Electron Classification in CMS Open Data Using a Neural Network

Jokūbas Mačiulis

December 15, 2025

Abstract.

A supervised neural network is used to classify CMS 2016 proton–proton collision events as Drell–Yan signal or background using high-level electron and jet features from NanoAOD open data. The model achieves a test accuracy of 87.7% and a ROC–AUC score of 0.948, demonstrating strong signal–background discrimination.

1 Method

CMS 2016 NanoAODSIM data were used to train a feed-forward neural network for signal–background classification using high-level electron and jet features. Model performance was evaluated using accuracy and ROC–AUC metrics.

2 Results and Performance

The training history shows stable convergence with no significant overfitting:

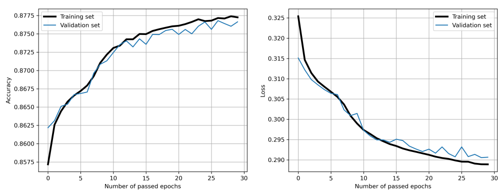

Figure 1: Training and validation accuracy and loss versus epoch.

The model converged within 30 epochs with no significant overfitting, as indicated by the stable training and validation curves.

The AUC evolution during training indicates consistent improvement in signal–background separation:

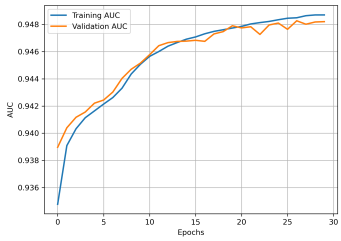

Figure 2: AUC evolution for training and validation datasets.

The final evaluation on the test set yields an accuracy of **87.7%** and a ROC-AUC score of **0.948**. The ROC curve demonstrates strong classification performance across a wide range of thresholds, significantly outperforming random guessing:

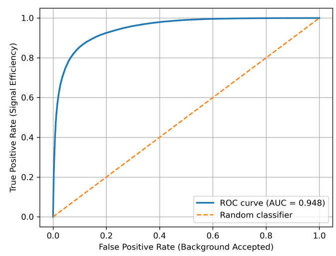

Figure 3: ROC curve for signal–background classification.

3 Conclusion

This study demonstrates that a relatively simple neural network trained on physically motivated electron and jet features can effectively distinguish Drell–Yan signal events from background in CMS open data. The achieved performance is competitive for a baseline model and highlights the potential of machine learning techniques in particle physics analyses. Further improvements could be achieved by optimizing hyperparameters, exploring deeper architectures, or incorporating additional physics-inspired variables.