

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea Triennale in Fisica

Tesi di laurea in Fisica delle Alte Energie

**Tecniche di apprendimento automatico per
la ricerca di eventi top-top con
l'esperimento ATLAS a LHC**

Relatore:
Prof. Maximiliano Sioli

Presentata da:
Virginia Mazza

Correlatore:
Dott.ssa Noemi Cavalli

Sommario

Il Modello Standard è attualmente la teoria che meglio spiega il comportamento della fisica subnucleare, includendo la definizione delle particelle e di tre delle quattro forze fondamentali presenti in natura; risulta però una teoria incompleta sulle cui integrazioni i fisici stanno lavorando in diverse direzioni: uno degli approcci più promettenti nella ricerca di nuova fisica risulta essere quello delle teorie di campo efficaci. Il vertice di interazione del processo di produzione di coppie di quark top dello stesso segno a partire da protoni è fortemente soppresso nel Modello Standard e deve quindi essere interpretato con le teorie di campo efficaci. La catena di decadimenti studiata è la seguente:

$$pp \rightarrow tt \rightarrow W^+bW^+b \rightarrow l^+\nu bl^+\nu b$$

Il presente elaborato si concentra su questo nuovo approccio per la ricerca di quark top *same-sign* e si focalizza sull'utilizzo di una rete neurale per discriminare il segnale dal fondo. L'obiettivo è capire se le prestazioni di quest'ultima cambino quando le vengono fornite in ingresso variabili di diversi livelli di ricostruzione.

Utilizzando una rete neurale ottimizzata per la discriminazione del segnale dal fondo, le si sono presentati tre set di variabili per l'allenamento: uno di alto livello, il secondo strettamente di basso livello e il terzo copia del secondo con aggiunta delle due variabili principali di *b*-tagging. Si è dimostrato che la performance della rete in termini di classificazione segnale-fondo rimane pressoché inalterata sia che si forniscano variabili precostruite sia che si diano variabili di livello inferiore. La curva ROC, un sensibile indicatore delle prestazioni di una rete neurale, presenta aree sottostanti le curve pressoché identiche.

Si è notato inoltre che nel caso del set di variabili di basso livello, la rete neurale classifica come input più importanti gli angoli azimutali dei leptoni nonostante questi abbiano distribuzioni identiche tra segnale e fondo: questo avviene in quanto la rete neurale è in grado di sfruttare le correlazioni tra le variabili come caratteristica discriminante. Questo studio preliminare pone le basi per l'ottimizzazione di un approccio multivariato nella ricerca di eventi con due top dello stesso segno prodotti a LHC.

Indice

Introduzione	1
1 Particelle, Modello Standard e nuovi orizzonti	3
1.1 Modello standard	3
1.1.1 Interazioni fondamentali	5
1.1.2 Fermioni	6
1.1.3 Bosoni	7
1.1.4 Limiti del Modello Standard	7
1.2 Teorie di Campo Efficace (EFT)	8
1.2.1 Un esempio di applicazione delle EFT: il decadimento β	10
1.2.2 La SMEFT	11
1.2.3 Produzione di coppie di quark top same sign	13
2 L'esperimento ATLAS	15
2.1 Il CERN	15
2.1.1 Catena di acceleratori	15
2.2 Il Large Hadron Collider (LHC)	17
2.2.1 Esperimenti principali: ATLAS, CMS, LHCb, ALICE	18
2.2.2 Altri esperimenti	18
2.3 Esperimento ATLAS	19
2.3.1 Struttura di un rivelatore	19
2.3.2 Struttura del rivelatore di ATLAS	20
2.3.3 Trigger e acquisizione dati in ATLAS	22
3 Ricerca di eventi top-top mediante tecniche di apprendimento automatico	25
3.1 Reti neurali	26
3.1.1 Struttura di una NN	26
3.1.2 Funzionamento di una NN	27
3.1.3 Applicazione di una NN alla fisica delle particelle	29
3.2 Analisi dei dati e risultati	32
3.2.1 Ricerca del segnale di <i>quark top same-sign</i> con ATLAS	32

3.2.2	Scelta dei set up di analisi	34
3.2.3	Confronti tra i vari livelli	35
	Conclusioni	50
	Bibliografia	51
	Elenco delle figure	53

Introduzione

Il Modello Standard (MS) è la teoria attualmente utilizzata per descrivere la fisica subnucleare: descrive le particelle introducendo il concetto di funzione d'onda associata ad ognuna di esse e spiega tre delle quattro interazioni fondamentali, precisamente quella elettromagnetica, forte e debole lasciando ancora incompleto il settore della gravità. Il Modello Standard suddivide le particelle in due gruppi, i fermioni e i bosoni, in base a antisimmetria o simmetria della funzione d'onda della particella rispetto all'operatore di scambio e introduce numeri quantici per identificarle, talvolta esistenti in spazi interni con dimensione opportuna. Per quanto riguarda le interazioni, vengono descritte attraverso teorie di campo quantizzato che prevedono la simmetria della lagrangiana del sistema rispetto a trasformazioni di gauge locali. Sono diverse le questioni aperte lasciate dal Modello Standard, per questo sono in corso ricerche che possano completarlo.

Un interessante approccio risulta quello proposto dalle Teorie di Campo Efficaci (EFT), un tipo di approssimazione che si applica a modelli esistenti: nel caso del Modello Standard, viene introdotta nella lagrangiana del sistema una correzione data da queste nuove teorie che consente nuovi processi e produce una piccola variazione in quelli già noti a determinate scale di energia (precisamente a valori tali in cui le EFT non possono più essere trascurate). Nel presente elaborato si analizza la produzione di quark top dello stesso segno a partire da coppie di protoni, processo soppresso nel Modello Standard ma che assume maggiore importanza in ambito EFT.

L'obiettivo dell'analisi è verificare le prestazioni di una rete neurale ottimizzata per la discriminazione del segnale dal fondo quando le vengono presentate variabili di diverso livello di ricostruzione. I dati utilizzati provengono da ATLAS, uno dei quattro principali rivelatori del Large Hadron Collider, l'acceleratore di particelle più potente e più grande al mondo situato nel sito di ricerca del CERN di Ginevra. Una rete neurale è composta da vari livelli ognuno costituito da più neuroni artificiali, imitando la struttura del cervello umano: input, livelli nascosti e output. La rete riceve in ingresso le variabili sotto forma di segnali, le elabora assegnando ad ogni neurone un peso in modo che riceva un'importanza diversa e infine calcola il valore di uscita del neurone: il risultato elaborato per ogni ingresso è modellato da un algoritmo, che viene migliorato ad ogni iterazione in modo che la rete fornisca

risultati sempre più precisi.

Nel caso in esame sono stati forniti alla rete neurale tre set di variabili: il primo di alto livello, il secondo di basso livello e il terzo copia del secondo con aggiunte le due principali variabili di *b*-tagging. I risultati ottenuti sono stati analizzati con lo scopo di capire se le prestazioni della rete variassero al variare del livello delle variabili proposte in ingresso.

Capitolo 1

Particelle, Modello Standard e nuovi orizzonti

1.1 Modello standard

Scoperte teoriche e sperimentali di centinaia di fisici a partire dagli anni '30 costituiscono un incredibile bacino di conoscenza per quanto riguarda la struttura fondamentale della materia: tutto nell'universo sembra essere composto di pochi elementi chiave, chiamati particelle fondamentali, governati da quattro forze fondamentali. La massima comprensione di come queste particelle si comportino tra loro e nei confronti di tre delle interazioni fondamentali è racchiusa nel Modello Standard (MS), fino ad oggi la migliore conciliazione tra relatività ristretta e meccanica quantistica.

Sviluppato nel XX secolo, il Modello Standard ha spiegato con successo molti risultati sperimentali e predetto precisamente un'ampia varietà di fenomeni utilizzando teorie di campo quantizzato (*Quantum Field Theories, QFT*), che prevedono l'esistenza di campi di materia e di campi associati a tre delle quattro interazioni fondamentali (elettromagnetica, debole e forte): le particelle vengono interpretate come i quanti dei rispettivi campi materiali; precisamente l'esistenza di una determinata particella implica che il campo corrispondente abbia cambiato una delle proprie variabili dinamiche. Data questa nuova natura delle particelle subatomiche, si è pensato di descriverle tramite *funzioni d'onda* la cui simmetria o antisimmetria le suddivide in due grandi gruppi: i fermioni (particelle di materia) e i bosoni (quanti scambiati tra particelle nelle interazioni) [Figura 1.1].

In fisica subnucleare le particelle sono identificate da nuove variabili che vivono in *spazi interni* di dimensioni differenti e a cui sono associati particolari numeri quantici: il *sapore* che determina la differenza nominale tra le particelle fondamentali, il *colore* collegato all'interazione forte, lo *spin* o momento angolare intrinseco, e l'*isospin* che consente di categorizzare determinate particelle come diversi stati

quantici di una stessa particella. Grazie a questi numeri quantici è possibile definire le *antiparticelle* come particelle subatomiche aventi stessa massa ma proprietà elettriche e magnetiche (quindi numeri quantici interni) opposte rispetto alla corrispondente particella. E' importante sottolineare che le particelle veramente neutre, ossia quelle che possiedono cariche interne tutte nulle come i fotoni e i neutrini, non presentano corrispondenti antiparticelle perchè la loro funzione d'onda è *autoaggiunta* cioè uguale a quella dell'antiparticella. Tutte le altre invece possiedono almeno una carica interna diversa da zero e di conseguenza la loro funzione d'onda risulta diversa da quella della rispettiva antiparticella.

Il Modello Standard attualmente costituisce la migliore teoria per descrivere il mondo subatomico ma presenta diverse criticità e lascia ancora molte domande senza risposta. I fisici sono al lavoro per proporre soluzioni a questi punti di discontinuità e sicuramente le nuove scoperte avranno una portata notevole come è già accaduto in passato [1].

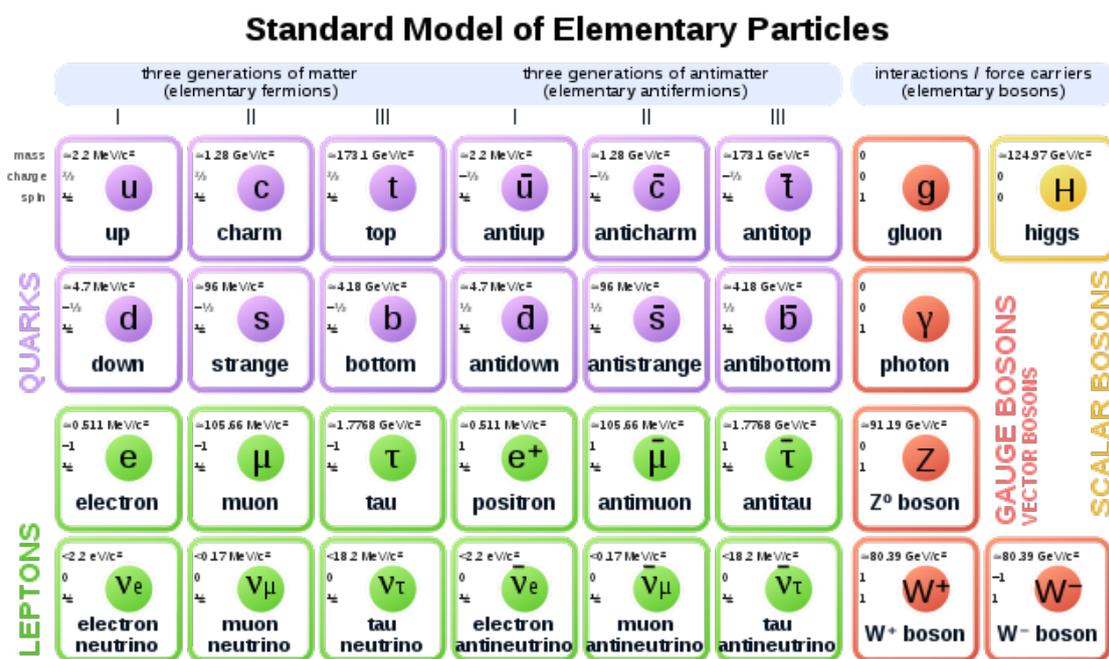


Figura 1.1: Suddivisione delle particelle e antiparticelle del Modello Standard. In viola quark e antiquark, in verde leptoni e antileptoni, in rosso mediatori di gauge delle interazioni e in giallo bosone di Higgs, unico mediatore scalare fino ad oggi conosciuto.

1.1.1 Interazioni fondamentali

Nell'universo ci sono quattro diverse forze fondamentali conosciute: l'interazione forte, l'interazione debole, l'interazione elettromagnetica e l'interazione gravitazionale. All'attuale scala di energia ognuna di queste ha una diversa intensità (in ordine dalla maggiore alla minore si hanno forte, elettromagnetica, debole e gravitazionale) e per riuscire ad unificarle si dovrebbe trovare un valore di energia al quale tutte e quattro si presentino circa con la stessa intensità. Ad oggi i livelli di energia raggiunti hanno consentito solo una parziale unificazione: la *QED* (*Quantum ElectroDynamics*) spiega l'interazione elettromagnetica, la teoria elettrodebole lega interazione elettromagnetica e debole mentre la *QCD* (*Quantum CromoDynamics*), costruita per analogia con la prima, include solo quella forte. Resta irrisolta la questione della gravità, la più debole tra le quattro tanto da diventare trascurabile a livello subatomico.

Il Modello Standard si fonda sulle teorie di campo quantizzato per l'interazione forte, debole ed elettromagnetica, le quali prevedono l'invarianza della funzione lagrangiana associata al sistema rispetto a trasformazioni di gauge locali. Queste teorie presentano un diretto legame tra le dimensioni del gruppo rispetto a cui ogni forza si rivela invariante e il numero di gradi di libertà (e quindi di mediatori) del campo corrispondente: per una generica dimensione N si hanno $(N^2 - 1)$ possibili mediatori. Per la precisione il Modello Standard è una teoria di gauge con gruppo di gauge $SU(3) \times SU(2)_L \times U(1)_Y$ dove $SU(3)$ rappresenta l'interazione forte e $SU(2)_L \times U(1)_Y$ quella elettrodebole.

L'invarianza di gauge locale richiede mediatori non massivi e ciò implicherebbe avere forze a lungo raggio eppure si è verificato che solo il fotone, quanto del campo elettromagnetico, soddisfa queste caratteristiche. I mediatori dell'interazione forte sono otto, vengono chiamati *gluoni* e sono associati a tre distinte cariche dette cariche di colore R, G, B (red, green, blue); si tratta di particelle non massive che generano però una forza a corto raggio a causa del *confinamento* dei quark. La teoria di campo quantizzato per l'interazione debole prevede l'esistenza di tre mediatori massivi W^+ , W^- , Z^0 che producono una forza a cortissimo raggio che non conserva il sapore delle particelle. Le "cariche" in questo caso sono le cariche deboli o di *isospin debole*. Le interazioni debole ed elettromagnetica sono state unificate tramite la teoria elettrodebole, sviluppata negli anni '60 da Weinberg, Salam e Glashow. Tale teoria prevede la conservazione dell'ipercarica, definita attraverso la relazione di Gell-Mann-Nishijima:

$$Y_W = 2(Q - I_3) \tag{1.1}$$

dove Q indica la carica elettrica in unità di carica elementare (pari a quella dell'elettrone in valore assoluto $e = 1.602 \times 10^{-19}$ C) e I_3 la terza componente dell'isospin.

Per risolvere il problema dei mediatori massivi e non massivi (e di conseguenza interazioni a lungo e corto raggio) è venuto in aiuto il *campo di Higgs*, associato al *bosone di Higgs* (teorizzato nel 1964 e rilevato per la prima volta nel 2012 dagli esperimenti ATLAS e CMS al CERN di Ginevra), che conferisce massa a tutte le particelle fondamentali (mediatori compresi) tramite il fenomeno della rottura spontanea di simmetria.

1.1.2 Fermioni

Tutta la materia presente in natura è costituita da sole dodici particelle elementari, generalmente suddivise in due gruppi detti *quark* e *leptoni*, ognuno dei quali è composto da sei particelle legate tra loro a due a due a formare le cosiddette *generazioni*.

Tutti i fermioni hanno funzione d'onda antisimmetrica e devono quindi necessariamente obbedire al *Principio di esclusione di Pauli*, inoltre per il *Teorema spin-statistica* hanno spin pari ad $1/2$.

Quark I quark più leggeri e più stabili sono il quark up e il quark down e formano la prima generazione; quelli più pesanti sono i quark *strange*, *charm*, *top* e *bottom* (o *beauty*) che formano rispettivamente la seconda e la terza generazione e data la loro massa consistente tendono a decadere facilmente con processi da cui si generano i quark più leggeri: è per questo che la maggior parte della materia (come protoni e neutroni) oggi è costituita da up e down. Questi fermioni sentono l'interazione elettromagnetica, quella forte e quella debole quindi possiedono carica elettrica, di colore e di isospin debole: hanno tutti cariche elettriche frazionarie rispetto a quella fondamentale, precisamente quark up, charm, top (u, c, t) pari a $+2/3$ la carica fondamentale mentre down, strange, bottom (d, s, b) $-1/3$ la carica fondamentale. Per quanto riguarda il sapore si definiscono per i quark pesanti delle variabili intere che possono assumere i valori -1, 0, 1 che identificano strangeness (S), charmness (C), topness (T) e bottomness (B) e risultano opposti per particelle e antiparticelle corrispondenti e nulli in caso di mancanza del determinato sapore. Il caso di up e down è più complesso a causa della loro massa molto simile e richiede l'introduzione di una variabile particolare, detta di *isospin*, che vive in uno spazio bidimensionale e va incontro a miscelamento: per fare un parallelismo con un noto esempio matematico, il piano cartesiano, si può pensare di indicare l'asse delle ordinate come "upness" e quello delle ascisse come "downness" e immaginare di avere la particella rappresentata da un vettore unitario che si muove in questo spazio e del quale possono essere individuate le proiezioni sugli assi x ed y che saranno rispettivamente la percentuale di up e quella di down che lo compongono. I quark si legano tra loro per interazione forte sempre in modo da formare *adroni* di *colore bianco* ossia aventi carica di

colore totale nulla. I casi che possono verificarsi sono due: mesoni, particelle costituite da un quark e un antiquark, e barioni, formati da tre quark o tre antiquark.

Leptoni I sei leptoni sono organizzati nello stesso modo a formare delle coppie: si hanno elettrone e neutrino elettronico, muone e neutrino muonico, tau e neutrino tauonico. Elettrone, muone e tau hanno carica elettrica negativa (pari a quella fondamentale) e carica debole quindi prendono parte ai processi elettromagnetici e deboli ma data la mancanza di carica di colore non a quelli forti; i tre neutrini possiedono solo carica debole e pertanto sono più difficili da rilevare, considerando anche il valore di massa molto piccolo. Al contrario dei quark, i leptoni esistono solo come particelle libere.

