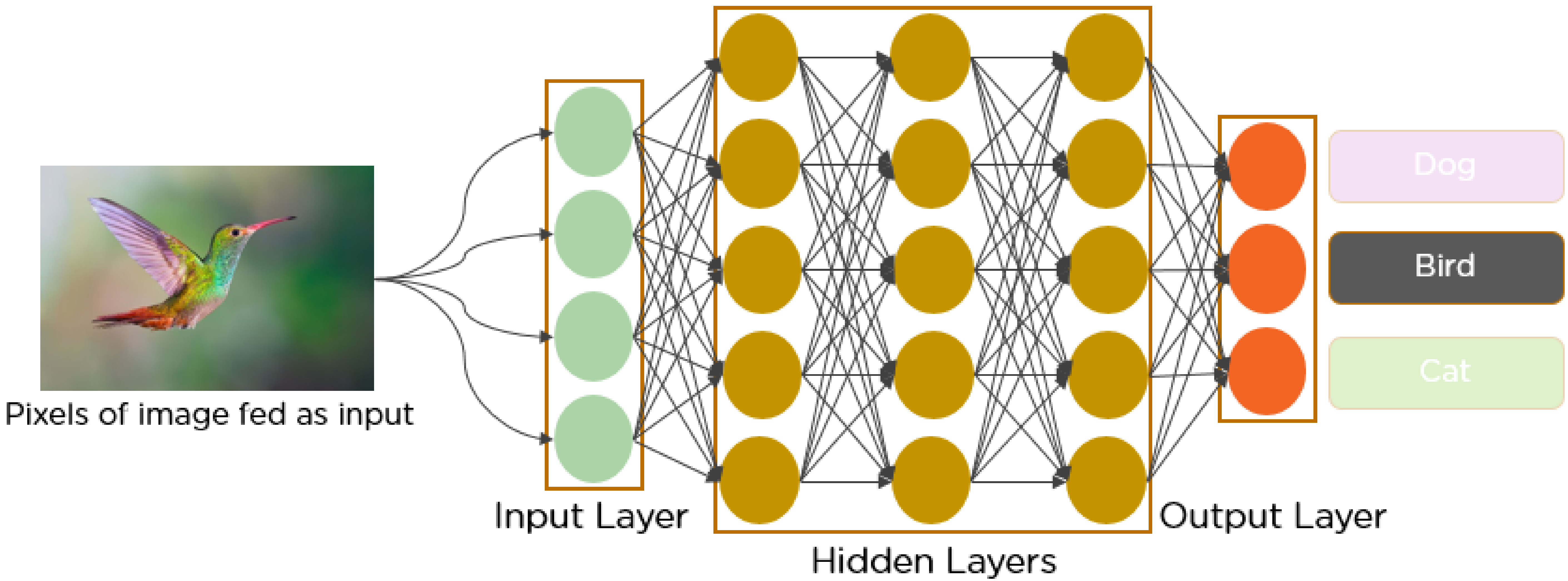
What about my Deep Neural Networks?

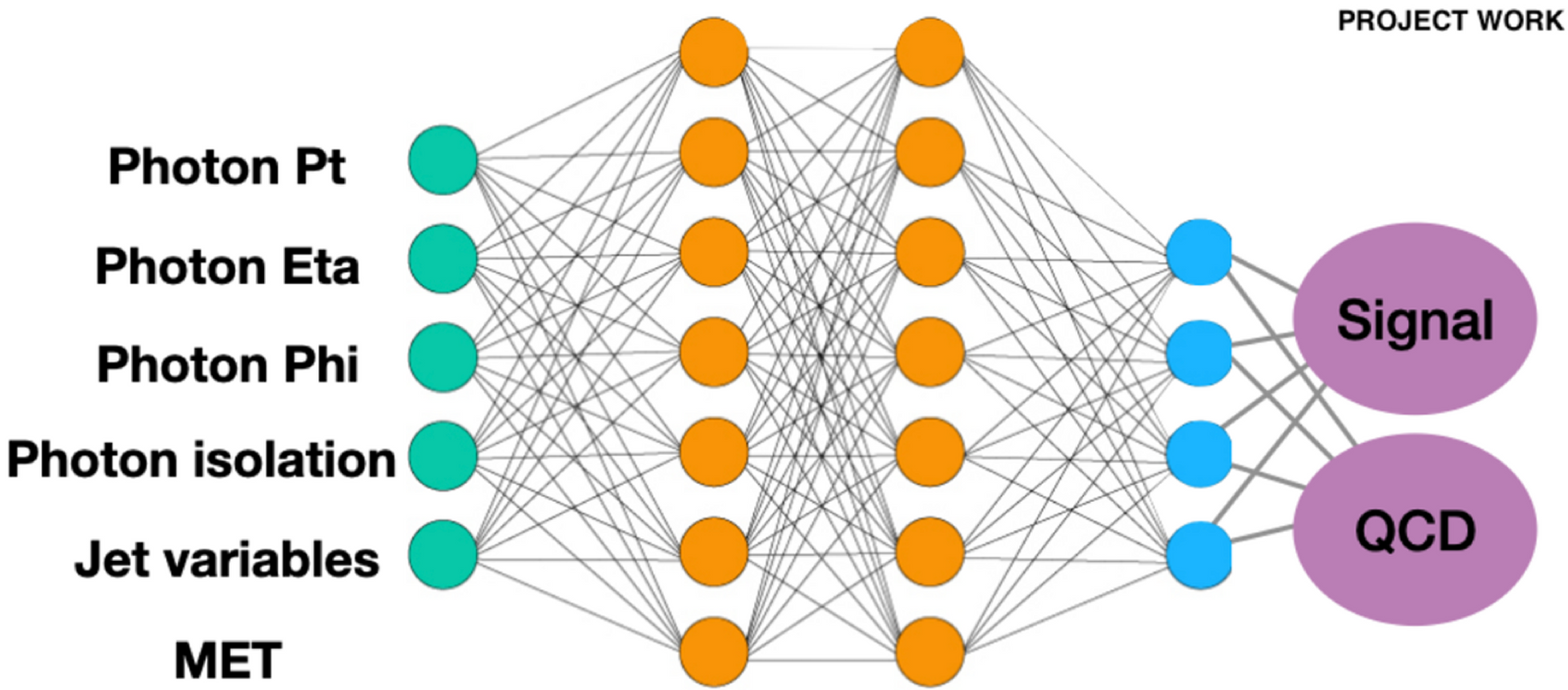
Geoffrey Hinton

Particle physics has traditionally been using physically-motivated classifiers

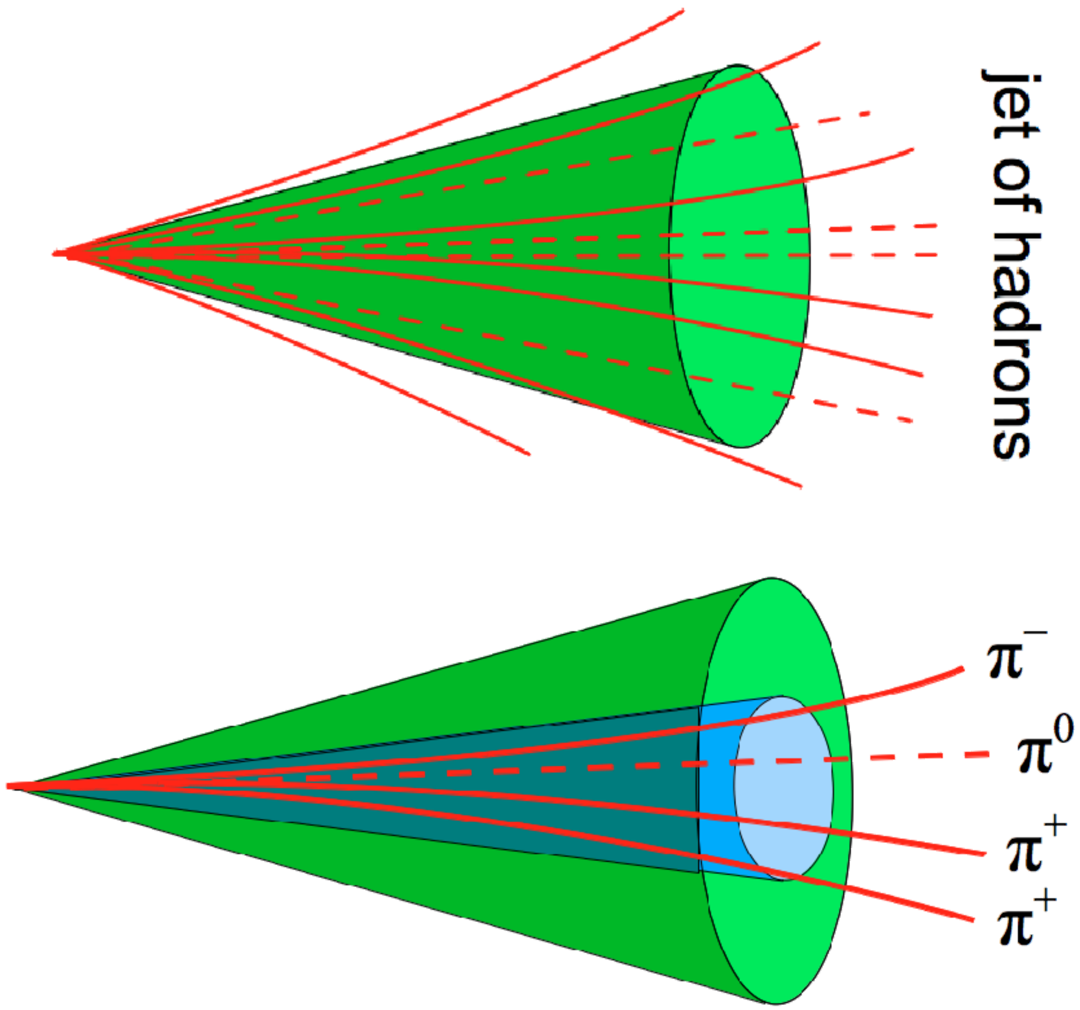
Over the last several years this has been replaced with "modern" machine learning

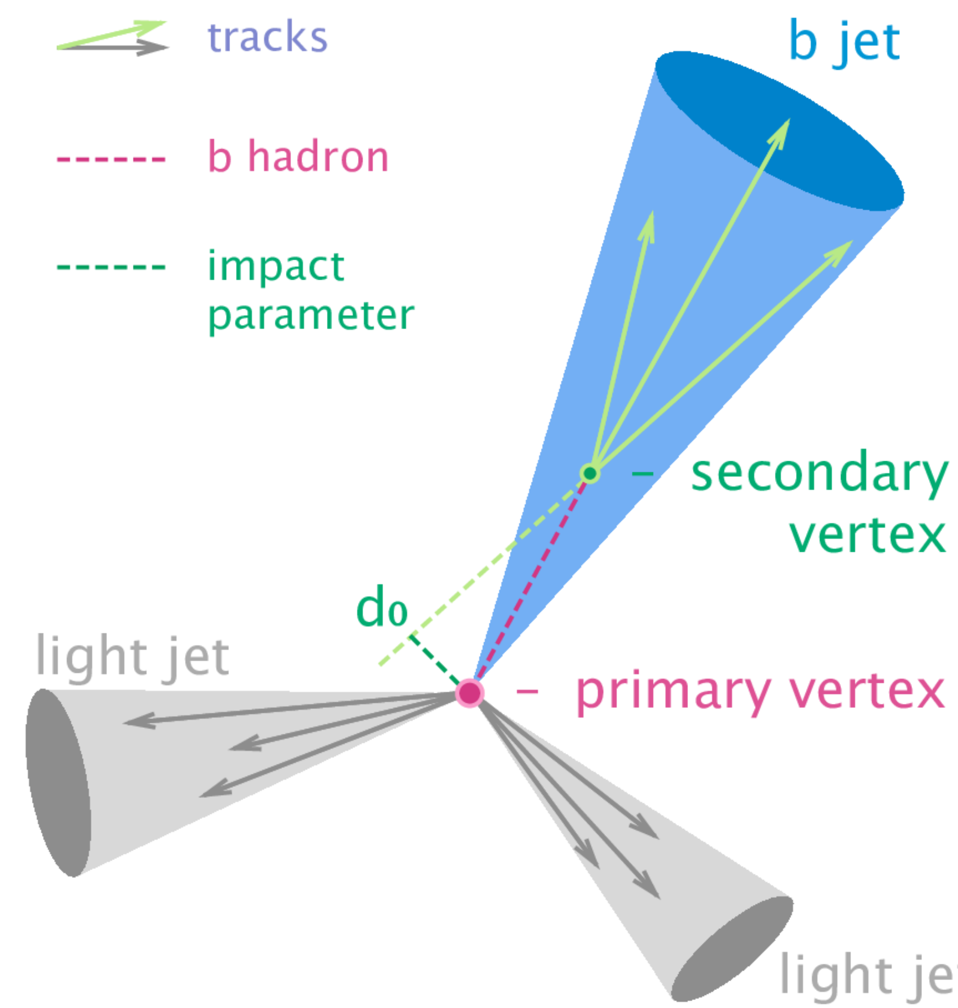
The modern approach is to feed raw, minimally-processed data, rather than high-level physically-motivated variables


Pixels of image fed as input
Input Layer
Hidden Layers
Output Layer
Dog
Bird
Cat


PROJECT WORK
Photon Pt
Photon Eta
Photon Phi
Photon isolation
Jet variables
MET
Signal
QCD

Deep Neural Network for jet categorisation?

