



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Soft Muon Tagging perfezionato basato su Machine Learning in ATLAS

Federico A.G. Corchia
Università di Bologna e INFN Bologna



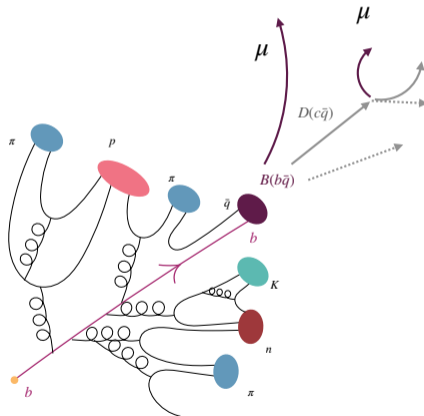
middle
PRIN 2022

PROJECT 20223N7F8K
PNRR M4.C2.1.1, D.D. n. 104 2/2/2022
CUP I53D23000820006



I muoni soft

- Gli **adroni pesanti** (associati a quark b/c) sono di **grande interesse** a LHC per molti studi di fisica: top, Higgs, W/Z +heavy flavour, b -physics...
- I **muoni soft** prodotti dai loro decadimenti semileptonici sono **correlati** alla loro cinematica e permettono di studiarne le proprietà;
- **Vantaggi**: segnatura chiara, osservabili leptoniche complementari a quelle basate sui jet;
- Studio svolto nel contesto dell'analisi di ATLAS *Measurement of the b -fragmentation in $t\bar{t}$ events using soft muon tagging*, che studia le proprietà del processo di frammentazione del quark b in eventi $t\bar{t}$;
- **Tutti i risultati che seguono sono generali, applicabili a ogni analisi che impieghi i muoni soft.**



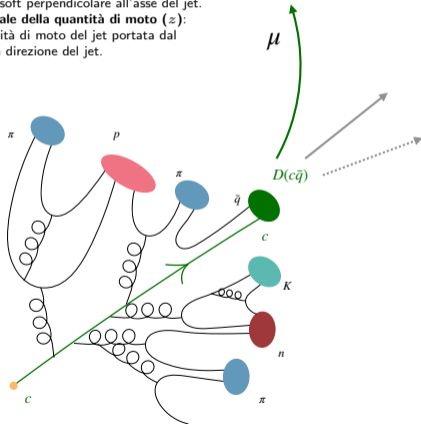


Misure coi muoni soft

- Si ricostruisce un jet adronico con algoritmo antiKt con parametro $R = 0.4$;
- Si identifica un muone associato al jet entro $\Delta R < 0.4$;
- Si possono ricostruire diverse osservabili:
 - cinematica del muone soft \rightarrow correlata a quella dell'adrone pesante;
 - $\Delta R(\text{soft} - \mu, \text{jet}), p_T^\mu/p_T^{\text{jet}}, p_T^{\text{rel}}, z \rightarrow$ sensibili alla frammentazione dell'adrone pesante;
 - in eventi $t\bar{t}$, osservabili sensibili a m_{top} , a violazioni CP, ecc...
- **Aspetto cruciale: i muoni devono effettivamente venire da un quark b o c !**

Momento trasverso relativo all'asse del jet (p_T^{rel}):
corrisponde al modulo della componente della quantità di moto del muone soft perpendicolare all'asse del jet.

Frazione longitudinale della quantità di moto (z):
frazione della quantità di moto del jet portata dal muone soft lungo la direzione del jet.