1.1.3 Bosoni

I bosoni sono le particelle con spin intero che mediano le interazioni fondamentali:

Fotone (γ) Mediatore dell'interazione elettromagnetica, ha massa e carica nulle, quindi non interagisce con se stesso. E' il bosone più facilmente osservabile a causa della moltitudine di processi in cui viene coinvolto.

Gluoni Sono gli otto mediatori dell'interazione forte. Non possiedono massa, carica nè carica debole ma solo carica di colore quindi interagiscono solo tramite la forza forte. Possono dare origine a processi di autointerazione.

W^+ , W^- , Z^0 Si tratta dei tre mediatori dell'interazione debole: hanno valori di masse notevoli (Tabella 1.1), possiedono carica elettrica (rispettivamente positiva, negativa e nulla) e carica debole tramite la quale è possibile si generi autointerazione.

Per quanto riguarda la forza di gravità, è stata ipotizzata l'esistenza di un unico mediatore detto *gravitone* di cui ancora non sono però arrivate le conferme sperimentali. Questa particella avrebbe massa nulla, come previsto dalle teorie di gauge, e spin pari a 2.

1.1.4 Limiti del Modello Standard

Il Modello Standard, seppure ad oggi risulti il miglior modello elaborato, non riesce a dare una completa giustificazione del panorama sperimentale, basti pensare al fatto che descriva solo il 4% dell'universo conosciuto.

In primo luogo una delle quattro forze fondamentali, la gravità, non viene incorporata all'interno della teoria. Inoltre restano aperte numerose altre domande quali: "Che cosa sono la materia oscura e l'energia oscura, che costituiscono il 96% della materia e dell'energia dell'universo?", "Cosa è accaduto all'antimateria dopo

Interazione	Mediatore	Massa	Carica elettrica	Spin
Elettromagnetica	γ	$< 10^{-18}$ eV	$< 10^{-35}$	1
Forte	g	0	0	1
Debole	W^+	80.385 ± 0.015 GeV	1	1
	W^-	80.385 ± 0.015 GeV	-1	1
	Z^0	91.1876 ± 0.0021 GeV	0	1
Gravitazionale	gravitone	0	0	2
Meccanismo di Brout-Engler-Higgs	H^0	125.09 ± 0.24 GeV	0	0

Tabella 1.1: Mediatori delle interazioni fondamentali con rispettivi valori identificativi.

il Big Bang e perchè ora costituisce la minoranza rispetto alla materia ordinaria?", "Perchè le tre generazioni di quark e leptoni hanno valori di massa così diversi tra loro?".

Nonostante il Modello Standard descriva accuratamente i fenomeni che rientrano all'interno del suo dominio di validità, risulta essere ancora incompleto. Un'ipotesi è che il MS sia una "teoria efficace", valida fino a una certa scala di energia, oltre la quale si aprono nuovi gradi di libertà che vengono comunemente denominati "nuova fisica".

1.2 Teorie di Campo Efficace (EFT)

Verrà ora presentato un particolare approccio nella ricerca di "nuova fisica" che risulta essere un promettente metodo d'indagine: il *BSM* (*Beyond Standard Model*), una teoria comprensiva di nuove particelle e con esse nuove interazioni.

I fisici sono consapevoli che questa nuova fisica esista da qualche parte e dovrebbero riuscire a rilevare qualche indizio di essa. Sorge a questo proposito un problema: le prove che potrebbero trovarsi si collocherebbero a livelli di energia superiori rispetto a quelli accessibili dagli odierni rivelatori. In questo caso non si potrebbe lavorare su rilevazioni dirette che confermino le teorie BSM ma si dovrebbe adottare un approccio differente, andare cioè ad osservare le correzioni apportate dalle nuove teorie rispetto all'andamento ormai consolidato del Modello Standard a valori di energia che siano raggiungibili dalla moderna strumentazione, come mostrato in Figura 1.2. Questo concetto è alla base delle *EFT* (*Effective Field Theories*).

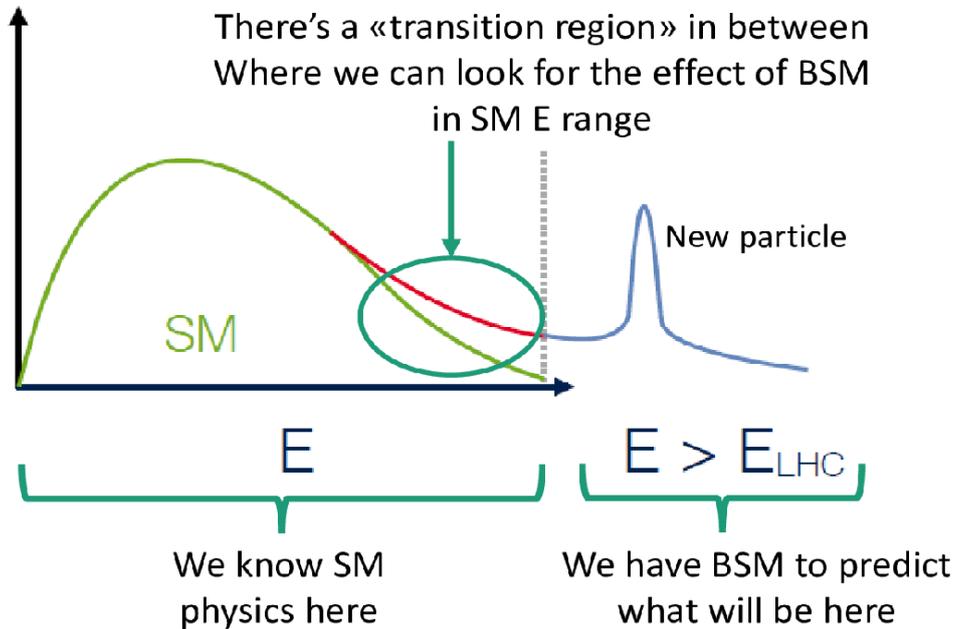


Figura 1.2: Ipotesi alla base delle EFT (Effective Field Theories).

In fisica una teoria di campo efficace è quindi un tipo di approssimazione che si applica ad un modello già esistente e verificato; essa include i gradi di libertà necessari per descrivere fenomeni fisici che avvengono a determinate scale di lunghezza ed energia mentre ignora la composizione interna e i gradi di libertà a distanze più corte e a più alte energie. Intuitivamente si cerca di capire il comportamento di una teoria a distanze minori per poi derivare quello che si spera sia un modello più semplice a scale maggiori. Le EFT solitamente lavorano meglio quando si ha una consistente separazione tra scala di energia di interesse e scala di energia valida per

i modelli già elaborati. Hanno già trovato largo impiego in diversi settori come in fisica delle particelle, meccanica statistica, fisica degli stati condensati, relatività generale e idrodinamica perchè semplificano i calcoli e permettono di esaminare anche effetti di dissipazione e radiazione.

1.2.1 Un esempio di applicazione delle EFT: il decadimento β

β

L'esistenza di fisica oltre il Modello Standard può essere investigata in due modi: usando un metodo "top-bottom" che si concentra sulla ricerca di nuovi segnali, o con un metodo "bottom-up" che indaga cercando nuove interazioni tra particelle già conosciute. In questo secondo caso, il metodo delle EFT propone un consistente modello per l'interpretazione dei risultati sperimentali [2]. L'esempio più noto di EFT è la teoria di Fermi del decadimento β , sviluppata durante i primi anni di studi dei decadimenti deboli quando ancora erano noti solamente gli adroni e i leptoni soggetti all'interazione debole. In particolare si analizza il decadimento del neutrone rappresentato dalla seguente equazione:

$$n \rightarrow p + e^- + \bar{\nu}_e \tag{1.2}$$

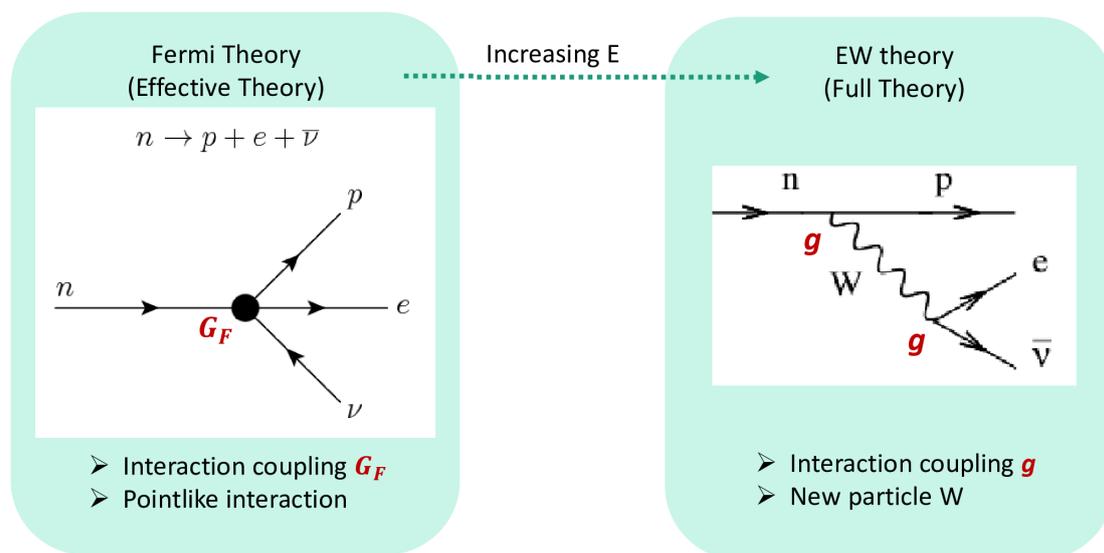


Figura 1.3: Esempio di applicazione delle EFT al decadimento beta. A sinistra si ha la rappresentazione del fenomeno attraverso teorie di campo effettive, a destra si trova la raffigurazione basata sulla teoria elettrodebole.

In questo caso la EFT, che coincide con il modello iniziale di Fermi, propone un'interazione puntiforme tra quattro fermioni con costante di accoppiamento G_F (che altro non è se non la *costante di Fermi* la cui determinazione più precisa

$G_F = 1.16638 \times 10^{-5} \text{GeV}^{-2}$ viene dalle misure del tempo di vita del muone che è inversamente proporzionale al quadrato di G_F). Questa interpretazione ha avuto un grande successo fenomenologico e venne in seguito ampliata fino all'elaborazione della teoria elettrodebole in cui l'interazione viene mediata da un bosone di gauge W^- che causa il cambiamento di sapore nei fermioni interagenti (si ricordi che l'interazione debole non conserva il sapore). Il grande successo riscosso dalla teoria di Fermi fu dovuto principalmente alla scala di separazione energetica che identificava i due modelli: nel primo caso (immagine a sinistra in Figura 1.3) gli esperimenti erano compiuti ad una energia pari a meno di 10 MeV mentre per rilevare i mediatori W^\pm era necessario salire a valori maggiori della loro massa, corrispondente a circa 80 GeV. La differenza tra le due scale, di circa tre ordini di grandezza, risulta enorme: fino al momento in cui non è stato possibile riprodurre tramite gli esperimenti energie così elevate, la EFT ha funzionato perfettamente. Con il progredire tecnologico e di conseguenza strumentale, i fisici si sono resi conto che la teoria di Fermi racchiudeva informazioni anche sulla teoria elettrodebole che costituisce una sua generalizzazione. La relazione tra i due modelli è esplicitata dalla seguente equazione:

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} \quad (1.3)$$

dove G_F è la costante di Fermi, g è la costante di interazione del diagramma della teoria elettrodebole e M_W rappresenta la massa dei mediatori W^\pm .

1.2.2 La SMEFT

Particelle e interazioni del Modello Standard sono descritte dalla *lagrangiana* L_{SM} del Modello Standard. Le teorie BSM aggiungono degli elementi a questa funzione ed è necessario capire come questi nuovi termini debbano essere inclusi. Ci sono due possibilità:

$L_{TOT} = L_{SM} + L_{BSM}$ In questo caso il modello teorico dipenderà direttamente dal secondo termine e si avranno particelle e interazioni proprie delle teorie BSM in aggiunta a quelle già note del Modello Standard quindi si dovrà elaborare un modello BSM preciso. Questo fatto implica che le nuove particelle siano entro il range raggiungibile sperimentalmente.

$L_{TOT} = L_{SM} + \delta L_{SM}$ Il secondo caso tratta le informazioni legate alle teorie BSM come fossero una perturbazione della lagrangiana del Modello Standard: vengono introdotte nuove interazioni che coinvolgono però particelle del Modello Standard. Così facendo il modello non richiede una specifica teoria e le nuove particelle sono ammesse solo a valori energetici superiori rispetto a quello raggiunto al giorno d'oggi.

La strada che si è scelto di percorrere è la seconda, in cui la lagrangiana totale è composta da quella del Modello Standard unita ad un termine perturbativo e prende il nome di L_{SMEFT} dove $SMEFT$ si traduce come *Standard Model Effective Field Theory*.

Sapendo che la lagrangiana del Modello Standard è costituita da operatori generati dai campi fermionici e bosonici e che ha dimensione 4, si dimostra che δL_{SM} prende fermioni e bosoni già noti e costruisce nuovi operatori di dimensione 6 con essi, rispettando le simmetrie proprie del Modello Standard. Matematicamente si ottiene la relazione:

$$L_{TOT} = L_{SM} + \delta L_{SM} = L_{SM} + \frac{1}{\Lambda^2} \sum c_i O_i^{(6)} \quad (1.4)$$

dove $O_i^{(6)}$ sono gli operatori di dimensione 6 e c_i sono i coefficienti di interazione detti *coefficienti di Wilson*. In Figura 1.4 vengono mostrati alcuni degli operatori che si possono avere, ognuno di questi con i rispettivi pedici identificativi del processo in cui viene coinvolto.

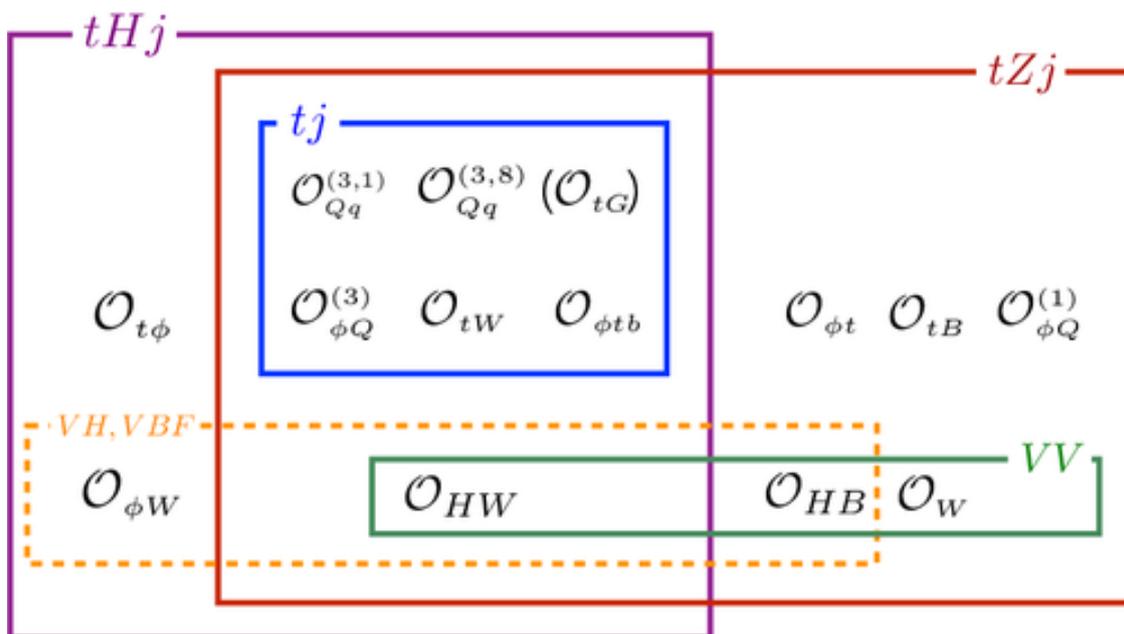


Figura 1.4: Alcuni dei possibili operatori di dimensione 6 che completano la lagrangiana L_{SMEFT} . I raggruppamenti colorati descritti dalle sigle dello stesso colore rappresentano i principali processi che questi operatori descrivono, la discussione di questi processi esula dall'obiettivo di questo elaborato.

L'effetto del termine correttivo δL_{SM} è duplice: consente nuovi processi (come la produzione di coppie di quark top same sign) e in quelli del Modello Standard crea delle lievi variazioni rispetto alle predizioni basate sulla presenza del solo termine L_{SM} , misurabili anche alle energie di oggi.

1.2.3 Produzione di coppie di quark top same sign

Tra tutti i fermioni del Modello Standard, il quark top è quello avente massa maggiore con valori pari a 171.77 ± 0.38 GeV (esperimento CMS dell'acceleratore LHC del 2022 con dati del 2016): questa caratteristica necessita di una grande energia per essere indagata e ciò lo rende un interessante candidato per lo studio delle teorie BSM.

Nel Modello Standard il processo $pp \rightarrow tt$, ossia la produzione di due quark top a partire da due protoni, risulta avere una sezione d'urto pari a $\sigma(pp \rightarrow tt) \approx 10^{-15}$ pb che è un valore molto piccolo: questo avviene a causa del *meccanismo di Glashow-Iliopoulos-Maiani (GIM)*, meccanismo che prevede la soppressione delle interazioni mediate da correnti neutre che cambiano il sapore (*Flavour-Changing Neutral Current, FCNC*) nei diagrammi di Feynman che presentano dei loop. Il motivo dello studio del meccanismo di GIM e delle interazioni FCNC risiede nell'importanza che questi potrebbero avere per lo sviluppo delle teorie BSM: mantenere o eliminare gli accoppiamenti FCNC cambia significativamente la strada da intraprendere per l'elaborazione di future teorie perchè prevede o impedisce lo sviluppo di determinati processi [3].

Si potrebbe pensare ad un processo in cui due quark up iniziali interagiscono tra loro scambiandosi un bosone e cambiando il loro sapore trasformandosi così in una coppia di quark top finali. Ad oggi sono stati proposti tre possibili modelli per questa interazione, ciascuno mediato da un bosone differente: il bosone Z^0 del Modello Standard, un nuovo bosone Z' oppure un bosone scalare che si accoppia fortemente con il quark top [4].

