Machine learning approach for vector boson fusion identification

Andrés Flórez, Alfredo Gurrola,
Raúl Ramos, Alexis Ruales*, Jose Ruiz

*** PhD Student**

UNIVERSIDAD DE ANTIOQUIA
1803

Grupo de Fenomenología e Interacciones Fundamentales (GFIF)
Instituto de Física - Universidad de Antioquia
2022

CMS

Outline

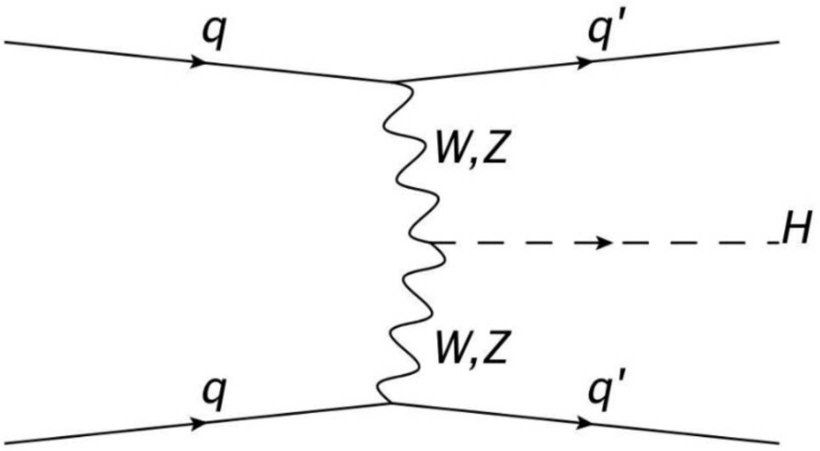
Introduction

- Vector boson fusion proposed initially as an alternative channel for finding heavy Higgs has now established itself as a crucial search scheme to probe different properties of the Higgs boson or for new physics.

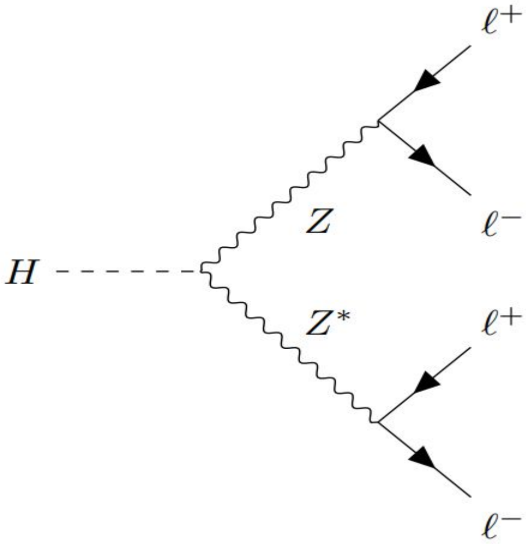
- The emergence of deep learning frameworks, a plethora of machine learning applications have gained immense importance in high-energy physics recently, in collider and neutrino physics.

- Deep learning applications have been widely explored to understand hadronic jets formation and properties, the most common structured object found in any event at LHC.

VBF and ML



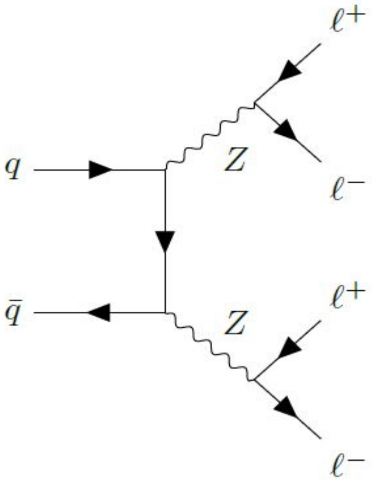
Higgs production via vector boson fusion