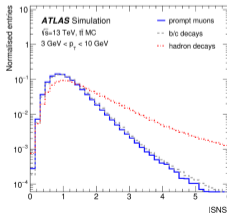
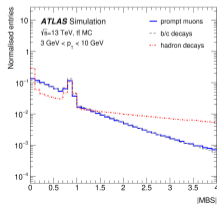
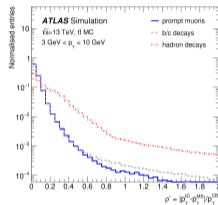
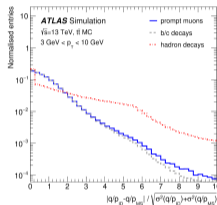




Soft Muon Tagging

- Muoni prodotti da π o $K =$ "fake";
- Decadimenti di π/K in μ con piccolo angolo: due tracce ricostruite come unica traccia muonica;
- Di solito rigettati in ricostruzione, ma comunque numerosi. Hanno peggiore χ^2 e notevole sbilanciamento tra le tracce nell'Inner Detector e nello spettrometro a muoni;
- Bisogna selezionare i muoni soft e scartare quelli fake! → **Soft Muon Tagging (SMT)**
- Approccio versatile complementare al tagging dei jet standard, ora basato sul punto di lavoro di identificazione dei muoni TIGHT da ATLAS Muon Combined Performance (basato su tagli);
- Si può migliorare SMT attraverso il Machine Learning (ML)? → TIGHTNN 1.0.

ATLAS Coll., Eur. Phys. J. C 81 (2021) 578





Architettura di TightNN 1.0

Layer di input

Layer di normalizzazione preprocessa le variabili (migliora la convergenza)

Layer fully connected a 64 neuroni. Funzione di attivazione ReLU. Seguono batch normalisation (stabilizza e accelera l'addestramento) e dropout (tasso=0.3, per mitigare effetti di overtraining).

Layer fully conn. a 128 neuroni. Attivazione ReLU, batch norm., dropout.

Layer fully conn. a 64 neuroni. Attivazione ReLU.

Layer fully conn. a 32 neuroni. Attivazione ReLU.

Layer fully conn. a 8 neuroni. Attivazione ReLU.

Layer di output a 1 neurone. Attivazione Sigmoid.

Implementato in Keras (backend Tensorflow), ottimizzatore Adam, loss binary cross-entropy, batch size = 2048

EarlyStopping ferma l'addestramento in assenza di miglioramenti rilevanti per 10 epoche. Migliora la generalizzazione del metodo ed evita l'overfitting.

ReduceLRonPlateau dimezza il learning rate se la loss di validazione non migliora per 5 epoche.



Addestramento e validazione

- #epoche max = 100 (di solito EarlyStopping interviene ben prima);
- Campioni di Run 2 (Run 3) prodotti col campione $t\bar{t}$ POWHEG BOX 2 + PYTHIA 8.230 (POWHEG BOX 2 + PYTHIA 8.308).
 - Selezione di base ℓ +jets, poi muoni soft con $p_T > 4$ GeV e $|\eta| < 2.5$ (entro $\Delta R < 0.4$ di un jet con $p_T > 25$ GeV e $|\eta| < 2.5$) e che passano il punto di lavoro per l'identificazione dei muoni *Medium*.
- Divisione 50%-50% addestramento-validazione (~ 11 M muoni ciascuno). Statistica sufficiente per l'addestramento; necessaria alta statistica in validazione per studi di prestazioni in diverse regioni dello spazio delle fasi. Provata anche divisione addestramento-validazione-test 60%-20%-20%, risultati analoghi;
- Addestramenti su dati di Run 2 (Run 3) su 1 **GPU** NVIDIA H100, con 6 s (10 s) a epoca.



Variabili di addestramento per Run 2 e Run 3

Hit in parti del rivelatore

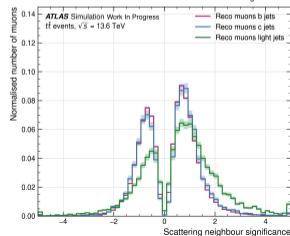
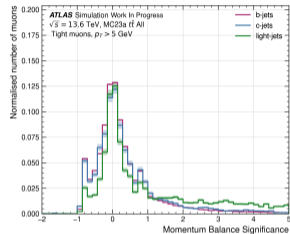
- MMHits (solo Run 3)
- etaLayer1STGCHits (solo Run 3)
- etaLayer2STGCHits (solo Run 3)
- middleLargeHits
- middleSmallHits
- numberOfPixelDeadSensors
- numberOfPixelHits
- outerLargeHits
- outerSmallHits
- phiLayer1STGCHits (solo Run 3)
- phiLayer2STGCHits (solo Run 3)