Nelle teorie BSM la produzione di coppie di quark top dello stesso segno attraverso FCNC è permessa e può essere descritta utilizzando le EFT e in particolare una interazione puntiforme come mostrato in Figura 1.5. Date queste premesse, la lagrangiana che descrive il sistema è quella di equazione 1.4 nella quale compaiono cinque operatori di dimensione 6, di cui si riporta solo per completezza la forma matematica di seguito:

$$\begin{aligned}
O_{RR} &= [\bar{t}_R \gamma^\mu u_R] [\bar{t}_R \gamma_\mu u_R] \\
O_{LL}^{(1)} &= [\bar{Q}_L \gamma^\mu q_L] [\bar{Q}_L \gamma_\mu q_L] \\
O_{LL}^{(3)} &= [\bar{Q}_L \gamma^\mu \sigma_a q_L] [\bar{Q}_L \gamma_\mu \sigma^a q_L] \\
O_{LR}^{(1)} &= [\bar{Q}_L \gamma^\mu q_L] [\bar{t}_R \gamma_\mu u_R] \\
O_{LR}^{(8)} &= [\bar{Q}_L \gamma^\mu T^A q_L] [\bar{t}_R \gamma_\mu T^A u_R]
\end{aligned}$$

I termini $O_{LL}^{(1)}$ e $O_{LL}^{(3)}$ non vengono inseriti nella sommatoria di equazione 1.4 perchè già inclusi all'interno di altri addendi. In modo estremamente semplificato,

si può pensare che ogni termine all'interno delle parentesi quadre rappresenti una coppia di quark (destrorsi o sinistrorsi, R o L) connessi da un particolare tipo di vertice.

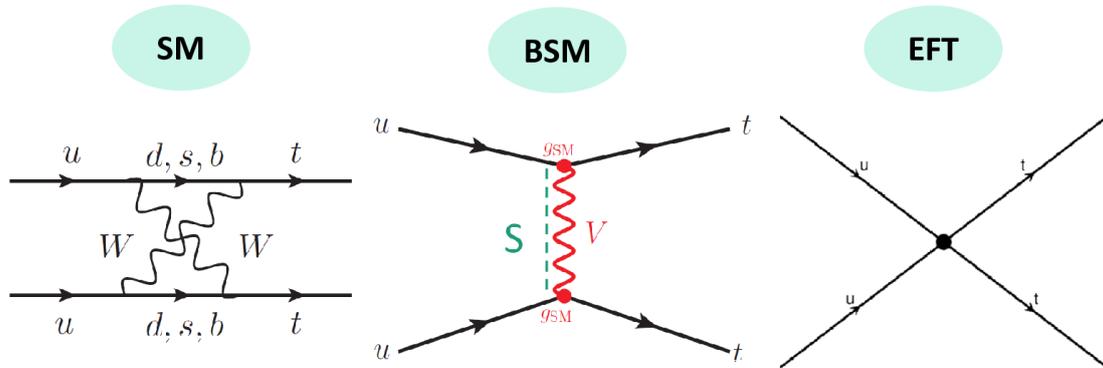


Figura 1.5: Confronto tra le rappresentazioni del processo $pp \rightarrow tt$ di produzione di coppie di quark top dello stesso segno proposte da diversi modelli teorici. Da sinistra: Modello Standard, teorie BSM e EFT.

Capitolo 2

L'esperimento ATLAS

2.1 Il CERN

L'*Organizzazione Europea per la Ricerca Nucleare*, comunemente conosciuta con la sigla *CERN*, è il più grande laboratorio al mondo di fisica delle particelle. Posto al confine tra la Francia e la Svizzera, alla periferia ovest della città di Ginevra, la convenzione che lo istituì fu firmata il 29 Settembre 1954 da 12 stati membri. Attualmente ne fanno parte 23 stati più alcuni osservatori, compresi stati extraeuropei.

Lo scopo principale di questa collaborazione è fornire ai ricercatori gli strumenti necessari per la ricerca in fisica delle alte energie: gli *acceleratori di particelle*, che portano nuclei atomici e particelle subnucleari ad energie molto elevate, e i *rivelatori*, che permettono di osservare i prodotti delle collisioni tra fasci di queste particelle.

2.1.1 Catena di acceleratori

Ad oggi il complesso degli acceleratori del CERN comprende sette acceleratori principali. Fin dal principio è stato previsto che ogni nuova e più potente macchina avrebbe utilizzato le precedenti come iniettori (cioè ogni macchina accelera le particelle un po' di più rispetto alle precedenti prima di iniettarle ad una velocità maggiore nella successiva), creando una catena di acceleratori che porta gradualmente un fascio di particelle ad energie sempre più elevate. Per consentire il funzionamento di questa catena, tutte le funzioni degli acceleratori sono coordinate da un unico segnale temporale di riferimento con una precisione dell'ordine del nanosecondo, generato da un sistema di orologi atomici e distribuito per tutta la struttura.

Gli acceleratori principali a disposizione del CERN sono, in ordine dalla sorgente iniziale delle particelle a bassa energia fino all'anello principale di collisione (Figura 2.1):

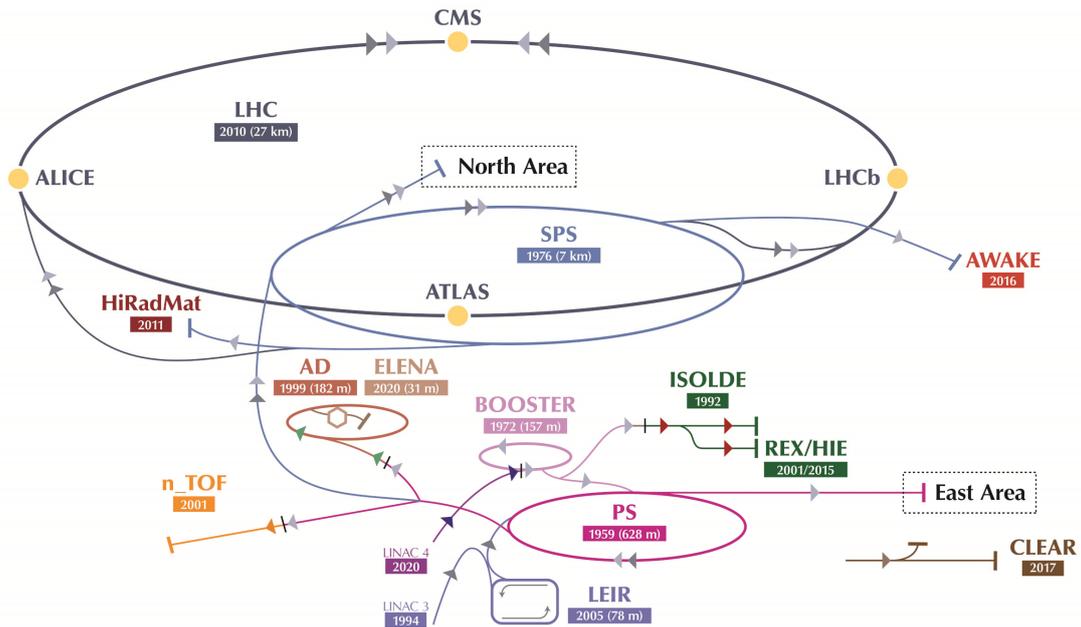


Figura 2.1: Rappresentazione del complesso di acceleratori che compongono il CERN, comprensiva anche dei principali esperimenti attualmente in atto sui vari anelli.

LINAC I LINAC, o acceleratori lineari, generano particelle a bassa energia successivamente immerse nel PS Booster. Sono noti come LINAC3 e LINAC4 e raggiungono i 50 MeV per protone. Tutta la catena di acceleratori successiva dipende da queste sorgenti.

LEIR Il *Low Energy Ion Ring* (LEIR), che accelera fasci di ioni di piombo fino a 72 MeV per nucleone, ha iniziato a lavorare nel 2010 nella catena di pre-accelerazione dell'LHC.

PS BOOSTER Il *Proton Synchrotron Booster*, costituito da quattro sincrotroni sovrapposti con un raggio di 25 m, aumenta l'energia delle particelle generate dai LINAC prima di iniettarle nel PS. Viene inoltre utilizzato per esperimenti separati, come ad esempio ISOLDE, che studia nuclei instabili di isotopi molto pesanti.

PS Il *Proton Synchrotron* (PS), costruito nel 1959, è un sincrotrone con una circonferenza di circa 628 m in grado di accelerare protoni e altre particelle fino a 26 GeV. In particolare riceve protoni dal *Proton Synchrotron Booster* e ioni piombo dal *Low Energy Ion Ring*.

SPS Il *Super Proton Synchrotron* (SPS) è un acceleratore circolare di 2 km di diametro che ha iniziato a funzionare nel 1976 e che alimenta l'LHC con protoni e ioni piombo. Oltre ad avere una propria linea di fascio rettilinea per esperimenti a bersaglio fisso, ha funzionato come collisore protone-antiprotone

e come stadio finale di accelerazione per gli elettroni e i positroni da iniettare nel LEP. Ha ripreso questo ruolo per i protoni e gli ioni piombo immessi nell'LHC.

LHC Il *Large Hadron Collider* (LHC), entrato in funzione il 10 Settembre 2008 dopo lo smantellamento del *Large Electron-Positron Collider* (LEP), si estende su una circonferenza di 27 km ed è stato progettato per accelerare le particelle fino ad un massimo di 6.8 TeV per fascio. A queste energie permette di studiare le particelle elementari in condizioni sperimentali paragonabili a quelle dei primi momenti di vita dell'Universo, immediatamente dopo il Big Bang.

L'intero percorso compiuto dalle particelle poi studiate nel centro di ricerca del CERN è il seguente: LINAC4 è diventato la sorgente di fasci di protoni nel 2020, accelera ioni negativi di idrogeno (H^- , che consistono in un atomo di idrogeno avente un elettrone in eccesso) fino a circa 160 MeV per prepararli ad entrare nel PSB. Durante il passaggio tra i due acceleratori vengono strappati via i due elettroni presenti in ogni ione: rimangono così solamente protoni che sono accelerati fino a 2 GeV prima di entrare nel PS, che aumenta l'energia dei fasci fino a 26 GeV. L'energia dei protoni viene aumentata ancora dal successivo acceleratore, il SPS, fino a valori di 450 GeV. Infine entrano nell'LHC, l'anello di dimensioni maggiori, dove vengono divisi in due fasci e fatti circolare rispettivamente in senso orario e antiorario. In circa 20 minuti i protoni raggiungono la loro massima energia, corrispondente a 6.8 TeV, ma sono necessarie diverse ore prima delle collisioni, che vengono fatte avvenire in quattro differenti punti della circonferenza, corrispondenti ai quattro esperimenti (*ATLAS*, *ALICE*, *CMS*, *LHCb*) all'interno dei quali l'energia raggiunge i 13.6 TeV.

I protoni non sono le uniche particelle accelerate nell'LHC: ioni piombo partono da una sorgente di piombo vaporizzato ed entrano in LINAC3 prima di essere inseriti in LEIR, per poi seguire lo stesso identico percorso descritto per i protoni [5].

2.2 Il Large Hadron Collider (LHC)

Il *Large Hadron Collider* (LHC) è l'acceleratore di particelle più potente e più grande al mondo, composto da un anello di magneti superconduttori di 27 km di lunghezza. La progettazione dell'LHC ha richiesto una precisione straordinaria, tanto da rendere necessario tenere conto dell'influenza della forza di attrazione gravitazionale esercitata dalla Luna sulla crosta terrestre e dei disturbi elettrici provocati dal passaggio dei treni in superficie ad un chilometro di distanza.

All'interno dell'acceleratore, due fasci di particelle ad alta energia scorrono in senso opposto dentro due tubi tenuti in un vuoto ultra-spinto per essere poi fatti

collidere. Affinchè i fasci seguano la curvatura dell'anello, viene utilizzato un forte campo magnetico mantenuto da magneti superconduttori costruiti con bobine di speciali cavi elettrici che operano nella condizione di superconduttività: in questo modo si assicura una grande conduttività elettrica senza perdita di energia dovuta alle resistenze. Questo sistema richiede un raffreddamento dei magneti a temperature di -271.3 °C, più basse di quelle registrate nello spazio, per questo gran parte dell'acceleratore è connesso con un sistema di distribuzione di elio liquido necessario al raffreddamento dei magneti e di altri sistemi di supporto.

Tutti i controlli per il funzionamento sono collocati all'interno del *CERN Control Centre*: da qui i fasci vengono fatti collidere all'interno dell'LHC all'energia del centro di massa pari a 13.6 TeV (per confronto il *TeVatron*, acceleratore protone-antiprotone che ha operato vicino a Chicago in Illinois negli Stati Uniti dal 1988 fino ai primi anni del LHC, lavorava ad un'energia massima di 1.8 TeV). Le collisioni avvengono in quattro diversi siti posizionati lungo l'anello, in corrispondenza dei quattro principali detector ATLAS, CMS, LHCb e ALICE [6].

2.2.1 Esperimenti principali: ATLAS, CMS, LHCb, ALICE

I rivelatori principali al LHC sono collocati in quattro grandi sale sotterranee: ATLAS e CMS sono progettati per una esplorazione generale della fisica protone-protone alle energie del TeV mentre LHCb è un rivelatore più piccolo dedicato allo studio degli adroni contenenti un quark bottom e, infine, ALICE è ottimizzato per gli esperimenti di collisioni ione-ione alle quali sarà dedicata una certa frazione del tempo di funzionamento della macchina. I più importanti obiettivi del LHC sono il completamento della verifica del Modello Standard della fisica delle particelle con il chiarimento del meccanismo di rottura della simmetria elettrodebole e la scoperta del bosone di Higgs, la ricerca di nuova fisica alle energie del TeV e, possibilmente, l'identificazione della particella (o delle particelle) responsabili della materia oscura nell'Universo. Inoltre LHCb esplora la dinamica del sapore e della violazione della simmetria CP nei decadimenti degli adroni contenenti un quark bottom e ALICE studia il comportamento della materia adronica ad alta temperatura e densità cercando di mettere in evidenza la transizione di fase di deconfinamento che fa passare dalla materia adronica al plasma di quark e gluoni [7].

2.2.2 Altri esperimenti

Lungo gli anelli del CERN sono in corso anche altri esperimenti oltre ai quattro principali già menzionati:

CLOUD (*Cosmics Leaving OUtdoor Droplets*) È un esperimento che ha lo scopo di studiare le relazioni tra raggi cosmici galattici e aerosol atmosferico in condizioni controllate. L'esperimento ha iniziato a operare nel Novembre 2009.

ISOLDE (*Isotope Separator On Line DEvice*) È un laboratorio in cui vengono prodotti nuclei atomici radioattivi di isotopi pesanti inviando protoni ad alta energia su targhette fisse di differenti elementi chimici per studiarne le proprietà atomiche e nucleari e i possibili decadimenti esotici.

AD (*Antiproton Decelerator*) È un anello di accumulazione che viene utilizzato per rallentare gli antiprotoni, successivamente inviati agli esperimenti che studiano l'antimateria e in particolare l'anti-idrogeno.

ALPHA Finalizzato allo studio delle simmetrie tra materia e antimateria sfruttando le caratteristiche spettrali dell'idrogeno; in Dicembre 2016 è stata ottenuta la prima riga spettrale di un atomo di anti-idrogeno.

AWAKE Si tratta di un esperimento che indaga l'accelerazione del plasma wake-field usando un insieme di protoni come guida.

2.3 Esperimento ATLAS

ATLAS è il più grande e uno dei più complessi rivelatori di particelle mai costruito: lungo 46 m e alto 25 m, è composto da un enorme sistema di bobine di magneti superconduttori, organizzati a formare un cilindro attorno al fascio posto al centro del rivelatore.

Gli strati concentrici di cui è composto il rivelatore misurano traiettoria, momento ed energia della maggior parte delle particelle cariche e neutre e ricostruiscono le tracce di quelle cariche. La curvatura della traiettoria nel campo magnetico permette di determinare l'impulso e il segno della carica elettrica delle particelle. Su un milione di collisioni, solo una ha caratteristiche interessanti per la ricerca di fenomeni nuovi: il sistema di trigger seleziona questi eventi e li registra, evitando così l'accumulo di enormi quantità di informazioni non necessarie. Le collisioni tra protoni che avvengono all'interno dell'apparato lasciano frammenti che riveleranno nuovi aspetti della fisica, dallo studio del bosone di Higgs ai quark top fino alla ricerca di nuove dimensioni e particelle che potrebbero essere i costituenti della materia oscura [8, 9].

2.3.1 Struttura di un rivelatore

L'obiettivo principale di un detector è quello di identificare le particelle prodotte durante le collisioni e di misurarne variabili dinamiche ed interne quali posizione, velocità, energia, massa e carica. Per fare ciò, i rivelatori sono costituiti da diversi strati o *subdetectors*, ognuno dei quali ha un ruolo particolare nella ricostruzione delle collisioni, e da un *magnete* che curva le traiettorie delle particelle consentendo la misurazione del loro momento [10].

Esistono due categorie di sotto-rivelatori:

Tracciatori Rivelano la "scia" di particelle cariche elettricamente attraverso la traccia che lasciano ionizzando la materia. In un campo magnetico possono essere utilizzati per misurare la curvatura delle traiettorie, da cui si ricava il momento che può aiutare ad identificare la particella in questione. I più moderni sistemi di tracciamento non rendono la "scia" direttamente visibile ma producono un segnale elettrico che può essere registrato e ricostruito in seguito da un computer. Due tipi di tracciatori sono i *vertex detectors* e le *camere a muoni*: i primi vengono posizionati vicino al punto di interazione (detto *vertice primario*) mentre le camere a muoni si trovano nello strato più esterno del detector perchè i muoni sono le uniche particelle cariche in grado di viaggiare attraverso strati di materiale denso.

Calorimetri Questi dispositivi calcolano l'energia delle particelle fermandole e misurandone l'energia liberata. Ci sono due tipi principali di calorimetri, che usano diversi materiali in base al loro scopo: i *calorimetri elettromagnetici* (*ECAL, Electromagnetic CALorimeter*) e i *calorimetri adronici* (*HCAL, Hadronic CALorimeter*). I primi assorbono elettroni e fotoni, che interagiscono attraverso la forza elettromagnetica: gli adroni, particelle che interagiscono fortemente quali protoni e pioni, iniziano a perdere energia nell'ECAL ma vengono fermati solo nell'HCAL; muoni e neutrini invece attraversano entrambi. I calorimetri permettono di identificare particelle neutre come fotoni e neutroni, che sfuggono ai tracciatori. Generalmente sono costituiti da strati di materiale assorbente (o *passivo*) ad alta densità (piombo ad esempio) intervallati da strati di materiale *attivo* come argon liquido.

2.3.2 Struttura del rivelatore di ATLAS

ATLAS presenta quattro componenti principali [9], ben visibili in Figura 2.2:

Rivelatore interno Costituisce la prima parte di ATLAS a rivelare i prodotti delle collisioni: è molto sensibile ed è formato da tre diversi sistemi di sensori immersi in un campo magnetico parallelo alla direzione di propagazione del fascio. Il rivelatore interno misura direzione, momento e carica delle particelle prodotte nelle collisioni protone-protone. I suoi componenti principali sono il *Pixel Detector*, il *Semiconductor Tracker* (SCT) e il *Transition Radiation Tracker* (TRT).