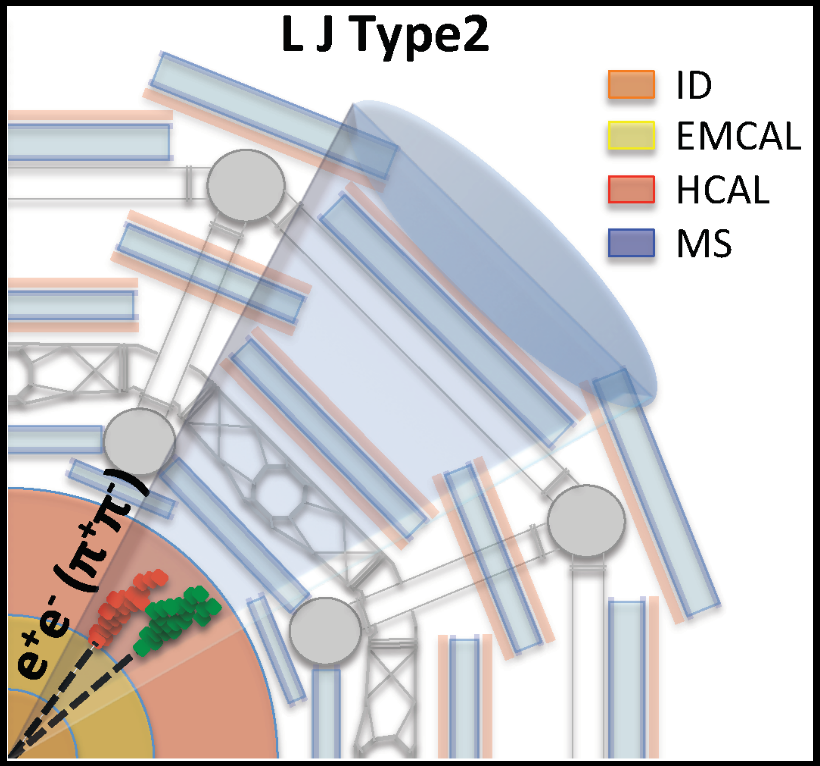


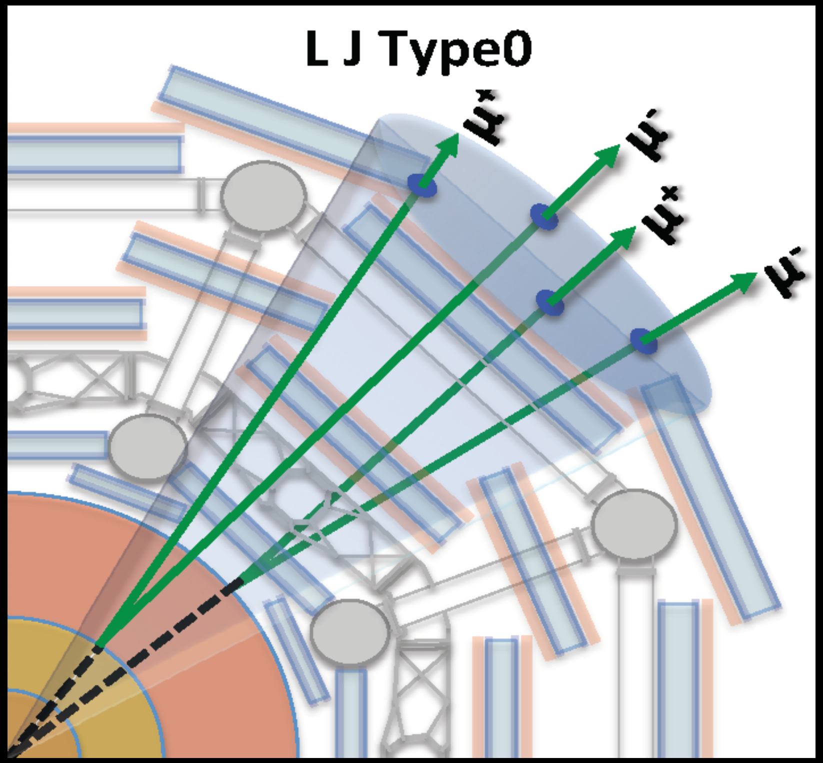


Dark Photons would typically be produced with **large boost** due to their small mass, forming collimated jet-like structures containing leptons and/or hadrons (lepton-jets)

Analysis divides signal **lepton-jets** (LJ) into categories

Selection is based on dark photon jets identification, its displaced vertex signatures and a **DNN** discriminant



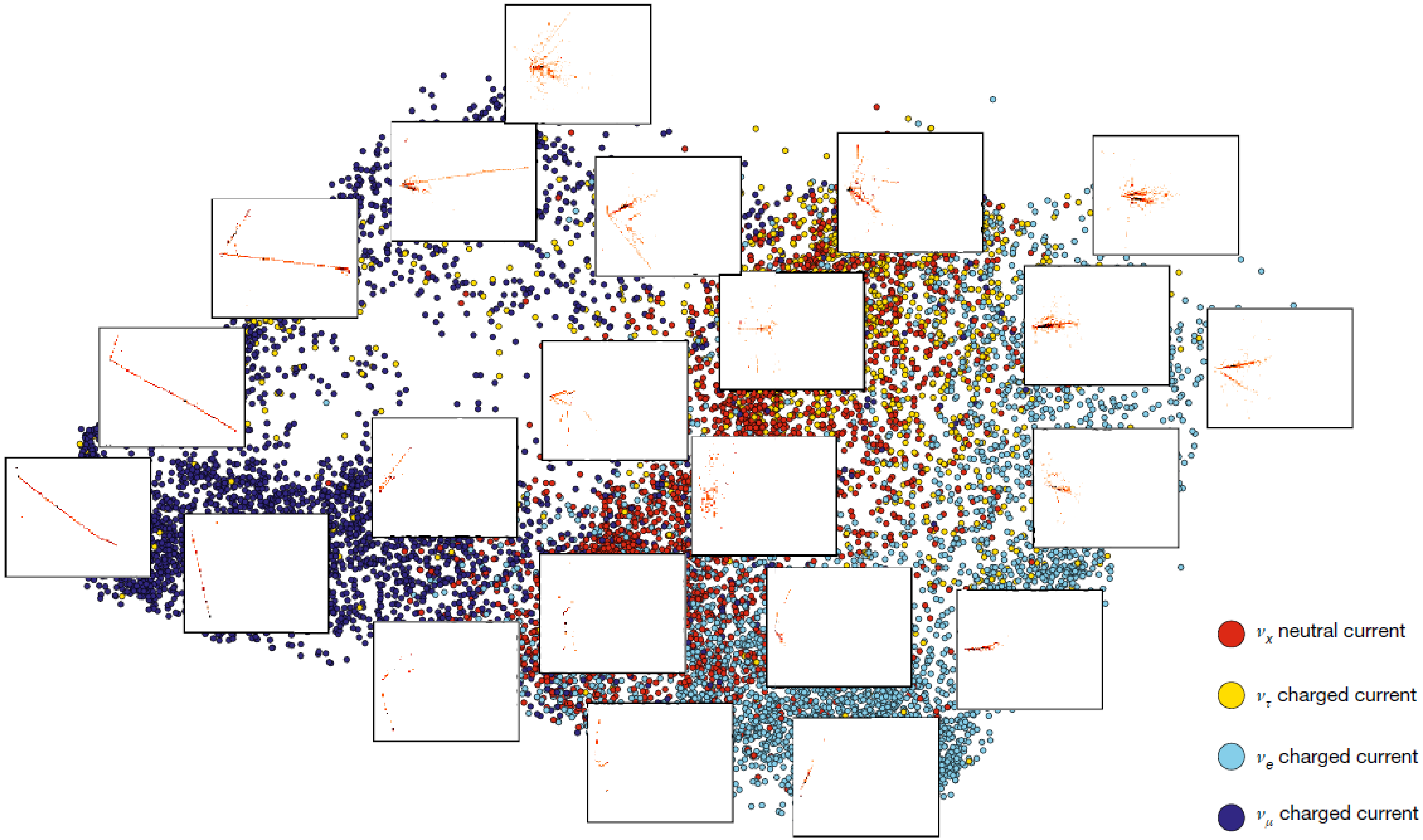


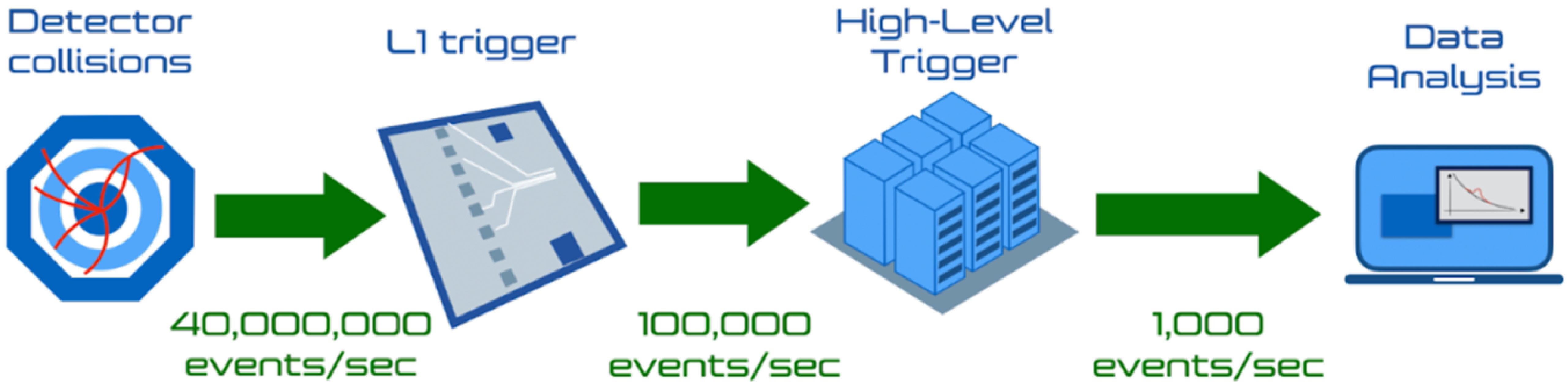
jet vertex tagger → Used the **displaced vertex** to reduce QCD bkg

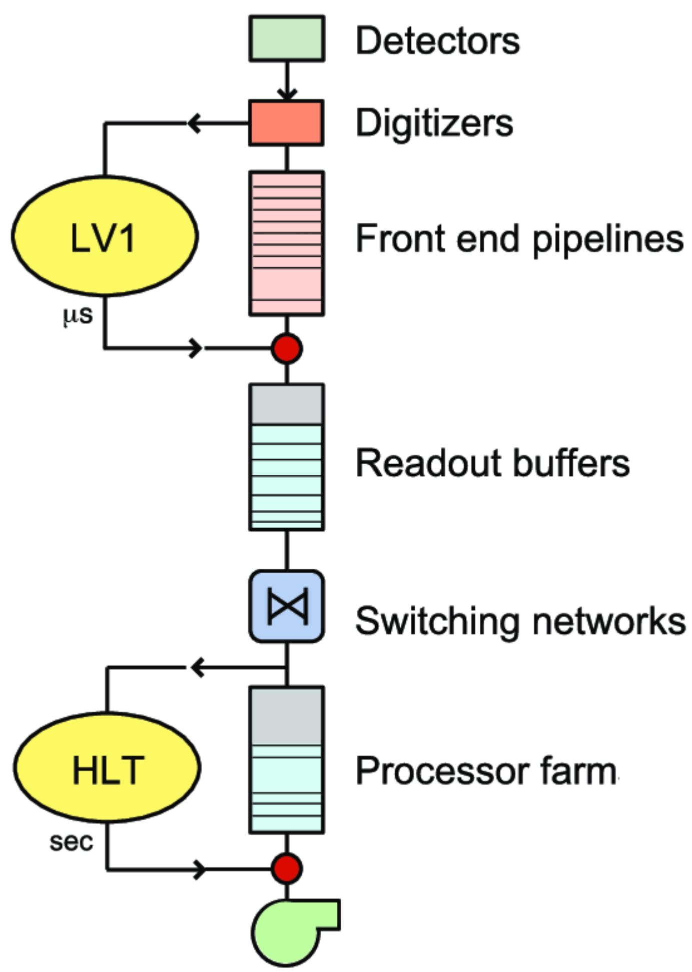
DPJ tagger → Discriminate **DPJ from SM jets**