- author (= algoritmo che ha creato dato oggetto ricostruito)
- spectrometerFieldIntegral (riferito al campo magnetico nello spettrometro a muoni)

Significance/differenze/agreement

- momentumBalanceSignificance
- qOverPSignificance
- reducedChi2
- reducedChi2_ID
- reducedChi2_ME
- rhoPrime
- scatteringCurvatureSignificance
- scatteringNeighbourSignificance
- segmentDeltaEta

Alcune var. solo in Run 3 per via delle differenti condizioni di questa presa dati (es. nuove componenti nel rivelatore)



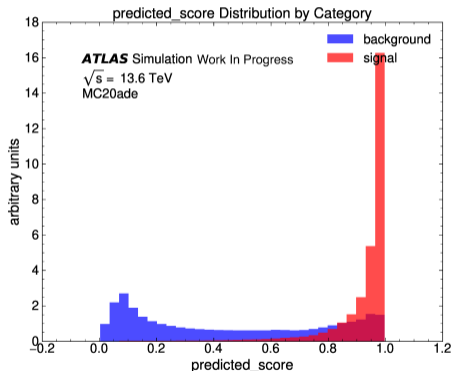


Prestazioni · Metodologia

- Addestramenti valutati su evoluzione della loss, accuratezza, curva ROC, AUC e rilevanza della singola variabile nelle prestazioni del modello.
 - Quest'ultima valutata selezionando le variabili una ad una e ricalcolando l'accuratezza dopo aver applicato permutazioni casuali sulla variabile selezionata per tutti i muoni nel campione. **Maggiore perdita di accuratezza quando si altera l'informazione contenuta in una delle variabili → maggiore importanza di quella variabile.**
- Ai muoni è attribuito un punteggio da 0 (muone simil-fondo) a 1 (simil-segnale). **Taglio di selezione applicato sul punteggio definisce un punto di lavoro del classificatore**, regolabile attraverso la curva ROC per l'efficienza di segnale vs rigetto del fondo desiderato.



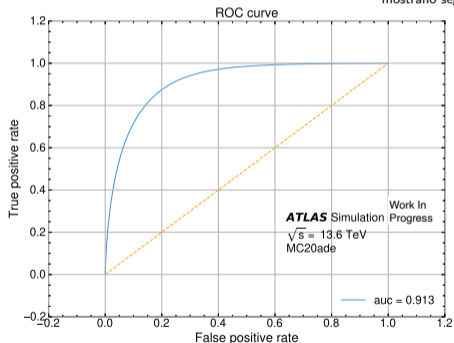
Prestazioni · Punteggio, curva ROC e AUC · Run 2



Curva ROC (Receiving Operating Characteristic): tasso di veri positivi vs tasso di falsi positivi per ogni possibile taglio tra segnale e fondo.

AUC (Area Under the (ROC) Curve): area sottesa dalla curva ROC.

Grafico fatto sul campione di validazione. Test recenti fatti separando un campione di validazione e uno di test non mostrano segni di overfit.

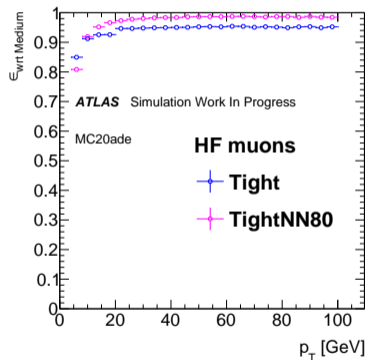
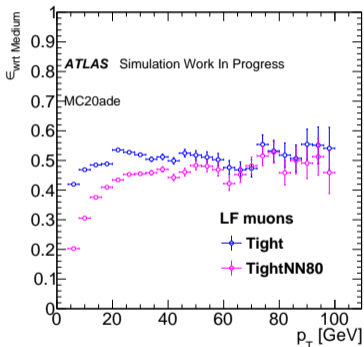


In TIGHTNN 1.0 taglio impostato a *punteggio* > 0.8, affinché l'efficienza di segnale ottenuta corrisponda a quella dell'implementazione basata su tagli del punto di lavoro di identificazione dei muoni TIGHT.