Calorimetro Misura l'energia che una particella perde attraversando il rivelatore e, sfruttando la struttura composta da un'alternanza di strati di materiali diversi, ferma tutte le particelle conosciute esclusi muoni e neutrini. Le sue parti principali sono il *Liquid Argon Calorimeter* (LAr) e il *Tile Hadronic Calorimeter*.

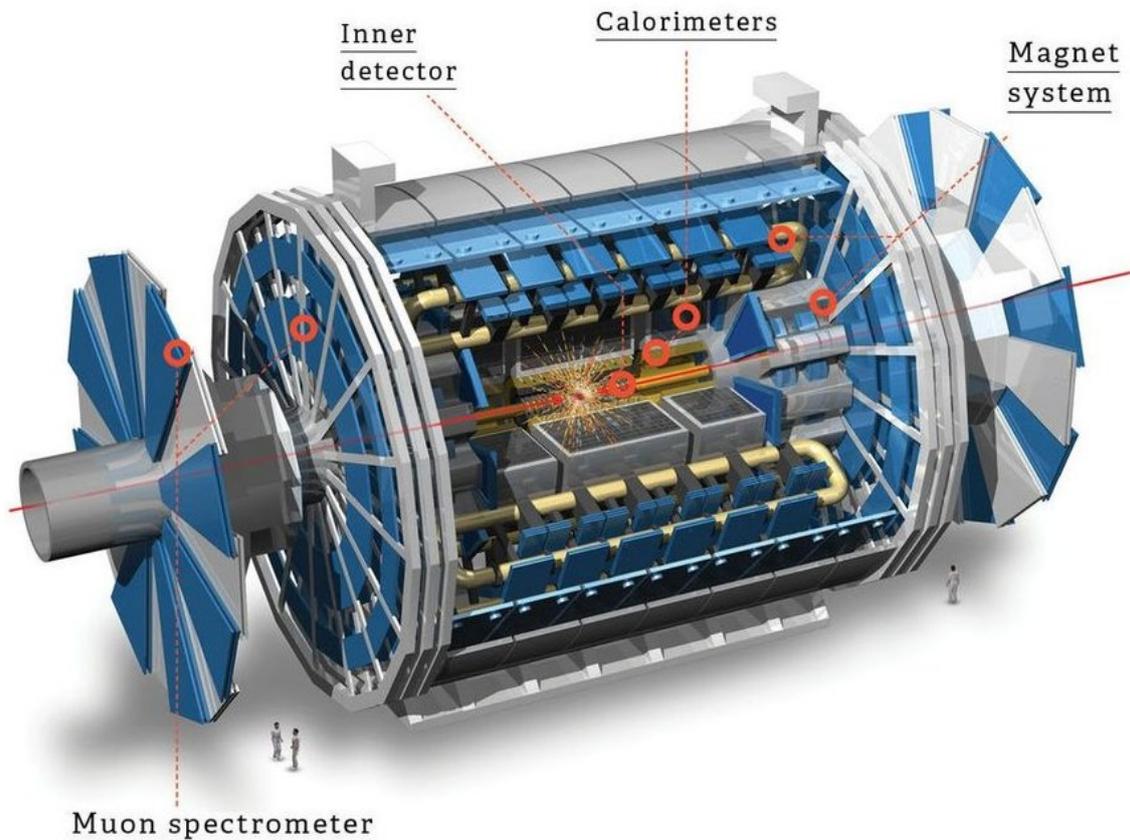


Figura 2.2: Schema del rivelatore ATLAS completo dei quattro componenti fondamentali: rivelatore interno, calorimetro, spettrometro a muoni e sistema di magneti superconduttori.

Spettrometro a muoni Identifica e misura il momento dei muoni che attraversano radialmente l'intero rivelatore. Si compone di 4000 singole camere a muoni e le sue parti fondamentali sono le *Thin Gap Chambers*, le *Resistive Plate Chambers*, i *Monitored Drift Tubes* e le *Cathode Strip Chambers*.

Sistema di magneti Devia le traiettorie delle particelle cariche attraverso i diversi strati del rivelatore, così che ATLAS possa misurarne momento e carica, usando due tipi di magneti superconduttori: solenoidali e toroidali. Questi sono raffreddati a temperature di circa 4.5 K affinché venga prodotto il campo magnetico necessario. I costituenti principali del sistema di magneti sono il *Central Solenoid Magnet*, il *Barrel Toroid* e i *End-cap Toroids*.

Il risultato di una struttura così complessa è il passaggio solo di determinate particelle attraverso alcuni strati e di altre attraverso parti diverse, come mostrato in Figura 2.3. Questo permette di ricavare dati fondamentali per la determinazione dei prodotti delle collisioni.

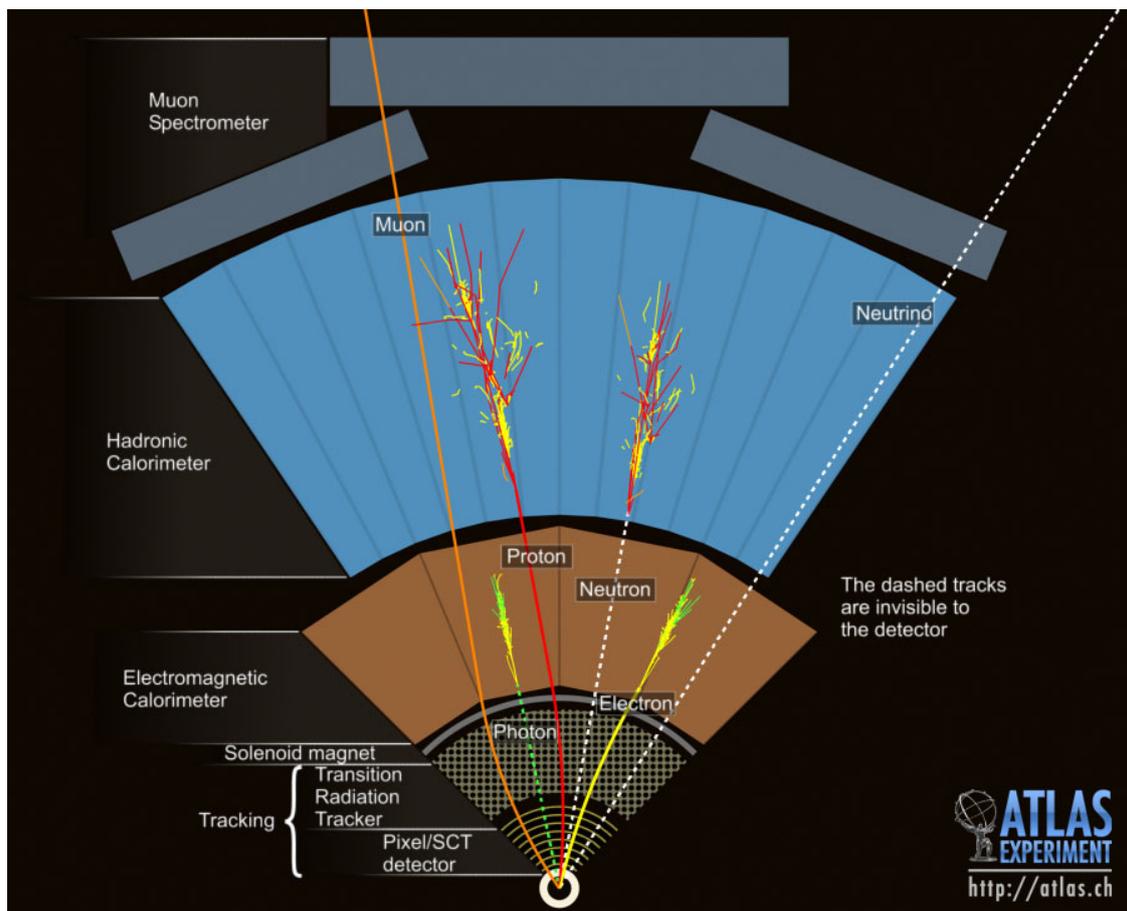


Figura 2.3: Rappresentazione del passaggio di particelle diverse attraverso le varie sezioni del detector ATLAS: fotoni ed elettroni vengono fermati dall'ECAL, protoni e neutroni si arrestano nell'HCAL mentre muoni e neutrini riescono a sfuggire anche allo spettrometro a muoni.

2.3.3 Trigger e acquisizione dati in ATLAS

ATLAS è studiato per osservare fino a 1.7×10^9 collisioni protone-protone al secondo, con un corrispondente volume di dati pari a 60×10^9 MB al secondo. Per ogni collisione l'obiettivo è contare, tracciare e caratterizzare tutte le particelle prodotte al fine di ricostruire l'intero processo. Le informazioni che si ricavano dal moto sono fondamentali: la carica ad esempio comporta la deviazione della traiettoria in un verso per le particelle positive e in quello opposto per le negative; il momento, legato alla velocità, è ricavabile dalla forma delle tracce lasciate [10].

Il sistema di trigger e acquisizione dati seleziona le collisioni più interessanti da studiare attraverso un procedimento in due fasi [11]:

1. Il primo livello di trigger è hardware e lavora con le informazioni di calorimetri e spettrometro a muoni. La decisione di mantenere un dato avviene meno di

2.5 secondi dopo l'evento analizzato e se viene selezionato passa direttamente al secondo livello di trigger che è in grado di accettare fino a 10^5 eventi al secondo.

2. Il secondo livello di trigger è software ed esamina dati ricevuti da specifiche zone del detector per poi selezionarli ed inviarli ad un sistema di archiviazione per l'analisi offline.

Capitolo 3

Ricerca di eventi top-top mediante tecniche di apprendimento automatico

Si intende utilizzare algoritmi di apprendimento automatico per la ricerca di un segnale costituito da eventi con due quark top nello stato finale. Tali algoritmi saranno adoperati per discriminare il segnale da eventi di fondo che possiedono caratteristiche molto simili (*problema di classificazione*). A tale scopo verranno individuate opportune variabili da utilizzare in ingresso per gli algoritmi, che poi forniranno in uscita un valore numerico utilizzato come classificatore. Le variabili in ingresso possono essere relative a dati sperimentali a vari livelli di ricostruzione: in linea di principio si può partire da un livello base di ricostruzione (“dati grezzi” provenienti direttamente dai rivelatori), oppure da variabili di basso livello in cui sono ricostruiti i quadrivettori delle particelle rivelate, fino a variabili di alto livello in cui sono ricostruite grandezze fisiche a partire da tali quadrivettori (come ad esempio masse invarianti, energia mancante, etc). Se da un lato le variabili di alto livello contengono un’informazione fisica direttamente collegata al processo in esame, dall’altro queste variabili, definite dall’utente, possono portare ad una perdita di informazione utile durante la ricostruzione. All’altro estremo, le variabili grezze contengono tutte le informazioni fisiche ma è più complesso in termini computazionali realizzare la costruzione di variabili complesse in modo automatico. Si intende investigare le prestazioni di una tecnica di riconoscimento automatico (*Artificial Neural Networks*) confrontando i risultati ottenuti fornendo dapprima variabili di alto livello, successivamente variabili strettamente di basso livello e infine le stesse del caso precedente con in aggiunta due necessarie per il b-tagging.

3.1 Reti neurali

Una *rete neurale* (*NN* dall'inglese *Neural Network*) è un modello matematico composto da neuroni artificiali ispirato alle reti neurali biologiche che costituiscono il cervello umano. Viene utilizzata per risolvere problemi ingegneristici di intelligenza artificiale legati a diversi ambiti tecnologici come l'informatica, l'elettronica o altre discipline.

3.1.1 Struttura di una NN

L'architettura della rete neurale si ispira al cervello umano. Le cellule del cervello umano, i neuroni, formano una rete complessa e altamente interconnessa e inviano segnali elettrici l'uno all'altro per aiutare gli esseri umani a elaborare informazioni. Analogamente, una rete neurale artificiale è composta di neuroni artificiali che collaborano per risolvere un problema. I neuroni artificiali sono moduli software, chiamati nodi, e le reti neurali artificiali sono programmi software o algoritmi che utilizzano sostanzialmente i sistemi di calcolo per risolvere problemi matematici.

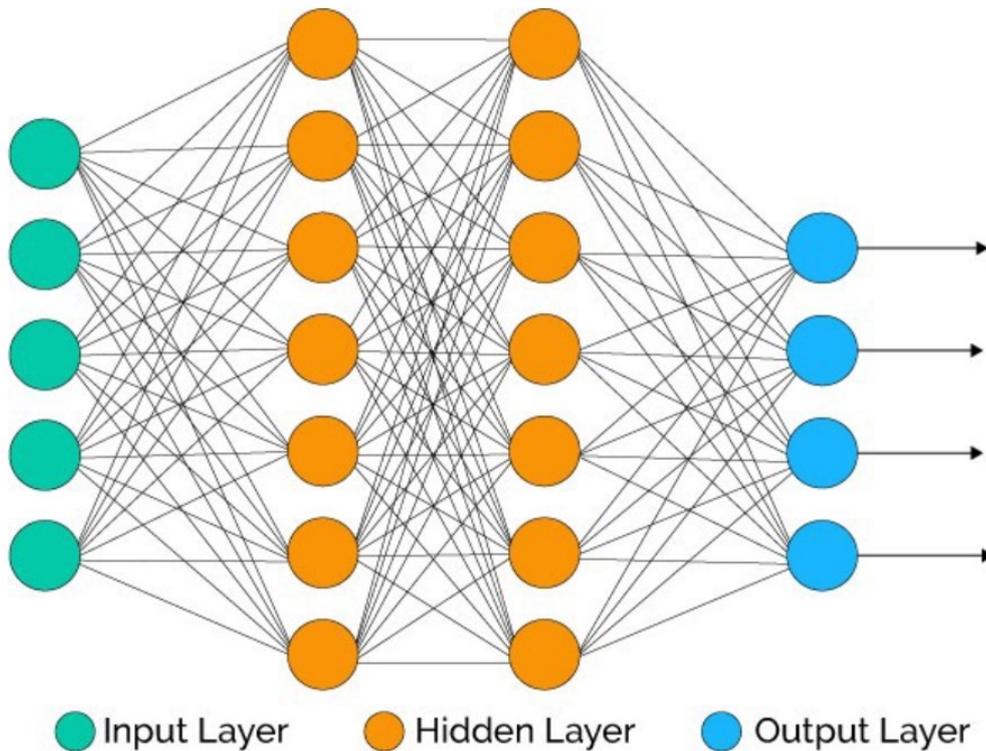


Figura 3.1: Rappresentazione schematica della struttura di una rete neurale: in verde il livello di input, in arancione i livelli intermedi e in azzurro il livello di output.

Architettura semplice di una rete neurale

Una rete neurale ha neuroni artificiali interconnessi su tre livelli (Figura 3.1):

Livello di input Le informazioni dal mondo esterno entrano nella rete neurale dal livello di input. I nodi di input elaborano i dati, li analizzano o li categorizzano, e li trasferiscono al livello successivo.

Livello nascosto I livelli nascosti prendono il proprio input dal livello di input o da altri livelli nascosti. Le reti neurali artificiali possono presentare un numero elevato di livelli nascosti. Ogni livello nascosto analizza l'output dal livello precedente, lo elabora ulteriormente e lo trasferisce al livello successivo.

Livello di output Il livello di output restituisce il risultato finale di tutta l'elaborazione dati attraverso la rete neurale artificiale. Può avere uno o più nodi: avendo ad esempio un problema a classificazione binaria (sì/no), il livello di output avrà un solo nodo di output che darà come risultato 1 o 0; se invece si presenta un problema multi-classe, il livello di output potrebbe consistere in più nodi.

Architettura profonda di una rete neurale

Le reti neurali profonde, o reti di *deep learning*, presentano molteplici livelli nascosti con milioni di neuroni artificiali connessi tra di loro. Teoricamente, le reti neurali profonde possono mappare qualsiasi tipo di input verso qualsiasi tipo di output, tuttavia hanno maggiormente bisogno di addestramento rispetto ad altri metodi di machine learning: necessitano di milioni di esempi di dati come allenamento rispetto alle centinaia o migliaia di cui una rete più semplice ha bisogno.

In generale esiste un numero, chiamato *peso*, che rappresenta le connessioni tra un nodo e l'altro: il peso è un numero positivo se un nodo ne stimola un altro, negativo se lo sopprime. I nodi con valori di peso più alti esercitano un'influenza maggiore sugli altri.

3.1.2 Funzionamento di una NN

Elaborazione delle informazioni nella rete neurale

L'elaborazione delle informazioni nella rete neurale segue sempre la stessa procedura: le informazioni sotto forma di modelli o segnali sono trasferite ai neuroni dello strato di ingresso, dove sono elaborate. A ogni neurone è assegnato un peso, in modo che i neuroni ricevano un'importanza diversa. Il peso, insieme a una funzione di trasferimento, determina l'ingresso, dove quindi il neurone è inoltrato.

Nella fase successiva, una funzione di attivazione e un valore di soglia calcolano e ponderano il valore di uscita del neurone. A seconda della valutazione delle

informazioni e della ponderazione, altri neuroni sono collegati e attivati in misura maggiore o minore.

Per mezzo di questi processi è modellato un algoritmo che produce un risultato per ogni ingresso. Ad ogni iterazione dell'addestramento, la ponderazione e quindi l'algoritmo vengono modificati in modo che la rete fornisca risultati sempre più precisi e migliori.

Training

Affinché le connessioni nelle reti neurali artificiali siano stabilite correttamente per risolvere il problema, le reti devono prima essere addestrate. A tal proposito, si possono distinguere due metodi fondamentali:

Apprendimento supervisionato Durante l'apprendimento supervisionato è definito un risultato concreto per diverse opzioni di input. Se, ad esempio, le immagini dei gatti devono essere riconosciute come tali dal sistema, le persone controllano il riconoscimento del sistema e danno un feedback su quale immagine è stata riconosciuta correttamente e quale in modo errato. In questo modo si modificano i pesi nella rete e si ottimizza l'algoritmo.

Apprendimento non supervisionato Nell'apprendimento non supervisionato il risultato dell'attività non è specificato. Il sistema impara esclusivamente sulla base delle informazioni in ingresso. A questo scopo sono utilizzate la *Regola di apprendimento di Hebb* o la *Teoria della risonanza adattiva*.

Un esempio di applicazione delle reti neurali

Le reti neurali possono essere utilizzate per il riconoscimento delle immagini. A differenza degli esseri umani, un computer non è in grado di riconoscere se un'immagine raffigura un essere umano, una pianta o un oggetto: deve esaminare le caratteristiche individuali della figura. Il computer sa quali caratteristiche sono rilevanti grazie all'algoritmo implementato oppure può scoprirlo da solo attraverso l'analisi dei dati.

