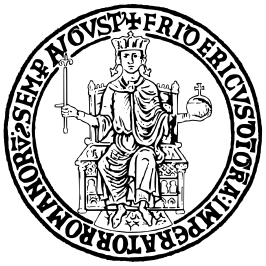
UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"



Scuola Politecnica e delle Scienze di Base

Area Didattica di Scienze Matematiche Fisiche e Naturali

Dipartimento di Fisica "Ettore Pancini"

Laurea Triennale in Fisica

Ricostruzione del bosone di Higgs in stati finali $b\bar{b}$ con tecniche di machine learning per ricerche di nuova fisica con il rivelatore CMS ad LHC

Relatore:
Prof. Alberto Orso Maria Iorio

Candidato:
Alexandro Martone
Matr. *N*85001068

Anno Accademico 2019/2020

À ceux qu'étaient là
et qui m'ont porté,
au propre comme au figuré.
Ceux qu'ont adapté leur vie
pour rendre la mienne
moins compliquée

Contents

Introduzione

Il Modello Standard (MS) è, ad oggi, la teoria di maggior successo nella descrizione delle particelle fondamentali e delle loro interazioni, sia in quanto descrive tre delle quattro forze fondamentali conosciute, sia per le numerose conferme sperimentali alle sue previsioni. Tutte le particelle da esso previste sono state scoperte ed i suoi parametri fondamentali sono stati misurati ad un altissimo livello di precisione. Il MS non è tuttavia esaustiva del comportamento della Natura, di fatti non spiega l'interazione gravitazionale o la materia oscura; per superare questi ed altri problemi insiti nel MS, sono state formulate diverse teorie dette Oltre il Modello Standard, o Beyond the Standard Model (BSM). Questo studio sarà incentrato su delle particelle la cui presenza è prevista da numerosi modelli BSM: i *Vector-Like Quarks* (VLQ). I VLQ hanno massa dell'ordine del TeV e differenze sostanziali dai quark del Modello Standard. Attualmente le ricerche di VLQ vengono portate avanti al *Large Hadron Collider*, in particolare agli esperimenti ATLAS e CMS. Il rivelatore CMS ha raccolto dati fino al 2018 con energia del centro di massa di $\sqrt{s} = 13$ TeV e luminosità di picco di $2 \times 10^{34} cm^2 s^{-1}$. La presa dati è prevista riavviarsi per il 2022 dopo un upgrade di LHC e degli esperimenti.

Oggetto del presente lavoro di tesi è la ricerca di un VLQ T' singolo, trattando in particolare il suo decadimento in un quark top ed un bosone di Higgs. Nello specifico viene studiato il caso in cui esso decade in una coppia quark/antiquark b. La segnatura del canale di decadimento del T' coincide inoltre con quello della produzione associata di quark top e bosoni di Higgs tHq, un processo molto interessante in quanto previsto dal MS ma non ancora osservato. Il lavoro è stato svolto utilizzando un campione simulato di segnali T' che riproduce le condizioni della presa dati a 13 TeV. Lo studio del canale tHq è stato effettuato utilizzando tecniche di Machine Learning - in particolare l'algoritmo XGBoost- con l'intento di migliorare l'efficienza dei metodi tradizionali, sviluppando tecniche che potrebbero essere utilizzate per le analisi all'inizio della nuova presa dati di LHC. Per tale scopo è stato, in prima istanza, implementato un classificatore binario, per poi approfondire lo studio creando un multiclassificatore.

La trattazione si articola in tre capitoli:

- Nel primo capitolo viene fatta una breve panoramica del Modello Standard e della fisica oltre il modello standard necessaria per inquadrare la ricerca;

- Il secondo capitolo consiste nella descrizione dell'acceleratore LHC e del rivelatore CMS;

- Il terzo ed ultimo capitolo, dopo una breve introduzione alle tecniche di Machine Learning utilizzate, riporterà i risultati ottenuti.

1 Il Modello Standard

La principale infrastruttura teorica della fisica delle particelle è il Modello Standard (**MS**), una teoria quantistica e relativistica [1] [2], che, ad oggi, riesce a mettere in relazione con ottima accuratezza le interazioni fondamentali: elettromagnetica, debole, forte e gravitazionale. [3]

Le prime due vengono unificate nella teoria delle interazioni elettrodeboli, o modello Glashow-Weinberg-Salam (GWS)[4]; mentre la cromodinamica quantistica (QCD) viene a completare il MS. L'interazione gravitazionale è dunque l'unica interazione fondamentale nota che non trova spiegazione nell'ambito del MS. Nel seguito della trattazione, verranno utilizzate le cosidette "unità naturali", in cui i valori numerici della velocità della luce, c, e della costante di Planck $\hbar$ sono 1.

1.1 Le particelle del MS

Le particelle che permettono di descrivere le interazioni del MS sono molteplici - come si evince in figura 1 - , e sono racchiuse in tre categorie principali: i Quark (fermioni con spin 1/2), i Leptoni (anch'essi fermioni di spin 1/2) e particelle mediatrici (bosoni). Alle particelle, inoltre, si associano le rispettive antiparticelle, che presentano numeri quantici opposti ma stessa massa.

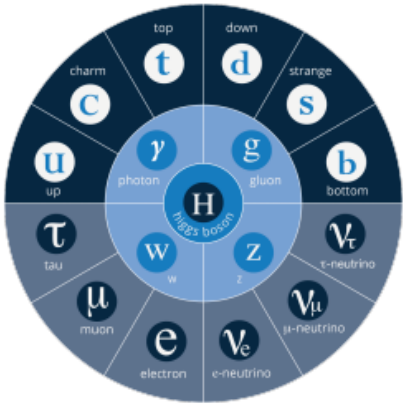


Figure 1: Le particelle del MS

1.1.1 I Quark e le loro interazioni

I Quark sono particelle divise in tre famiglie, descritte in tabella 1. La loro carica elettrica è espressa in termini della carica unitaria dell'elettrone e e

vale +2/3 per i quark u, c e t, mentre per d, s e b vale -1/3. Sono gli unici fermioni che risentono della forza forte e, conseguentemente sono le uniche ad essere caratterizzate da una carica aggiuntiva associata proprio a tale interazione: il colore. Ogni quark, infatti, può essere associato al colore verde (g), rosso (r) o blu (b). La carica di colore tuttavia non è mai osservabile in nessuna particella formata da quark, a causa di uno dei postulati della cromodinamica quantistica, il confinamento del colore. Per questo, nei mesoni avremo un quark di un colore e l'altro che trasporta il corrispondente anti-colore, mentre in un barione avremo i tre quark di colori diversi, in modo che il risultato sia "bianco", ovvero sia neutro in termini di carica di colore.

Table 1: Le famiglie di quark e le loro caratteristiche

Famiglia	Quark	Carica [Q/e]	Massa [GeV]
Prima	u	+2/3	$< 2.3 \times 10^{-3}$
	d	-1/3	$< 4.8 \times 10^{-3}$
Seconda	c	+2/3	1.28
	s	-1/3	95×10^{-3}
Terza	t	+2/3	173.5
	b	-1/3	4.18

Per lo stesso motivo, non è mai possibile osservare un quark isolato: agendo su un adrone allontanandone i due quark, si aumenterà l'energia potenziale fino al punto in cui essa sarà tale da creare una nuova coppia quark-antiquark, i quali potranno proseguire nella creazione di nuove coppie. Il processo si arresta quando tutte le particelle con carica di colore si sono ricombinate in adroni.

Questo si verificherà anche in alcuni dei fenomeni in esame in questo lavoro, in cui i quark nello stato finale di un processo di *scattering* produrranno jet adronici (sciami di particelle) che verranno poi ricostruiti e analizzati.

1.1.2 I leptoni e le loro interazioni

I Leptoni (tabella 2) sono suddivisi in tre classi o famiglie, ognuna formata da una leptone carico e dal corrispettivo leptone neutro, detto neutrino.

I Leptoni carichi, i quali risentono sia dell'interazione elettromagnetica che di quella debole, sono: elettroni e^-, muoni μ^- e tauoni τ^-. Queste particelle hanno una massa di 0,511 MeV per l'elettrone, 105,7 MeV per il muone e i 1777 MeV per il tauone, il più massivo. Esse, inoltre, hanno la medesima carica elettrica ($1,602\times\ 10^{19}$ C).

Table 2: Leptoni e le loro caratteristiche

Famiglia	Leptone	Carica [Q/e]	Massa [MeV]
Prima	e	-1	0.511
	ν_e	0	$< 0.22 \times 10^{-3}$
Seconda	μ	-1	105.7
	ν_μ	0	< 0.19
Terza	τ	-1	1777
	ν_τ	0	< 18.2

I Leptoni privi di carica, invece, sono i rispettivi neutrini (elettronico ν_e, muonico ν_μ e tauonico ν_τ), i quali sono soggetti alla sola forza debole e possono quindi attraversare lunghi tratti di materia senza interagire, pertanto rilevarli risulta molto complesso. Queste particelle sono di massa molto inferiore a tutte le altre e a tutt'oggi il valore esatto di tale massa risulta ancora non definito.

1.1.3 Particelle mediatrici

Dette anche bosoni di gauge o vettoriali, in quanto hanno spin uguale a 1. Queste particelle sono suddivise in base alla forza che mediano: l'interazione elettromagnetica ha il suo mediatore nel fotone, la forza forte si estrinseca mediante il gluone e la forza debole viene esercitata attraverso i bosoni $W^{\pm}$ e Z_0

Table 3: Bosoni mediatori e le loro caratteristiche

Interazione	Bosone	Carica [Q/e]	Spin	Massa [GeV]
Elettromagnetica	γ	0	1	0
Debole	$W^{\pm}$	± 1	1	80.4
	Z_0	0	1	91.2
Forte	g	0	1	0

Quest'ultima suddivisione avviene in virtù delle caratteristiche dell'interazione: se si ha lo scambio di un bosone $W^{\pm}$ si ha la cosiddetta interazione di corrente carica, mentre se avviene lo scambio di un bosone Z_0 si ha una interazione di corrente neutra.

Per quanto riguarda il gluone, essendo esso il mediatore della forza forte deve essere dotato di una carica di colore: si dice pertanto che il gluone sia

bicolorato, ossia presenti il colore del quark che lo ha emesso e l'anticolore del quark target dell'interazione.

1.2 Il bosone di Higgs

Sia nella teoria elettrodebole che nella cromodinamica quantistica le interazioni tra particelle vengono descritte da teorie di campo, ognuna delle quali costruita a partire da una particolare simmetria detta invarianza di gauge.[5] [6] Una simmetria di gauge prevede bosoni di campo di spin 1 e massa 0, quali il fotone per le interazioni elettromagnetiche e i gluoni per le interazioni forti. L'invarianza di gauge gioca un ruolo fondamentale nella teoria unificata elettrodebole ove, per altro, è richiesta per cancellare divergenze che compaiono nelle ampiezze relative a singoli grafici di Feynman.