Prestazioni · Efficienza su p_T · Run 2

$$\text{Efficienza} = \frac{\text{Eventi che passano la selezione}}{\text{Eventi totali}}$$

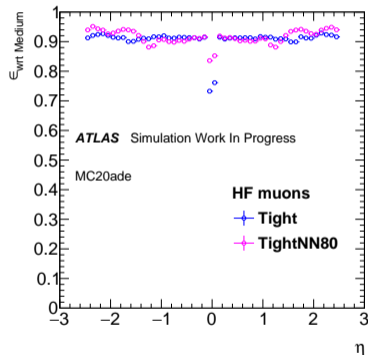
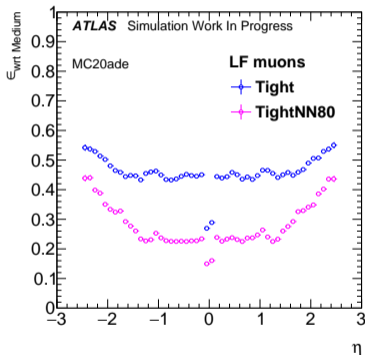


Studio sul campione di validazione con eventi $t\bar{t}$ di POWHEG BOX 2 + PYTHIA 8.230. Prestazioni simili anche con POWHEGHERWIG7 e SHERPA → dipendenza molto bassa dalla modellizzazione MC.



Prestazioni · Efficienza su η · Run 2

$$\text{Efficienza} = \frac{\text{Eventi che passano la selezione}}{\text{Eventi totali}}$$

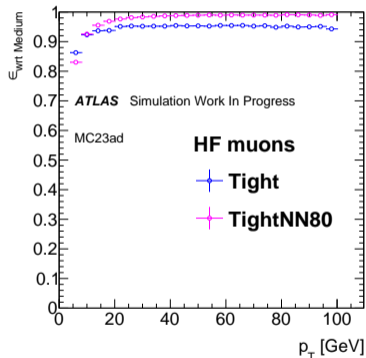
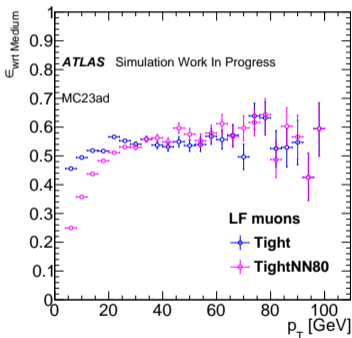


Acceptance del fondo di TIGHTNN 1.0 **inferiore** in maniera consistente su tutto lo spazio delle fasi → **separazione segnale-fondo più efficace!** Dipendenza molto bassa da modellizzazione MC.



Prestazioni · Efficienza su p_T · Run 3

$$\text{Efficienza} = \frac{\text{Eventi che passano la selezione}}{\text{Eventi totali}}$$

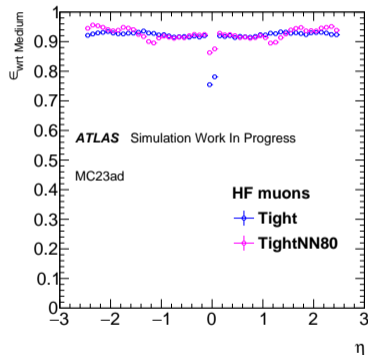
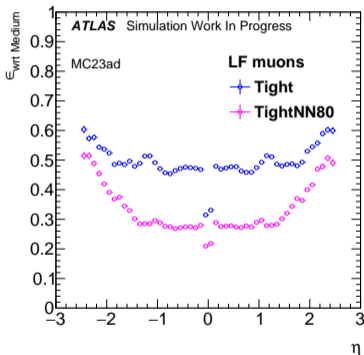


Studio sul campione di validazione con eventi $t\bar{t}$ di POWHEG BOX 2 + PYTHIA 8.308. Prestazioni simili anche con POWHEGHERWIG7 e SHERPA → dipendenza molto bassa dalla modellizzazione MC.