In ogni strato della rete il sistema controlla i segnali di ingresso, cioè le immagini, in base a criteri individuali come il colore, gli angoli e le forme. Dopo ogni controllo, il computer può valutare meglio ciò che è rappresentato nell'immagine.

Inizialmente, i risultati saranno relativamente soggetti ad errori. Se la rete neurale non riceve un feedback da un addestratore umano ed è in grado di modificare l'algoritmo, si parla di apprendimento automatico. Nel deep learning l'addestramento umano può essere omesso: in questo caso il sistema impara dalla propria esperienza e diventa tanto migliore quanto più materiale visivo ha a disposizione.

Idealmente, il risultato finale è un algoritmo in grado di identificare il contenuto delle immagini senza errori, indipendentemente dal fatto che queste immagini siano in bianco e nero o in una determinata posa o prospettiva.

3.1.3 Applicazione di una NN alla fisica delle particelle

Gli esperimenti di fisica delle alte energie (*High Energy Physics*, HEP) utilizzano spesso l'apprendimento automatico per imparare complicate funzioni inverse, cercando di dedurre qualcosa sul processo fisico studiato a partire dalle informazioni misurate nel rivelatore. Lo schema di funzionamento è mostrato in Figura 3.2. Il deep learning viene spesso applicato a funzionalità di alto livello derivate da dati di fisica delle particelle: questo procedimento può migliorare i metodi classici di analisi ma non utilizza tutto il potenziale del deep learning, che risulta più efficace quando si opera su informazioni di livello inferiore.

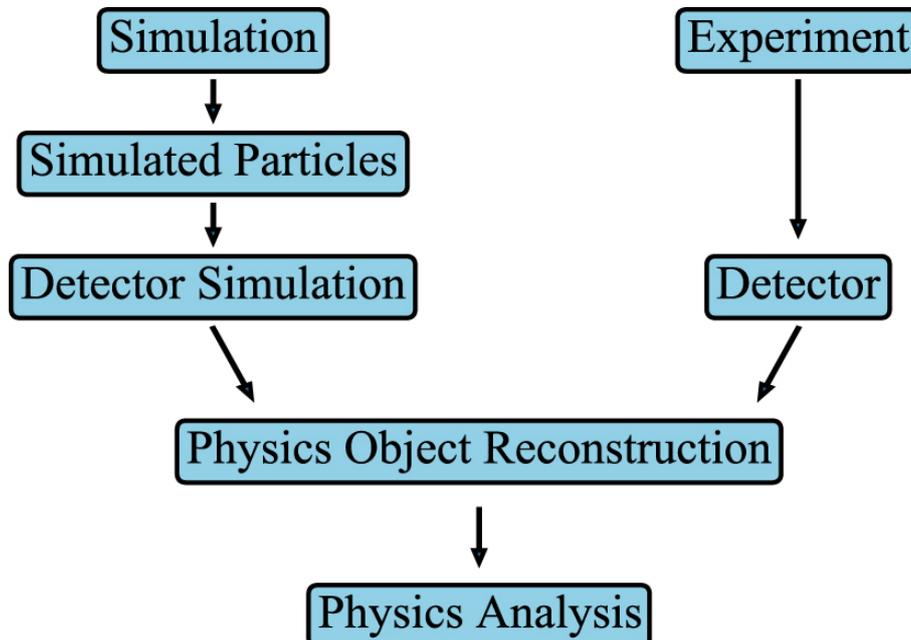


Figura 3.2: La simulazione è usata negli esperimenti di HEP per creare una registrazione veritiera dell'evento fisico che ha causato una determinata risposta del rivelatore. Questa registrazione viene utilizzata per allenare algoritmi di apprendimento per invertire la simulazione del rivelatore e dedurre qualcosa sulla fisica dei dati osservati. Questi algoritmi vengono quindi applicati ai dati reali che sono stati misurati dal rivelatore.

Alcuni dati nella fisica delle particelle possono essere interpretati in modo frazionario come immagini per poi essere sottoposti a tecniche di visione artificiale, tuttavia la rappresentazione di immagini deve affrontare alcune limitazioni principalmente per la geometria irregolare dei rivelatori. A causa della perdita intrinseca di informazioni, le rappresentazioni delle immagini possono limitare la quantità di dati estratta dagli esperimenti. I dati sono insiemi (raccolte non ordinate) di elementi: se in un insieme si considera la relazione tra i suoi elementi (geometrici o fisici), questo, con l'aggiunta di appropriate introduzioni matematiche, può trasformarsi in un grafico. Su questa rappresentazione intrinseca dei dati vi è a priori

meno limitazione nell'applicazione del deep learning. Viene illustrata una varietà di dati HEP e la loro formulazione come grafici nella Figura 3.3.

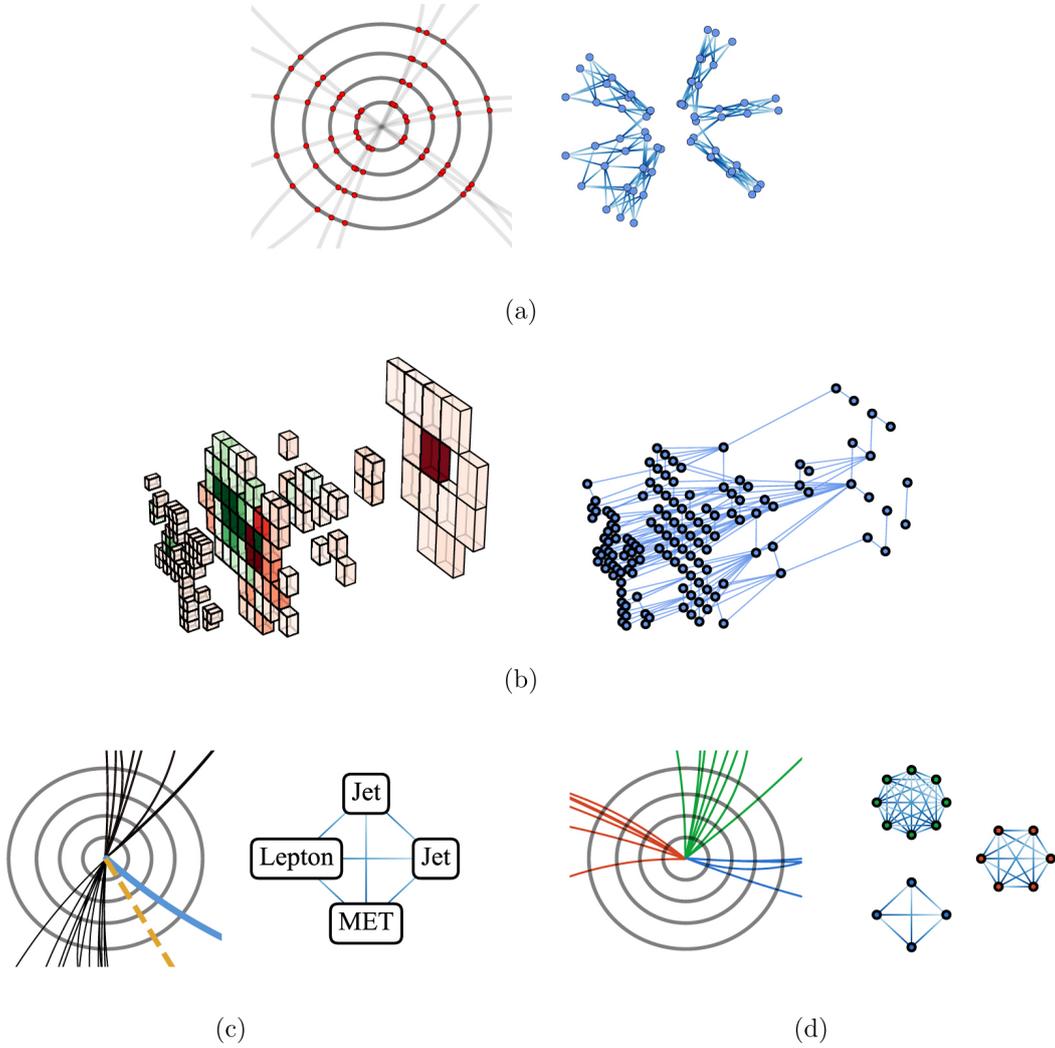


Figura 3.3: I dati di HEP si prestano ad essere rappresentati come grafici in molte applicazioni: (a)insiemi di rilevazioni convertiti in tracce, (b)segmentazione di celle calorimetriche, (c)classificazione di eventi con diversi tipi di oggetti fisici, (d)classificazione dei jets basata sulle particelle associate al singolo jet.

Ad oggi, la maggior parte degli algoritmi di apprendimento basati su grafici in HEP si concentra sui compiti di ricostruzione ed identificazione. Nonostante siano obiettivi diversi, i corrispondenti flussi di lavoro basati su *GNN* (*Graph Neural Network*) hanno somiglianze strutturali: si inizia con un insieme non ordinato di dati, ad esempio posizioni spaziali o caratteristiche cinematiche di particelle, in seguito si procede con la costruzione del grafico che incorpora i vari punti come fossero dei nodi legandoli tra di loro per rappresentare informazioni relazionali.

Questo processo può verificarsi in modo esplicito prima dell'attivazione della rete neurale o altrimenti essere ripetuto dinamicamente come parte dell'algoritmo di apprendimento [12, 13].

Di seguito alcune importanti tipologie di applicazioni di una rete neurale ad esperimenti di HEP:

Classificazione dei jets Un jet è uno stretto cono di adroni o altre particelle prodotte dall'adronizzazione di un quark o di un gluone in un esperimento di fisica delle particelle. Le particelle che trasportano una carica di colore, come i quark, non possono esistere in forma libera a causa del confinamento della cromodinamica quantistica (QCD) che consente solo stati incolore. Quando un oggetto contenente una carica di colore si frammenta, ogni frammento trasporta con sé parte della carica di colore: per rispettare il confinamento, questi frammenti creano altri oggetti colorati intorno a loro per formare oggetti incolore. L'insieme di questi oggetti è chiamato jet poiché i frammenti tendono tutti a viaggiare nella stessa direzione formando uno stretto jet di particelle. I jets vengono misurati nei rivelatori e studiati per determinare le proprietà dell'oggetto originale da cui sono stati generati. La definizione di un jet include un algoritmo e uno schema di ricombinazione: il primo definisce come alcuni input, ad esempio particelle, sono raggruppati in jets mentre il secondo specifica come viene assegnata la quantità di moto ad un determinato jet.

Negli esperimenti di fisica delle particelle, i jets sono solitamente individuati a partire dalle tracce di energia che lasciano nel calorimetro del rivelatore: quando si studiano processi simulati, i jets calorimetrici possono essere ricostruiti sulla base di una risposta simulata del rivelatore oppure direttamente da particelle stabili che emergono dai processi di frammentazione. Un buon algoritmo solitamente consente di ottenere insiemi simili di jets a diversi livelli nell'evoluzione dell'evento. Tipici algoritmi di ricostruzione dei jets sono l'*algoritmo anti- k_T* , l'*algoritmo k_T* , l'*algoritmo del cono*. Un tipico schema di ricombinazione è lo *schema E* , o *schema a quadrivettori*, in cui il quadrivettore di un jet è definito come la somma dei quadrivettori di tutti i suoi costituenti. Una nota a parte meritano i *b-jets* ossia i jets generati da quark bottom. A causa della vita media piuttosto lunga dei b-adroni e quindi di un vertice di decadimento significativamente spostato rispetto al vertice di produzione, l'identificazione di b-jets (comunemente chiamata *b-tagging*) utilizzando i metodi classici ha avuto un discreto successo. Con l'avvento dei metodi di deep learning, le informazioni di livello inferiore sono state utilizzate per migliorare le prestazioni del b-tagging e hanno aperto la possibilità di identificare i jets provenienti da altre particelle (c-adroni, top-quark, tau, etc.). I jets provenienti dall'interazione adronica pura (interazione forte), coprono uno spazio

delle fasi estremamente ampio e costituiscono un fondo irriducibile per altre classi di jets.

Classificazione degli eventi Si utilizza il termine evento in riferimento all'acquisizione da parte di un esperimento dell'intera storia di un processo fisico. In astrofisica, ad esempio, corrisponde all'insieme di segnali che descrive l'interazione di una particella ad alta energia con l'atmosfera. L'attività di etichettatura dei jets presentata nella sezione precedente fa parte di un'identificazione completa degli eventi nella fisica dei collisionatori. La classificazione degli eventi è il compito di prevedere o dedurre il processo fisico all'origine dei dati registrati.

Ricostruzione del flusso di particelle I rivelatori nella fisica delle particelle sono composti da più sotto-rivelatori con varie tecnologie di acquisizione, ognuno dei quali misura caratteristiche specifiche della particella: l'insieme di tutte le misurazioni consente la caratterizzazione delle proprietà delle particelle. La ricostruzione del *flusso di particelle*, o *flusso di energia*, è un algoritmo che mira ad assegnare ad una particella candidata tutte le misurazioni ad essa associate avvenute in ciascun sotto-rilevatore.

Parametrizzazione dell'efficienza L'analisi dei dati in fisica delle particelle richiede l'applicazione di criteri di selezione sull'ampio volume di informazioni al fine di aumentare la proporzione di segnali interessanti. È fondamentale determinare con la minor incertezza possibile la frazione di segnale che supera queste selezioni. Viene prestata molta attenzione alla determinazione di queste efficienze di selezione, poiché svolgono un ruolo significativo nella misurazione delle sezioni d'urto di processi noti e stabiliscono limiti alla produzione di segnali sconosciuti. Le efficienze possono essere misurate da dati o simulazioni e possono riguardare un intero evento oppure solo una singola componente di esso. Accade spesso che l'efficienza di una selezione specifica su una componente di un evento dipenda anche dalle altre componenti dello stesso: prendere in considerazione la correlazione tra tutti i costituenti di un evento è un compito difficile per cui l'apprendimento automatico può essere d'aiuto.

3.2 Analisi dei dati e risultati

3.2.1 Ricerca del segnale di *quark top same-sign* con ATLAS

Il vertice di interazione del processo di produzione di coppie di quark top dello stesso segno a partire da protoni è rappresentato in Figura 1.5: il grafico qui di interesse è il terzo, in cui si interpreta il processo usando un approccio di tipo EFT. La

produzione dello stato finale avviene attraverso la seguente catena di decadimenti:

$$pp \rightarrow tt \rightarrow W^+ b W^+ b \rightarrow l^+ \nu b l^+ \nu b \quad (3.1)$$

dove W^+ è il mediatore dell'interazione debole, b il quark bottom, l^+ un leptone che può essere positrone e^+ o muone μ^+ e ν il neutrino elettronico o muonico. Si ricorda che si possono avere i seguenti decadimenti: per il quark top $t \rightarrow W^+ b$ e per il W^+ o $W^+ \rightarrow l^+ \nu$ o $W^+ \rightarrow q\bar{q}$.

Dati questi decadimenti, il segnale che si cerca deve avere determinate caratteristiche, in particolare dovrà presentare: una coppia di leptoni *same-sign*, due *b-jets* (cioè jets adronici prodotti dall'adronizzazione di un quark bottom) e *MET* (*Missing Transverse Energy*). Si tratta di un segnale "pulito", cioè caratterizzato da un fondo molto limitato, costituito da eventi che potrebbero per vari motivi nascondere. Durante il funzionamento del LHC, avvengono molte collisioni tra protoni e le interazioni che si generano possono essere di tipo diverso.

La ricerca del segnale avviene attraverso quattro passaggi:

1. Simulazione del segnale e dei processi di fondo (*Monte Carlo samples*): nel caso del segnale, si assumono diverse configurazioni sugli operatori che compaiono nella lagrangiana che descrive il sistema, così da poter generare diversi possibili set di dati.
2. Selezione sul campione per rimuovere la maggior quantità di fondo possibile mantenendo un buon numero di eventi di segnale: nel caso del processo di produzione di quark top preso in esame si cercano nello specifico segnali che presentano due leptoni isolati dello stesso segno, la quantità di moto trasversa dei leptoni deve essere superiore a 20 GeV. Per quanto riguarda i jets dovranno essere in numero maggiore o uguale a due, almeno uno o due devono essere *b-jets* e la loro quantità di moto trasversa superiore a 25 GeV. A questo punto si allena una *rete neurale* affinché riesca a discriminare nel miglior modo possibile il segnale cercato dai vari fondi.
3. Applicazione della stessa selezione agli eventi reali (non simulati) e confronto con i risultati attesi dalla simulazione.
4. Nel caso in cui si noti una discrepanza dei dati raccolti con la simulazione è necessario quantificare quanto è probabile che le informazioni ricavate dagli esperimenti siano effettivamente il segnale cercato; altrimenti, se non si trovano differenze, è possibile porre limiti sui parametri della teoria. Questa procedura equivale a supporre che se in quel range di parametri il segnale non si osserva esso può trovarsi a valori più bassi (i coefficienti di Wilson hanno un estremo superiore quindi si cerca il limite massimo).

3.2.2 Scelta dei set up di analisi

L'obiettivo di questa analisi è capire come la rete neurale si comporta quando le si danno in input variabili di basso o di alto livello.

Variabili di analisi		
Prima analisi	Seconda analisi	Terza analisi
MLepMet	lep_E_0	sumPshtag
MtLepMet	lep_E_1	sumPshtag77
MtLep0Met	lep_Eta_0	lep_E_0
nJets_OR_DL1r_60	lep_Eta_1	lep_E_1
nJets_OR_DL1r_70	lep_Phi_0	lep_Eta_0
nJets_OR_DL1r_77	lep_Phi_1	lep_Eta_1
jet_pseudoscore_DL1r0	lep_Pt_0	lep_Phi_0
jet_pseudoscore_DL1r1	lep_Pt_1	lep_Phi_1
sumPshtag	met_met	lep_Pt_0
sumPshtag77	nJets_OR	lep_Pt_1
DeltaR_max_lep_bjet77	jet_eta0	met_met
DeltaR_min_lep_bjet77	jet_eta1	nJets_OR
HT_lep	jet_pt0	jet_eta0
DRll01	jet_pt1	jet_eta1
DPhill_SS		jet_pt0
DEtall_SS		jet_pt1
Ptll01		
Mll01		
met_met		
mjj_lead		
nJets_OR		
jet_pt0_nofwd		
DRjj_lead		
HT		
HT_jets		

Tabella 3.1: Tre set di variabili in input alla rete neurale nella fase di analisi

Per raggiungere questo scopo, è stato necessario allenare la rete neurale presentandole i tre set scelti per l'analisi: si è iniziato con variabili di alto livello, che corrispondono a variabili che ricostruiscono grandezze fisiche a partire dai quadrivettori delle singole particelle, per poi passare a quelle di basso livello, ossia quelle legate ai quadrimomenti dei leptoni e dei jets (nel caso in esame si sono inserite anche l'energia trasversa mancante e il numero di jets per completezza), per concludere con il terzo set che è una copia del secondo con aggiunte due variabili di b-tagging.