Tuttavia, nella teoria elettrodebole, mentre il fotone continua ad avere massa zero, gli altri tre bosoni di campo, $W^{\pm}$ e Z_0, hanno massa, per di più significativamente diversa da zero e maggiore di tutti i fermioni del MS, eccezion fatta per il quark top. Questo problema - ed altri - si superano assumendo che le particelle del MS interagiscano con un nuovo tipo di campo, il campo di Higgs, la cui esistenza ha due conseguenze:

- i bosoni di gauge acquistano massa senza violare l'invarianza di gauge, mentre il fotone resta privo di massa
- ci sono quanti neutri di spin 0 (H_0, bosone di Higgs) associati al campo di Higgs

Pur non predicendone la massa, la teoria dice come il bosone di Higgs si accoppi alle altre particelle, ovvero tramite interazione di contatto con costante di accoppiamento proporzionale alla loro massa. Si accoppia, dunque, molto debolmente con le particelle leggere (ν; e; μ; u; d; s) e fortemente a particelle pesanti quali τ, b, $W^{\pm}$, Z_0 e top.

Per poterlo quindi osservare con sufficiente probabilità, occorre innanzitutto essere in grado di produrre energia sufficiente a raggiungere la sua massa. La ricerca dell'Higgs inizia nei primi anni Settanta e, col progredire delle tecnologie, ha consentito di imporre limiti inferiori alla massa, escludendo masse via via crescenti nel corso degli anni. I risultati più significativi, prima di LHC, vengono da LEP e TEVATRON ed hanno permesso di porre un limite inferiore $m(H) > 114.4$ GeV con un confidence level del 95% . Successivamente, nel 2012, gli esperimenti ATLAS e CMS ad LHC hanno annunciato la scoperta di una risonanza interpretabile come bosone di Higgs, di massa 125 GeV.

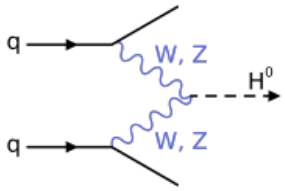
1.2.1 Processi di formazione

Grazie agli esperimenti di LHC, sono stati osservati diversi meccanismi di produzione del bosone di Higgs ed in particolare la fusione di bosoni vettori (*Vector Boson Fusion* VBF), la fusione di gluoni (*gluon gluon Fusion* ggF), la produzione associata VH (*Higgsstrahlung*) e la fusione di quark Top ($t\bar{t}$).

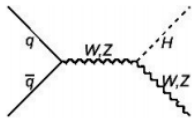
Il processo VBF consiste nella fusione di due bosoni vettoriali ($W^{\pm}$ e Z_0) emessi da una coppia di quark. La produzione del bosone H è associata a quella di due getti adronici (fig. 2a).

L'Higgsstrahlung (fig. 2b) prevede una coppia quark antiquark che può fondere e produrre un bosone massivo che decade producendo un Higgs per irraggiamento. In questo processo la produzione dell'Higgs è accompagnata a quella di un altro bosone massivo ($W^{\pm}$, Z_0).

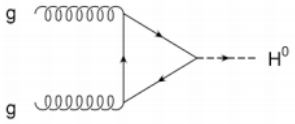
Il processo ggF è il meccanismo dominante di produzione (difatti ha una sezione d'urto all'incirca di un ordine di grandezza maggiore a quella del processo VBF) e si basa sulla fusione di due gluoni provenienti dalla collisione di protoni. Il bosone si accoppia attraverso uno stadio (loop) intermedio di quark top o bottom. (fig.2c) La produzione top/antiquark-top, infine, prevede che l'Higgs sia generato dalla fusione di due quark ed antiquark top prodotti da gluoni (fig. 2d).



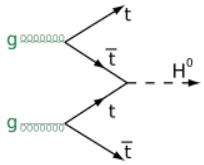
(a) Processo VBF



(b) Higgsstrahlung



(c) Processo ggF



(d) Processo tt

Figure 2: Diagrammi di Feynman per i quattro processi

1.2.2 Decadimento

Analogamente a quanto avviene per i processi di formazione, il bosone di Higgs ha diversi canali di decadimento, come riportato in figura 3.

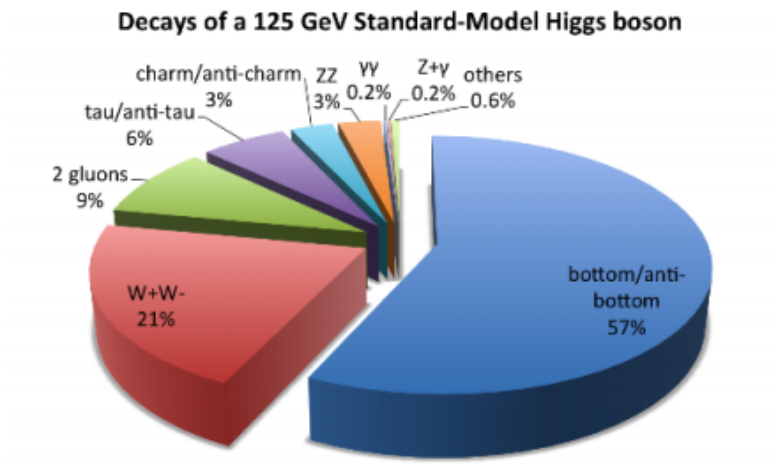


Figure 3: Canali di decadimento dell'Higgs e rispettivi branching ratio

Di questi, principali canali di decadimento sono:

- H $\rightarrow$ $q\overline{q}$: Il processo dominante di questo tipo è il decadimento in bottom quark-antiquark. Sebbene sia il favorito, questo canale presenta un di fondo dovuto ai numerosi di jet in seguito alla collisione tra i due protoni.

- H $\rightarrow \gamma\ \gamma$ /gg : Fotoni e gluoni, entrambi privi di massa, possono essere prodotti mediante loop che coinvolgono particelle cariche o colorate. Tuttavia, il canale di decadimento in 2 gluoni è inutilizzabile per lo stesso motivo del canale in due quarks; il decadimento in 2 fotoni, invece, è di grande importanza sperimentale, nonostante il basso Branching Ratio, a causa delle caratteristiche distintive dovute alla presenza di 2 fotoni molto energetici nello stato finale.

- H $\rightarrow$ Z_0 Z_0 oppure W^+W^-: i bosoni vettoriali in cui può decadere l'Higgs non vengono rivelati direttamente, ma decadono a loro volta in leptoni o in quark. Di particolare interesse per lo studio dell'Higgs è il canale di decadimento H $\rightarrow$ Z_0 Z_0 $\rightarrow$ 4l, con l=($\mu\pm$;e $\pm$). Tale canale è detto *Golden Channel*, per il caratteristico stato finale in 4 leptoni, che permette di ricostruire completamente il decadimento; inoltre, grazie alla grande accuratezza con cui è possibile ricostruire i leptoni, si può

ottenere un'ottima risoluzione nella misura della massa del bosone di Higgs. Le due Z potrebbero anche decadere in due leptoni e due quark, ma questo canale di decadimento non permette la stessa risoluzione nella misura della massa, a causa della minore accuratezza nella ricostruzione dei jet prodotti dai quark in cui decade uno dei bosoni Z. Sebbene il Branching Ratio del canale $H \rightarrow W^+W^-$ sia maggiore di quello del canale $Z_0\ Z_0$, quando l'Higgs decade in due bosoni W, la sua massa non è ricostruita con un'alta risoluzione, poichè i possibili stati finali di questo decadimento sono: lνqq e l$\nu\nu$, caratterizzati entrambi da problematiche a livello di studio sperimentale: il primo per la presenza di jet difficili da ricostruire accuratamente, mentre il secondo a causa della grande energia mancante dovuta alla non rivelazione dei neutrini.

1.3 Fisica oltre il modello standard

Nonostante il suo grande successo, il modello standard presenta diverse problematiche:

- non discende da un unico principio di simmetria;
- dipende da molti parametri da determinare sperimentalmente;
- la stabilità del protone non è fatta risalire a ragioni dinamiche;
- non riesce a spiegare l'asimmetria materia-antimateria;
- è in contraddizione con le recenti misurazioni della massa del neutrino;
- non include la gravitazione;
- il valore basso della massa del bosone di Higgs (problema della gerarchia);
- la grande distanza tra scala di Planck e scala elettrodebole
- non prevede l'esistenza della materia oscura

Per superare tali difficoltà, sono state proposte diverse teorie dette *Beyond Standard Model* (BSM). Ne sono un esempio quella dell'Higgs Composito e quella delle Extra Dimensions, che prevedono l'esistenza di nuove particelle dette *Vector Like Quark* (VLQ). [7] La teoria dell'Higgs Composito prevede che l'Higgs sia uno stato composito di una nuova interazione forte. La dimensione dell'operatore di massa dell'Higgs potrebbe essere maggiore di 4, ciò

spiegherebbe il perché dei bassi valori misurati della massa dell'Higgs. La teoria sulle Extra Dimensioni invece, per spiegare il problema dell'unificazione delle interazioni fondamentali, predice che esista un'ulteriore dimensione oltre quella del tempo. Per spiegare entrambe le teorie sono fondamentali i VLQ.[7]

1.3.1 Vector Like Quark

I *Vector Like Quark* (VLQ) sono previsti da diverse teorie di estensione del modello standard, pur tuttavia senza ancora riscontri sperimentali. [8] Si tratta di fermioni non chirali con la stessa carica di colore dei quark del MS. I VLQ possono accoppiarsi con i quark del modello standard modificando il loro accoppiamento con i bosoni W, Z e di Higgs e possono rompere il meccanismo di GIM (Glashow–Iliopoulos–Maiani), permettendo quindi cambiamento di sapore dei quark in interazioni a corrente neutra.

Meccanismi di produzione

Nelle collisioni protone protone, le sezioni d'urto della produzione dei VLQ dipendono dal loro accoppiamento con i quark del Modello Standard e anche dall'intensità del loro accoppiamento con i bosoni W e Z. I Meccanismi di produzione dei VLQ possono essere divisi in:

- produzione singola tramite processi di interazione debole. Essa dipende dalla massa dei fermioni, dai parametri di mescolamento con i quark del MS e dall'accoppiamento tra i nuovi quark e i bosoni W e Z;
- produzione in coppia regolata da processi di QCD. La sezione d'urto di questo processo diminuisce all'aumentare della massa dei nuovi fermioni e rispetto alla produzione singola è necessaria più energia per produrre due particelle. Questi processi sono simili alla produzione di coppie di quark del Modello Standard.

Canali di decadimento

I Vector Like Quark, rompendo il meccanismo di GIM, possono decadere in maniera elettrodebole e neutra in quark del MS o in altri VLQ. I principali canali di decadimento permessi in particelle del MS sono:

- T $\rightarrow W^+$b, Zt, Ht
- B $\rightarrow W^-$t, Zb, Hb
- X $\rightarrow W^+$t
- Y $\rightarrow W^-$b

2 L'acceleratore LHC e l'esperimento CMS

Il Large Hadron Collider (LHC) è, con i suoi 27 km di circonferenza, il più grande acceleratore di particelle al mondo. E' situato al confine Franco-Svizzero ad una profondità di circa 100m ed è in grado di far collidere particelle con un'energia nel centro di massa di progetto pari a 14 TeV e luminosità di $10^{35}\ cm^{-2}\ s^{-1}$.

2.1 Funzionamento di LHC

Prima di essere introdotte nell'LHC, le particelle sono iniettate in acceleratori più piccoli e raggiungono l'energia finale gradualmente. Esse passano prima attraverso il LINAC2 acquistando un'energia di 50 MeV; sono poi trasferite all'interno di un sistema di 4 sincrotroni sovrapposti - il PSB (Proton Syncrotron Booster) - che le accelera fino a raggiungere 1.4 GeV. Il passaggio finale avviene attraverso il SPS (Super Proton Syncrotron), dove vengono portate fino ad un'energia di 450 GeV e poi iniettate in LHC, dove dovranno circolare per circa 20 minuti prima di raggiungere le energie stabilite. Un'immagine schematica del sistema di acceleratori è mostrata in figura 4.