Beam-induced-background tagger → Reduce particles resulting from interactions between proton bunches and **residual beam elements**

t-SNE Nova event-selection using CNN





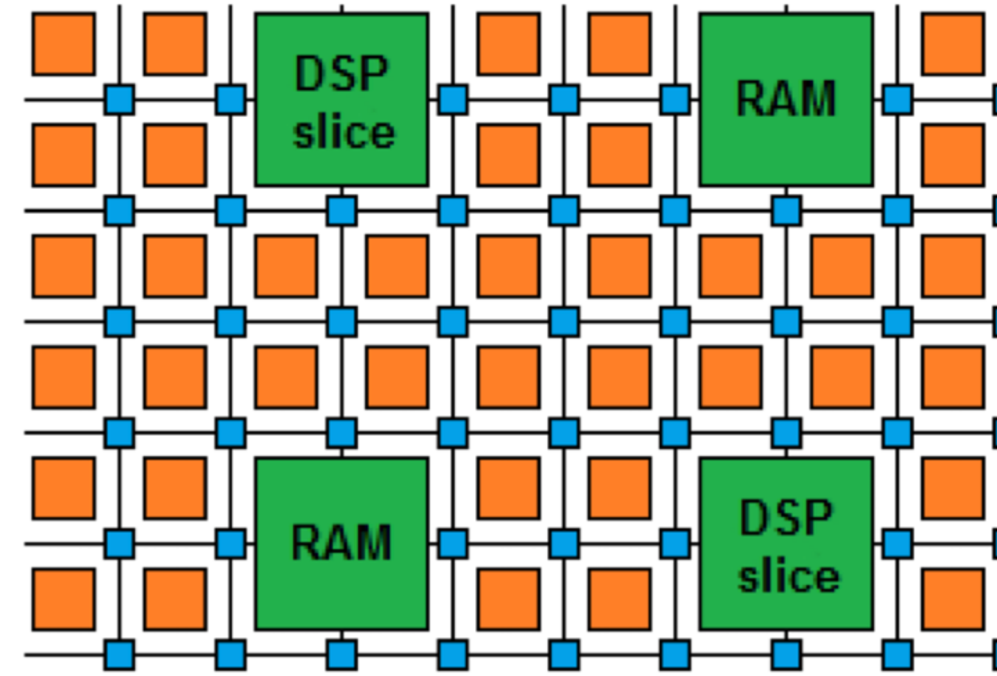


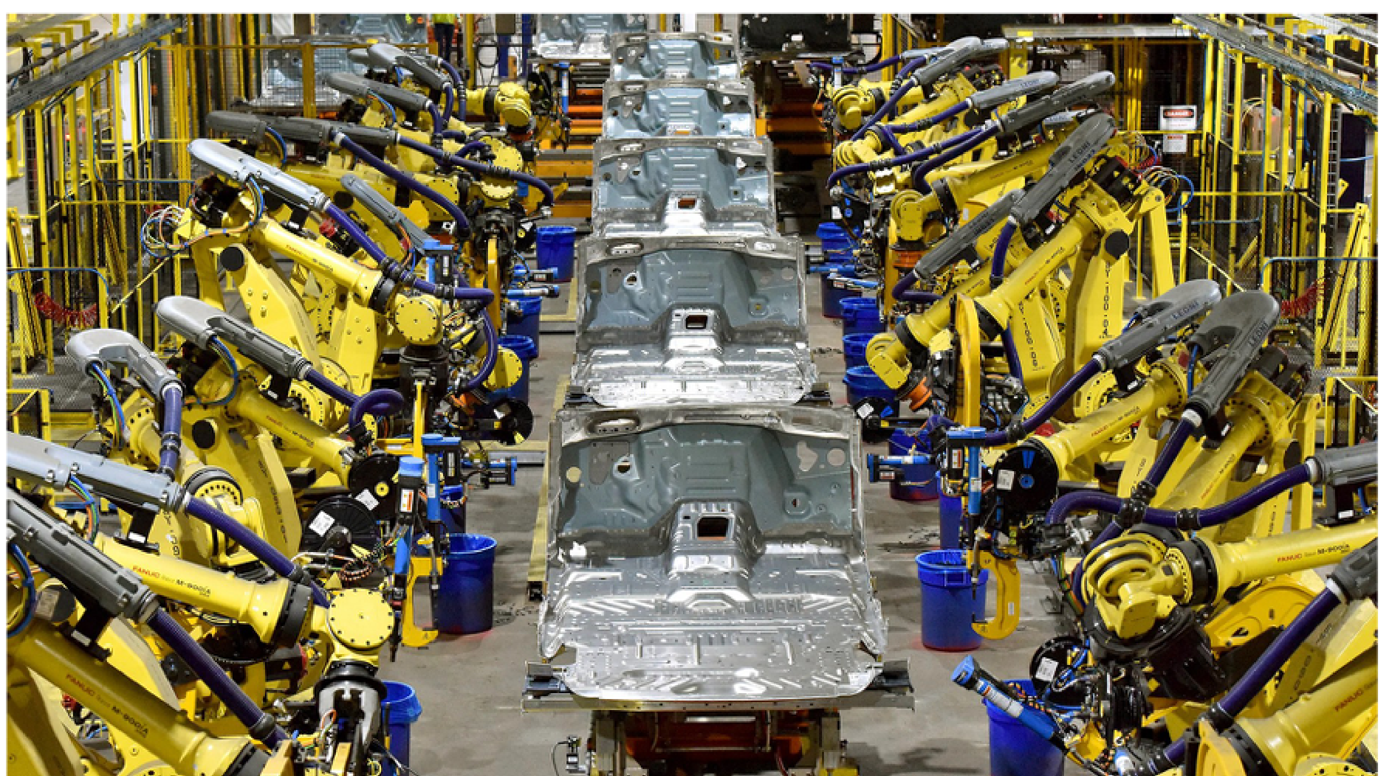
L1 trigger rejects 99.75% of events in 10 micro seconds

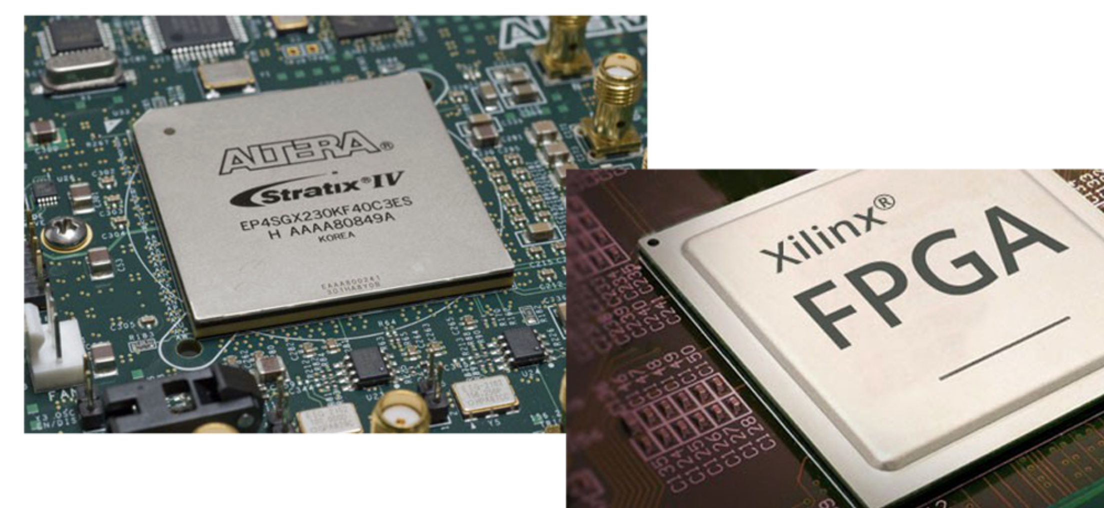
Field Programmable Gate Arrays (FPGA) are used as fast programmable integrated circuits for event selection

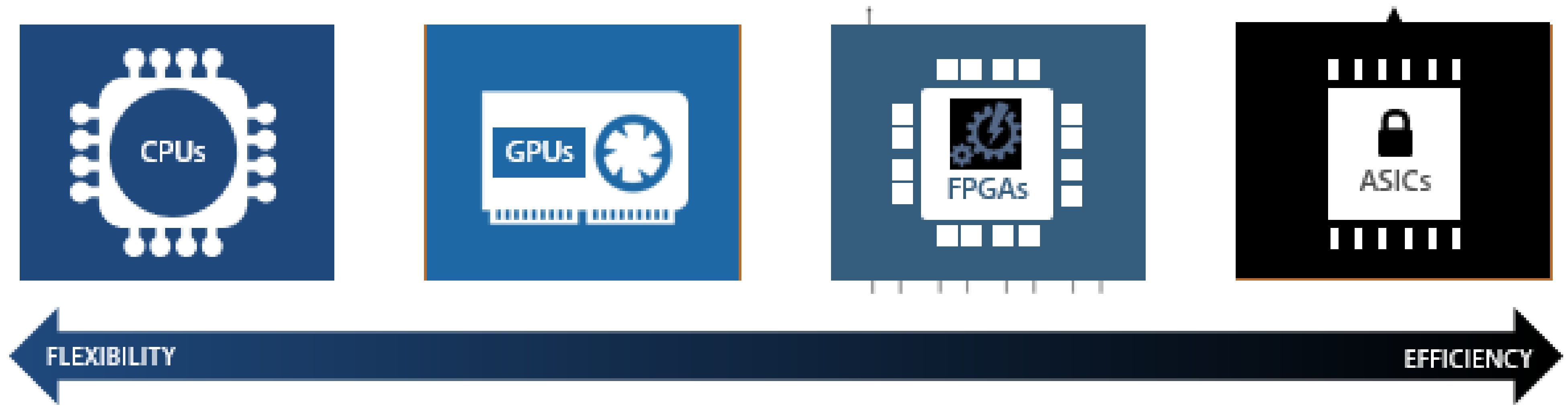
FPGA contains many different building blocks that are connect according to requirements

It uses all resources to work on different features of the selection simultaneously, achieving low latency









Since FPGA are fast parallelizable programmable integrated circuits, can we integrate ML algorithms to improve data categorisation?

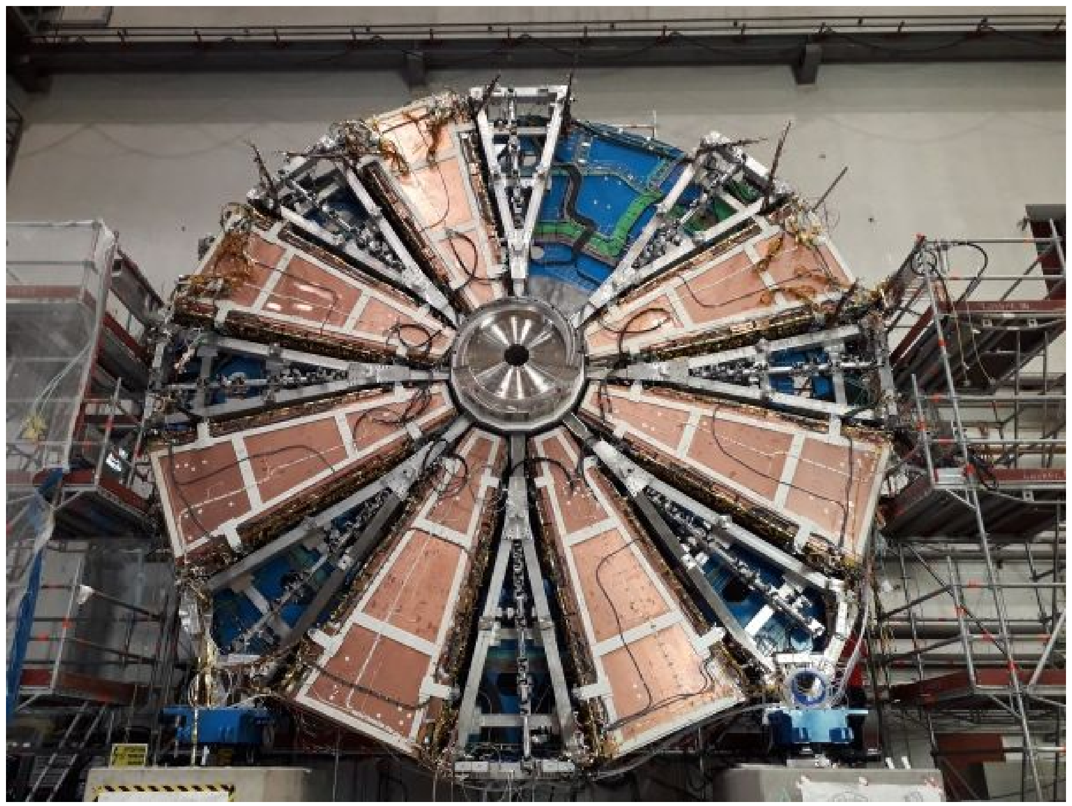
Can we combine ML with FPGA?