L'obiettivo di questa analisi è verificare il comportamento della rete neurale quando vengono utilizzati dati con livelli diversi di ricostruzione: confrontando le performance dei set di variabili di tabella 3.1, si suppone che il secondo gruppo lavori peggio del primo, perché alcune informazioni importanti per la discriminazione del segnale dal fondo non sono contenute nei quadrivettori, e che il terzo sia invece più performante del secondo, perché comprendente anche le principali variabili di b-tagging.

3.2.3 Confronti tra i vari livelli

Per ogni set sono stati elaborati i plot mostrati da Figura 3.4 a Figura 3.14.

I grafici in Figura 3.4 e in Figura 3.5 mostrano le curve di accuratezza e di perdita della rete neurale: la loro utilità consiste nel mostrare come il modello proposto lavori rispetto ai dati di validazione. Per realizzarli si plottano i dati relativi alla fase di training (memorizzati dalla rete neurale) e quelli nuovi, che sono i dati di validazione, per ogni epoca. Quando si parla di epoca, si fa riferimento al fatto che la rete neurale viene allenata con tutti i dati di training per un ciclo, quindi in un'epoca vengono utilizzati tutti i dati per una sola volta. La rete neurale in esame prevede un limite massimo di epoche pari a 125 che però generalmente non viene raggiunto in quanto l'ottimizzazione dei pesi necessita di un numero inferiore di iterazioni: nel caso dei tre set di dati presi in esame si sono raggiunte rispettivamente le epoche numero 77, 102, 94.

Il grafico di accuratezza cresce molto rapidamente per le prime 15 epoche in tutte le tre immagini di Figura 3.4: ciò sta ad indicare che la rete neurale sta imparando velocemente. Dopodiché la curva si appiattisce a mostrare che sono necessarie sempre meno epoche affinché la rete neurale risulti ben allenata. I dati di validazione, mostrati con una curva arancione, seguono l'andamento dell'allenamento e da ciò risulta un accordo significativo tra i due: se la curva dell'allenamento (in azzurro) continuasse a crescere mentre quella di validazione (in arancione) iniziasse a calare, significherebbe trovarsi nel caso cosiddetto di *overfitting* che si ha quando il modello inizia a memorizzare i vari dati.

I grafici di perdita mostrati in Figura 3.5 hanno un andamento inverso rispetto ai precedenti: la curva di training (in azzurro) decresce rapidamente per le prime

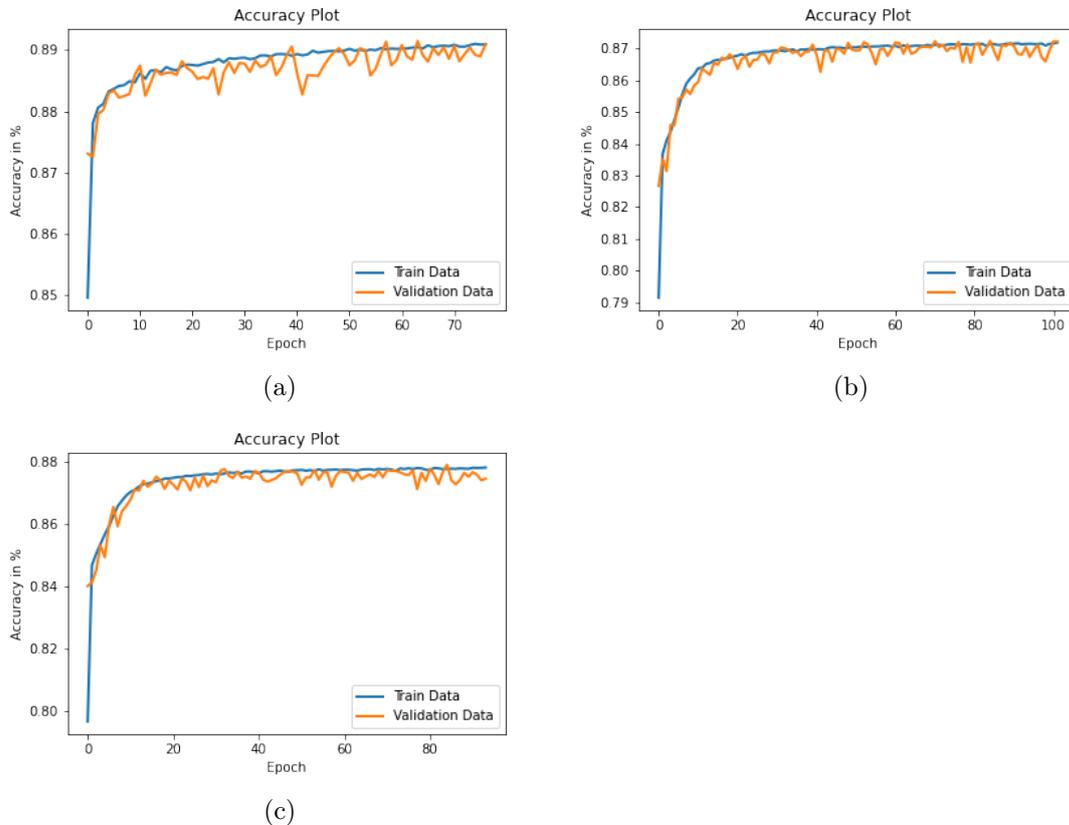


Figura 3.4: Grafici raffiguranti le curve di accuratezza: in azzurro i dati derivanti dal training e in arancione quelli legati alla validazione della rete neurale. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.

15 epoche e continua nella sua discesa ma in maniera meno consistente per le epoche successive; quella di validazione (in arancione) ha un comportamento simile ma presenta un leggero spostamento per un numero elevato di epoche. Questa discrepanza è maggiormente evidente nel grafico 3.5a relativo al primo set di variabili ma non raggiunge valori considerevoli quindi si può affermare che generalmente il modello si comporta bene nei confronti di "nuovi dati".

La rete neurale propone in output, oltre ai grafici descritti, anche valori numerici a rappresentare efficienza e perdita del modello. Si riportano di seguito quelli relativi alla penultima epoca, quindi al penultimo ciclo di lavoro della rete neurale: in ordine per i tre set di variabili si hanno valori di accuratezza e di perdita rispettivamente pari a 89% e 21% nel primo caso, 87% e 26% nel secondo e 88% e 24% nel terzo. Questi numeri mostrano una leggera differenza tra l'aver in input variabili di basso livello o di alto livello: in quest'ultimo caso, la rete neurale performa in maniera migliore. Per le variabili di basso livello, la rete performa meglio se vengono inserite le variabili relative al b-tagging.

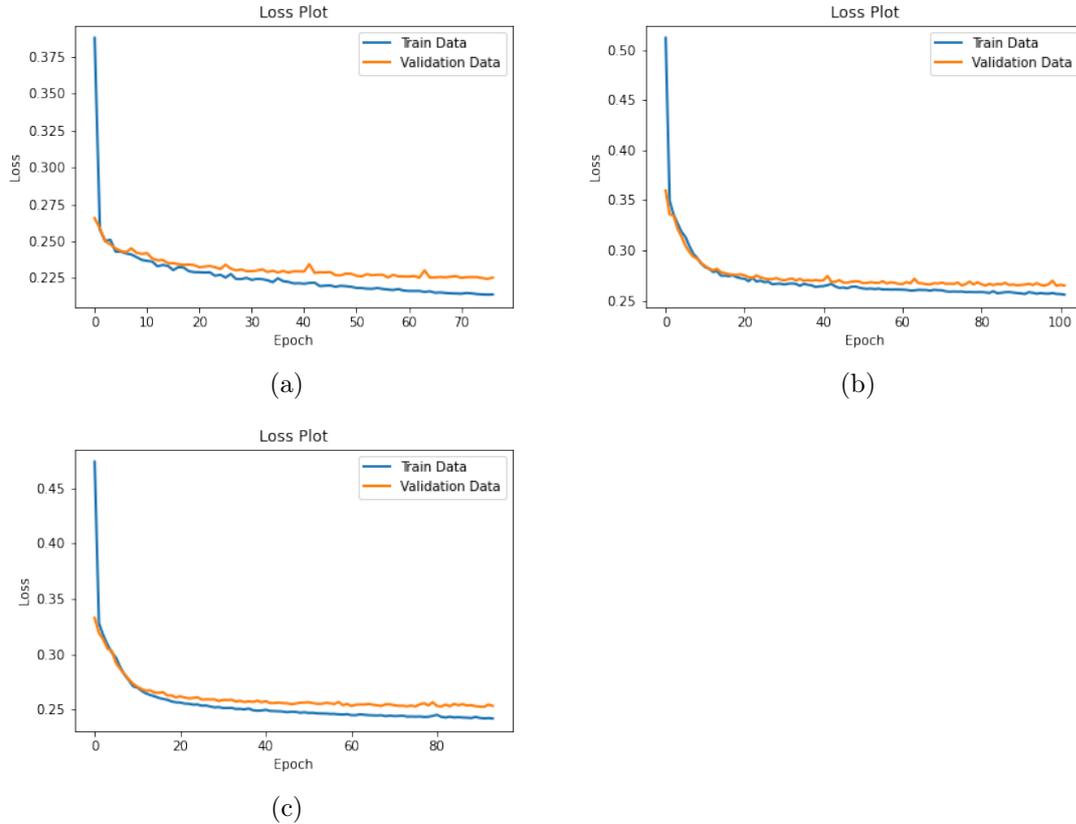


Figura 3.5: Grafici raffiguranti le curve di perdita: in azzurro i dati derivanti dal training e in arancione quelli legati alla validazione della rete neurale. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.

Dopo avere svolto la validazione della rete neurale attraverso i grafici di accuratezza e perdita per i tre diversi set di analisi, sono stati prodotti altri plot utilizzando dati di test per svolgere l'analisi. Un plot interessante è quello raffigurante la curva ROC, da cui si può ricavare il valore dell'area sottostante il grafico che è un indice della performance del modello: formalmente detta *Receiver Operating Characteristic curve* cioè Curva Caratteristica Operativa del Ricevitore, è un grafico che illustra la capacità diagnostica di un sistema di classificazione binario al variare della sua soglia di discriminazione e viene creata tracciando il tasso di veri positivi (*TPR*, *True Positive Rate*) rispetto al tasso di falsi positivi (*FPR*, *False Positive Rate*) a varie impostazioni di soglia. Si noti che questo grafico è intrinsecamente legato alla matrice di confusione, infatti ogni istanza della suddetta matrice corrisponde al risultato di una previsione e può essere rappresentato come un punto del plot. La diagonale rappresenta le uscite di un classificatore random e divide lo spazio ROC in due regioni: i punti sopra la diagonale rappresentano buoni risultati di classificazione (meglio che casuale) mentre i punti sotto la linea rappresentano

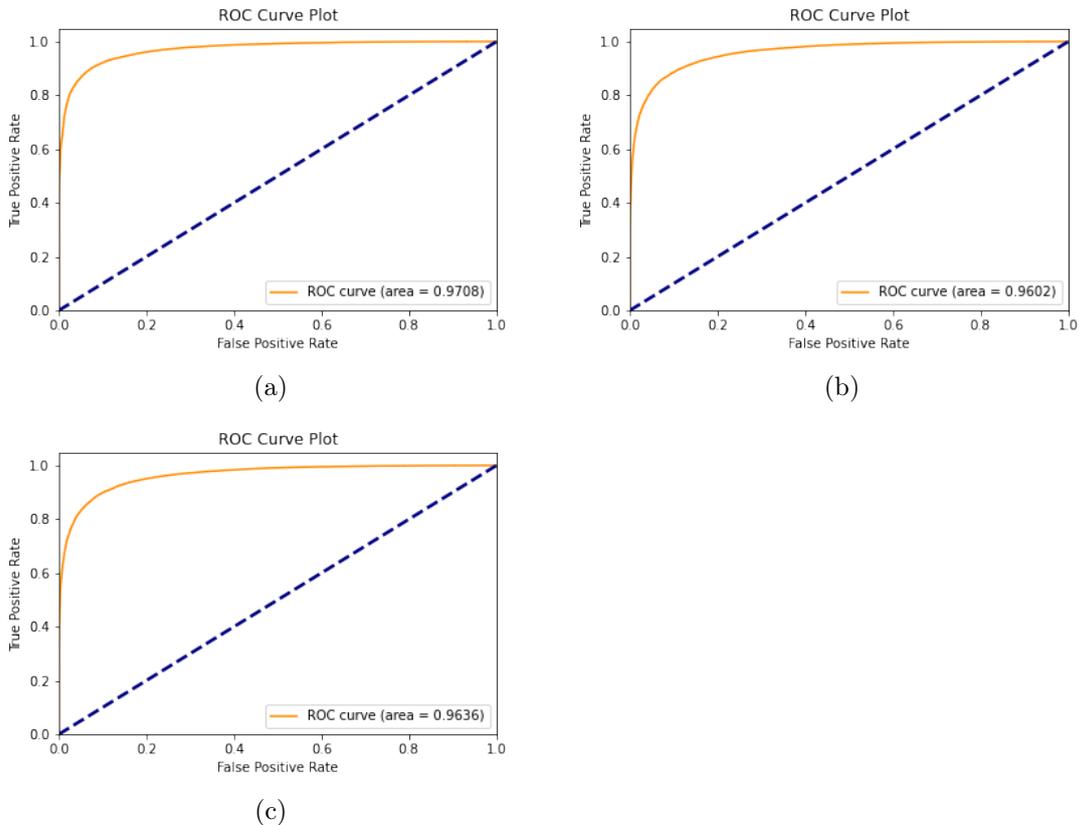


Figura 3.6: Grafici raffiguranti le curve ROC: l'area sottostante il grafico è pari a 0.9708 per (a), 0.9602 per (b) e 0.9636 per (c). I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.

risultati negativi (peggio che casuali). Il miglior metodo di previsione possibile produrrebbe un punto nell'angolo in alto a sinistra con coordinate (0, 1) nello spazio ROC, che rappresenta il 100% di sensibilità (nessun falso negativo) e il 100% di specificità (nessun falso positivo). I grafici relativi ai tre set di variabili scelti per l'analisi sono mostrati in Figura 3.6. Le aree sottostanti le curve assumono i valori 0.9708 nel primo caso, 0.9602 nel secondo e 0.9636 nel terzo: le variabili di alto livello hanno una performance migliore di quelle di basso livello mentre queste ultime risultano più performanti se vengono aggiunte le variabili di b-tagging.

In Figura 3.7 vengono mostrate le distribuzioni prodotte in output dalla rete neurale, precisamente quattro curve di colori diversi che rappresentano rispettivamente segnale e fondo del training (in verde e rosso) e segnale e fondo dell'analisi (in blu e arancione). Si ha un andamento simile per le due coppie di segnali e di fondi, a significare che il lavoro della rete neurale svolto per l'analisi dei dati non si discosta di molto da ciò che essa ha imparato nella fase di allenamento. Piccolo aspetto da sottolineare riguarda il grafico 3.7c: i valori rappresentati dalle curve

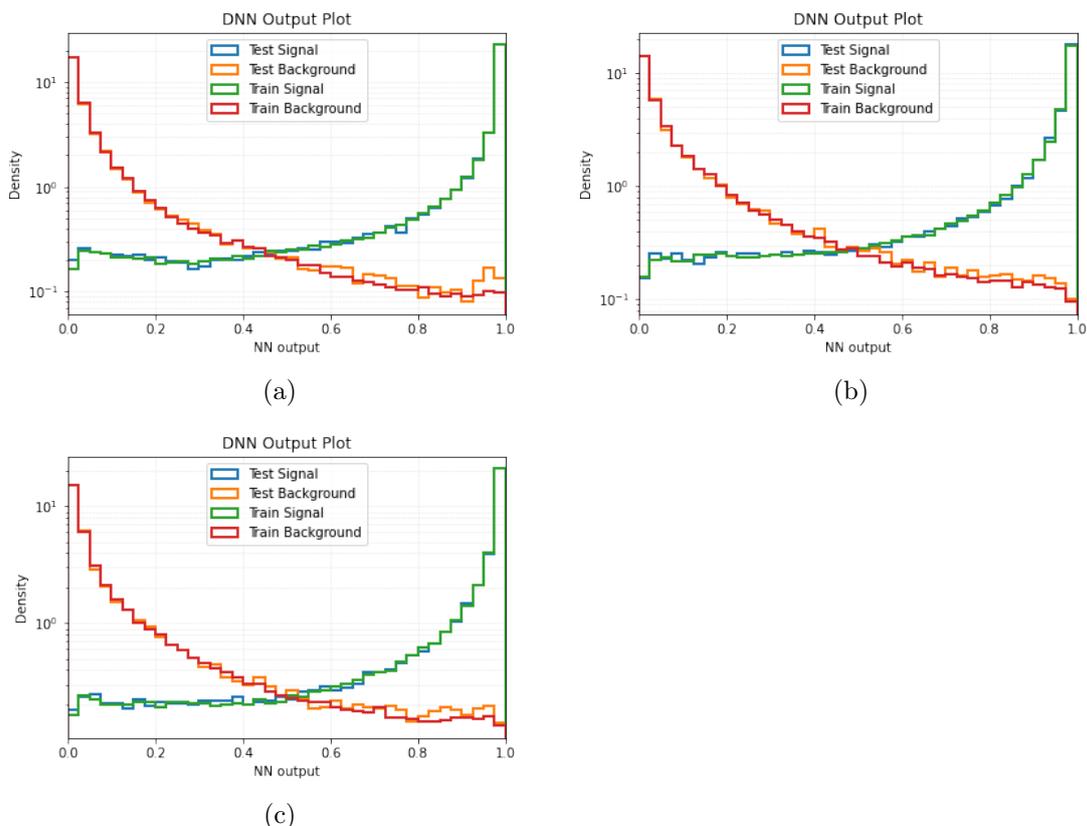


Figura 3.7: Grafici raffiguranti segnale e fondo in output dalla rete neurale sia per la fase di analisi che per quella di training: in azzurro e verde si rappresenta il segnale rivelato rispettivamente nella fase di test e di training, in arancione e in rosso il fondo registrato rispettivamente nella fase di test e di training. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.

nell'estremo destro del plot sono più alti rispetto a quelli raggiunti nelle altre due immagini: questa differenza è visibile se si osserva la suddivisione di asse y, che non necessita di rappresentare sottomultipli di 10^{-1} come invece accade per gli altri due. In riferimento al segnale, in tutti e tre i casi analizzati risulta un evidente accordo tra training e analisi.