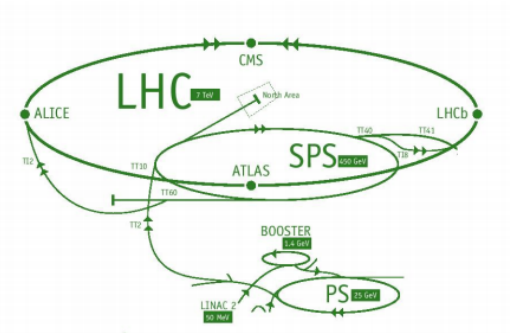


Figure 4: La struttura di acceleratori di LHC

La scelta di utilizzare adroni anziché leptoni è dettata da una convenienza in termini energetici: le particelle cariche infatti, nel descrivere traiettorie circolari, subiscono una perdita di energia (radiazione di sincrotrone) proporzionale alla quarta potenza dell'inverso della propria massa, oltre che al raggio della traiettoria (molto grande nel caso dell'LHC). Essendo gli adroni molto più massivi dei leptoni, questo garantisce all'acceleratore una minore spesa energetica per mantenere le particelle in moto all'energia desiderata. Favorevole dal punto di vista energetico è anche la scelta del tipo di collisioni generate all'interno dell'LHC, ovvero del tipo particella-particella e non

particella-bersaglio fisso, in quanto nel primo caso l'energia prodotta nella collisione è la somma delle energie trasportate dai singoli fasci, mentre nel secondo si genera un'energia proporzionale alla radice quadrata dell'energia della particella incidente.

Ad LHC ogni fascio di protoni è strutturato in circa 2800 pacchetti, detti *bunch*, con una separazione spaziale di 25 ns l'uno dall'altro e contenenti 10^{11} protoni. Dal rate di eventi per un dato processo di scattering si ha:

$$L = \sigma \cdot R$$

dove

$$L = \frac{dN_{eventi}}{dSdt}$$

è la luminosità istantanea (eventi per unità di tempo e superficie) e σ rappresenta la sezione d'urto del processo espressa in femtobarn [1]. Un problema che si riscontra nei processi con un tale numero di particelle è il cosiddetto *pile-up*, ovvero il sovrapporsi delle tracce dovute al verificarsi di eventi diversi dall'argomento di studio e che andranno a costituire il fondo da cui si dovrà cercare di estrarre il segnale desiderato.

Rispetto alla dimensione del bunch, solo poche particelle interagiranno e il loro punto di collisione sarà detto *vertice primario*. Lungo la traiettoria di LHC sono previsti quattro punti di collisione per i bunch, presso i quali sono situati i principali esperimenti:

- ALICE, specializzato nella fisica degli ioni pesanti e nello studio del plasma di quark e gluoni
- LHCB, che si concentra sullo studio della fisica del quark bottom e della simmetria tra particelle e anti-particelle
- ATLAS e CMS, infine, sono due grandi esperimenti per la ricerca di nuove risonanze pesanti e di nuova fisica in generale.

2.2 L'esperimento CMS

Il Compact Muon Solenoid (CMS) è un esperimento cosidetto *general puprose*, ossia è in grado di rivelare - con grande efficienza e precisione - fotoni, elettroni, muoni, leptoni tau, jet ed energia mancante indice della presenza di particelle neutre elusive. [9] Nonostante condivida gli stessi obiettivi di ATLAS, è stato sviluppato con tecniche e configurazioni diverse. Costituito in un cilindro di lunghezza 22m e diametro 15m, racchiude quello che, ad

[1] $10^{-39} cm^2$

oggi, è il più grande solenoide superconduttore, capace di generare un campo di 3.8T, la qual cosa gli permette di curvare le traiettore delle particelle, permettendone una misura accurata del momento.

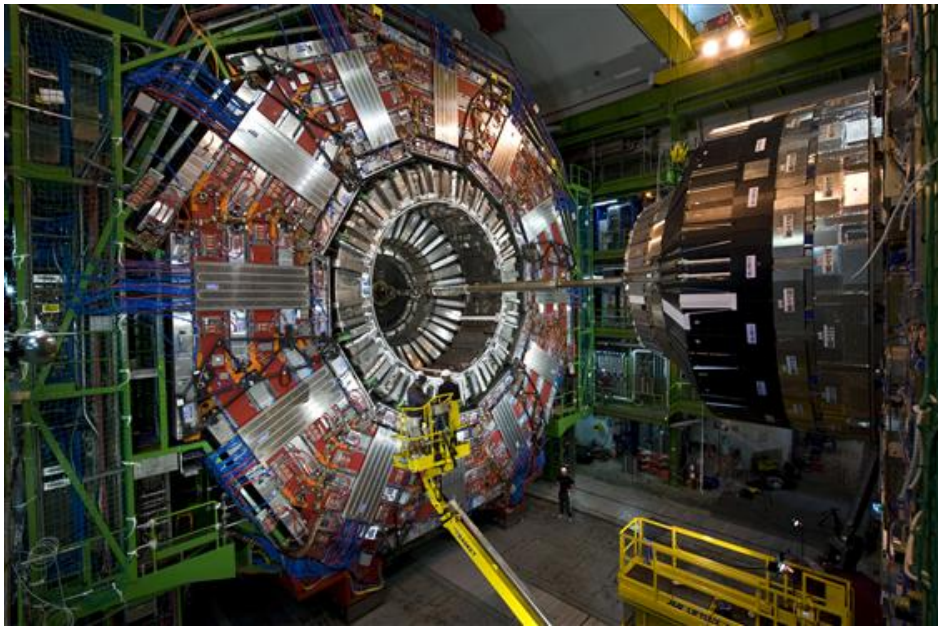


Figure 5: Una foto di CMS

2.2.1 Il sistema di coordinate

Il rivelatore è centrato nel punto di collisione per sfruttare al meglio la simmetria cilindrica dei processi di collisione. Possiamo quindi identificare due tipi di coordinate (figura 6):

1. Una terna cartesiana destrorsa:
    - **L'asse x** diretto verso il centro dell'anello di LHC
    - **L'asse y** diretto verso l'alto, ortogonalmente all'anello
    - **L'asse z** diretto tangenzialmente all'anello in senso antiorario
2. Un sistema di coordinate cilindriche:
    - La distanza radiale **r** dall'asse z
    - L'angolo azimutale ϕ attorno all'asse z
    - L'angolo polare θ attorno all'asse x

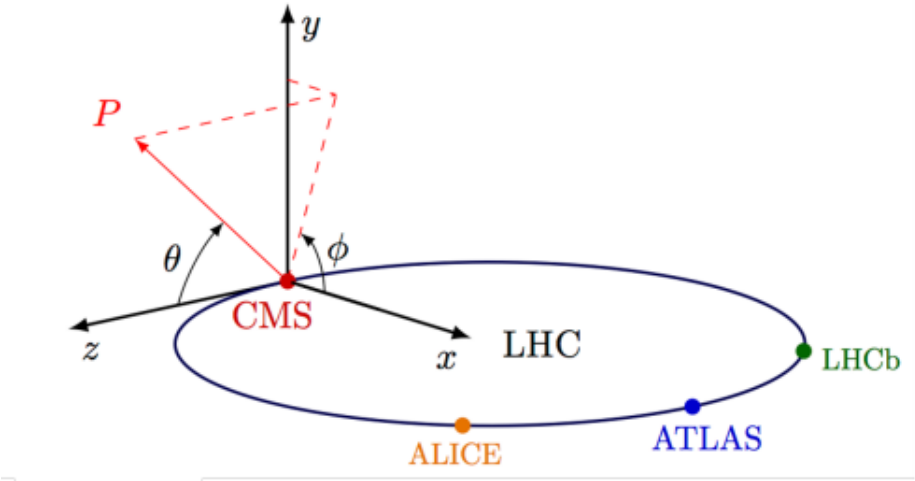


Figure 6: Il sistema di coordinate

Una variabile alternativa a θ è la pseudorapidità η:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$$

Questo implica che la differenza tra pseudirapidità tra due particele ultra-relativistiche è costante sotto boost lungo z. Altre grandezze invarianti per boost di Lorentz, sono:

- **La distanza nel piano eta-phi**

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$$

- **L'impulso trasverso** p_t ed il suo modulo

$$p_t = \sqrt{p_x^2 + p_y^2}$$

- **L'energia trasversa**

$$E_t = E \sin\theta$$

Dove p_x, p_y ed E sono, rispettivamente, le componenti x e y dell'impulso e l'energia della particella.

2.2.2 Il sistema di sottorivelatori

Data la vastità dei suoi scopi di fisica, CMS ha necessità di una strategia di ricostruzione ed identificazione delle particelle che garantisce versatilità ed accuratezza. A tale scopo è composto da un sistema di vari sottorivelatori, come si può vedere dalla figura 7:

- Il sitema di tracking interno
- Il calorimetro elettromagnetico (ECAL)
- Il calorimetro adronico (HCAL)
- Solenoide superconduttivo
- Il sistema a Muoni
- Il trigger

Ogni sottorivelatore è composto da un *barrel*, uno strato cilindrico coassiale al beam pipe, e due *endcap*: strati piani posti alle estremità del cilindro, al fine di garantire la rivelazione delle particelle che viaggiano vicine all'asse z.

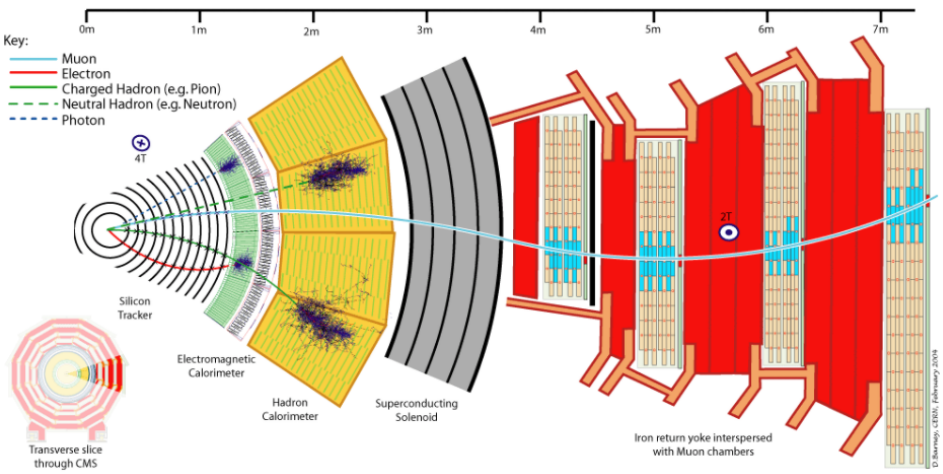


Figure 7: Le sezioni di CMS

Il tracker

Il tracker è il sottorivelatore più vicino al punto di interazione, al fine di garantire l'accuratezza maggiore nella ricostruzione dei vertici secondari e delle tracce di particele cariche. E' concepito per misurare la posizione di particelle cariche provenienti dalle collisioni ed ha un diametro di 2.5 m, con una lunghezza di 5.8 m ed un'area attiva di 200 m^2. Tale struttura fornisce un' acceptance in η, ovvero la regione dello spazio delle fasi che il rivelatore è in grado di osservare, corrispondente a $|\eta| < 2.4$. Il tracker è composto da un rivelatore a pixel più interno, fondamentale per la rivelazione di particelle a brevissima vita media, e da un rivelatore esterno a microstrip. Il rivelatore interno ha una precisione nella misura della posizione radiale e trasversa di 10 µm e 20 µm rispettivamente; dopo un upgrade nel 2017, ha raggiunto una struttura da 127 milioni di pixel distribuiti in quattro strati nel barrel e tre nell' endcap. [10] Per quanto riuarda il tracciatore più esterno a microstrip, invece, ha precisione sulla posizione radiale compresa tra 35 e 52 µm, mentre quella trasversa è di 530 µm ed è composto da quattro strati di rivelatori nel barrel e tre negli endcap.