Feynman diagram of the $H \to ZZ^* \to 4\ell$ decay channel

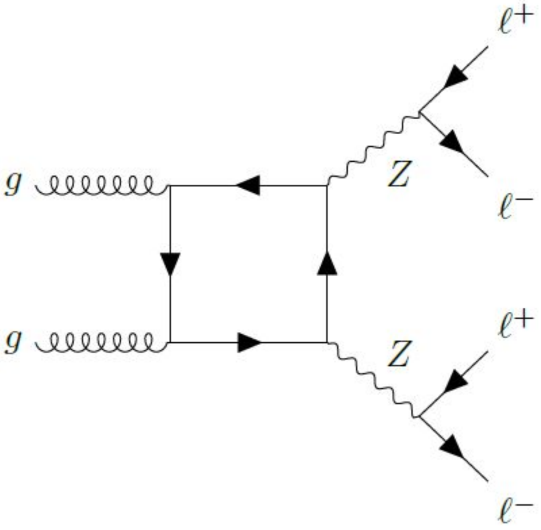
Background

The main background



$$q\bar{q} \to ZZ \to 4l$$

Making up about 80% of all the background



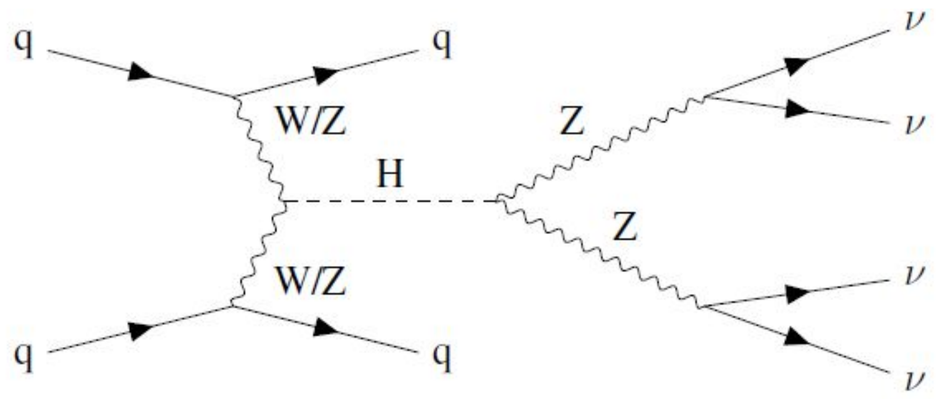
$$gg \to ZZ \to 4l$$

Making up about 3% of all the background

https://repository.upenn.edu/edissertations/3124/

The reference

Machine learning approach towards an improved vector boson fusion selection by ***Janna Vischer***



Leading order Feynman diagram of a Higgs created via vector boson fusion and decaying into undetectable particles

VBF and ML

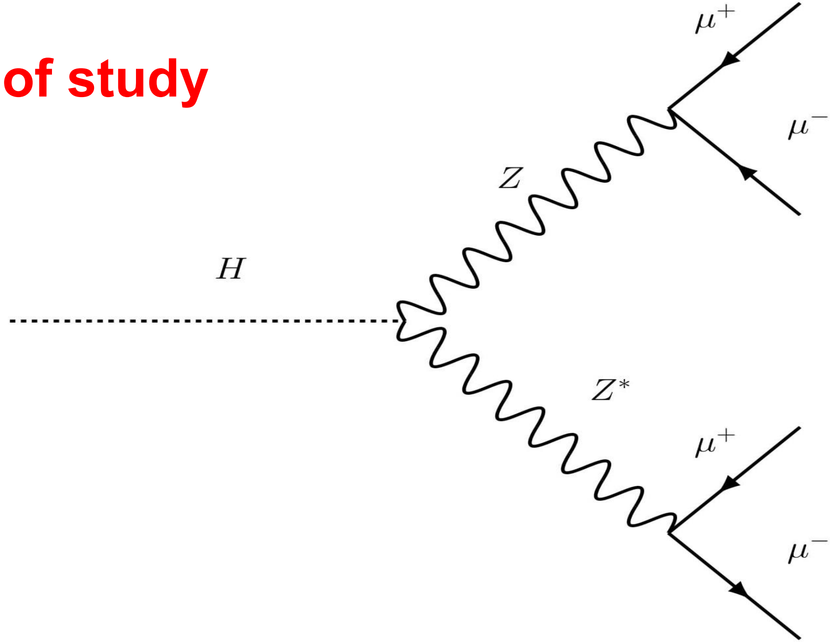
In the reference work, three models are proposed.