Connesse alla curva ROC sono le matrici di confusione poiché ogni istanza di esse corrisponde ad un punto della curva. Questa matrice è una misurazione delle prestazioni per un problema di classificazione dell'apprendimento automatico in cui l'output può essere di due o più classi. Nel caso in esame le classi sono nove: il segnale e otto casi di fondo (ttZ , tty , ttW , $tttt$, ttH , $other$, $diboson$, $fakes$) che possono essere rilevati a partire dagli stessi elementi iniziali che compaiono nel processo di interesse e potrebbero quindi essere confusi con il segnale vero e proprio per vari motivi, sia perché alcuni processi del Modello Standard producono coppie di leptoni

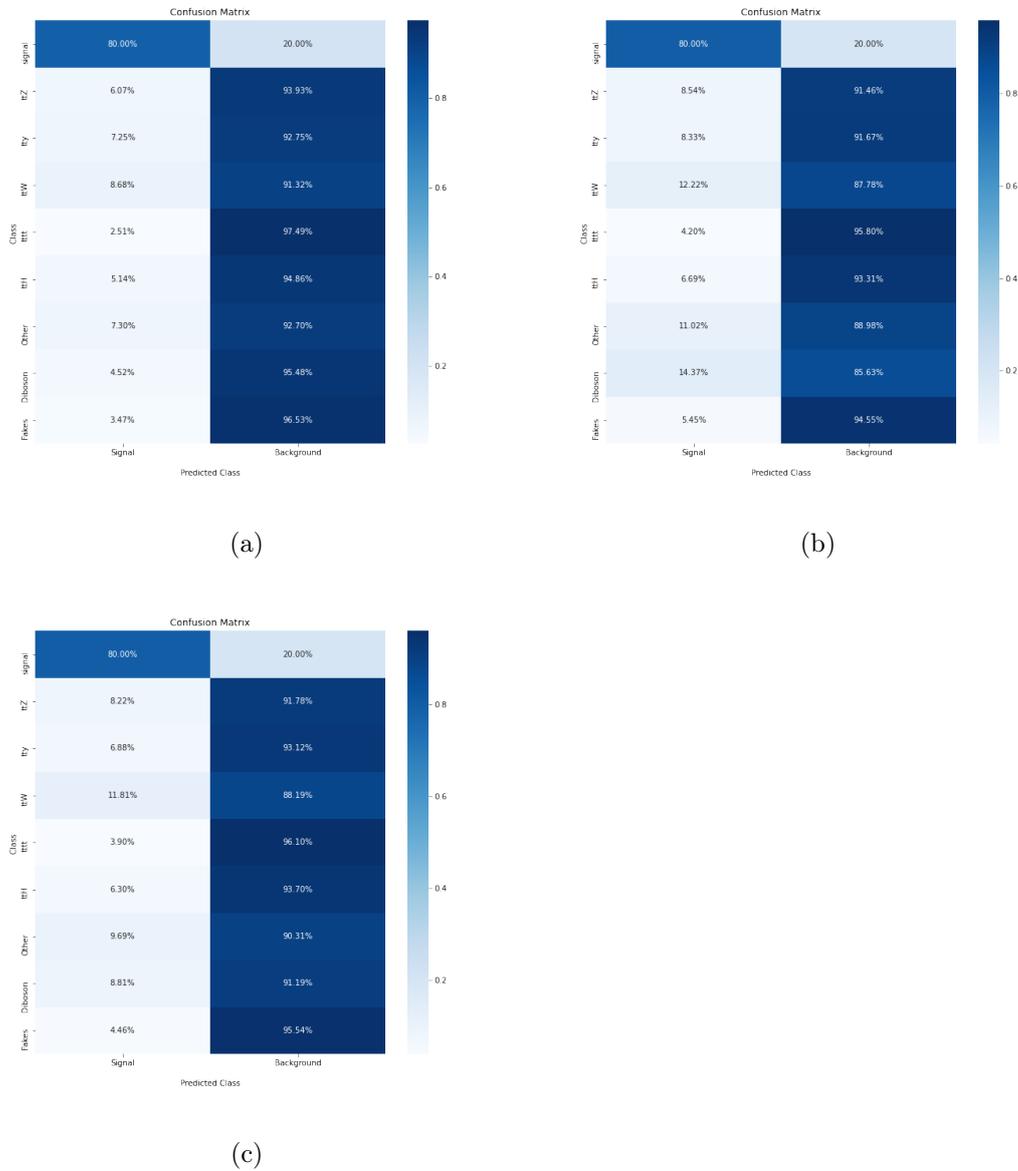


Figura 3.8: Grafici raffiguranti le matrici di confusione: in blu si rappresentano percentuali pari a 100% mentre al tendere al bianco si indicano percentuali dello 0%. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.

same-sign e b-jets e jets (come ttW, ttZ) sia perché alcuni oggetti possono essere identificati in modo errato dal rivelatore (come le categorie chiamate fakes). Ogni classe può essere rilevata e classificata come segnale o come fondo, come mostrato in Figura 3.8.

I vari tipi di fondo sono riportati insieme nel plot di Figura 3.9. La previsione è

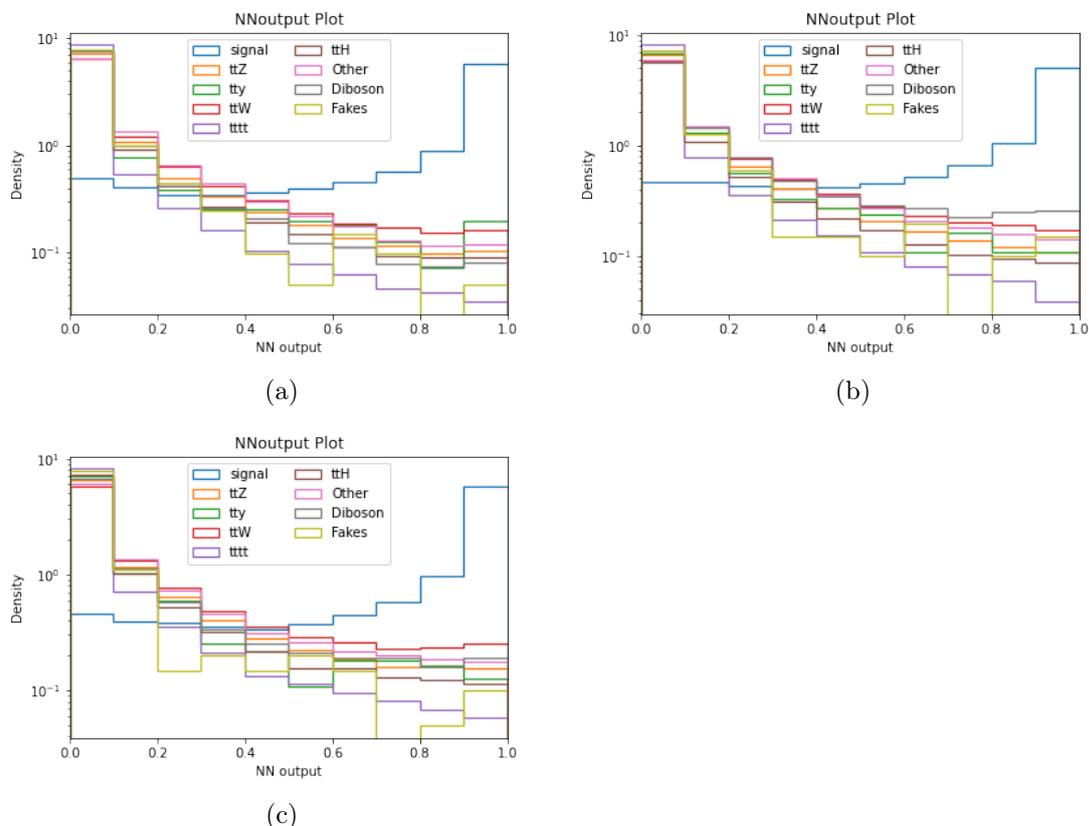


Figura 3.9: Grafici raffiguranti tutti i fondi che la rete neurale può registrare a partire dalle stesse particelle iniziali (*problema di classificazione*). I vari fondi sono presentati tramite una determinata classificazione cromatica: in azzurro il *segnale*, in arancione il fondo ttZ , in verde tty , in rosso ttW , in viola $tttt$, in marrone ttH , in violetto la classe *altro*, in grigio quella *diboson* e infine in giallo quella *fakes*. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.

che il grafico 3.9b presenti il fondo VV (categoria diboson) più alto rispetto a 3.9a mentre in 3.9c vi sia un basso fondo diboson: la stessa cosa è osservabile anche confrontando le matrici di confusione nella categoria diboson. Nei tre grafici presentati, il segnale risulta avere un andamento comparabile; i fondi si comportano in modo simile, una piccola differenza si nota nelle ultime due colonne degli istogrammi: nel secondo e terzo caso i valori più alti vengono assunti rispettivamente dal fondo diboson (in grigio) e ttW (in rosso) con i fakes leggermente più alti nel secondo grafico e bassi nel terzo. Si noti che in accordo con le previsioni si ha un basso fondo diboson in Figura 3.9c rispetto a Figura 3.9b.

Un altro output interessante prodotto dalla rete neurale sono i grafici di Figura 3.10. Si tratta di un plot che raffigura attraverso un istogramma l'importanza delle

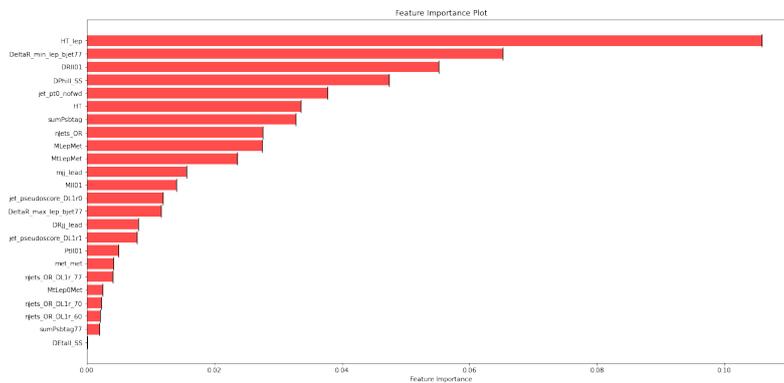
variabili in input (indicate in asse y) per un determinato modello: i punteggi associati ad ognuna rappresentano la loro "importanza", un punteggio più alto significa che la variabile specifica avrà un effetto maggiore nella discriminazione tra segnale e fondo sul modello utilizzato. Nel caso in esame le variabili principali tra quelle proposte per l'alto livello sono HT_{lep} , $\Delta R_{min_{lep_{bjet}}}$ e DR_{ll01} ; tra le variabili strettamente di basso livello sono più influenti lep_{Phi_1} , lep_{Phi_0} e $njets_{OR}$ che rimangono nelle stesse posizioni di classifica anche dopo l'introduzione delle due variabili di b-tagging. Queste ultime cambiano gli output della rete neurale, come si è mostrato nei plot precedenti, ma non in maniera così sensibile: la più influente risulta essere $sumPshtag$ che appare quinta nel grafico di importanza.

In Figura 3.11 e in Figura 3.12 si propone l'andamento delle variabili più importanti per il modello analizzato: in azzurro viene mostrato il comportamento della variabile se registrata come segnale mentre in arancione la curva che rappresenta il fondo.

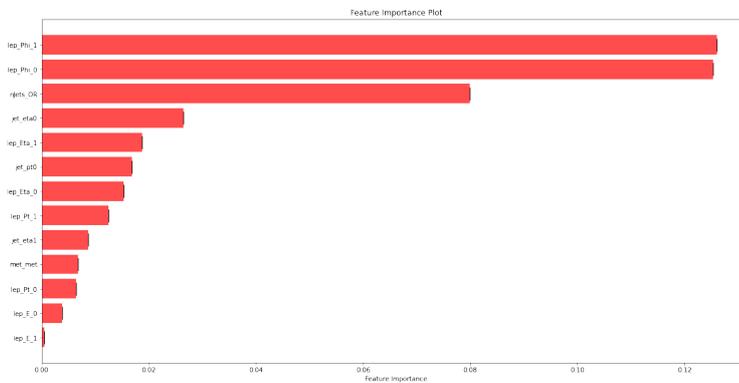
Si noti come sia evidente dai grafici in Figura 3.11 il motivo dell'importanza delle variabili mostrate: le curve di segnale e fondo sono chiaramente diverse ed è quindi logico che la rete neurale le consideri come variabili fortemente discriminanti nell'analisi. Nel caso di Figura 3.12 invece, la discrepanza non è così consistente, specialmente in riferimento a lep_{Phi_1} e lep_{Phi_0} . Il motivo di andamenti così simili per segnale e fondo risiede nella capacità della rete neurale di capire quali variabili (definite a partire da quelle in input) siano importanti ed eventualmente di ricostruirsele sulla base delle informazioni ricevute. Un esempio pratico di questo concetto potrebbe essere un rivelatore che ha come informazione le tracce di due particelle: a partire da queste potrebbe riuscire a ricavarci l'angolo che le separa e capire che si tratta di una informazione rilevante per l'analisi. È ciò che succede per i grafici di Figura 3.12a e Figura 3.12b: la somiglianza delle curve è data dal fatto che la rete neurale capisca l'importanza delle variabili, nel caso specifico della distanza ΔR , ma non le abbia a disposizione subito e le debba quindi ricostruire. A conferma di ciò nel primo set di dati le variabili legate alla distanza (DR_{ll01} e $DPhi_{ll_{SS}}$) occupano il terzo e il quarto posto nella classifica di importanza.

Infine, è stato scelto di ottenere tra gli output della rete neurale anche un grafico che rappresenti la correlazione tra le singole variabili fornite in input, di modo che siano chiare le relazioni tra di esse. Una matrice di correlazione è semplicemente una tabella che mostra i coefficienti di correlazione per diverse variabili: la matrice rappresenta la correlazione tra tutte le possibili coppie di valori ed è un potente strumento per riassumere un grande set di dati e per identificare e visualizzare i modelli nei dati forniti; è composta da righe e colonne che mostrano le variabili, la cui intersezione è rappresentata da una cella di una tabella che contiene il coefficiente di correlazione. Nel caso in esame si mostra il grado di correlazione delle variabili attraverso una scala cromatica che parte da un giallo intenso, ad indicare la massima correlazione avente coefficiente 1.0, passando per un verde e un azzurro fino a giungere ad un blu che indica la minima correlazione possibile con coefficiente

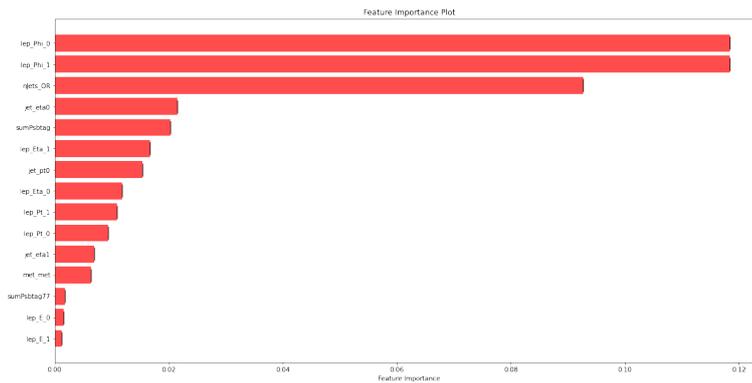
0.0. Naturalmente le celle poste sulla diagonale principale sono tutte gialle poichè rappresentano l'intersezione tra due variabili che di fatto sono la stessa.



(a)

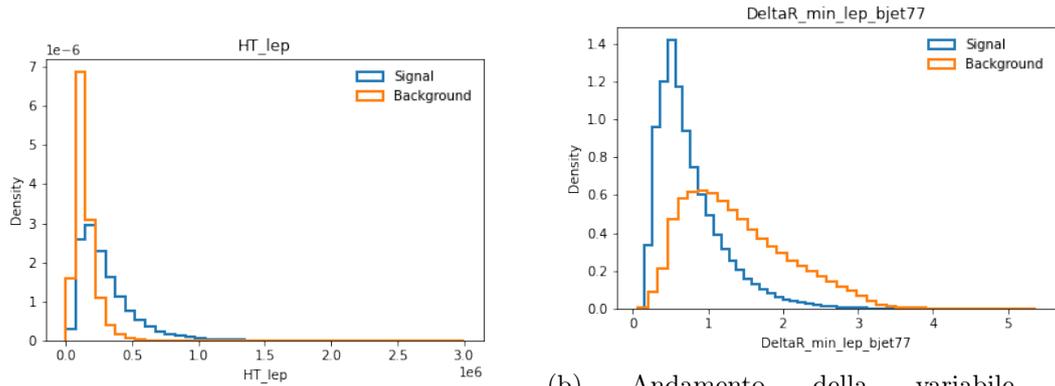


(b)



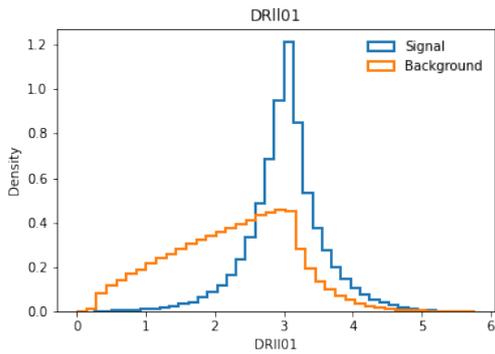
(c)

Figura 3.10: Istogrammi raffiguranti l'importanza delle variabili in input (indicate in asse y) per il modello studiato. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.



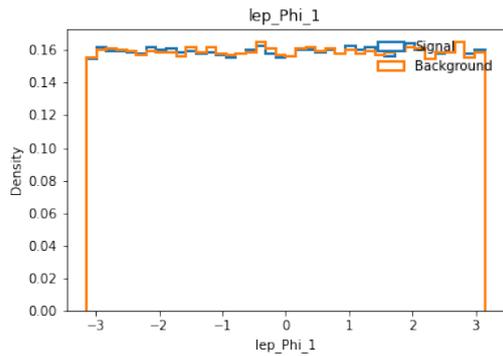
(a) Andamento della variabile HT_{lep} .

(b) Andamento della variabile $\Delta R_{min_lep_bjet77}$.

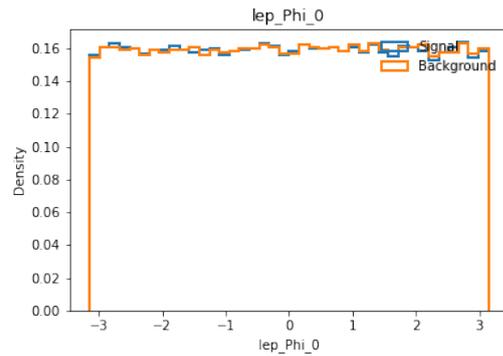


(c) Andamento della variabile $DRll01$.