Il calorimetro elettromagnetico

Il calorimertro elettromagnetico o ECAL, è un calorimetro ermetico omogeneo costituito da cristalli di tungstanato di piombo (PBWO_4), posto ad una distanza compresa tra 1.25m e 1.8m e con un'acceptance $|\eta| < 3$. Il suo scopo è misurare energia di elettroni e fotoni tramite scintillazione, con un raggio di Molière pari a 2.2 cm, la qual cosa gli consente di assorbire completamente gli sciami elettromagnetici e distinguerli, emettettendo l'85% della luce in 25 ns, proprio il tempo che intercorre tra la collisione di 2 bunch. I fotoni emessi per scintillazione sono raccolti da fotodiodi a valanga (APD) nel barrel e fototriodi a vuoto negli endcap, nei quali gli APD non sono utilizzabili a causa della radiazione troppo elevata in quell'area.

Il calorimetro adronico

Il calorimetro adronico o HCAL, è un calorimetro a campionamento, con lo scopo di misurare l'energia degli adroni prodotti dalle collisioni, all'interno del quale si alternano strati di materiale attivo (ottone) e materiale assorbente. L'ottone ha una lunghezza di radiazione X_0= 1.49 cm ed una lunghezza di interazione nucleare λ_I = 16.42 cm, che permettono al calorimetro di contenere lo sciame adronico entro le sue dimensioni. Il calorimetro è costituito da quattro sezioni:

- Il Barrel (HO) e l'endcap (HE) con acceptance $|\eta| < 3$
- Forward Section (HF), posta a 11.2 m dal punto di collisione lungo l'asse z, sfrutta fenomeni Cherenkov e ha un'acceptance $3 < \eta < 5.2$
- Outer Barrel section (HO), un sistema di ulteriori scintillatori posti al di fuori del solenoide, al fine di garantire un'adeguata profondità di campionamento e di misurare eventuali sviluppi tardivi dello sciame.

Un parametro che caratterizza entrambi i calorimetri è la *risoluzione in energia*:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2$$

dove a, b e c sono parametri che tengono conto, rispettivamente, di fluttuazioni statistiche nello sciame, rumori elettronici e pile-up ed errori sistematici.

Il sistema a muoni

Il sistema a muoni si trova fuori dal magnete ed è una componente essenziale di CMS, in quanto i muoni sono prodotti da una vasta gamma di processi e possono attraversare diversi strati di materiale senza arrestarsi, la qual cosa rende difficile la misurazione della loro energia. Lo scopo di questo rivelatore è di identificare i muoni e misurarne la posizione al fine di ricavare la tracia all'esterno del magnete. Per questo, il rivelatore di muoni presenta una superficie attiva molto estesa, di circa 25000 m^2. Per il suo scopo sono impiegati tre tipi di rivelatori a gas:

1. Camere a drift: situate nel barrel, coprono un range di pseudorapidità compreso tra 0 e 1.2. Sono poste in quattro stazioni, tre delle quali misurano le coordinate r - ϕ, mentre l'ultima misura la coordinata z.
2. Camere a strip: sono poste negli endcap e spaziano un range $0.9 < |\eta| < 3$. Il motivo per cui sono poste agli endcap è che hanno una risposta veloce ed una grande distanza di radiazione, tutte proprietà necessarie dove il tasso di muoni è più alto, ovvero agli endcap. La striscia catodica di ogni camera fornisce una misura di r - ϕ, mentre il filo anodico dà una misura della pseudorapidità.
3. Camere a piastre resistive: si trovano sia nel barrel che agli endcap ed hanno tempi di risposta molto brevi ed ottima risoluzione temporale, caratteristiche che li rendono ideali alla funzione di trigger.

Il magnete

Il magnete è un solenoide superconduttivo, capace di creare un campo magnetico di 3.8T lungo l'asse z - allo scopo di curvare la traiettoria delle particelle nel rivelatore. Attorno ad esso è piazzato un *iron return yoke* (lett. giugo metallico) che ha il compito di evitare effetti di bordo e di curvare le linee di campo per avere un campo magnetico quasi costante di 1.8T anche all'esterno della cavità del solenoide.

Trigger e acquisizione dati

Questo sistema serve a registrare ed immagazzinare i dati provenienti dalle collisioni. A causa dell'elevato numero di dati, tuttavia, non è possibile - allo stato attuale delle tecnologie - di salvarli tutti. Sorge quindi la necessità di selezionare i dati più significativi e a tale scopo è pensato il sistema di trigger. Esso è strutturato in due fasi: il Level-1 trigger (figura 8), che fornisce una selezione rapida e automatica basata sui depositi di energia nei calorimetri e tracce dei muoni, mentre la decisione finale sull'accettare o meno un evento spetta all'High Level Trigger (HLT), il quale esegue una selezione in base ad un software.

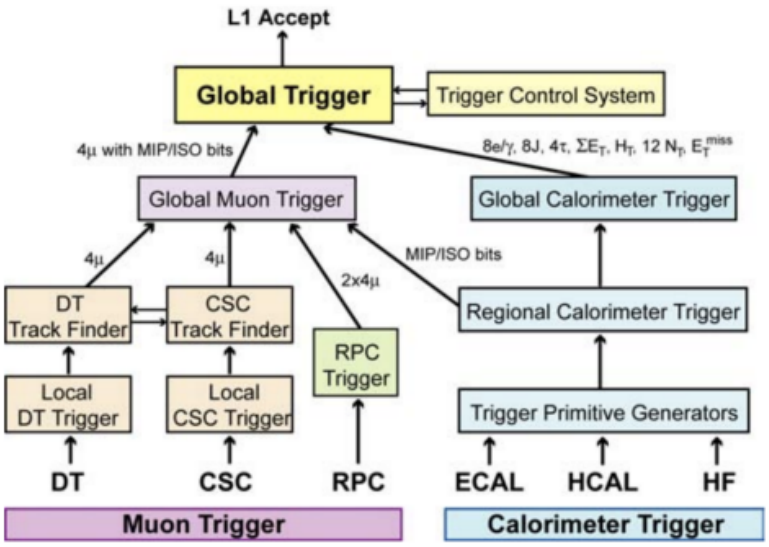


Figure 8: Architettura del trigger L1

3 Ricostruzione del Bosone di Higgs tramite Machine Learning

3.1 Il fenomeno in esame

Questo lavoro di tesi è incentrato sullo studio di un processo raro di formazione dell'Higgs di cui ancora non c'è ancora evidenza diretta: la produzione di Vector Like Quark T' attraverso interazione elettrodebole. Come spiegato nel capitolo 1, tale processo, illustrato in figura 9, è previsto da numerosi scenari di fisica oltre il Modello Standard. Questo stato finale ha un'ulteriore ragione di interesse: il corrispondente processo di produzione singola nel MS, in cui il bosone di Higgs emerge da un W interno o da un top esterno, con una sezione d'urto piccola a causa dell'interferenza distruttiva dei due processi. Il lavoro è stato svolto utilizzando un campione simulato di segnali T' che riproduce le condizioni della presa dati a 13 TeV, con masse di 700 GeV, 1000 Gev, 1200 Gev e 1800 GeV.

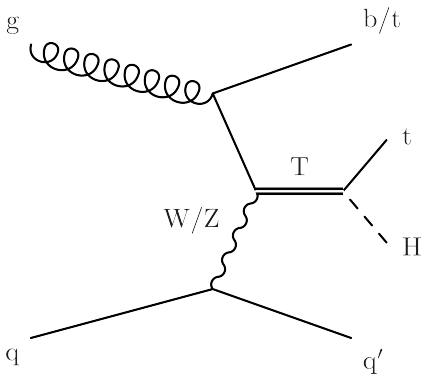


Figure 9: Il processo di produzione tHq BSM

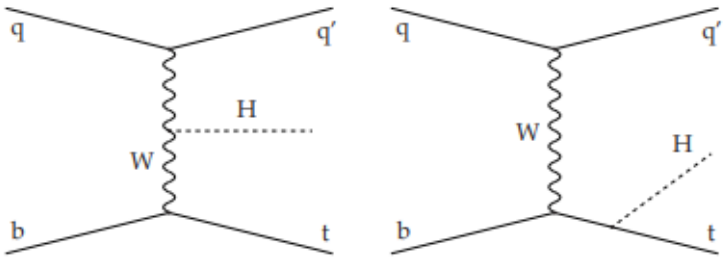


Figure 10: I due processi di produzione tHq nel MS

I principali stati finali dell'Higgs sono:

- $\gamma\gamma$
- $WW/$
- $ZZ/$
- $\tau\tau$
- $b\bar{b}$

In questo lavoro, ci concentriamo sul canale di decadimento $b\bar{b}$: esso, nell'ambito della ricerca tHq, è molto promettente grazie all'alto branching ratio. La principale difficoltà della ricerca in questo canale è dovuta alla presenza di un consistente fondo $t\bar{t}$, che riproduce in buona parte la segnatura sperimentale del segnale. A causa della difficoltà nell'individuare e ricostruire esattamente le catene di decadimento adroniche dei quark top, i corrispondenti prodotti di decadimento possono riprodurre la segnatura del bosone di Higgs.

Lo scopo di questo lavoro è migliorare la ricostruzione del bosone di Higgs tramite stati finali adronici alle energie caratteristiche dei processi di nuova fisica sfruttando le tecniche di ML. [11]

3.1.1 I jet

Quark e gluoni derivanti dalle collisioni non sono osservabili come particelle libere: propagandosi dal punto di collisione al rivelatore effettuano "adronizzanzione", creando *jet adronici*, ossia cluster di adroni non colorati. Dall' analisi dei jet si possono ricavare informazioni sul partone da cui sono originati. I jet sono ricostruiti usando l'algoritmo di clustering *anti-k_t* [12]; quelli con parametro di raggio R = 0.4 in metrica $\eta - \phi$ sono chiamati *narrow jets* (AK4), mentre quelli con parametro R = 0.8 sono detti *fat jet* (AK8)

Questo studio sfrutta il canale di decadimento dell'Higgs in una coppia quark/antiquark, in particolare $b\bar{b}$; non ci siamo concentrati sul decadimento del quark top, i cui jet tuttavia sono una possibile fonte di fondo combinatorio. Occorre quindi identificare i b-jet, ossia i jet originati da un quark b; questa operazione va sotto il nome di b-tagging [13] e viene fatta utilizzando l'algoritmo *DeepFlavour* di CMS, attraverso il quale viene generata una variabile continua, detta score, che viene usata nella selezione [14]. Algoritmi di ML sono applicati anche alla ricostruzione dei fat jet, producendo un discriminatore analogo capace di discriminare dai jet provenienti da processi di cromodinamica quantistica [11].