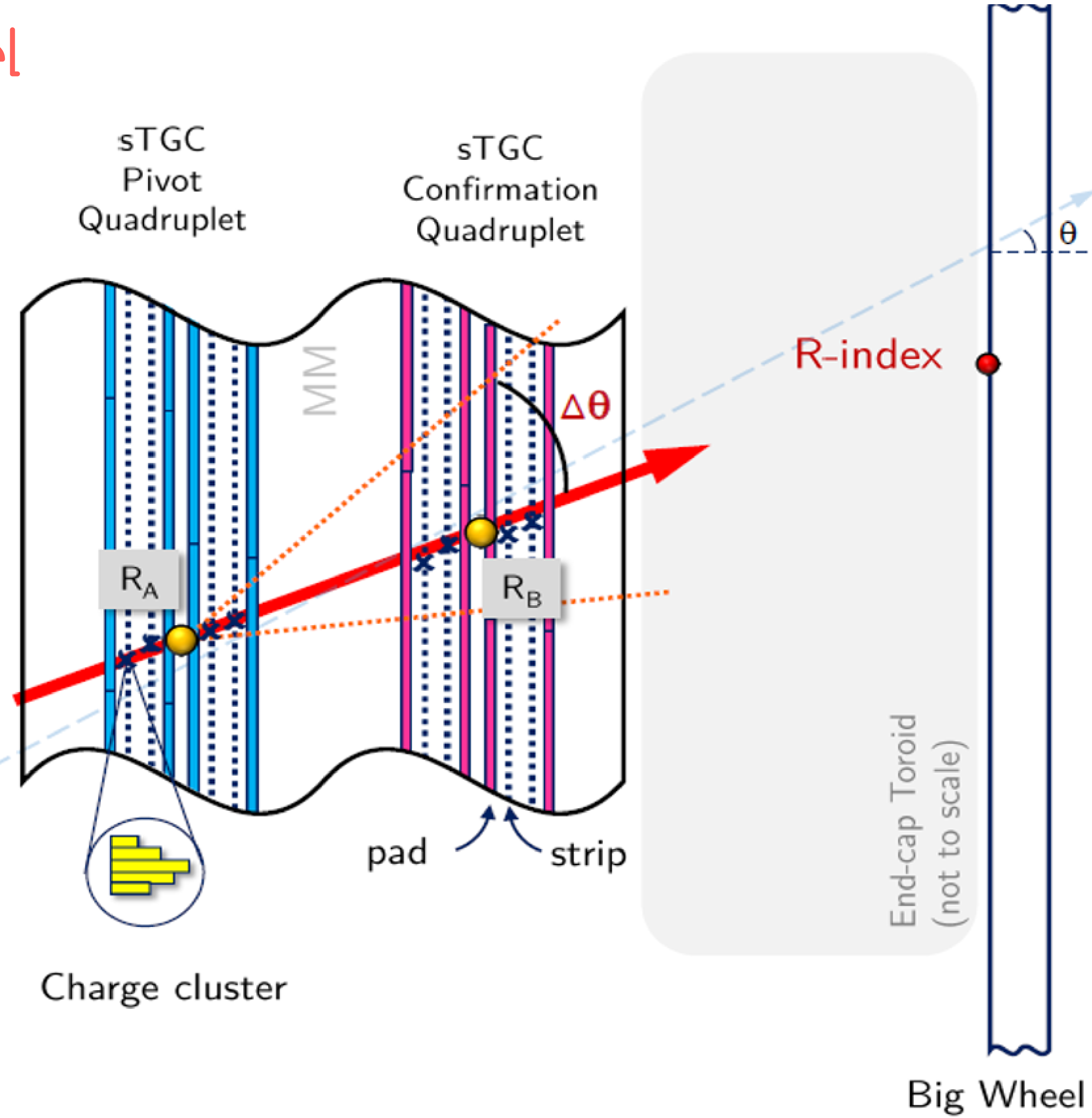
Can we apply other ML algorithms?

Nowadays industry-based FPGAs (e.g. Xilinx) can be programable with ML algorithms such as RNN, LSTM, CNN (VGG16), and even attention-augmented LSTM

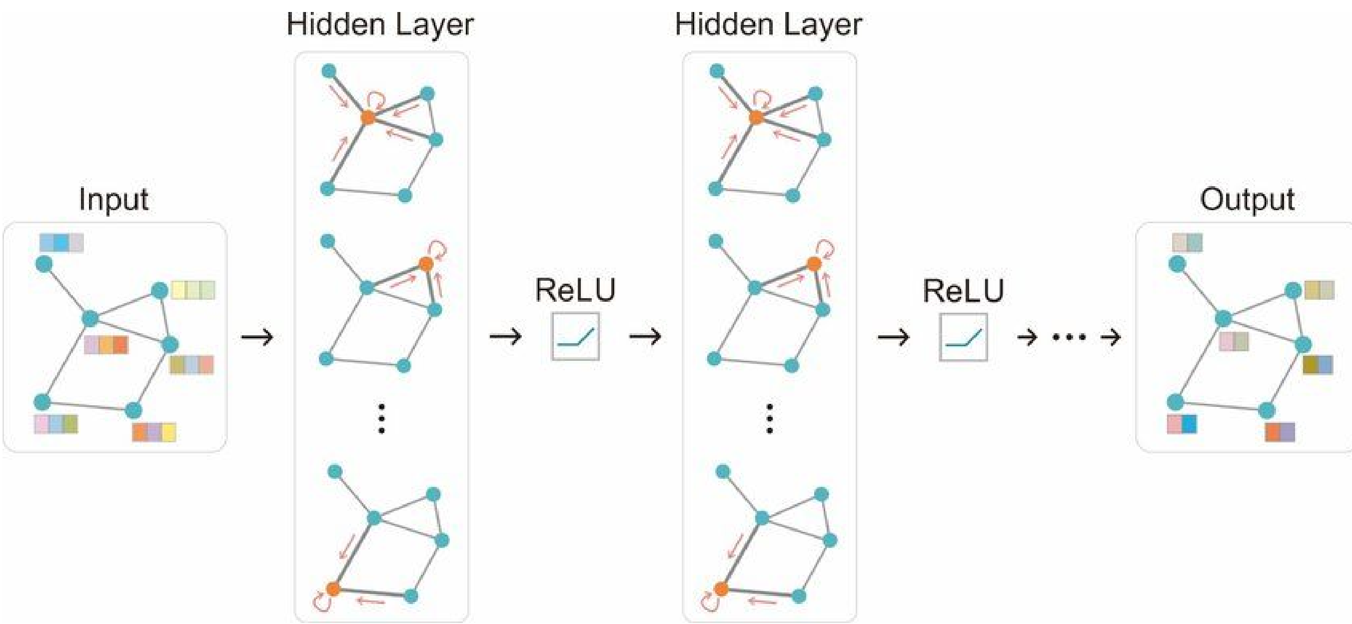
But do we have the flexibility to develop our own algorithms to be included in the FPGA?

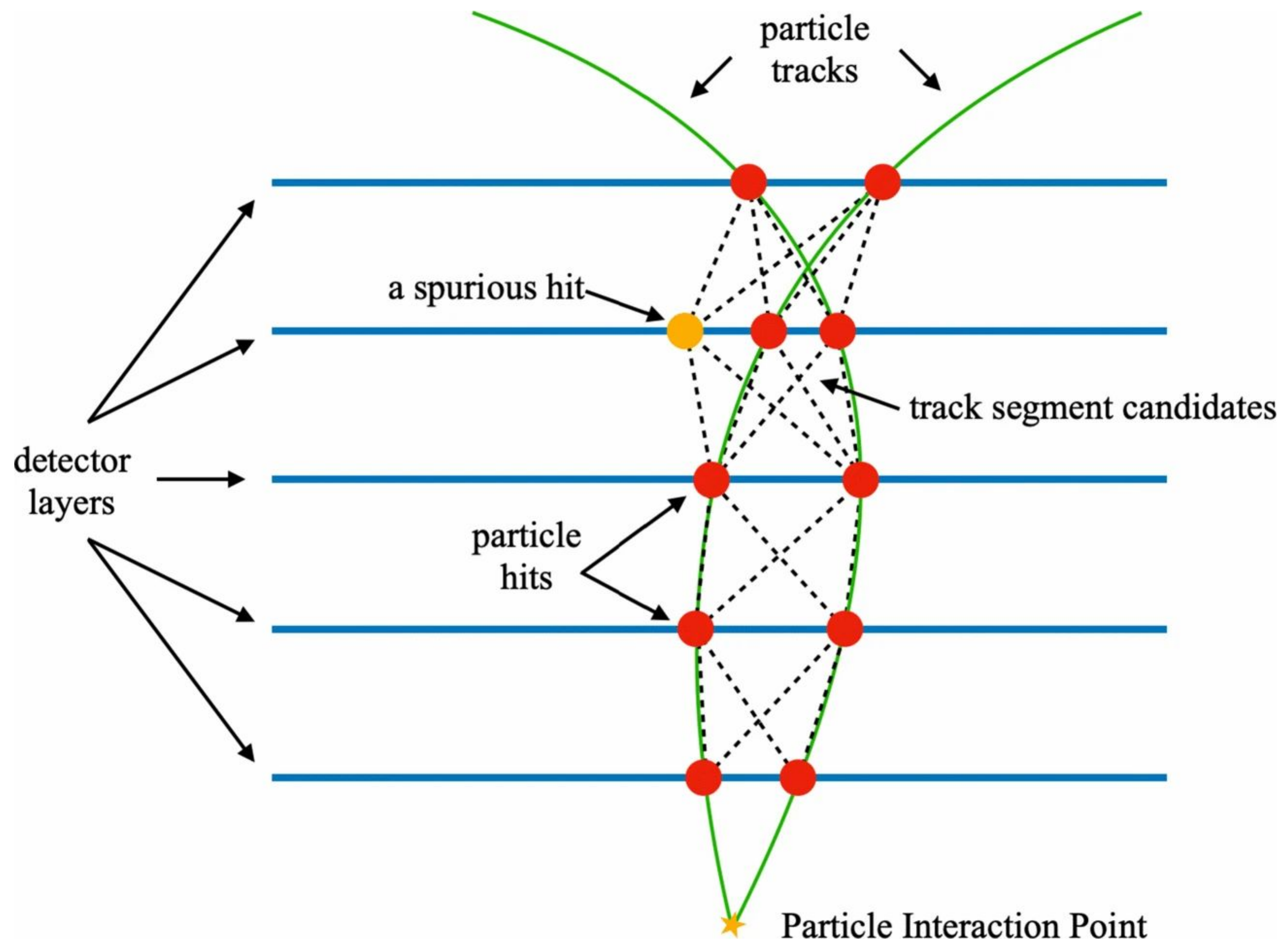
ATLAS New Small Wheel





Can we take advantage of the NSW for L1 Muon trigger?

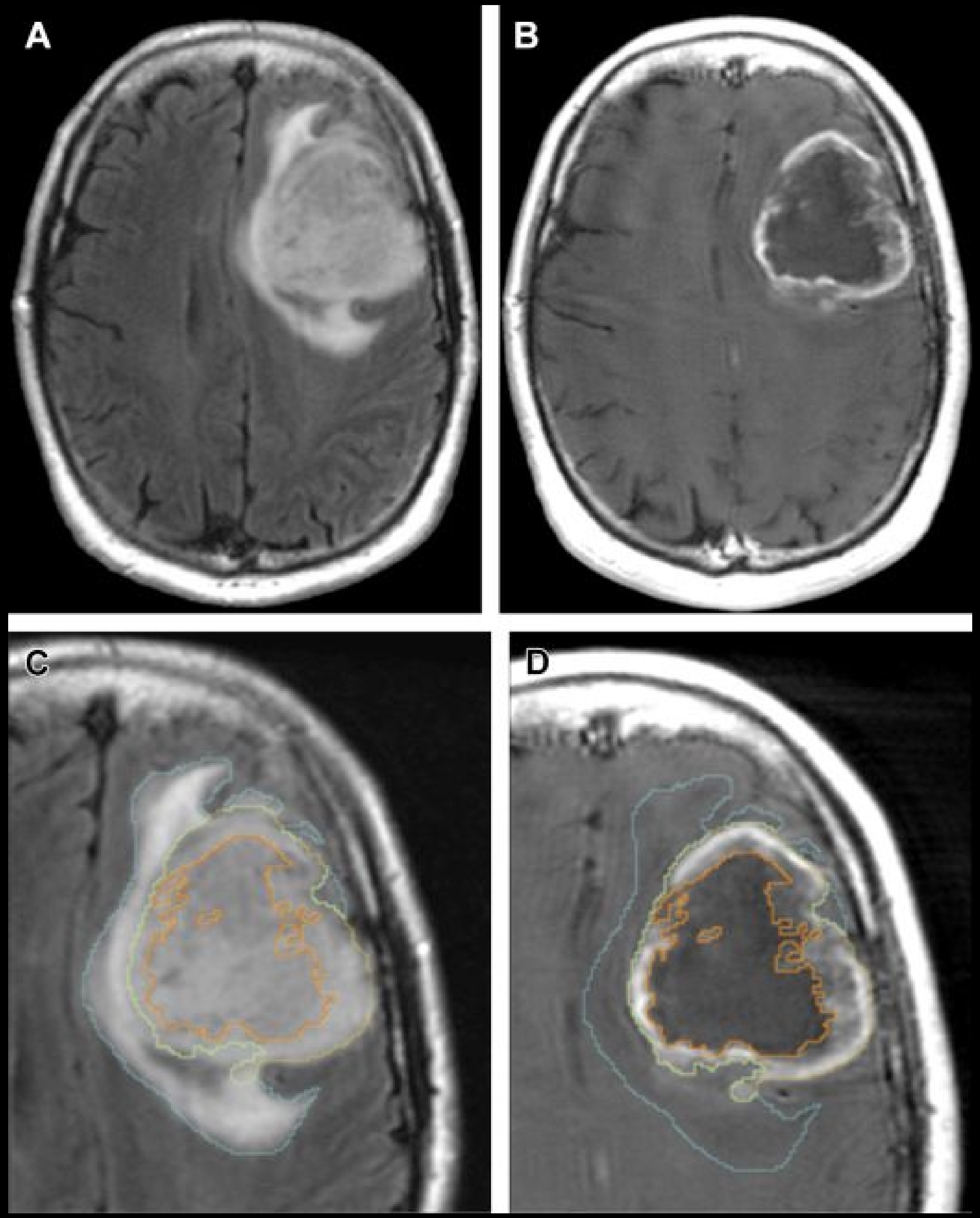




In the past decade, the field of **medical image analysis** has grown exponentially

Conversion of images into mineable data and the subsequent analysis of these data for decision support

Radiomics



Reuniones, alianzas y trabajo conjunto son primordiales para el éxito de la relación entre ciencia y empresa.