Network	Learning Rate	AUC-Score	Training Duration	Inference Duration
FullCon-ff	0.0005	0.8591	1 h 35 min	6 min
FullCon-mf	0.0050	0.8757	1 h 40 min	6 min
Particle-Net	0.0050	0.8956	66 h	47 min

Comparison of the three investigated neural networks in terms of AUC-Score and approximate training and inference duration

VBF (Vector Boson Fusion) and ML

❖ Process of study



Feynman diagram of the $H \to ZZ^* \to 4\mu$ decay channel

Work in progress

Simulation

Monte Carlo samples

- MADGRAPH5 : Simulates and calculates the effective sections at partonic level
- PYTHIA 8 : Hadronization and Showering processes
- DELPHES 4 : Emulates the detector’s response

Work in progress

Input model

34 input features for the model.

```
['missinget_met', 'missinget_phi', 'jet_pt0', 'jet_pt1', 'jet_pt2',
 'jet_pt3', 'jet_eta0', 'jet_eta1', 'jet_eta2', 'jet_eta3',
 'jet_phi0', 'jet_phi1', 'jet_phi2', 'jet_phi3', 'jet_mass0',
 'jet_mass1', 'jet_mass2', 'jet_mass3', 'muon_pt0', 'muon_pt1',
 'muon_pt2', 'muon_pt3', 'muon_eta0', 'muon_eta1', 'muon_eta2',
 'muon_eta3', 'muon_phi0', 'muon_phi1', 'muon_phi2', 'muon_phi3',
 'muon_charge0', 'muon_charge1', 'muon_charge2', 'muon_charge3'],
```

Work in progress

ML approach

Data processing

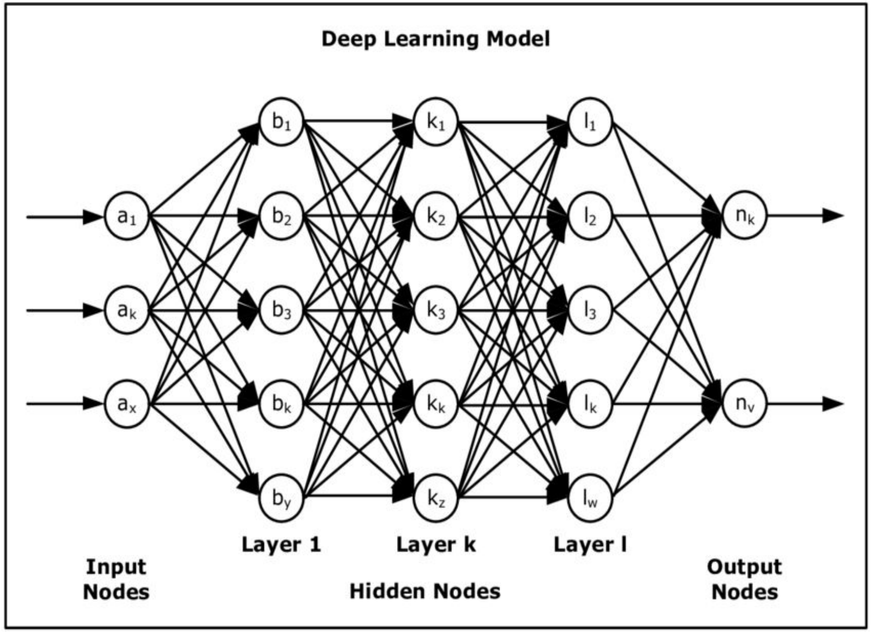
1. "n_Jet" >= 2
2. "Pt_Jet" > 30
3. The background is bigger that the Signal
4. Scalar data with method StandardScaler()
5. Resample signal
6. 70% train, 30% Val
7. Batch_size and epochs are defined depending on the performance of the model.

