UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Sviluppo di algoritmi di identificazione di Jet per l'upgrade del trigger di primo livello dell'esperimento CMS

Relatore

Prof. Jacopo Pazzini

Laureando

Raffaele D'Agostino

Anno Accademico 2023/2024

Abstract

L'esperimento CMS (Compact Muon Solenoid) è uno dei rivelatori di particelle al Large Hadron Collider (LHC) del CERN. L'esperimento ha tra i suoi obiettivi quello di esplorare la fisica nella scala energetica del TeV, approfondendo sia lo studio del Modello Standard sia la ricerca di possibile fisica oltre il Modello Standard. La futura Fase 2 dell'aggiornamento del rivelatore produrrà un miglioramento del trigger di primo livello (L1) e consentirà l'accesso alle informazioni del tracciatore per la prima volta a questo stadio, consentendo l'uso di algoritmi sofisticati per la ricostruzione dei jet adronici e la riduzione del rumore causato dalle collisioni multiple (pile-up).
L'obiettivo del progetto di tesi è studiare le simulazioni della risposta del futuro trigger di L1 e realizzare algoritmi basati su tecniche di machine learning mirati a distinguere i jet originati da b-quark dai jet provenienti dall'adronizzazione di quark "leggeri" o gluoni al fine di migliorare la capacità del detector di rilevazione di fenomeni fisici comprendenti un jet da quark b. Il dataset utilizzato proviene da simulazioni di collisioni protone-protone ad HL-LHC, e contiene tutte le informazioni che saranno disponibili a livello di L1 trigger dopo l'aggiornamento. Questi dati sono stati utilizzati per l'addestramento di una rete neurale binaria feed-forward, che distingue jet da adronizzazione di quark b da jet leggeri. I risultati ottenuti, misurati per mezzo di ROC curve, la cui area è pari a AUC=0.95, mostrano che il classificatore è in grado di distinguere efficacemente le due classi. Questo ha permesso di definire tre possibili punti di lavoro al fine di selezionare jet da b con diversa purezza, in base alla frazione di falsi positivi ammessa.
Come punto di partenza per sviluppi futuri del lavoro, è stato sviluppato in via preliminare un secondo algoritmo basato su una rete neurale convoluzionale (CNN), i cui risultati saranno discussi e confrontati con quelli prodotti dalla rete feed-forward.

Indice

Capitolo 1

Introduzione

1.1 L'esperimento CMS ad LHC

1.1.1 Il rivelatore CMS

L'esperimento Compact Muon Solenoid (CMS) al Large Hadron Collider (LHC) è uno dei principali rivelatori di particelle del CERN. [1] [2]
Il risultato più celebre e significativo ottenuto dall'esperimento CMS è la scoperta del *bosone di Higgs*: il 4 luglio 2012 venne annunciata la scoperta di una nuova particella con una massa di circa 125 GeV/c^2, compatibile con la massa prevista del *bosone di Higgs*. Questo risultato valse l'assegnazione del premio Nobel ai fisici Peter Higgs e François Englert, che ne avevano previsto le proprietà.

Analizziamo di seguito la struttura di CMS (figura 1.1) per comprenderne il meccanismo di funzionamento. Esso presenta una struttura cilindrica a strati (*layer*) disposta attorno all'asse delle collisioni dei fasci di LHC: ogni layer del rivelatore svolge un ruolo nella rivelazione di particelle.
L'intero sistema di rivelazione è stato concepito con una sezione cilindrica a barile (*barrel*) e due regioni di chiusura planari (*endcap*). Muovendoci dallo strato più interno a quello più esterno (figura 1.2), troviamo innanzitutto un tracciatore al silicio composto da due parti: un detector a pixel con oltre 120 milioni di pixel di dimensione $100 \times 150 \mu m^2$ che consente il tracciamento della traiettoria di particelle cariche con elevata precisione, anche a brevissime distanze (minori di 16 cm) dal centro della collisione dove la densità di particelle per unità di superficie è maggiore; un detector a strisce (*strip*), posizionate a partire da dove terminano gli strati di pixel, fino ad una distanza di circa 130 cm dal centro dell'interazione, regione dove la minore densità di particelle permette di usare rivelatori di dimensioni più grandi. Questo tipo di tracciatore è fondamentale nell'ambito della rivelazione dei vertici di decadimento e del flavour tagging, che verrà discusso più in dettaglio in una sezione successiva.
Successivamente troviamo due calorimetri: un calorimetro elettromagnetico (ECAL), formato da cristalli di tungstato di piombo, che ha lo scopo di misurare l'energia di fotoni ed elettroni, ed un calorimetro adronico (HCAL), caratterizzato da un'alternanza di strati passivi di ottone, e strati attivi di materiale scintillatore plastico, che rivela l'energia depositata dal passaggio di adroni.
Il cuore di CMS è rappresentato da un solenoide superconduttore dal diametro interno di circa 6 metri, in grado di mantenere acceso un campo magnetico costante di 3.8 T con la finalità di curvare la traiettoria di particelle cariche ottenute come prodotti delle collisioni protone-protone (pp) in LHC e permetterne quindi la misura della quantità di moto nel piano trasverso del rivelatore.
L'ultimo insieme di layer è composto da un alternarsi di camere a muoni e gioghi di ferro. Le camere a muoni rivelano il passaggio di particelle con lunghe vite medie e una bassa tendenza ad interagire con la materia, mentre i gioghi di ferro hanno lo scopo di stabilizzare il campo magnetico al di fuori del solenoide la cui intensità si attesta all'incirca a 2T.