POR FRANCISCA ORELLANA

ACADEMIA E INDUSTRIA AVANZAN HACIA UNA MEJOR VINCULACIÓN

La vinculación ciencia y empresa es un desafío que no ha estado exento de dificultades, en un contexto donde Chile invierte sólo el 0,36% de su PIB en I+D, según la última Encuesta sobre Gasto y Personal en I+D año 2017 (publicada en 2019), convirtiéndose así en el país de la OCDE que menos invierte en este ítem.

No obstante, lo positivo es que el gasto en el área aplicada creció 10% en comparación al año anterior. "La intensidad de la transferencia de tecnología se relaciona con la madurez del ecosistema de innovación", explica el director de Innovación de la U. Andrés Bello (UNAB), Francisco Chiang.

Y en ese sentido, "en Chile existe un alto potencial, por la alta calidad de nuevo conocimiento generado, el nivel de desarrollo de la industria nacional y el creciente apoyo del gobierno a través de programas orientados a la generación de innovaciones", acota Chiang, enfatizando que hay que aumentar la conexión y capital social entre las universidades y centros tecnológicos con la industria. "Según cifras de 2014, sólo 8,4% de las empresas realizaba colaboraciones con universidades o centros tecnológicos, a diferencia de Finlandia, Noruega o Austria, que superaban el 60%", sostiene.

A su juicio, un factor clave para potenciar esta conexión es el Programa Ciencia e Innovación para el 2030, algo en lo que coincide el decano de la Facultad de Física de la U. Católica (UC), Max Bañados: "El programa aparece en el momento correcto para una nueva etapa en que la ciencia y la industria colaboran".

Con un beneficio que es mutuo, añade, pues "el sector industrial sabe que la sofisticación de sus procesos es una necesidad inminente (por desafíos ambientales, competencia internacional, globalización), y la ciencia sabe que no puede vivir para siempre sólo de fondos estatales. Esta relación será un *win-win*".

Mayor interés

El vicepresidente ejecutivo de Corfo, Pablo Terrazas, dice que para lograr una adecuada transferencia es primordial "que se creen alianzas y surjan experiencias conjuntas que pavimenten el camino". Para ello, "las universidades requieren conocer los

CIENTÍFICOS PARA EL SALTO AL DESARROLLO

POR CONSTANZA GARÍN L.

Que los profesionales del ámbito de las ciencias básicas den respuestas innovadoras al mercado y la sociedad, es el objetivo del programa Ciencia e Innovación para el 2030, impulsado por Corfo.

Transformar las facultades de ciencias en entidades que realicen actividades de transferencia tecnológica, innovación y emprendimiento de base científica-tecnológica, son los objetivos principales del programa Ciencia e Innovación para el 2030, liderado por Corfo, en colaboración con el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, que busca implementar acciones y planes estratégicos para avanzar en este sentido.

Para llevar a cabo este proyecto, en 2017 la gerencia de Capacidades Tecnológicas de Corfo hizo un diagnóstico en base a las tecnologías que se generaban en las facultades de ciencias en los diferentes planteles universitarios, un proceso que arrojó que sólo se estaban aplicando cuatro tecnologías, todas asociadas a la transferencia tecnológica.

"El análisis también contempló el hecho de que existe poca proporción de egresados de estas carreras que emigran al sector empresarial, ya que la mayoría se queda en la academia", explica el vicepresidente ejecutivo de Corfo, Pablo Terrazas, haciendo hincapié en que en los países de la OCDE, la proporción es mucho más alta. En Australia, por ejemplo, "el 16% del total de los investigadores en ciencias naturales, trabaja en la industria, a diferencia del 2% de Chile".

En ese sentido, el decano de la Facultad de Física de la Universidad Católica (UC), Máximo Bañados, valora la iniciativa pues "una clave para el desarrollo de Chile" es la conexión ciencia-industria y ciencia-sector público, ya que en la actualidad la industria genera productos y servicios, mientras que la universidad realiza investigación. "El siguiente paso es conectar estos mundos mediante innovación en investigación en la industria (...) la conversación ciencia e industria es un círculo virtuoso que da oportunidades para mejorar la vida de las personas", sostiene.

Para ello se requiere preparar a todos los ámbitos universitarios, dice Terrazas, añadiendo que es necesario mejorar programas de estudio, incluyendo metodologías innovadoras de enseñanza para que los alumnos estén preparados para innovar y emprender, junto con "generar nuevo conocimiento que dé respuesta a los desafíos de innovación que presentan las empresas y la sociedad".

En la misma línea, Francisco Chiang, director de Innovación de la Universidad Andrés Bello (UNAB), destaca la iniciativa señalando que "hoy la innovación es el principal precursor del crecimiento económico y la competitividad, dado que permite generar más valor hacia los clientes y diferenciarse de la competencia".

Los pasos a seguir

La primera etapa de este programa se inició en 2018, cuando se definió un plan estratégico y un *benchmarking* internacional, proceso que culminó en 2019 con siete planes generados de forma compartida y en consenso dentro de las facultades de ciencias.

Sólo el **2%** de los investigadores en ciencias naturales de Chile trabaja en la industria.

$10 MIL MILLONES es el presupuesto para los proyectos ganadores en la segunda etapa del programa.

✓ Precise

✗ Slow (weeks)

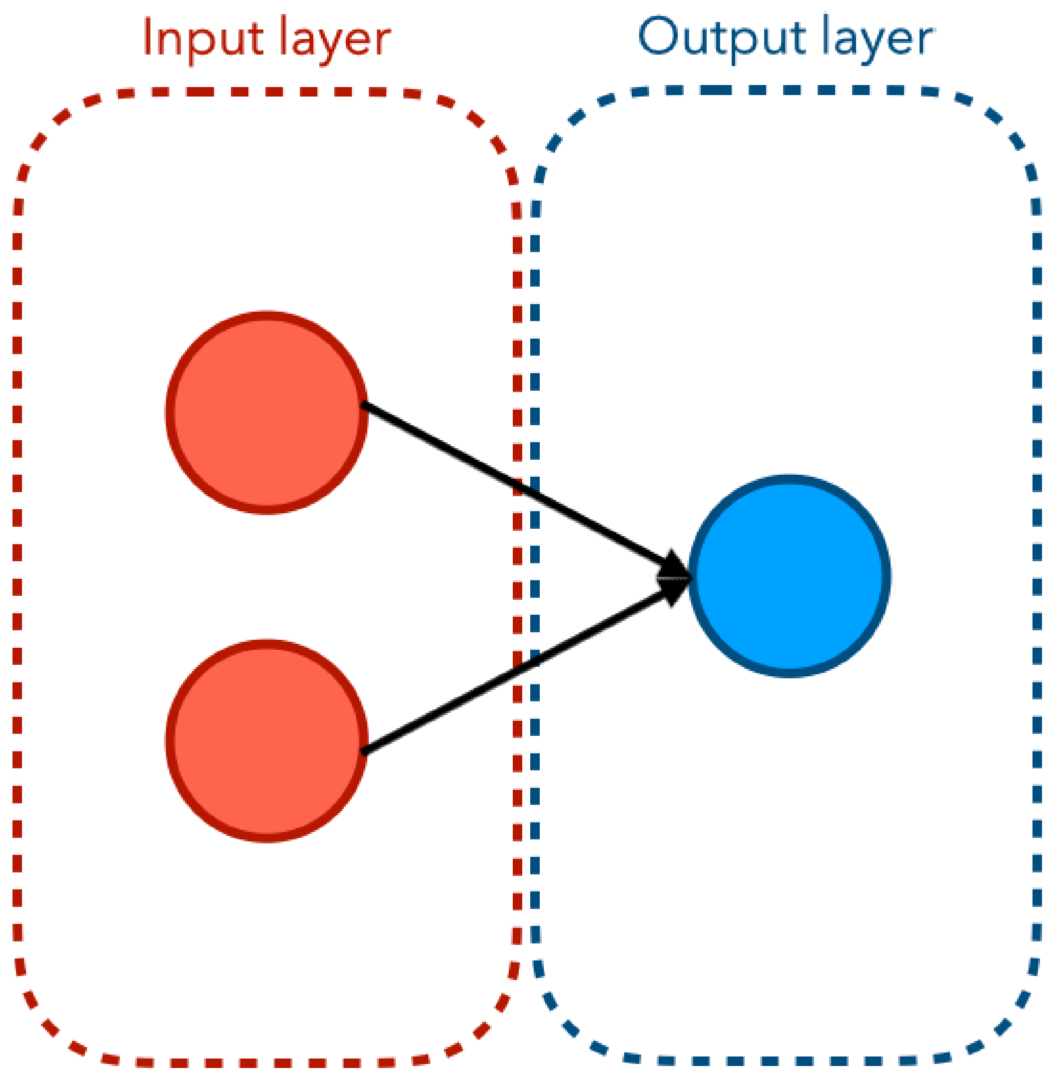
- ✓ Precise
- ✓ Fast (seconds)

Machine Learning is a method to teach computers to learn from data, without being explicitly programmed

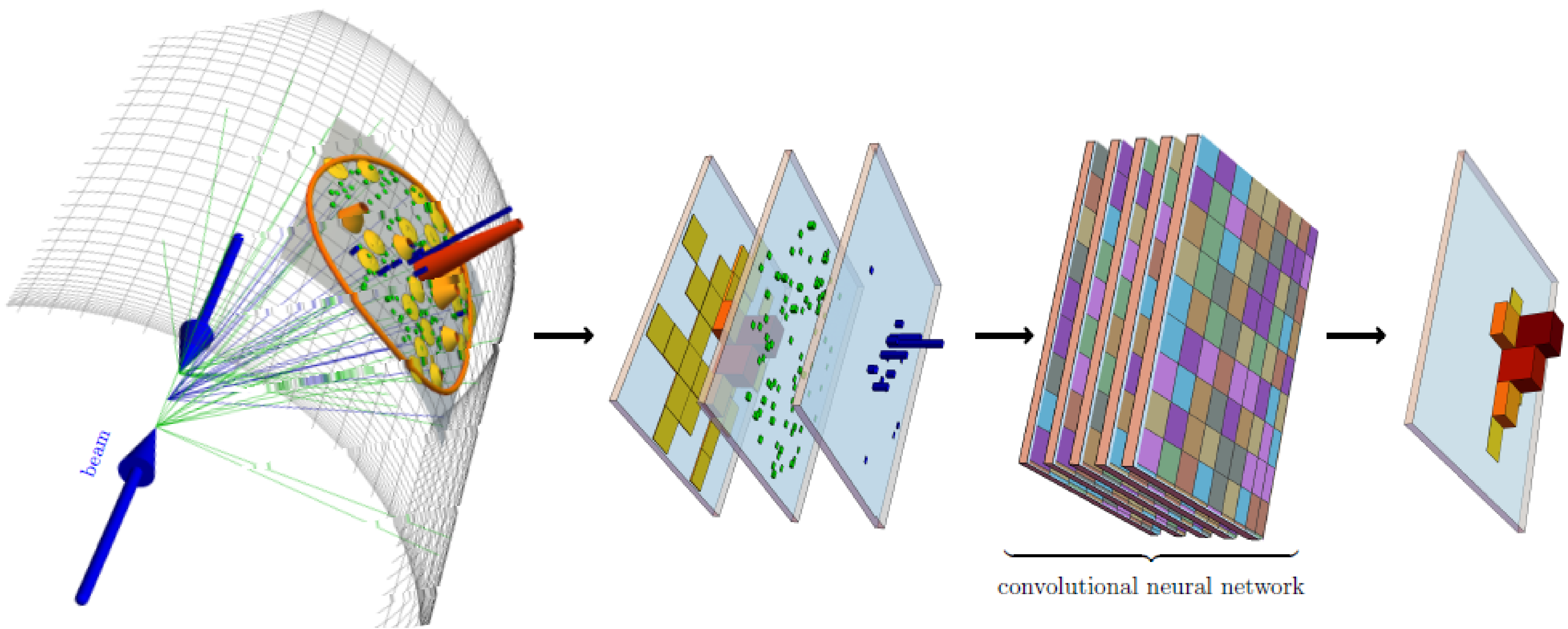
"I feel more optimistic about the possibility of transcendental progress in fundamental physics now that at any other time in my career, even if I myself may not be able to comprehend the final theory."