Prestazioni · Efficienza su η · Run 3

$$\text{Efficienza} = \frac{\text{Eventi che passano la selezione}}{\text{Eventi totali}}$$



Acceptance del fondo di TIGHTNN 1.0 **inferiore** in maniera consistente su tutto lo spazio delle fasi → **separazione segnale-fondo più efficace!** Dipendenza molto bassa da modellizzazione MC.



Ripesatura

- È possibile che, anche se il p_T dei muoni non è tra le variabili del modello, il modello **impari a scartare più muoni a basso p_T per via di correlazioni di altre variabili con p_T , causando una selezione biased**;
- Si può rendere TIGHTNN 1.0 **più robusto** contro la dipendenza da p_T ? → **ripesatura degli eventi di segnale e fondo** usati per l'addestramento per averli con distribuzione del p_T dei muoni simile;
- Tre approcci di ripesatura studiati per TIGHTNN 1.0:
 - **Appiattimento**: distribuzioni di segnale e fondo appiattite;
 - **Reshaping del segnale**: trasforma il segnale affinché corrisponda alla shape della distribuzione del fondo;
 - **Reshaping del fondo**: trasforma il fondo affinché corrisponda alla shape della distribuzione del segnale.
- Modello non ripesato: forte dipendenza da p_T del muone e identificazione del segnale che si deteriora notevolmente a p_T molto basso. Modelli ripesati: comportamento più stabile per tutti i p_T . Comunque, **per l'acceptance del fondo (da tenere il più bassa possibile), il modello non ripesato ha prestazioni migliori, quindi TIGHTNN 1.0 al momento non adopera ripesatura.**

Architetture alternative studiate

- **Reti neurali più sofisticate**, incluse reti neurali ricorrenti con Long Short-Term Memory (**LSTM**) e reti basate su **transformer**.
 - Prestazioni simili alla Deep Neural Network (DNN) di riferimento, ma con **tempi di addestramento ed esecuzione notevolmente maggiori** → **approcci non studiati ulteriormente**.
- **Gradient-Boosted Decision Trees (GBDTs)**, basati su XGBoost. Altamente ottimizzati per compiti di classificazione relativamente semplici e generalmente più semplici da addestrare e interpretare rispetto alle reti neurali. Queste ultime invece possono modellizzare relazioni non lineari e adattarsi a una vasta gamma di complessità di problemi.
 - Prestazioni paragonabili a TIGHTNN 1.0, quindi **solitamente sarebbero preferiti**, ma TIGHTNN 1.0 **mostra tempo di addestramento ridotto e impatto trascurabile sul tempo di esecuzione del software di analisi di ATLAS** per la valutazione del punteggio per singolo muone. Inoltre, **l'algoritmo deve essere estendibile a compiti più complessi di classificazione di eventi, attualmente in considerazione** → **DNN come scelta finale**.

Conclusioni e prospettive

- TIGHTNN 1.0 è un **nuovo algoritmo basato su ML per l'identificazione dei muoni soft** in jet adronici prodotti dal decadimento di adroni a sapore pesante;
- TIGHTNN 1.0 è basato su una **DNN** e adopera una selezione di variabili dei muoni per rigettare il fondo dei muoni prodotti dai decadimenti di adroni a sapore leggero;
- Il confronto delle prestazioni con quelle dei punti di lavoro di identificazione dei muoni esistenti basati su tagli mostra un **miglioramento fino a un fattore 2 nel rigetto del fondo per un'efficienza di segnale simile**;
- **Miglioramento significativo** per analisi influenzate significativamente dai muoni di fondo prodotti da adroni a sapore leggero! Sviluppato specificamente per il tagging SMT, ma può essere impiegato in generale da analisi interessate a tassi di muoni fake più bassi;
- Studio in corso per ulteriori miglioramenti.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