3.1.2 Ricostruzione standard

Il primo passo dell'analisi standard, ovvero quella usualmente eseguita per la ricostruzione dello stato finale in esame, è ricostruire l'Higgs a partire dai jet prodotti. L'approccio standard è detto *cut-based*, in cui si impongono criteri di selezione successiva (detti "tagli") atti a rimuovere i candidati di Higgs originati dalla combinazione di due jet provenienti da altre fonti, come ad esempio il decadimento di un quark top. Nel caso a minore energia, definito caso *resolved*, si hanno due jet AK4 distinti, provenienti ognuno dall'adronizzazione di un quark b. Essi sono combinati in un *dijet*, di cui si ricostruisce il quadrimomento come la somma vettoriale dei quadrimomenti dei jet e, di conseguenza, la massa invariante da usare nella selezione. Nel caso ad energia maggiore, detto caso *merged*, invece, i due quark b sono più vicini e, quindi, si ha un unico jet AK8 a cui applicare i tagli. Il lavoro presentato, invece, sfrutterà tecniche di machine learning al fine di eseguire un'analisi multivariata, basata su diverse caratteristiche dinamiche oltre alla massa invariante e all'informazione sul b-tagging.

3.2 Machine Learning

Il machine learning, o apprendimento automatico, è una branca dell'informatica il cui obiettivo è quello di rendere una macchina in grado di effettuare ragionamenti induttivi a partire dalla propria esperienza, ovvero dalle informazioni dategli in input, per ricavare un'informazione più generale sul fenomeno di cui le informazioni sono un'espressione. Dato un campione di dati che sottostà ad un modello o una legge ignota, l'algoritmo vuole ricavare informazioni su tale legge. L'algoritmo parte da un'ipotesi induttiva sul modello sottostante i dati, e tramite un procedimento iterativo adatta l'ipotesi per meglio stimare detto modello. Affinché la generalizzazione offra le migliori prestazioni possibili, la complessità dell'ipotesi induttiva deve essere pari alla complessità della funzione sottostante i dati: se l'ipotesi è meno complessa della funzione, allora il modello manifesta *underfitting*; al contrario, se l'ipotesi è troppo complessa, allora il modello manifesta *overfitting* e la generalizzazione sarà più scarsa. [15] Al di là delle prestazioni, va considerata anche la fattiblità dell'apprendimento stesso: una computazione è considerata fattibile se può essere svolta in tempo polinomiale. I compiti dell'apprendimento automatico possono essere divisi in tre diversi paradigmi:

- L'apprendimento supervisionato, in cui vengono forniti degli esempi di input e i rispettivi output al fine di dedurre una regola generale per associarli

- L'apprendimento non supervisionato, che differisce dal primo per la mancanza di categorie date all'output
- Apprendimento rinforzato, nel quale il modello interagisce con un ambiente dinamico allo scopo di raggiungere un obiettivo, avendo come aiuto solo l'informazione sul raggiungimento o meno di tale obiettivo.

Possiamo categorizzare i compiti del machine learning anche in funzione del suo output:

- Classificazione: abbiamo output divisi in due o più classi ed il modello deve assegnare nuovi input mai visti alla giusta categoria.
- Regressione: output e modello sono continui
- Clustering: un insieme di input viene diviso in gruppi, ma, a differenza della classificazione, tali gruppi non sono definiti a priori

3.2.1 Decision Tree

Un *Decision Tree* (DT) è un metodo di apprendimento supervisionato non parametrico, utilizzato per classificazioni e regressioni. Consiste in una sequenza di tagli sul dataset, applicati in un determinato ordine. Ogni taglio divide il dataset in nodi, che corrispondono ad un certo numero di campioni classificati come segnale o fondo; ogni nodo può essere ulteriormente diviso da altri tagli. Quando, in un nodo, ci sono troppi pochi campioni di una categoria per essere diviso, esso prende il nome di foglia. Affinchè un DT possa fare predizioni sui dati va allenato con un dataset - noto - di allenamento.

Formalmente, data una variabile di output y ed un vettore di variabili input x descritti da una distribuzione di probabilità P(x,y) e, utilizzando un set di allenamento $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, l'obiettivo è trovare un'approssimazione $\hat{F}(x)$ di una funzione F(x) che minimizzi il valore atteso di una specifica *loss function* $L(y, F(x))$:

$$\hat{F} = \arg\min_{F} \mathbb{E}_{x,y}[L(y, F(x))]$$

Un DT ha diversi vantaggi: è semplice da visualizzare, richiede poca preparazione dei dati, consente di valutare un modello utilizzando test statistici, ha un basso costo computazionale per la CPU - che va come n·log(N), con n variabili e N eventi di training ed utilizza un modello a "scatola bianca" (al contrario delle reti neurali che sono " a scatola nera"). Ciononostante, presenta un grosso svantaggio: è molto sensibile al rumore e all'overfitting.

3.2.2 XGBoost

Per superare i limiti di un decision tree, si ricorre ad una tecnica chiamata *boosting*, che nel nostro caso è incarnata dall'algoritmo XGBoost, acronimo di **eXtreme Gradient Boosting** [16]. Tale metodo consiste nell'assumere una y a valori reali e cercare un'approssimazione nella forma:

$$\hat{F}(x) = \sum_{i=1}^{M} \gamma_i h_i(x) + \text{const}$$

ossia una somma pesata, dove le funzioni h_i appartengono ad una classe H, chiamata *weak learners*. L'approssimazione sarà ricercata minimizzando ricorsivamente la loss function applicando uno step di discesa quanto più ripido possibile (discesa del gradiente); nel caso continuo, dove H è l'insieme delle funzioni differenziabili su R, si avrà:

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^{n} \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) - \gamma \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))\right)$$

in cui le derivate sono calcolate rispetto alle funzioni F_i, con $i \in \{1, .., m\}$, e γ_m è la lungezza dello step. Operativamente, questo si traduce nei seuenti step:

1. Inizializzazione del modello ad un valore costante (tipicamente 0,5):

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$$

2. Per m=1 fino ad M:

   (a) Calcolo degli pesudo-residui:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad \text{per } i = 1, \ldots, n$$

   (b) Fit di un weak learner $h_m(x)$ agli pseudo-residui, e.g allenamento con il set $(x_i, r_{im})_{i=1}^{n}$

   (c) Calcolo del moltiplicatore γ_m:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)\right)$$

(d) Aggiornamento del modello:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

3. Output di $F_M(x)$.

3.2.3 Variabili di allenamento

Il lavoro inizia implementando un classificatore binario, in particolare un Boosted Decision Tree (BDT) con la tecnica del gradient boosting sopra descritta. Lo scopo è di discriminare il segnale, in questo caso un bosone di Higgs, dal fondo proveniente da jet adronici ed altre fonti. Il tipo di fenomeno studiato è quello, precedentemente descritto, in cui l'Higgs decade in una coppia di jet. Prima di iniziare l'allenamento della BDT, si è effettuata una pre-selezione dei dati, scartando quelli direttamente riconducibili al fondo: sono stati eliminati tutti i jet il cui impulso trasverso (p_t) fosse minore di 30 GeV e, per la configurazione resolved, quelli la cui massa fosse minore di 80 e maggiore di 160 GeV. Per ottimizzare l'algoritmo, inoltre, il campione è stato diviso in tre fasce in base al p_t: la prima fascia con $0 < p_t < 100$, la seconda con $100 < p_t < 250$ e l'ultima con $\mathrm{p}_t > 250$. Questa divisione è stata effettuata al fine di evitare di allenare il modello in maniera troppo specifica al campione, in cui i bosoni di Higgs - povenendo da oggetti molto energetici - hanno impulso molto più elevato ripetto a quelli ricostruiti nel fondo.

Il set di variabili x per l'allenamento è costituito, al primo step, da massa invariante, impulso trasverso e coordinate angolari ϕ ed η del candidato Higgs. Successivamente sono state aggiunte anche le caratteristiche cinematiche dei singoli jet AK4 assieme ad i rispettivi score di b-tagging.

3.2.4 Allenamento in canali resolved e merged e risultati

Completato il processo di allenamento, si ottiene un set di dati output in cui ad ogni evento è stato assegnato un valore tra 0 e 1. L'allenamento è stato eseguito in tutte le configurazioni discusse prima e di seguito riportiamo gli output, in cui in blu si ha il fondo ed in rosso il segnale.

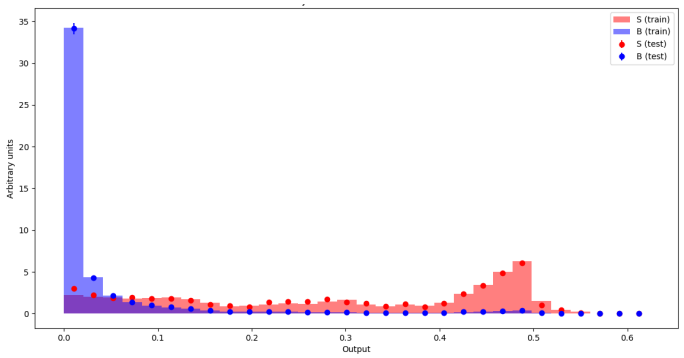


Figure 11: Output della BDT in configurazione resolved

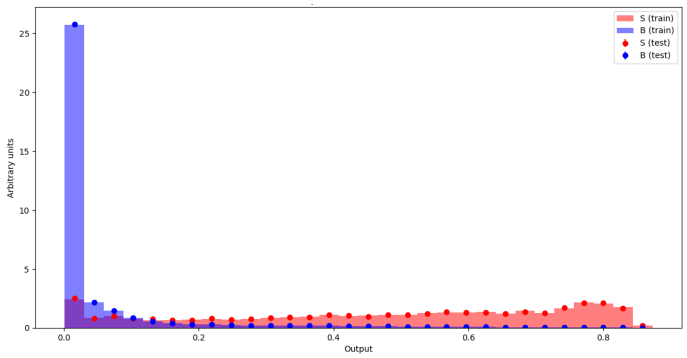


Figure 12: Output della BDT in configurazione merged

La discriminazione è buona e non è presente overtraining, la qual cosa è conferamata dall'andamento della logloss:

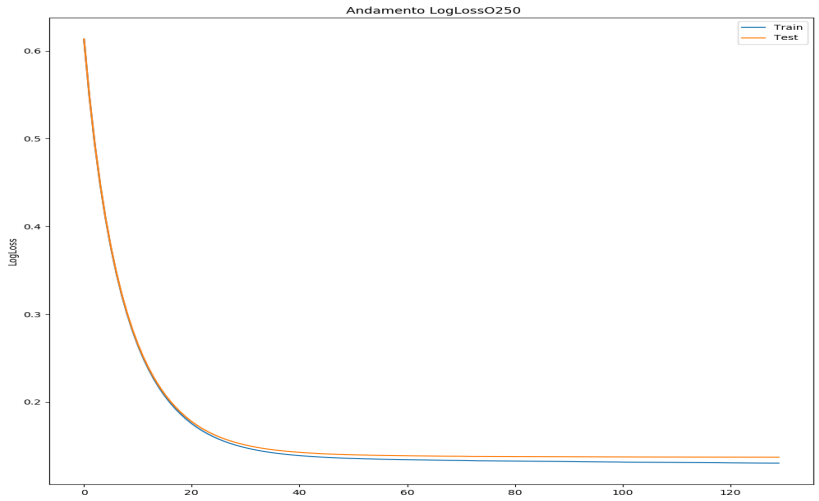


Figure 13: Logloss in configurazione resolved

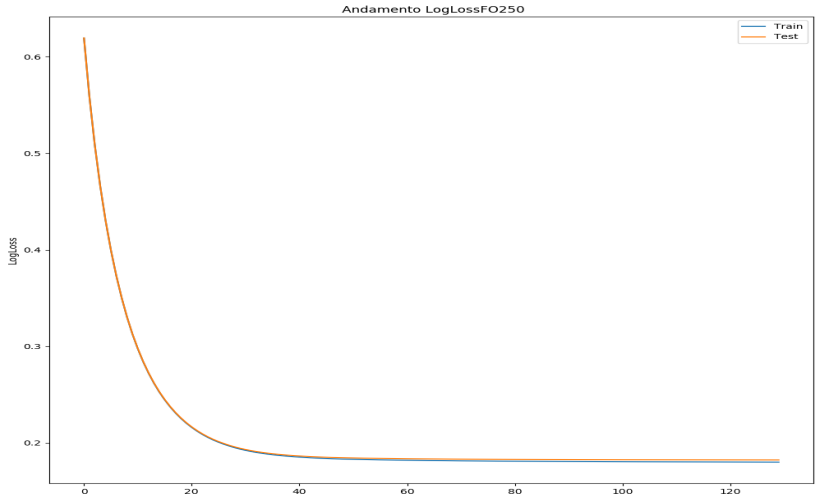


Figure 14: Andamento della logloss in configurazione merged

Avendo utilizzato una BDT, è possibile vedere quali delle variabili sono più significative per la discriminazione tramite un parametro chiamato *F-Score*:

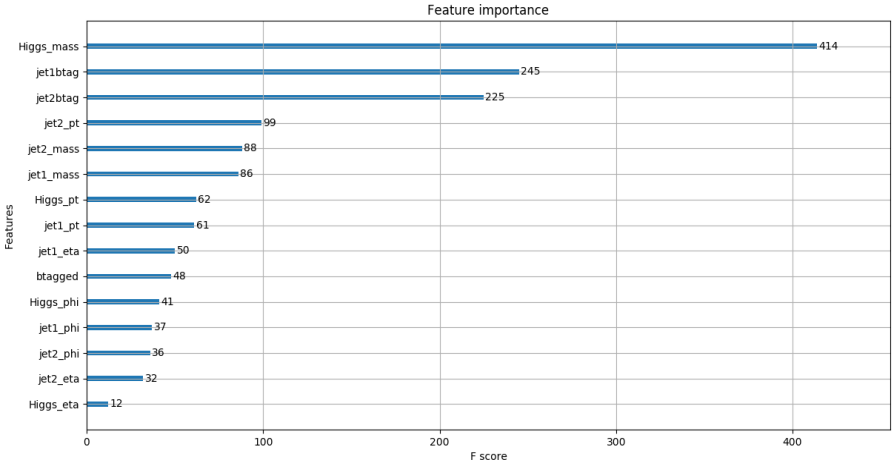


Figure 15: F-score in configurazione resolved

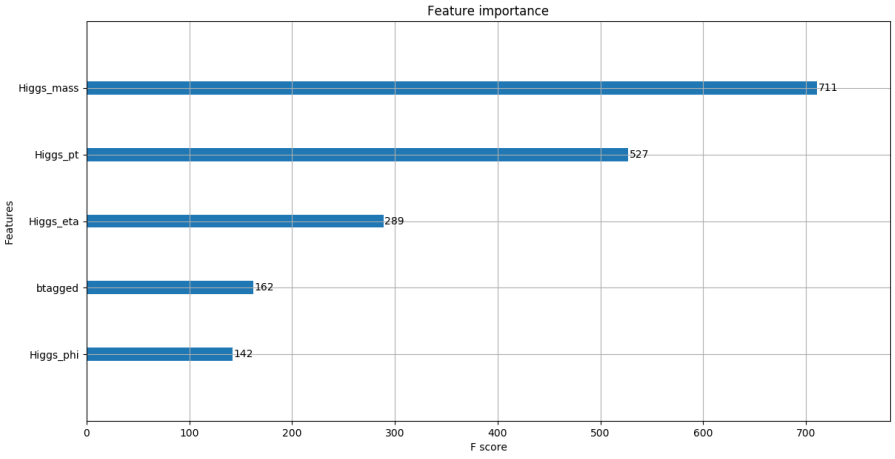


Figure 16: F-score in figurazione merged

Uno strumento utile a quantificare l'efficienza è la *Receiver Operating Characteristic* o ROC, in cui sull'asse x viene riportato il *false posivie rate* (FPR), mentre sull'asse y il *true positive rate* (TPR). Queste quantità rappresentano, rispettivamente, la frazione di fondo classificata come segnale e la frazione di segnale correttamente classificata.

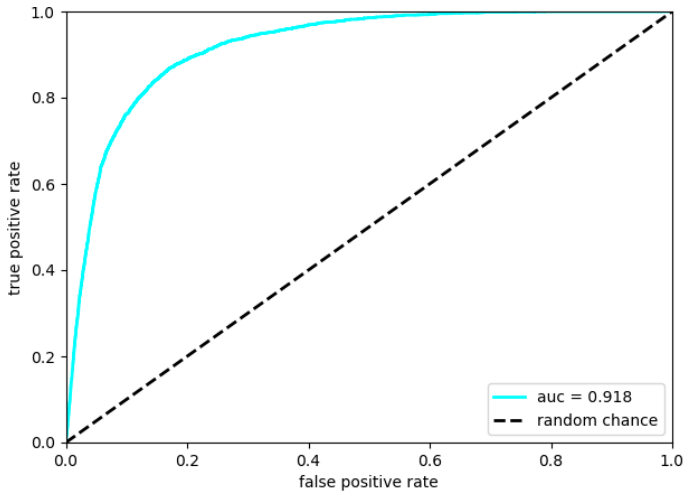
$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

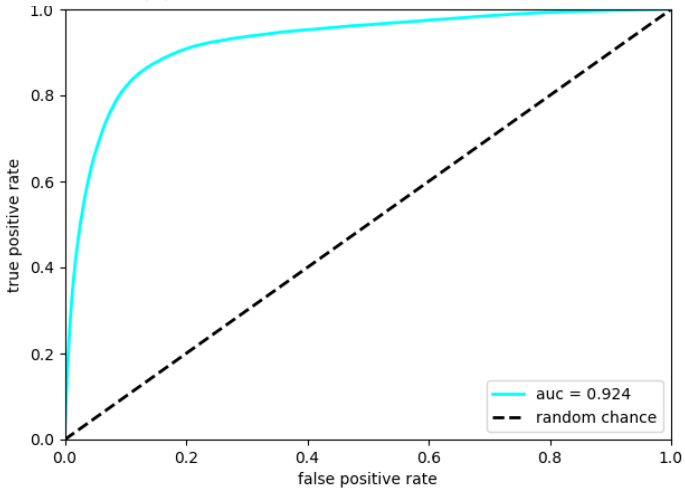
True Positive Rate e False Positive Rate sono anche definite come le *efficienze* nel contesto della selezione degli eventi usando l'output del discriminatore. In questo sistema, una retta con coefficiente angolare 1 rappresenta uguali efficienze di segnale e di fondo, mentre un gradino rappresenta un taglio perfetto. L'efficienza di taglio è data dall'area sottesa alla curva ROC, detta *Area Under the Curve* (AUC). Tale quantità è un criterio di qualità del discriminatore, una maggiore AUC in genere indica un discriminatore globalmente con migliori proprietà. Per scopi di selezione tipicamente sono identificati dei valori di benchmark, detti punti di lavoro o "*Working Points*" (WP), da essere usati poi per discriminare i segnali dal fondo in una selezione successiva. Nel nostro caso abbiamo considerato i WP corrispondenti ad un FPR del 10 e 1%, riportando i corrispondenti TPR in tabella 4

Table 4: Performance della BDT nelle diverse configurazioni

Regime e massa del campione ([GeV])	P_t bin [GeV]	AUC	WP 10%	WP 1%
Resolved 700	$p_t < 100$	0,92	0,75	0,25
	$100 < p_t < 250$	0,91	0,79	0,29
	$p_t > 250$	0,93	0,75	0,14
Resolved 1200	$p_t < 100$	0.91	0.77	0.25
	$100 < p_t < 250$	0.93	0.81	0.33
	$p_t > 250$	0.92	0.77	0.22
Merged 1200	$p_t < 100$	0,81	0,44	0,06
	$100 < p_t < 250$	0,88	0,65	0,16
	$p_t > 250$	0,87	0,56	0,12
Merged 1800	$p_t < 100$	0,85	0,56	0,08
	$100 < p_t < 250$	0,93	0,82	0,35
	$p_t > 250$	0,92	0,82	0,29



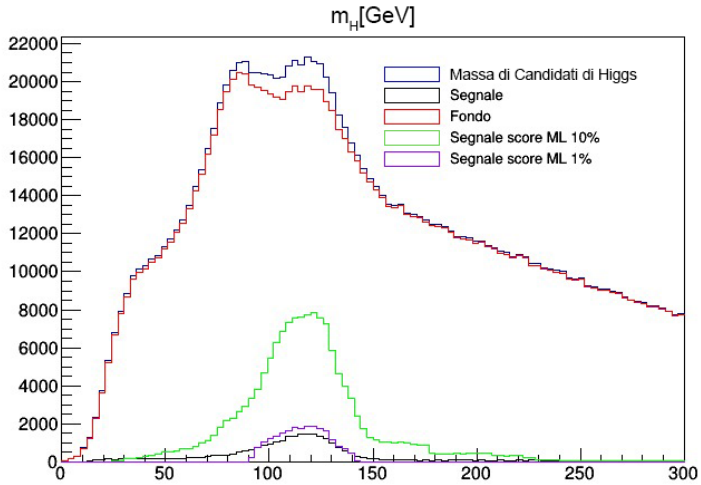
(a) ROC in configurazione resolved



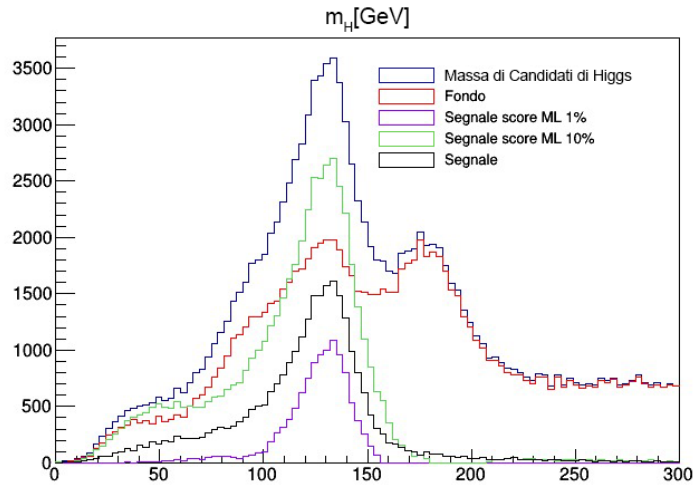
(b) ROC in configurazione merged

3.2.5 Ricostruzione della massa invariante

E' possibile ricostruire il bosone di Higgs prendendo i punteggi della BDT corrispondenti alle efficienze di fondo al 10% ed all'1% di tabella 4. La massa dell'Higgs ricostruita per massa del T' 700 GeV in configurazione resolved è mostrata in Figura 18a, per masssa del T' 1800 GeV in configurazione merged è mostrata in Figura 18b