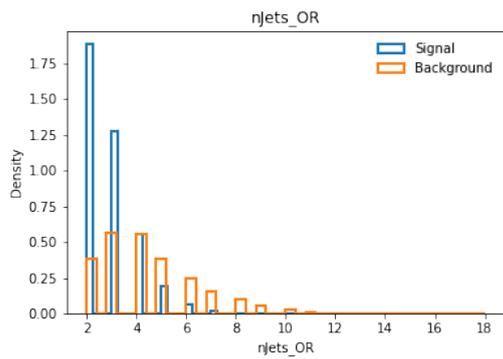
Figura 3.11: Andamenti singoli delle tre variabili identificate come più importanti dalla rete neurale, ossia le prime tre che appaiono nel grafico di importanza proposto in Figura 3.10: HT_{lep} , $\Delta R_{min_lep_bjet77}$ e $DRll01$. Si fa riferimento in particolare al plot di Figura 3.10a, sviluppato a partire da variabili di input di alto livello. In azzurro si ha il comportamento della variabile per il segnale, in arancione quello della stessa ma per il fondo.



(a) Andamento della variabile lep_Phi_1 .



(b) Andamento della variabile lep_Phi_0 .



(c) Andamento della variabile $njets_OR$.

Figura 3.12: Andamenti singoli delle tre variabili identificate come più importanti dalla rete neurale, ossia le prime tre che appaiono nel grafico di importanza proposto in Figura 3.10: lep_Phi_1 , lep_Phi_0 e $njets_OR$. Si fa riferimento in particolare ai plot di Figura 3.10b e di Figura 3.10c, sviluppati a partire da variabili di input rispettivamente di basso livello e di basso livello con l'aggiunta delle due variabili necessarie al b-tagging ($sumPshtag$ e $sumPshtag\gamma\gamma$). In azzurro si ha il comportamento della variabile per il segnale, in arancione quello della stessa ma per il fondo.

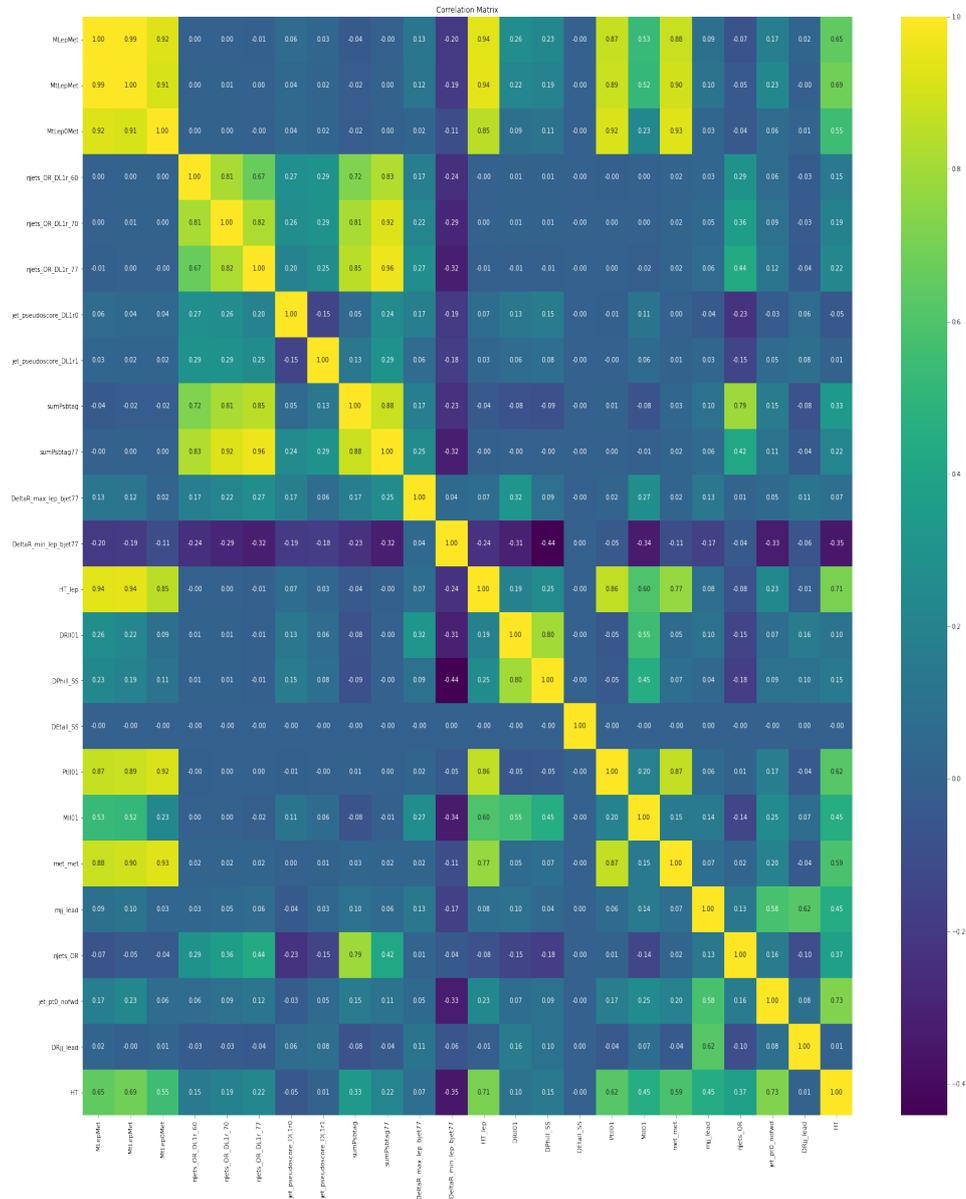


Figura 3.13: Grafico raffigurante la matrice di correlazione: la scala cromatica utilizzata indica per il giallo valori del coefficiente di correlazione pari a 1.0 e per il blu pari a 0.0, le cifre intermedie sono rese attraverso una progressiva scala di verde e azzurro. Il plot rappresenta il set di variabili di alto livello cioè la prima colonna di tabella 3.1.

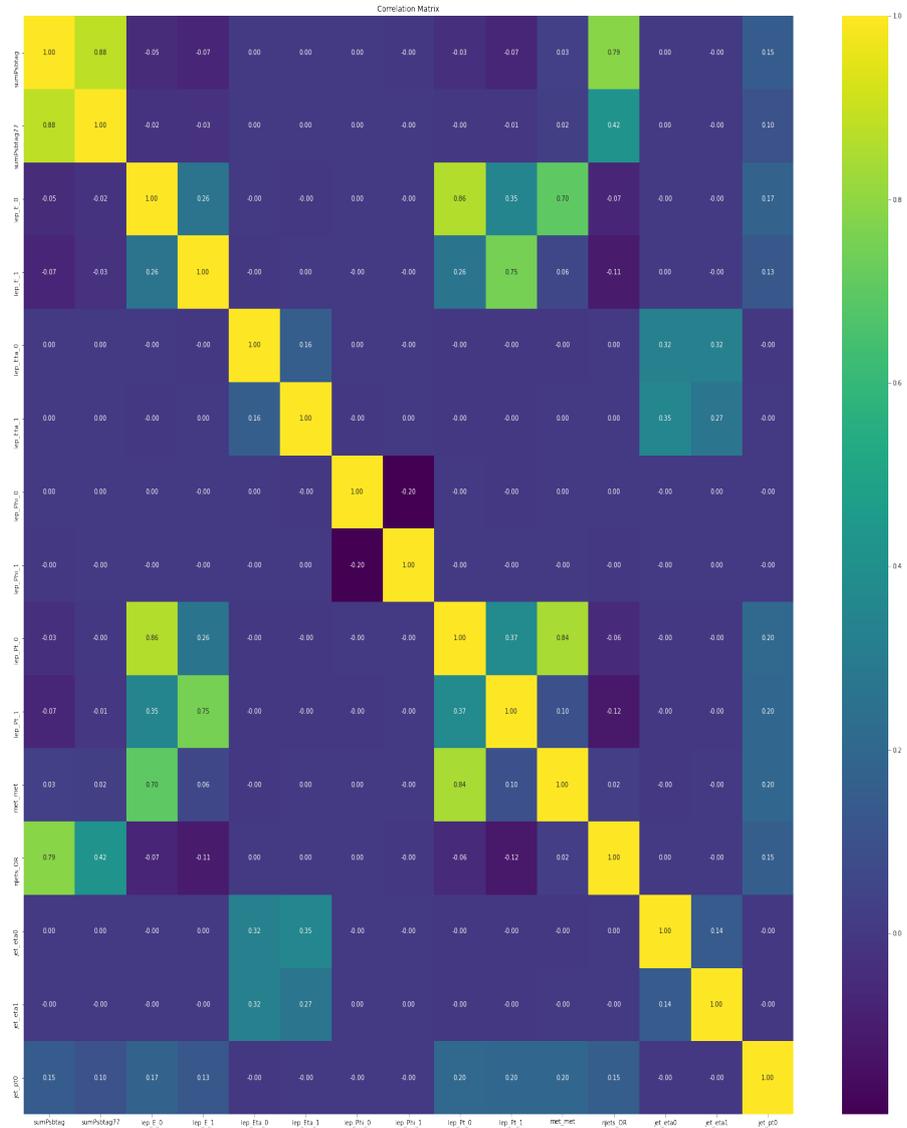


Figura 3.14: Grafico raffigurante la matrice di correlazione: la scala cromatica utilizzata indica per il giallo valori del coefficiente di correlazione pari a 1.0 e per il blu pari a 0.0, le cifre intermedie sono rese attraverso una progressiva scala di verde e azzurro. Il plot rappresenta il set di variabili di basso livello unito a due delle variabili di b-tagging, cioè la seconda e la terza colonna di tabella 3.1.

Conclusioni

In questo elaborato è stata presentata un'analisi sulle prestazioni di una rete neurale che riceve in input variabili di diverso livello. La rete neurale ha come scopo la discriminazione del segnale associato a coppie di quark top *same-sign* dal fondo.

Utilizzando una rete neurale ottimizzata per la discriminazione del segnale dal fondo, le si sono presentati tre set di variabili per l'allenamento rispettivamente di alto livello, basso livello e copia del secondo caso con due variabili di *b*-tagging. Studiando i grafici prodotti in output dalla rete neurale è emerso che la performance nei tre casi risulta essere molto simile: precisamente le aree sottostanti le curve ROC assumono rispettivamente i valori 0.9708, 0.9602 e 0.9636. Questo significa che fornire alla rete neurale variabili di alto o di basso livello non altera le prestazioni in termini di capacità discriminante.

Si è anche dimostrato che la rete neurale sia in grado di riconoscere autonomamente le variabili più importanti ed eventualmente di ricostruirle. Un esempio di ciò è il caso delle variabili di basso livello lep_Phi_1 e lep_Phi_0 che presentano un andamento praticamente identico per segnale e fondo: queste due variabili vengono classificate come importanti dalla rete neurale nonostante le curve non presentino alcuna differenza, in quanto l'algoritmo è capace di sfruttare le correlazioni tra variabili per discriminare segnale e fondo. Questo studio pone le basi per l'utilizzo di una rete neurale con variabili di basso livello per la ricerca di eventi top-top a LHC in un approccio EFT.

Bibliografia

- [1] CERN. Physics: The Standard Model.
- [2] UNIBO. Phenomenology of fundamental interactions.
- [3] O. M. Ozsimek, V. Ari, and O. Cakir. Studying same-sign top pair production in flavor changing scalar models at the hl-lhc. 5 2021.
- [4] F. Larios and F. Penunuri. Flavour changing neutral current production of same sign top quark pairs at the LHC. *J. Phys. G*, 30:895–904, 2004.
- [5] CERN. CERN’s accelerator complex.
- [6] CERN. The Large Hadron Collider.
- [7] Enciclopedia Treccani. LHC (Large Hadron Collider).
- [8] ATLAS. L’esperimento ATLAS. Technical report, CERN, 2010.
- [9] ATLAS. Detector and technology.
- [10] CERN. LHC: The guide. Technical report, CERN, 2021.
- [11] ATLAS. Trigger and Data Acquisition.
- [12] Peter Battaglia Jonathan Shlomi and Jean-Roch Vlimant. Machine learning: Science and technology. 2(2), 29 December 2020.
- [13] Savannah Thais, Paolo Calafiura, Grigorios Chachamis, Gage DeZoort, Javier Duarte, Sanmay Ganguly, Michael Kagan, Daniel Murnane, Mark S. Neubauer, and Kazuhiro Terao. Graph neural networks in particle physics: implementations, innovations, and challenges, 2022.

Elenco delle figure

1.1	Suddivisione delle particelle e antiparticelle del Modello Standard. In viola quark e antiquark, in verde leptoni e antileptoni, in rosso mediatori di gauge delle interazioni e in giallo bosone di Higgs, unico mediatore scalare fino ad oggi conosciuto.	4
1.2	Ipotesi alla base delle EFT (Effective Field Theories).	9
1.3	Esempio di applicazione delle EFT al decadimento beta. A sinistra si ha la rappresentazione del fenomeno attraverso teorie di campo effettive, a destra si trova la raffigurazione basata sulla teoria elettrodebole.	10
1.4	Alcuni dei possibili operatori di dimensione 6 che completano la lagrangiana L_{SMEFT} . I raggruppamenti colorati descritti dalle sigle dello stesso colore rappresentano i principali processi che questi operatori descrivono, la discussione di questi processi esula dall'obiettivo di questo elaborato.	12
1.5	Confronto tra le rappresentazioni del processo $pp \rightarrow tt$ di produzione di coppie di quark top dello stesso segno proposte da diversi modelli teorici. Da sinistra: Modello Standard, teorie BSM e EFT.	14
2.1	Rappresentazione del complesso di acceleratori che compongono il CERN, comprensiva anche dei principali esperimenti attualmente in atto sui vari anelli.	16
2.2	Schema del rivelatore ATLAS completo dei quattro componenti fondamentali: rivelatore interno, calorimetro, spettrometro a muoni e sistema di magneti superconduttori.	21
2.3	Rappresentazione del passaggio di particelle diverse attraverso le varie sezioni del detector ATLAS: fotoni ed elettroni vengono fermati dall'ECAL, protoni e neutroni si arrestano nell'HCAL mentre muoni e neutrini riescono a sfuggire anche allo spettrometro a muoni.	22
3.1	Rappresentazione schematica della struttura di una rete neurale: in verde il livello di input, in arancione i livelli intermedi e in azzurro il livello di output.	26

3.2	La simulazione è usata negli esperimenti di HEP per creare una registrazione veritiera dell'evento fisico che ha causato una determinata risposta del rivelatore. Questa registrazione viene utilizzata per allenare algoritmi di apprendimento per invertire la simulazione del rivelatore e dedurre qualcosa sulla fisica dei dati osservati. Questi algoritmi vengono quindi applicati ai dati reali che sono stati misurati dal rivelatore.	29
3.3	I dati di HEP si prestano ad essere rappresentati come grafici in molte applicazioni: (a)insiemi di rilevazioni convertiti in tracce, (b)segmentazione di celle calorimetriche, (c)classificazione di eventi con diversi tipi di oggetti fisici, (d)classificazione dei jets basata sulle particelle associate al singolo jet.	30
3.4	Grafici raffiguranti le curve di accuratezza: in azzurro i dati derivanti dal training e in arancione quelli legati alla validazione della rete neurale. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.	36
3.5	Grafici raffiguranti le curve di perdita: in azzurro i dati derivanti dal training e in arancione quelli legati alla validazione della rete neurale. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.	37
3.6	Grafici raffiguranti le curve ROC: l'area sottostante il grafico è pari a 0.9708 per (a), 0.9602 per (b) e 0.9636 per (c). I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.	38
3.7	Grafici raffiguranti segnale e fondo in output dalla rete neurale sia per la fase di analisi che per quella di training: in azzurro e verde si rappresenta il segnale rivelato rispettivamente nella fase di test e di training, in arancione e in rosso il fondo registrato rispettivamente nella fase di test e di training. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.	39
3.8	Grafici raffiguranti le matrici di confusione: in blu si rappresentano percentuali pari a 100% mentre al tendere al bianco si indicano percentuali dello 0%. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.	40

3.9	Grafici raffiguranti tutti i fondi che la rete neurale può registrare a partire dalle stesse particelle iniziali (<i>problema di classificazione</i>). I vari fondi sono presentati tramite una determinata classificazione cromatica: in azzurro il <i>segnale</i> , in arancione il fondo <i>ttZ</i> , in verde <i>tty</i> , in rosso <i>ttW</i> , in viola <i>tttt</i> , in marrone <i>ttH</i> , in violetto la classe <i>altro</i> , in grigio quella <i>diboson</i> e infine in giallo quella <i>fakes</i> . I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi.	41
3.10	Istogrammi raffiguranti l'importanza delle variabili in input (indicate in asse y) per il modello studiato. I tre plot sono rappresentati rispettando l'ordine delle colonne di tabella 3.1 che costituiscono i vari set di variabili forniti in input nei tre casi studiati nell'analisi. .	44
3.11	Andamenti singoli delle tre variabili identificate come più importanti dalla rete neurale, ossia le prime tre che appaiono nel grafico di importanza proposto in Figura 3.10: <i>HT_lep</i> , <i>DeltaR_min_lep_bjet77</i> e <i>DRll01</i> . Si fa riferimento in particolare al plot di Figura 3.10a, sviluppato a partire da variabili di input di alto livello. In azzurro si ha il comportamento della variabile per il segnale, in arancione quello della stessa ma per il fondo.	45
3.12	Andamenti singoli delle tre variabili identificate come più importanti dalla rete neurale, ossia le prime tre che appaiono nel grafico di importanza proposto in Figura 3.10: <i>lep_Phi_1</i> , <i>lep_Phi_0</i> e <i>njets_OR</i> . Si fa riferimento in particolare ai plot di Figura 3.10b e di Figura 3.10c, sviluppati a partire da variabili di input rispettivamente di basso livello e di basso livello con l'aggiunta delle due variabili necessarie al b-tagging (<i>sumPshtag</i> e <i>sumPshtag77</i>). In azzurro si ha il comportamento della variabile per il segnale, in arancione quello della stessa ma per il fondo.	46
3.13	Grafico raffigurante la matrice di correlazione: la scala cromatica utilizzata indica per il giallo valori del coefficiente di correlazione pari a 1.0 e per il blu pari a 0.0, le cifre intermedie sono rese attraverso una progressiva scala di verde e azzurro. Il plot rappresenta il set di variabili di alto livello cioè la prima colonna di tabella 3.1.	47
3.14	Grafico raffigurante la matrice di correlazione: la scala cromatica utilizzata indica per il giallo valori del coefficiente di correlazione pari a 1.0 e per il blu pari a 0.0, le cifre intermedie sono rese attraverso una progressiva scala di verde e azzurro. Il plot rappresenta il set di variabili di basso livello unito a due delle variabili di b-tagging, cioè la seconda e la terza colonna di tabella 3.1.	48