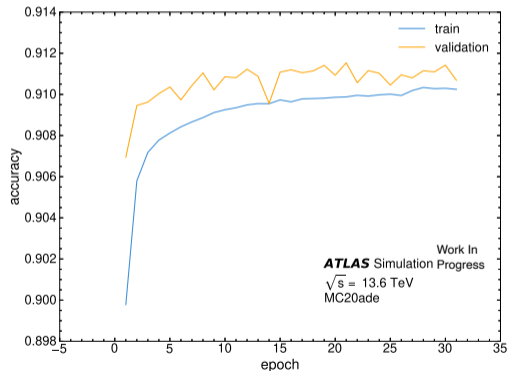
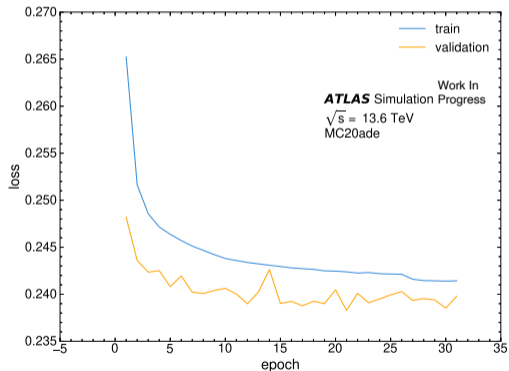
Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

BACKUP

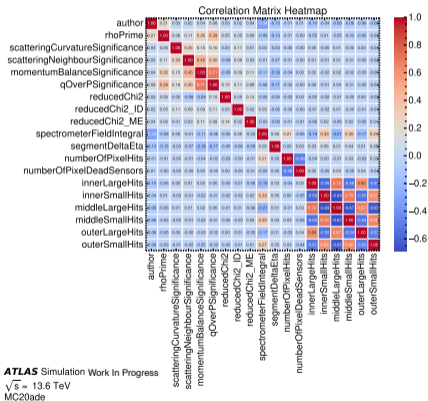
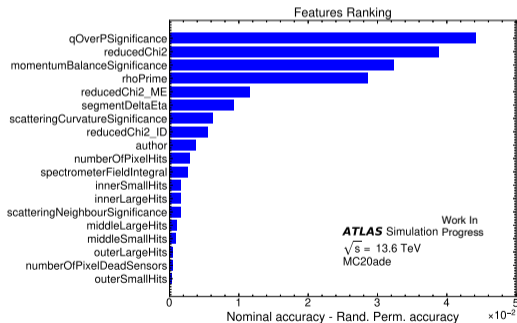
Prestazioni · Loss e accuratezza · Run 2



Prestazioni di validazione leggermente migliori di quelle di addestramento per via del dropout.

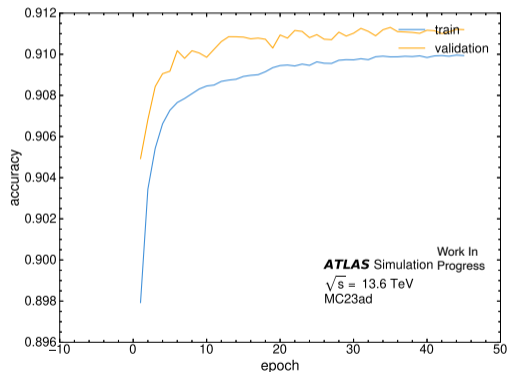
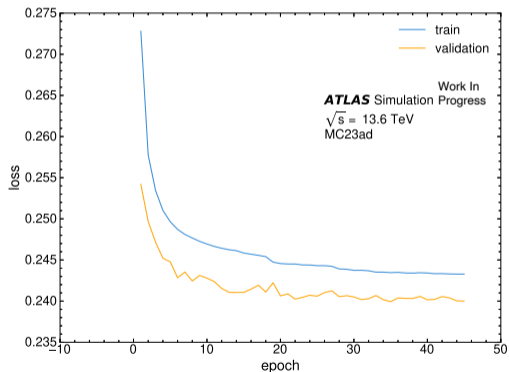


Prestazioni · Feature Ranking e matrice di correlazione · Run 2



Solo alcune variabili hanno la maggiore influenza sulla selezione.