M. D. Schwartz (2021)

Backup

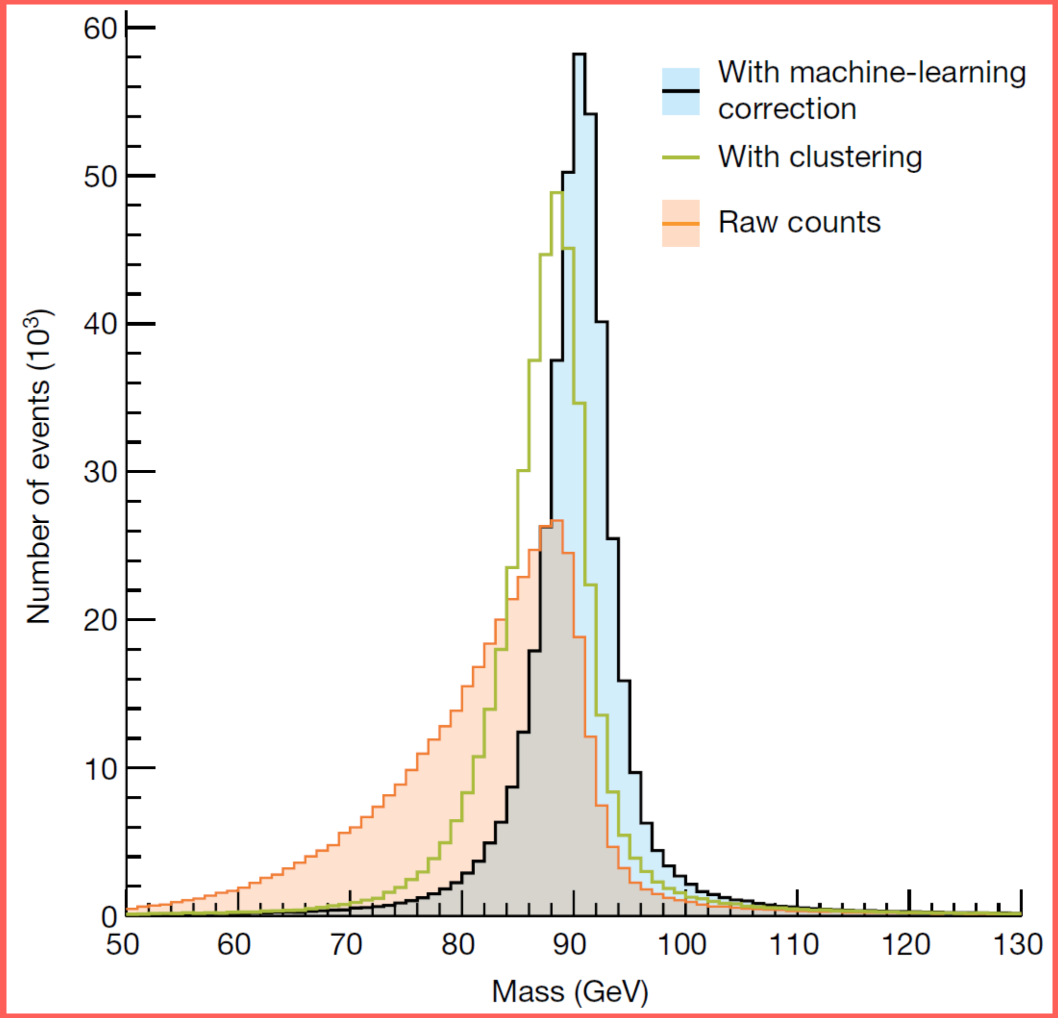

Input layer
Output layer

PUMML (PileUp Mitigation with Machine Learning) algorithm



- For the first years, LHCb primary algorithm for classification was BDT
- The sensitivity achieved by a LHCb dark matter searches analysis with BDT using data collected just in 2016, would have required 10 years of data collection without the use of machine learning

- BDTs are used to increase the resolution of the CMS electromagnetic calorimeter
- Deposited energy in different sensors are clustered together to recover the original energy of the particle
- Applying this energy correction to the Z boson decay into electron-positron

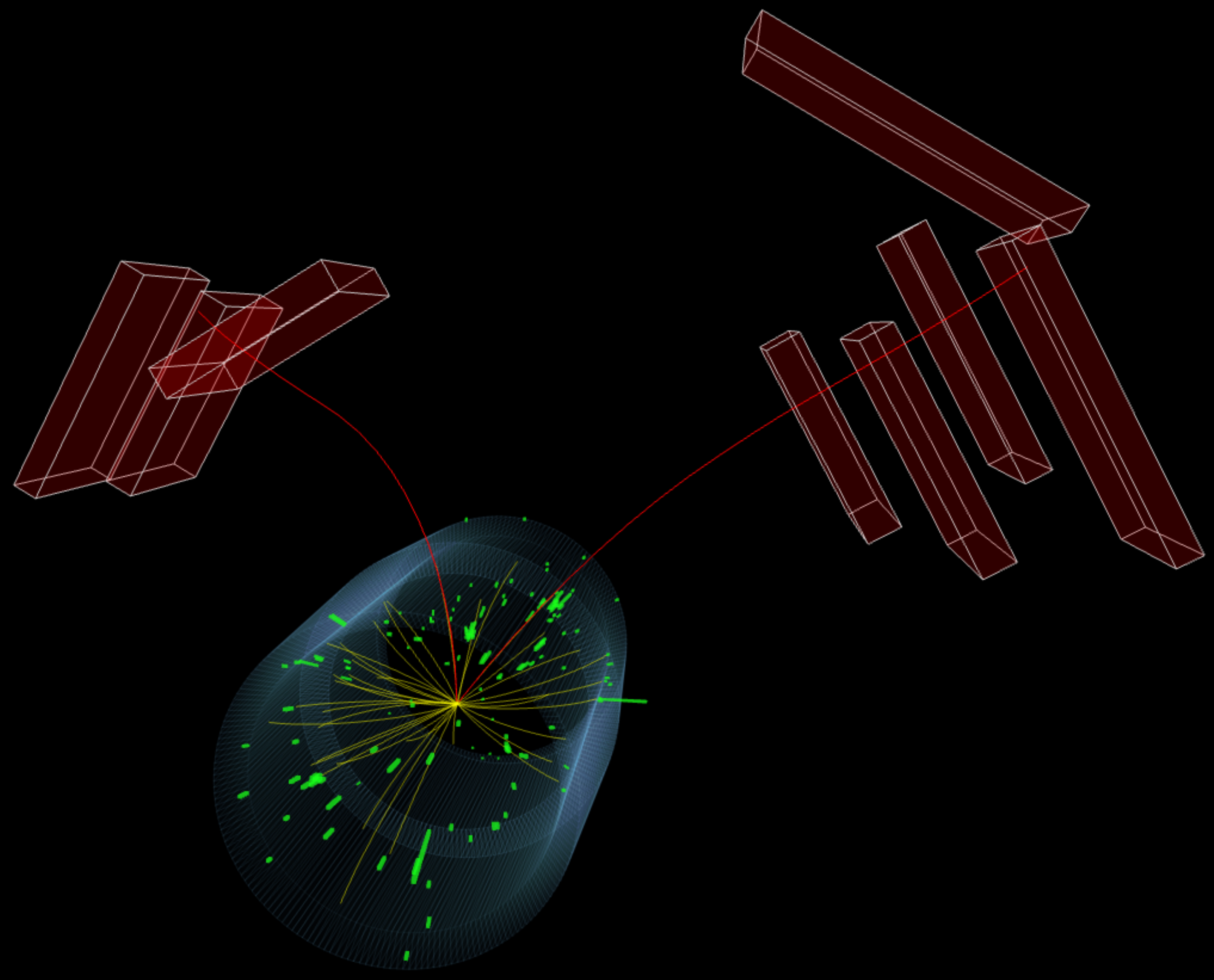


SM predicts that **only three out of every billion** Bs mesons decay into muon-antimuon final state

CMS and LHCb were the first to find evidence of this decay

They used a **human-designed** algorithm which first reconstructs the paths taken by the muon and the antimuon, to then calculate di-muon masses and angles which are used by a **BDT**

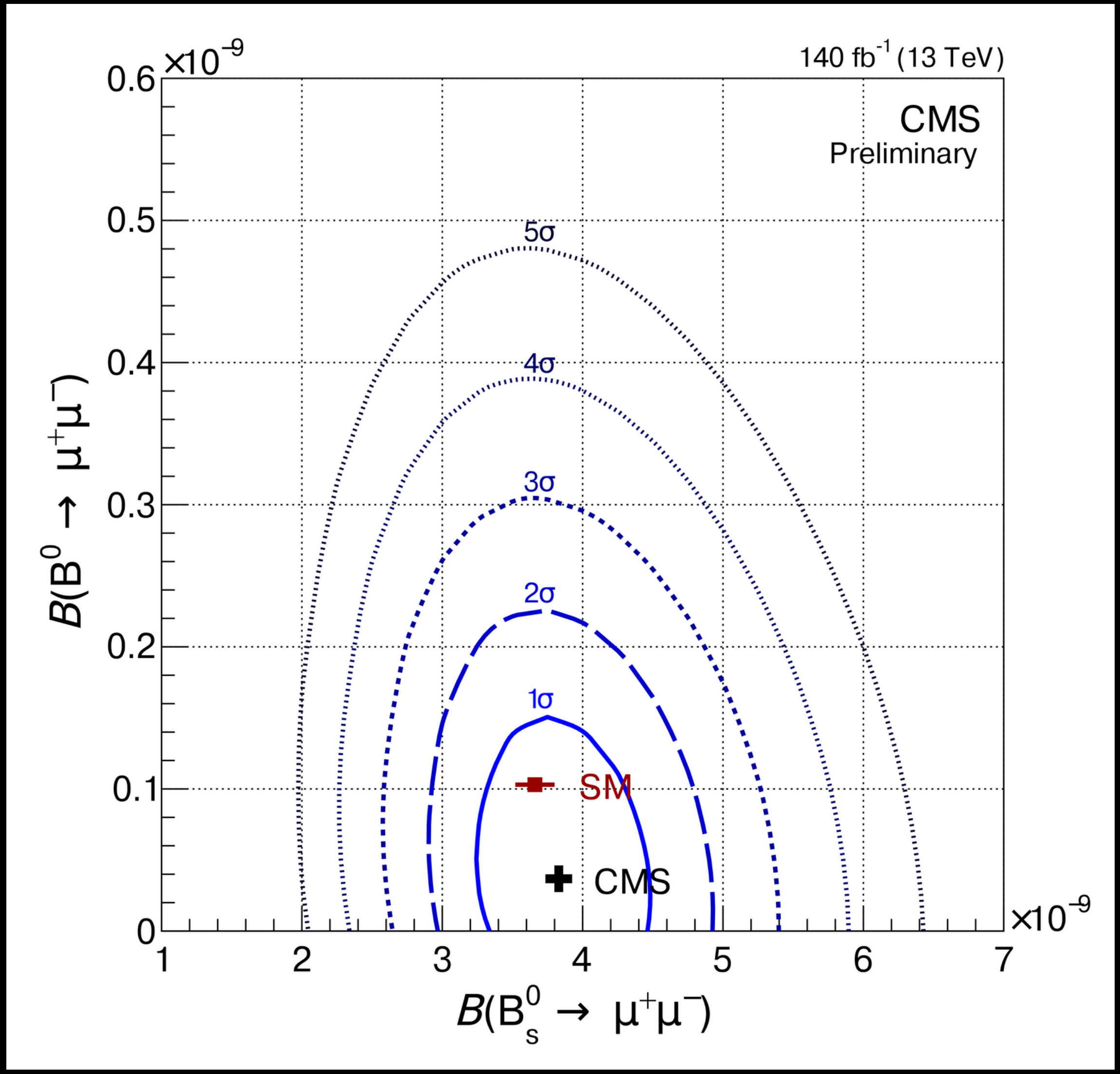
Bs-meson decay into muon-antimuon



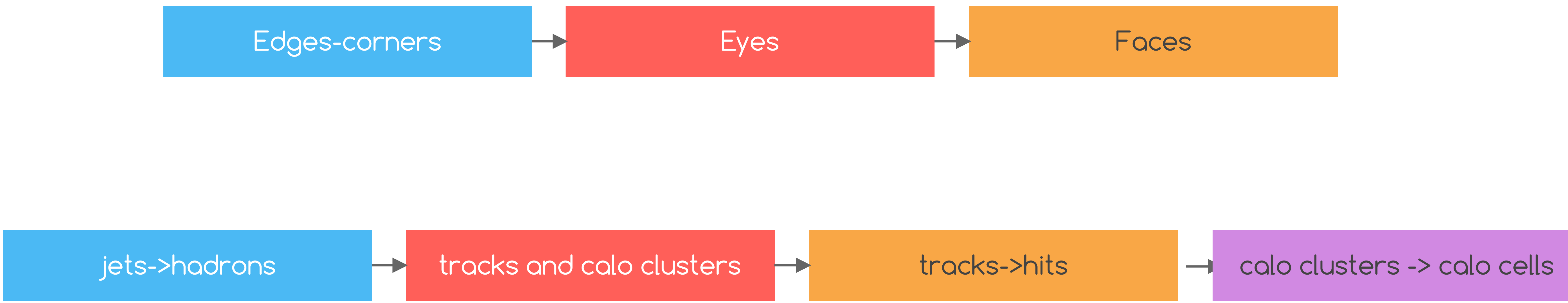
However, we might be missing information when human-designed tools are used to extract features that fail to fully pacture the complexity of the problem