(a) Configurazione resolved



(b) Configurazione merged

Figure 18: Massa Higgs

Abbiamo osservato, inoltre, la particolare conformazione del campione dati. Prendendo, ad esempio, l'output della BDT in configurazione merged, notiamo che questo - al crescere della massa del campione - migliora le sue performance, come evidente nella figura 19

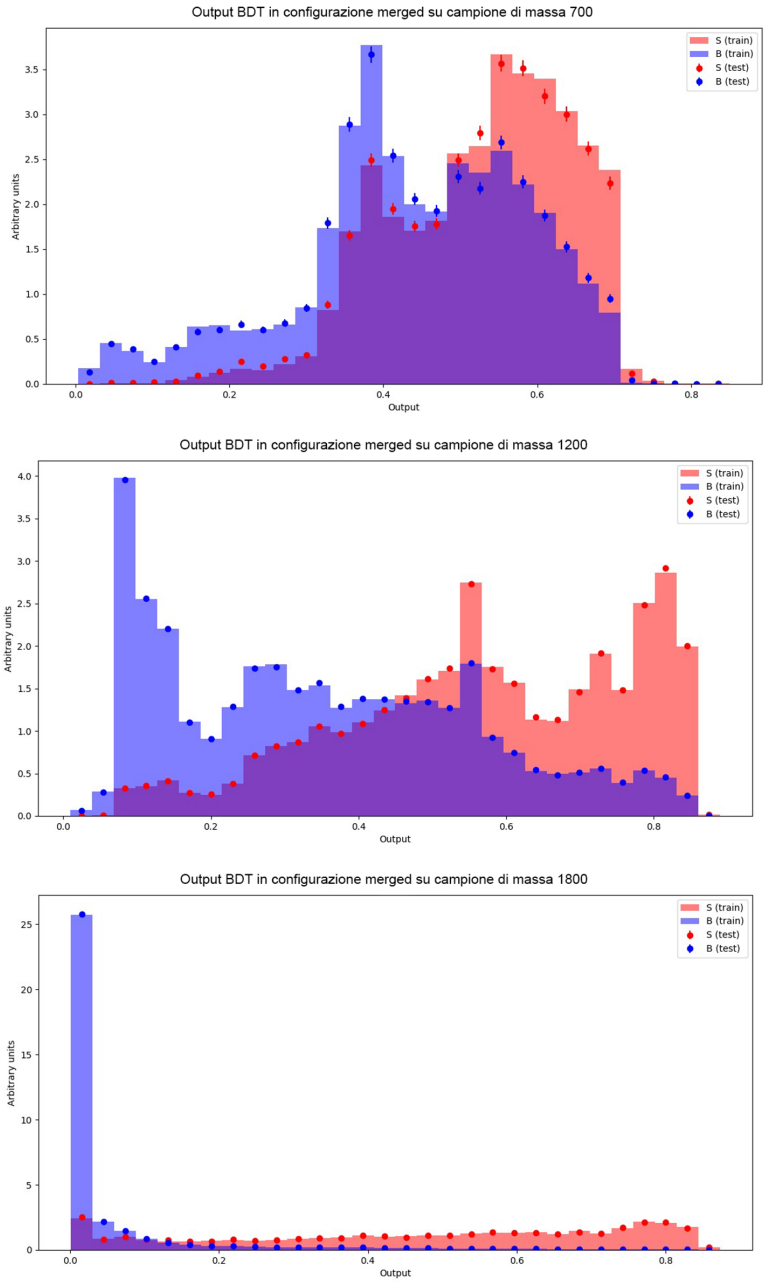


Figure 19: Output della BDT in configurazione merged sui campioni a massa 700 GeV, 1200 GeV e 1800 GeV

3.2.6 Multi-classificazione

Dopo aver analizzato i risultati di questo modello, si è proceduto a studiare un'eventuale sovrapposizione delle due configurazioni, il cosidetto "caso mixed", ossia quelo in cui il jet AK8 include i due jet AK4. A tale scopo abbiamo implementato un multi-classificatore che opera sulle seguenti classi:

1. Fondo
2. Higgs che decade in un jet AK8
3. Higgs che decade in due jet AK4
4. Higgs che decade in due jet AK4 contenuti in un jet AK8

Per valutare le performance del multi-classificatore utilizziamo un'indice detto *recall*, che rappresenta la frazione di veri positivi classificata correttamente:

$$\frac{TP_i}{TPi + FNi}$$

dove $\mathrm{TP_i} + \mathrm{FN_i}$ sono gli eventi di sengnale scartati (falsi negativi) nell'ipotesi i-esima. Tale quantità corrisponde alla nostra definizione di efficienza per la categoria i-esima.

Dai risultati, riportati nella tabella sottostante, sono consistenti con quanto visto con la BDT e, inoltre, si nota un'ottima efficienza per il segnale di tipo mixed.

classe	$100 < p_t < 250$	$p_t > 250$
fondo	0,59	0,65
merged	0,50	0,64
resolved	0,71	0,56
mixed	0,91	1
media	0,64	0,65

Table 5: Performance del multiclassificatre con massa 1000

Riportiamo infine le performance del multi-classificatore rappresentate come *confusion matrix*, dove è indicato il numero di eventi classificati nelle diverse categorie. Si nota la struttura perlopiù diagonale, segno di una buona discriminazione; in particolare nella categoria mixed si ha un campione molto più puro rispetto al fondo. Siccome mixed, resolved e merged sono ortogonali, gli elementi di matrice corrispondenti a tale confusione sono zero per costruzione.

classe	$100 < p_t < 250$	$p_t > 250$
fondo	0,64	0,69
merged	0,50	0,48
resolved	0,57	0,565
mixed	1	1
media	0,63	0,64