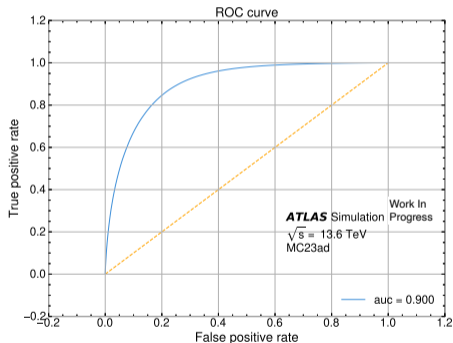
Prestazioni · Loss e accuratezza · Run 3



Prestazioni di validazione leggermente migliori di quelle di addestramento per via del dropout.



Prestazioni · Curva ROC e AUC · Run 3



Curva ROC (Receiving Operating Characteristic): tasso di veri positivi vs tasso di falsi positivi per ogni possibile taglio tra segnale e fondo.

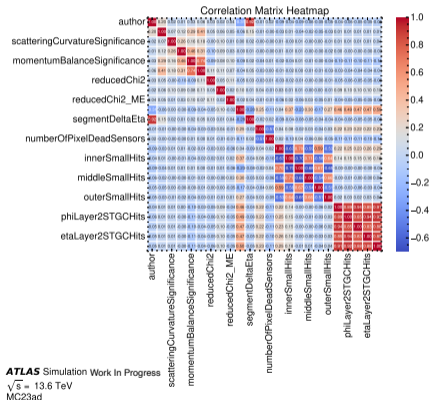
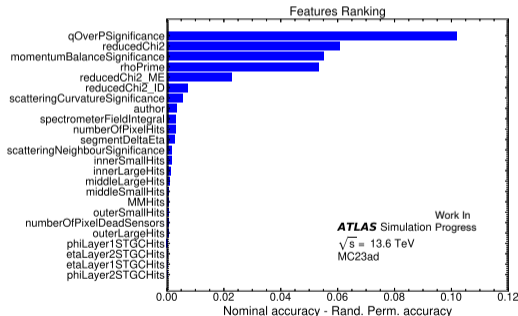
AUC (Area Under the (ROC) Curve): area sottesa dalla curva ROC.

Grafico fatto sul campione di validazione. Test recenti fatti separando un campione di validazione e uno di test non mostrano segni di overfit.

In TIGHTNN 1.0 taglio impostato a *punteggio* > 0.8, affinché l'efficienza di segnale ottenuta corrisponda a quella dell'implementazione basata su tagli del punto di lavoro di identificazione dei muoni TIGHT.



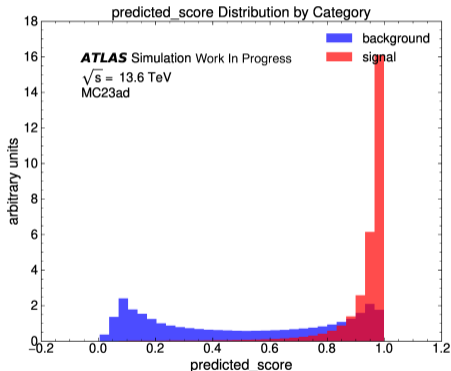
Prestazioni · Feature Ranking e matrice di correlazione · Run 3



Solo alcune variabili hanno la maggiore influenza sulla selezione.



Prestazioni · Punteggio · Run 3



In TIGHTNN 1.0 taglio impostato a *punteggio* > 0.8 , affinché l'efficienza di segnale ottenuta corrisponda a quella dell'implementazione basata su tagli del punto di lavoro di identificazione dei muoni TIGHT.