Work in progress

ML approach

Proposed model architecture

- Input layer with **34** input features.
- Dense layer with **200** neurons, **relu** activation function.
- Dense layer with **100** neurons, **relu** activation function.
- Dense layer with **50** neurons, **relu** activation function.
- Dense layer with **25** neurons, **relu** activation function.
- Dense layer with **1** neurons, **sigmoid** activation function.
- Output with VBF identification score



Model architecture Preliminary

Conclusions

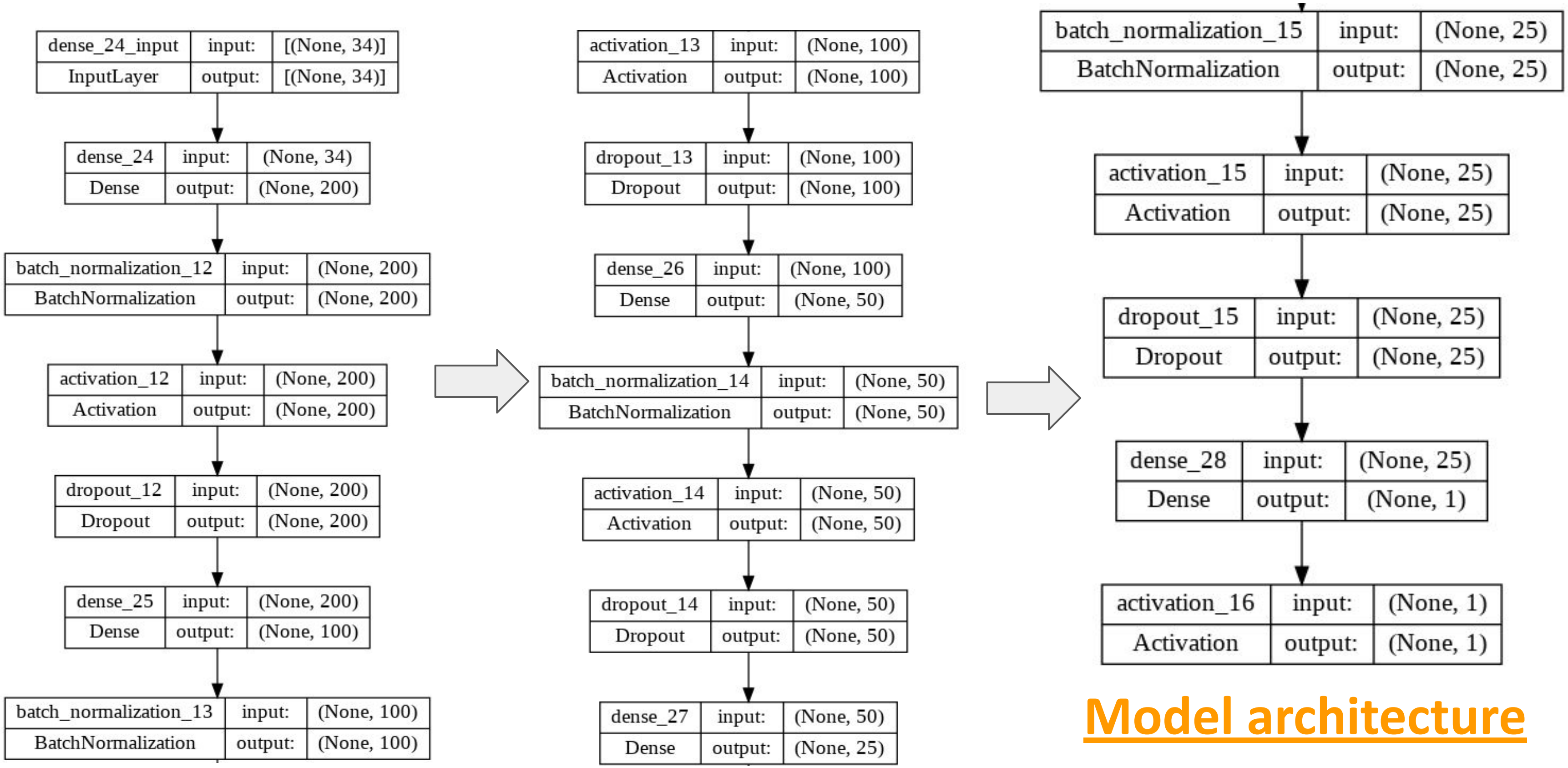
- A new ML architecture for VBF identification is proposed, it is expected to be more efficient in execution time and performance.