Table 6: Performance del multiclassificatre con massa 1200

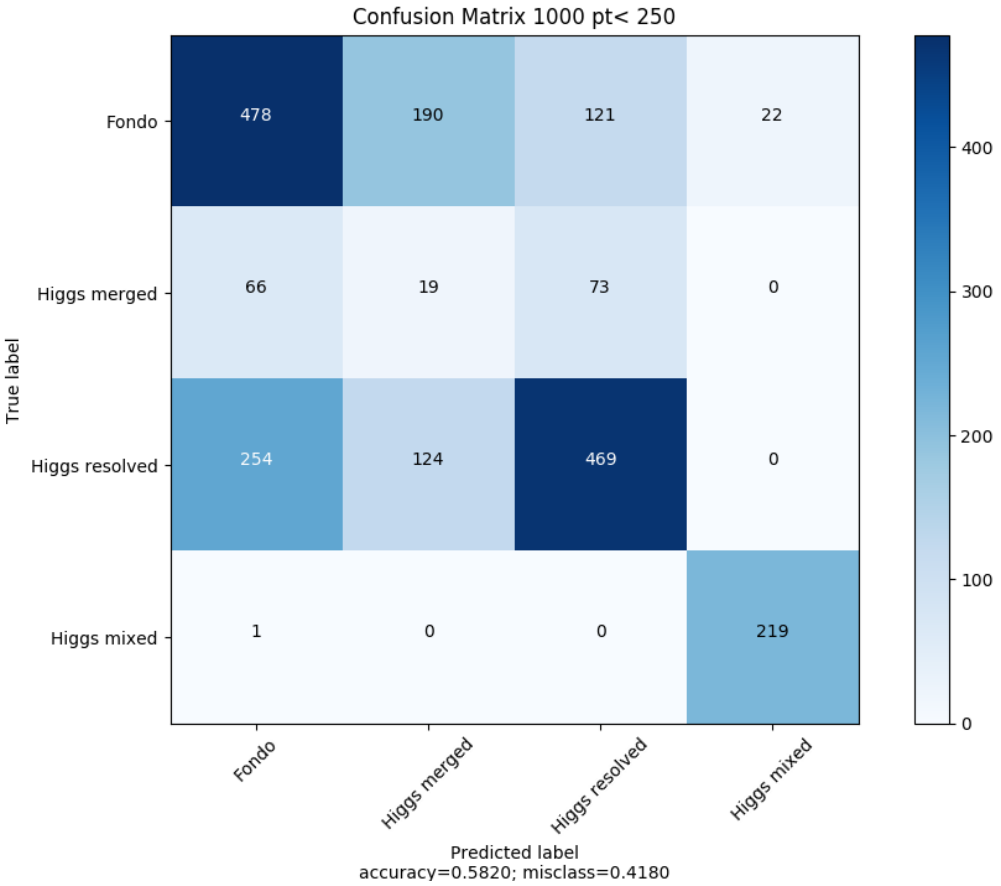


Figure 20: Confusion matrix per massa 1000 GeV e p_t under 250 GeV

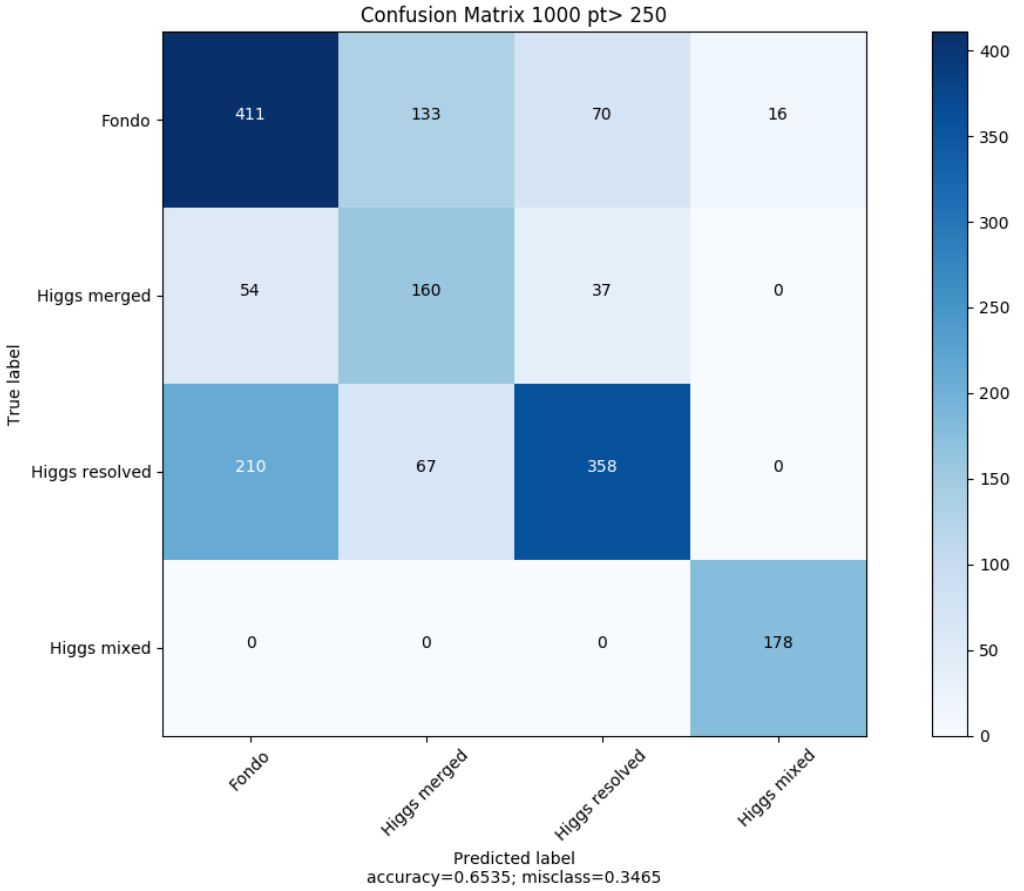


Figure 21: Confusion matrix per massa 1000 GeV e p_t over 250 GeV

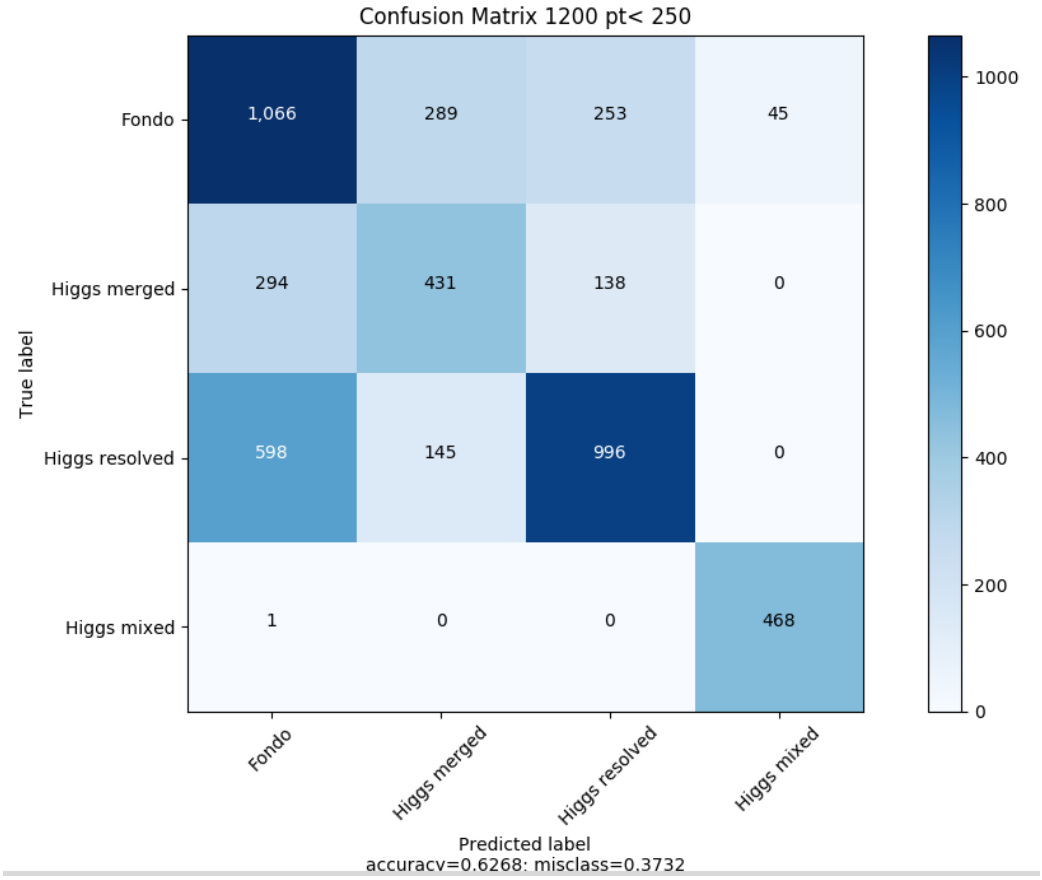


Figure 22: Confusion matrix per massa 1200 GeV e p_t under 250 GeV

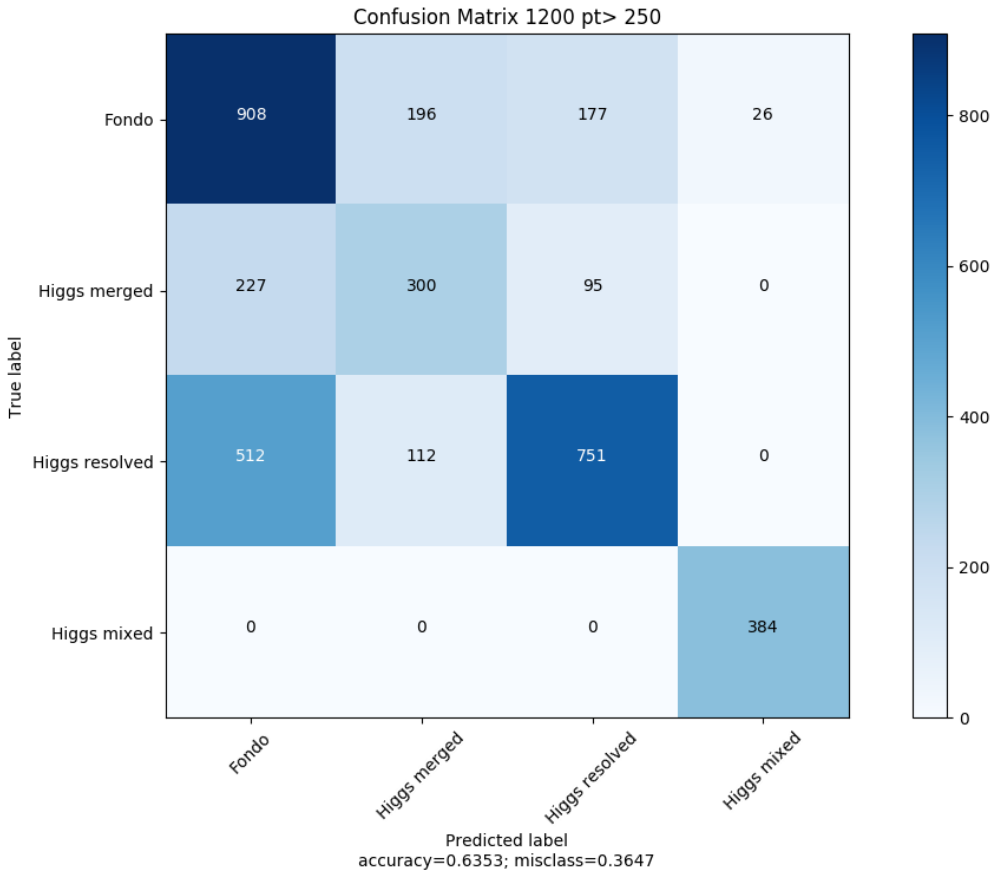


Figure 23: Confusion matrix per massa 1200 GeV e p_t over 250 GeV

4 Conclusioni

In questo lavoro di tesi è presentata resentata una tecnica di ricostruzione di bosoni di Higgs prodotti in processi di nuova fisica in collisioni protone-protone ad LHC, effettuata a partire dagli stati finali costituiti da una coppia quark/antiquark b attraverso l'uso di tecniche di Machine Learning.

La teoria che, ad oggi, descrive al meglio le particelle elementari e le interazioni tra esse è il Modello Standard. Essa ha ricevuto numerose conferme sperimentali negli ultimi anni, soprattutto grazie all'esperimento LHC di cui ricordiamo, tra l'altro, la scoperta del Bosone di Higgs nel 2012. La teoria del Modello Standard non è però esaustiva di tutti i fenomeni in natura e di conseguenza sono stati proposti diversi modelli oltre il MS, le teorie Beyond Standard Model. Una previsione comune a molteplici ipotesi BSM consiste in nuove particelle chiamate Vector Like Quark (VLQ). L'esperimento Compact Muon Solenoid è uno dei quattro esperimenti a Large Hadron Collider che grazie all'energia del centro di massa raggiunta potrebbe essere in grado di provare l'esistenza dei VLQ.

Il presente lavoro si è incentrato sullo studio di un processo di produzione di un singolo VLQ T' in collisioni protone protone ad LHC all'energia nel centro di masssa di 13 TeV, considerando stati finali in cui il bosone di Higgs decade in una coppia quark/antiquark b che produrrà jet adronici. Tali campioni presentano un interessante caso di studio in quanto il decadimento in questione si manifesta in due diversi stati finali, la configurazione *merged* e quella *resolved*, la cui abbondanza relativa è legata alla massa della particella T',

Per poter ricostruire il bosone di Higgs è stato fatto uso dell'algoritmo di machine learning *eXtreme Gradient Boosting*, xGBoost, grazie al quale è stato implementato un Boosted Decision Tree con lo scopo di discriminare il fondo dal segnale, analizzando quali delle variabili cinematiche dei quark b fossero importanti al fine della discriminazione. I risultati ottenuti col machine learning sono stati comparati con l'analisi standard, rilevando un miglioramento nell'efficienza di selezione.

E' stato poi fatto un ulteriore passo in avanti implementando un multi-classificatore al fine di distinguere sia il fondo dal segnale, sia di distinguere il segnale delle diverse tipologie di jet, ipotizzando un terzo stato finale oltre ai due preventivati: il caso *mixed*, in cui in realtà le due configurazioni precedenti coesistono.

L'efficienza nel discriminare segnale dal fondo è particolarmente importante nel fenomeno in esame, in quanto esso, pur essendo il favorito tra i processi di decadimento dell'Higgs, soffre anche di un - in genere - maggiore contributo di fondo proveniente principalmente da processi di cromodinamica

quantistica. Possibili passi successivi sono l'applicazione del multiclassificatore in una selezione per la ricostruzione del T' e nell'analisi vera e propria, nonchè il potenziale confronto con altri algoritmi.

In vista dell'upgrade di LHC e della costruzione di nuovi acceleratori in grado di raggiungere energie e, soprattutto, luminosità più elevate, lo sviluppo di applicazioni di tecniche di ML specifiche per tali condizioni sperimentali si rivelerà un importante strumento per la ricerca degli anni a venire

References

[1] I. J. R. Aitchison and A. J. G. Hey. Gauge Theories in Particle Physics Vol I: From Relativistic Quantum Mechanics to QED. *CRC Press/Taylor and Francis*, 2013 (4th ed).

[2] I. J. R. Aitchison and A. J. G. Hey. Gauge Theories in Particle Physics Vol II: QCD and the Electroweak Theory. *CRC Press/Taylor and Francis*, 2013 (4th ed).

[3] Paul Langacker. Introduction to the Standard Model and Electroweak Physics. *arXiv preprint arXiv:0901.0241*, 2009.

[4] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[5] P. W. HIGGS. Broken symmetries and the masses of gauge bosons. page 508, 1964.

[6] F. ENGLERT and R. BROUT. Broken symmetry and the mass of gauge vector mesons. page 321, 1964.

[7] et al. Aguilar-Saavedra, J. A. Handbook of vectorlike quarks: Mixing and single production. *Phys. Rev. D, 88:094010*, 2013.

[8] L. Corpe D. Huang P. Sun A. Buckley, J. M. Butterworth. New sensitivity of current LHC measurements to vector-like quarks. 2020.

[9] CMS: CMS Collaboration. The CMS experiment at the CERN LHC. *JINST3(2008) S08004*, 2008.

[10] The Tracker Group of the CMS Collaboration. The CMS Phase-1 Pixel Detector Upgrade. 2020.

[11] CMS Collaboration. Identification of heavy, energetic, hadronically decaying particlesusing machine-learning techniques. *JINST15(2020) P06005*, 2008.

[12] Gregory Soyez Matteo Cacciari, Gavin P. Salam. The anti-$k_t jetclusteringalgorithm. JournalofHighEnergyPhysics$, 2008.

[13] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.

[14] Mauro Verzetti. Machine learning techniques for jet flavour identification at CMS. *EPJ Web Conf., 214:06010*, 2019.

[15] Ethem Alpaydin. Introduction to Machine Learning. *The MIT Press*, 2010.

[16] Carlos Guestrin Tianqi Chen. XGBoost: A Scalable Tree Boosting System. 2016.