By relaxing the preselection requirements, developing new discriminating observables, and using a more advance ML algorithm, a new CMS analysis achieved the same sensitivity as the old analysis with just 60% of the previous data

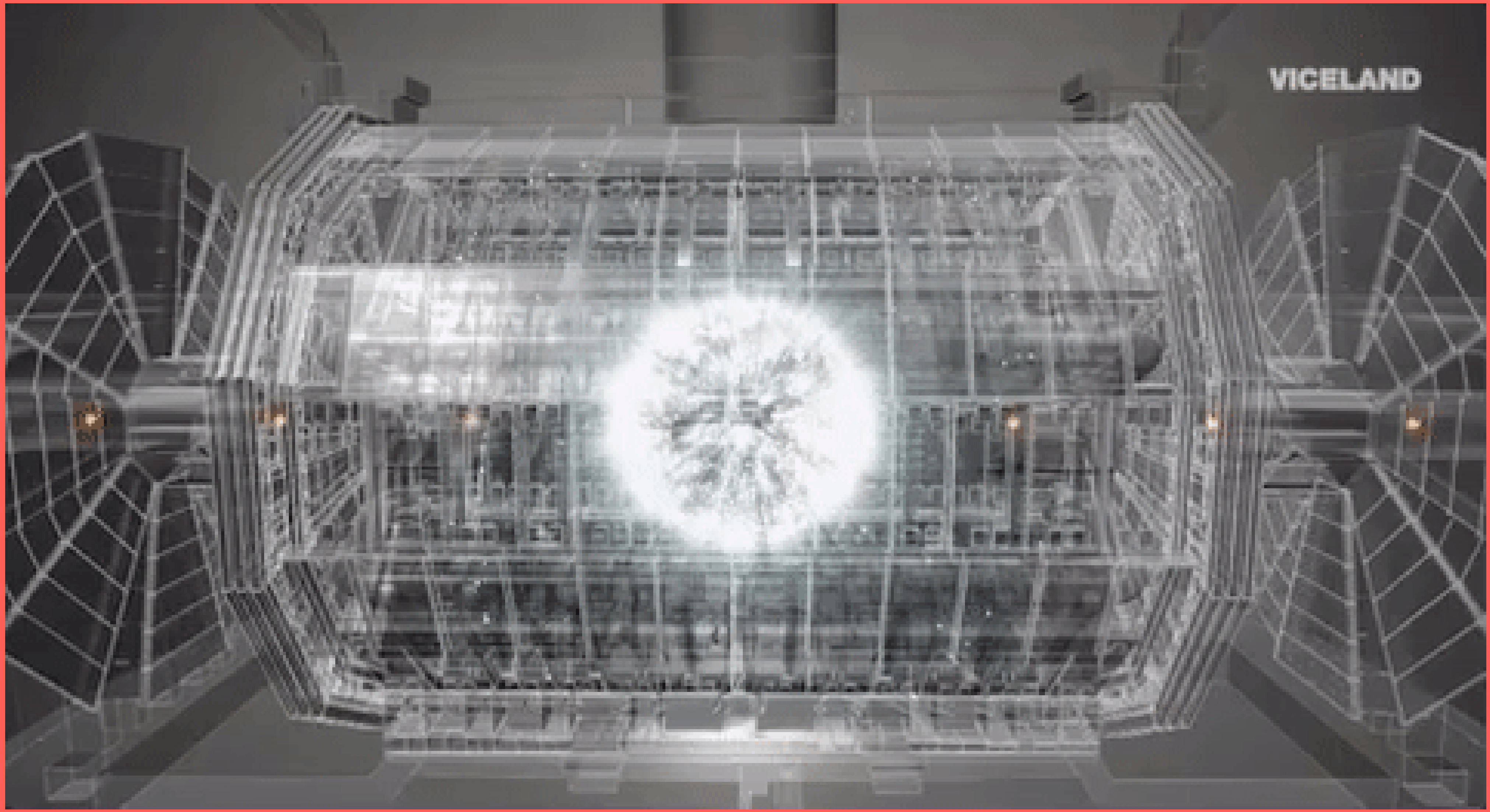
With full Run-2 data the significance improved by 1.2 standard deviations



Deep networks not only have more expressive capacity, but also the layers can be interpreted as building up a **hierarchical** representation of the data



Let's go back to the LHC data-taking...



The CMS Endcap Muon Track Finder (EMFT) is part of the L1 Muon trigger that has only 500 ns to reconstruct a muon through the Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC)

BDT

CMS discretise the ML evaluation into a single operation using a LUT (trade time for memory)

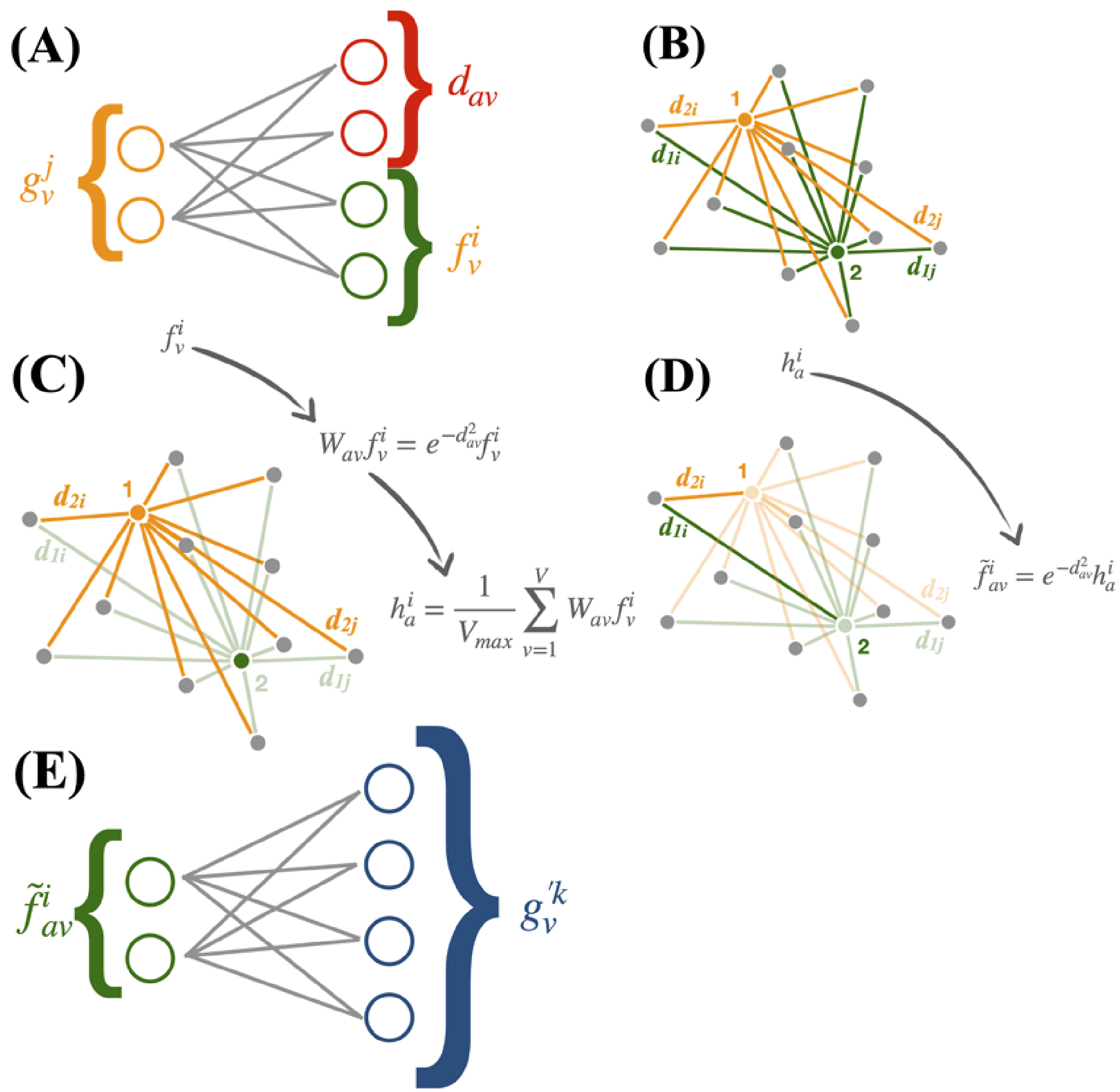
500 trees, 5 decisions = 2500 operations per p_T

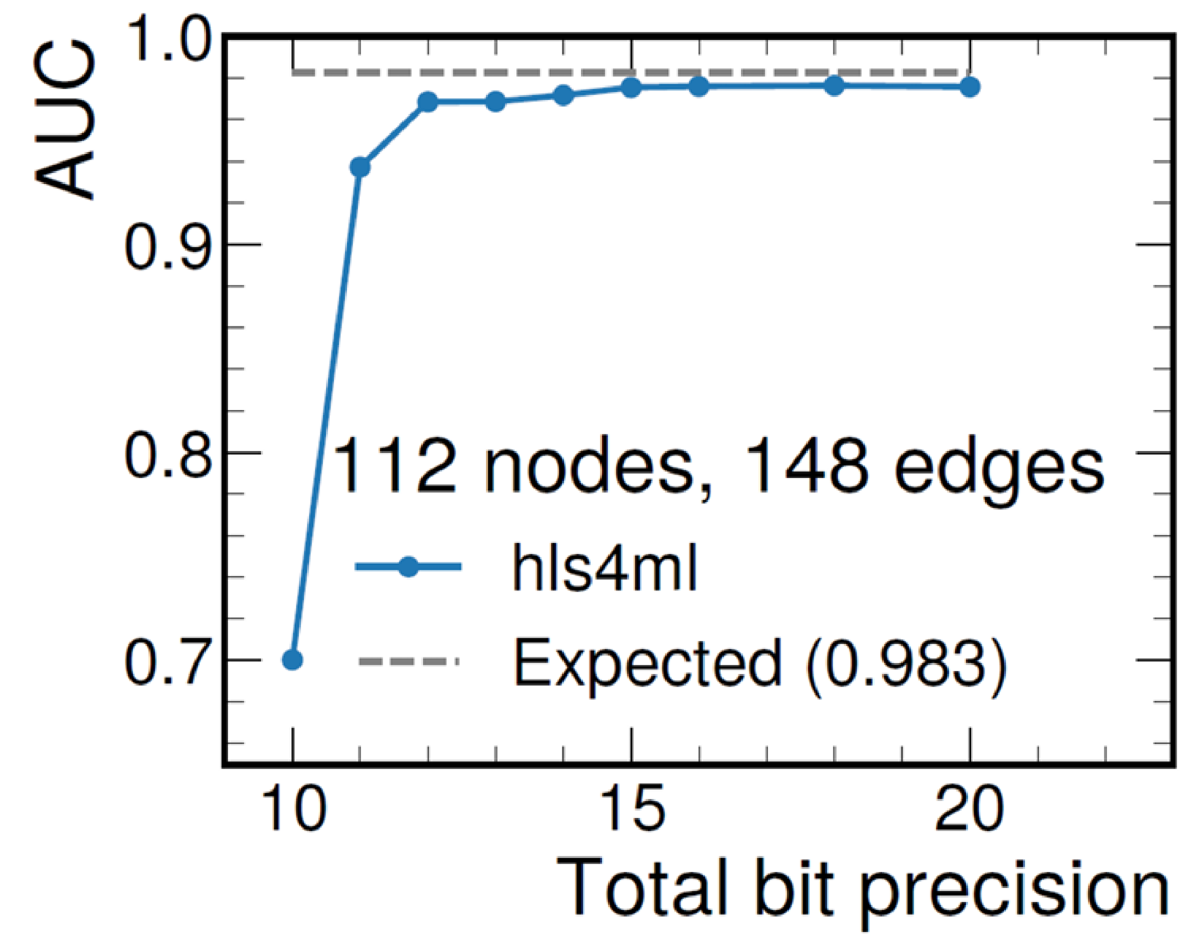
Assuming 1 operation per ns (1 GHz)

2500 ns >> 25 ns

The design was implemented in 2016/2017 data-taking, improving the trigger by a factor of 2!

Can we take advantage of the NSW for L1 Muon trigger?