- The methodology for the VBF identification with machine learning is defined.

- Different sources that apply ML models for VBF were studied, the performance of these models in AUC-Score is greater than 0.85.

Thank you!!

Backup



Model architecture

Preliminary

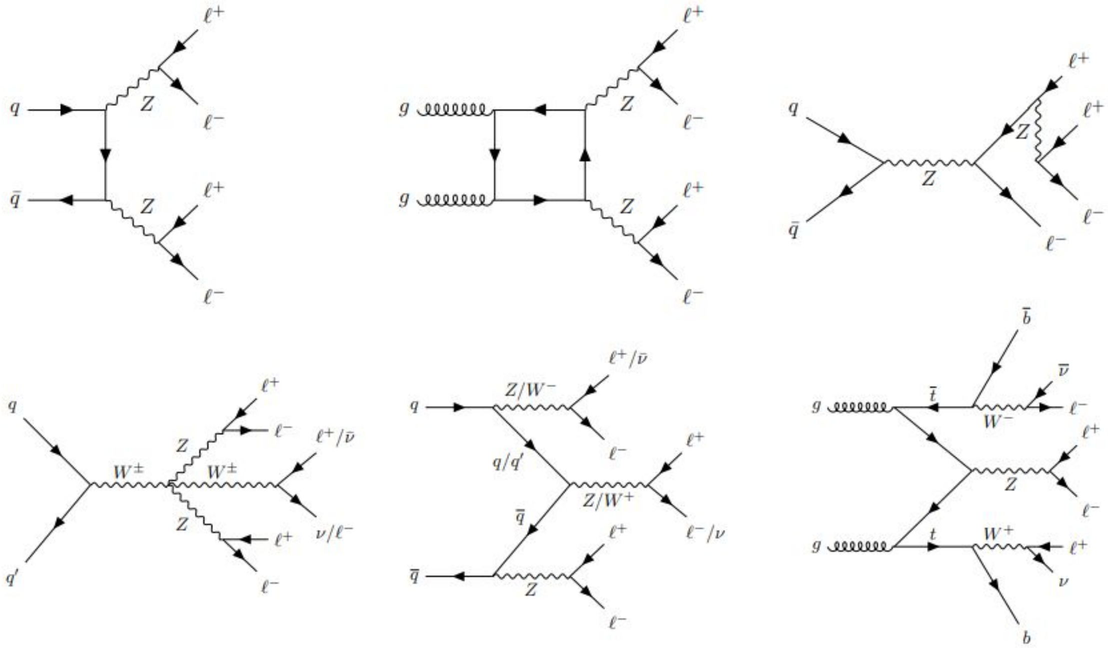


Figure 62: Irreducible background contributions to the 4ℓ signal selection. $q\bar{q}ZZ$ (top left) contributes the most. Next is $ggZZ$ (top middle) and single resonant $Z \to 4\ell$ (top right). Triboson (ZZW,WWZ,ZZZ) (bottom left and middle) and $V + t\bar{t}$ all-leptonic (bottom right) are minor irreducible backgrounds.

Model architecture - Preliminary

```
def MODEL_VBF(n_features):
    model = Sequential()
    model.add(Dense(200, input_shape=(n_features,), kernel_initializer="glorot_normal"))
    model.add(BatchNormalization())
    model.add(Activation("relu"))
    model.add(Dropout(0.5))
    model.add(Dense(100, kernel_initializer="glorot_normal", use_bias=False))
    model.add(BatchNormalization())
    model.add(Activation("relu"))
    model.add(Dropout(0.25))
    model.add(Dense(50, kernel_initializer="glorot_normal", use_bias=False))
    model.add(BatchNormalization())
    model.add(Activation("relu"))
    model.add(Dropout(0.15))
    model.add(Dense(25, kernel_initializer="glorot_normal", use_bias=False))
    model.add(BatchNormalization())
    model.add(Activation("relu"))
    model.add(Dropout(0.1))
    model.add(Dense(1, activation="sigmoid"))
    model.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"])
    return model
```