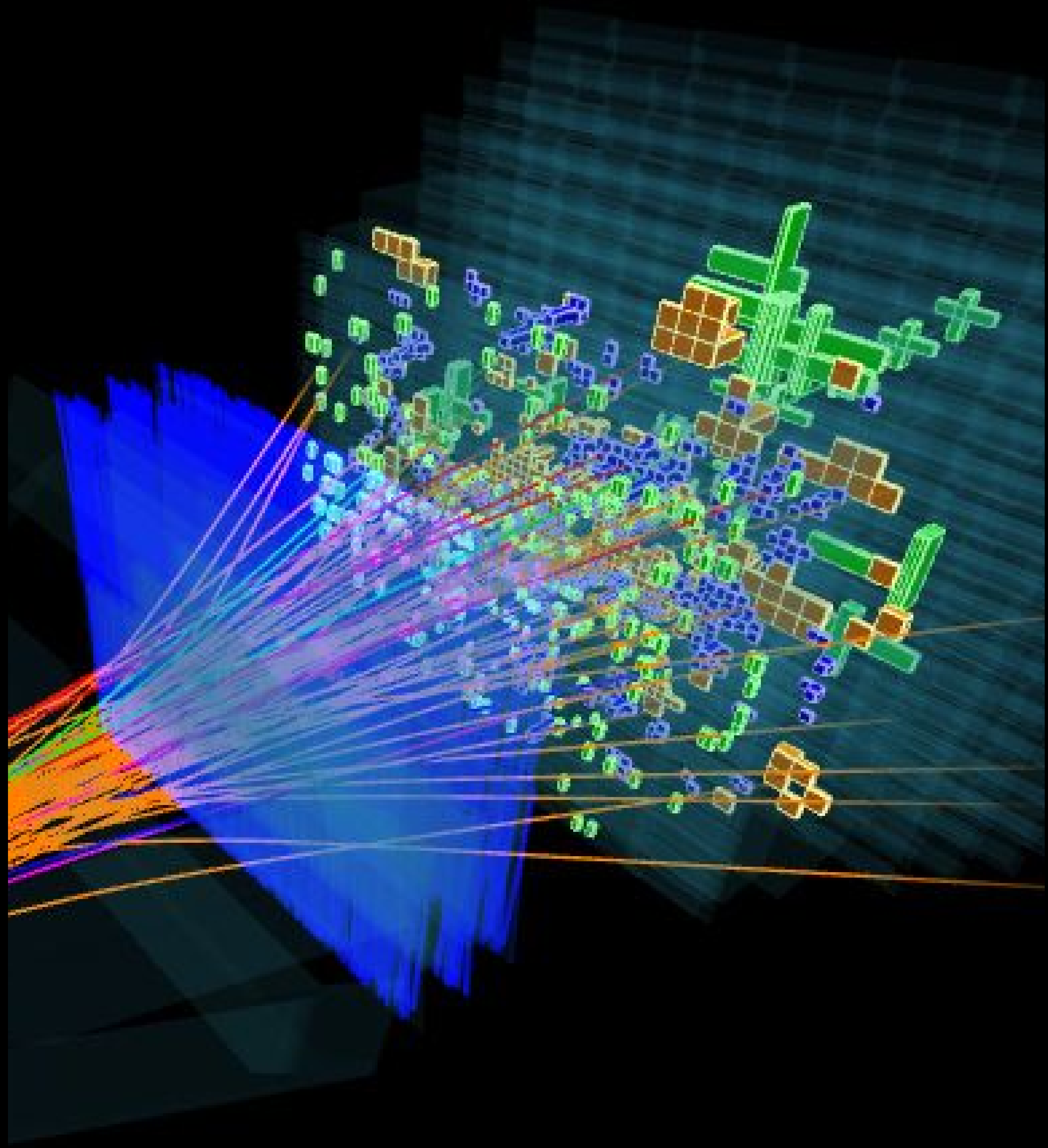


Conclusion

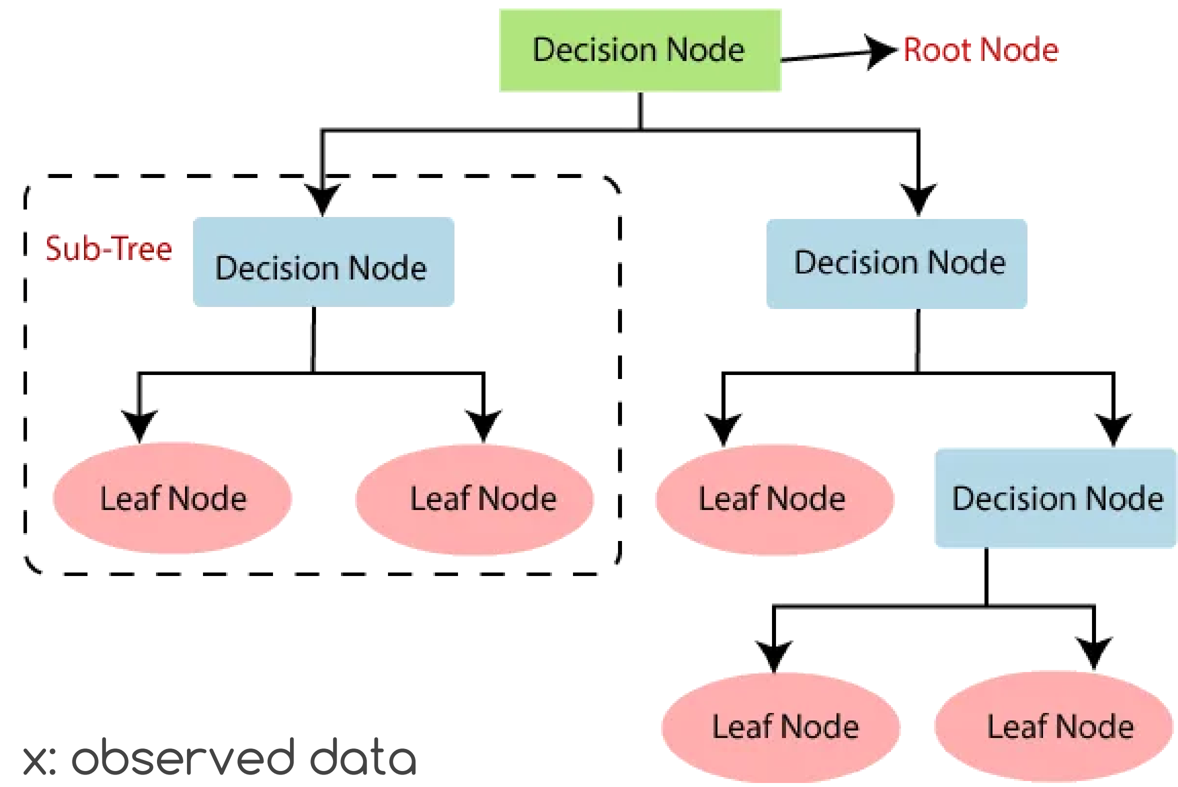
- Over the last years machine learning were heavily used in physics analyses
- Modern machine learning algorithms used in industries can be succesfully applied to different analyses
- Machine learning algorithms should be applied from the first data quality selection in order to not loose useful information
- Field programable gate arrays are ideal to implement fast machine learning algorithms since the first trigger level

LHCb experiment analyses as many events every six seconds as the Big European Bubble Chamber recorded in its entire 11 years of operation (1973-1983)

Big data era!



For several years the status quo of machine learning in HEP was to use boosted decision trees (BDT) implemented in the software package TMVA

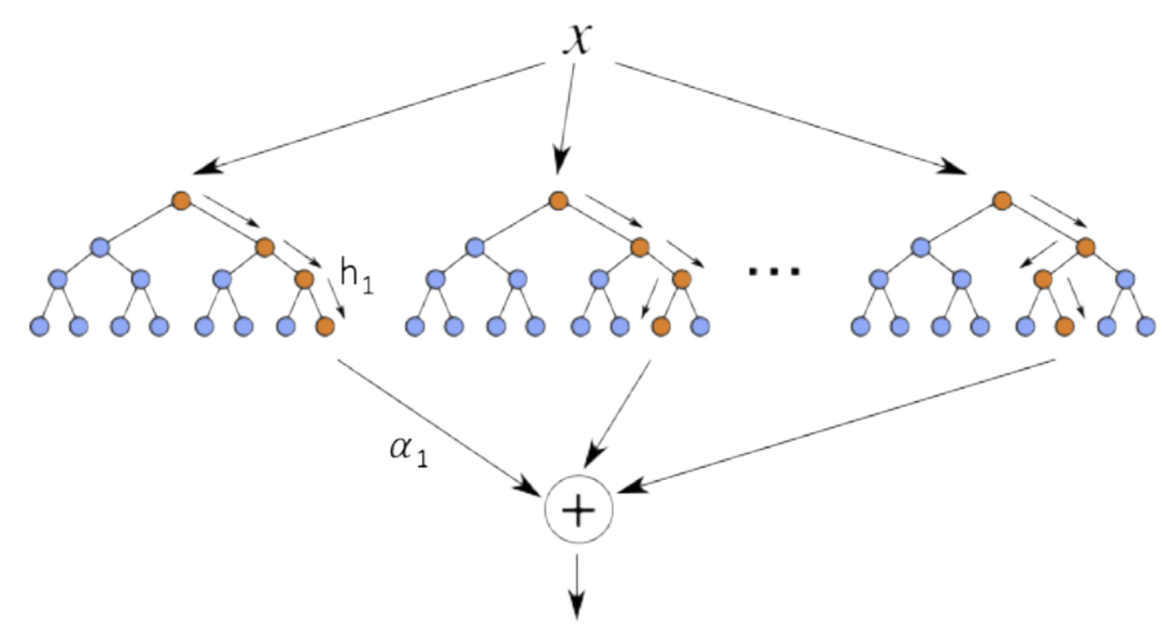


Ensemble output: $\hat{y}(x) = \sum_t w_t h_t(x)$

x: observed data
h(x): tree's output
w: weight
y: output

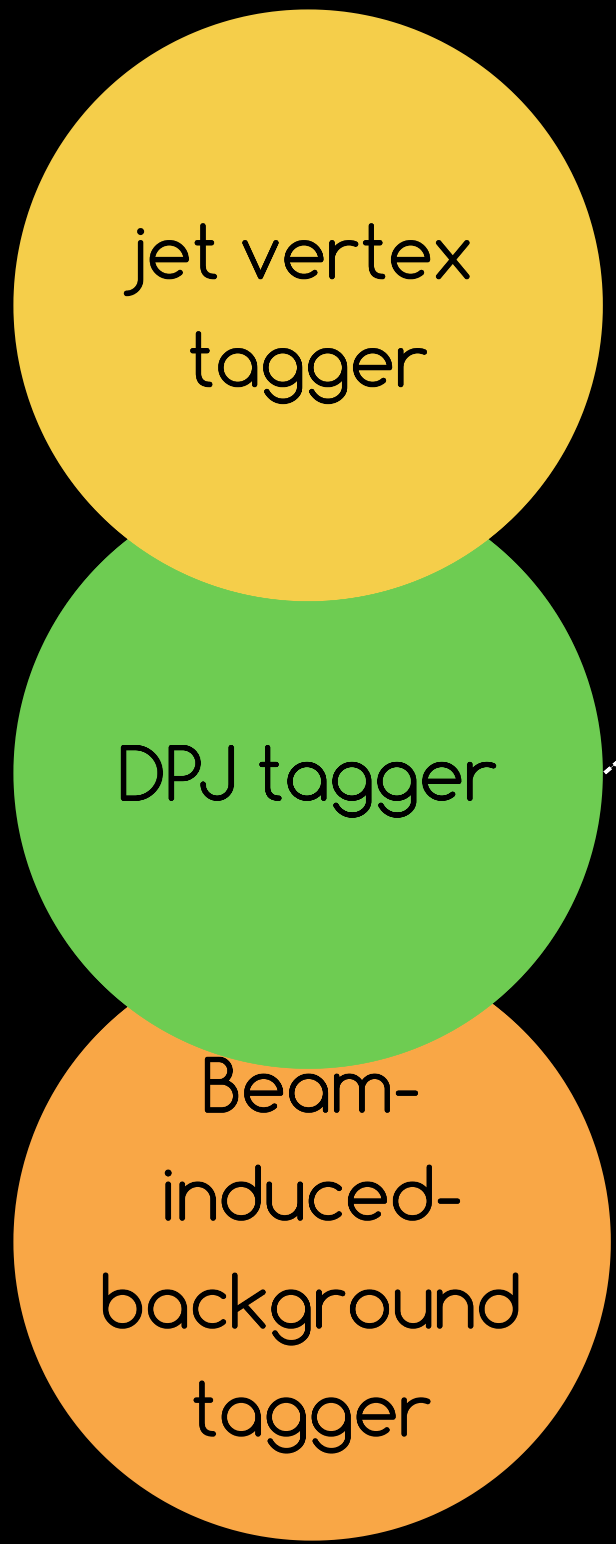
Objective function (goal is to minimise it):

$$O(x) = \sum_i l(\hat{y}_i, y_i) + \sum_t \Omega(f_t)$$



$l(\hat{y}_i, y_i)$ Loss function, which is the distance between the truth and the prediction

$\Omega(f_t)$ Regularisation function, which penalises the complexity of the tree



Discriminate DPJ from SM jets