Come intuibile del nome stesso dell'esperimento, la rivelazione di muoni svolge un ruolo fondamentale a CMS: essi sono spesso prodotti in decadimenti di processi fisici rari, come ad esempio nel decadimento

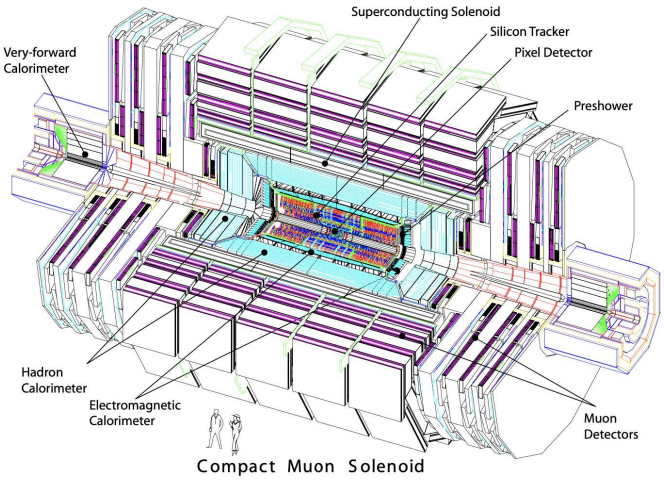

Figura 1.1: Vista longitudinale tridimensionale del rivelatore CMS. Crediti immagine [3]

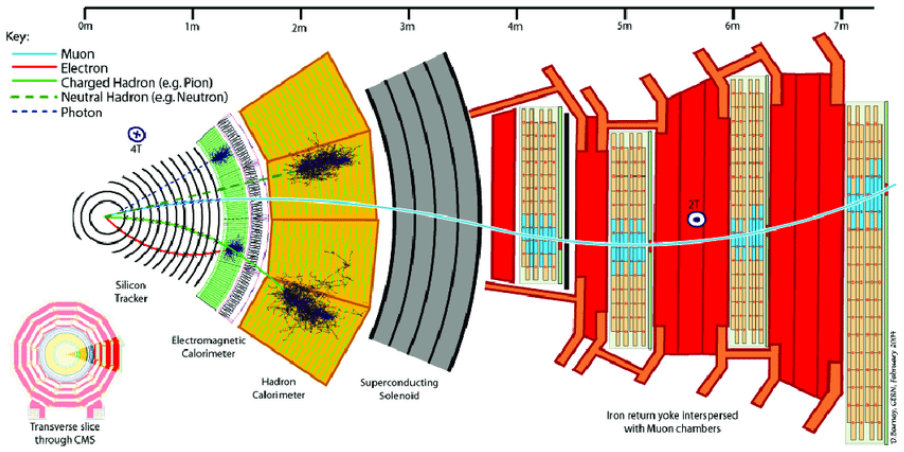

Figura 1.2: Sezione trasversale di CMS. Crediti immagine [4]

del *bosone di Higgs* $H \rightarrow ZZ \rightarrow 4l$, e sono inoltre facilmente tracciabili con elevata precisione a causa della loro bassa tendenza ad interagire con i materiali di rivelazione. Questo consente di rivelare segnali significativi anche in presenza di un elevato rumore di fondo che è generalmente presente ad LHC in condizioni di elevata luminosità.
CMS utilizza tre tipi di rivelatori a gas per l'identificazione dei muoni. Le sezioni a barile e le regioni planari delle camere a muoni presentano due diverse tecnologie per la rivelazione: la prima consiste in una una camera a tubi di deriva, per la rivelazione di muoni in un range $|\eta| < 1.2$ [1], le altre due invece consistono in camere a strisce catodiche e agiscono in un range $0.9 < |\eta| < 2.4$.
Esiste inoltre un sistema ulteriore per la rivelazione dei muoni, composto da camere a piastre resistive (RPC), sia nelle regioni del *barrel* che degli *endcap* che viene utilizzato in ridondanza con le camere a tubi di deriva e le camere a strisce catodiche.

1.1.2 Il sistema di trigger di CMS

LHC è progettato per poter produrre il passaggio dei fasci di protoni nei punti di interazione degli esperimenti con una frequenza di 40MHz (detta frequenza di *bunch crossing*). A seconda dello stato di riempimento dei fasci, si può arrivare a circa 30MHz di collisioni protone-protone ai punti di collisione, fatto che rende impossibile per i computer moderni il processing di tutti i dati provenienti

[1] La variabile η, o pseudorapidità, è legata all'angolo polare di emissione θ. Il suo significato verrà approfondito nella sezione 3.1

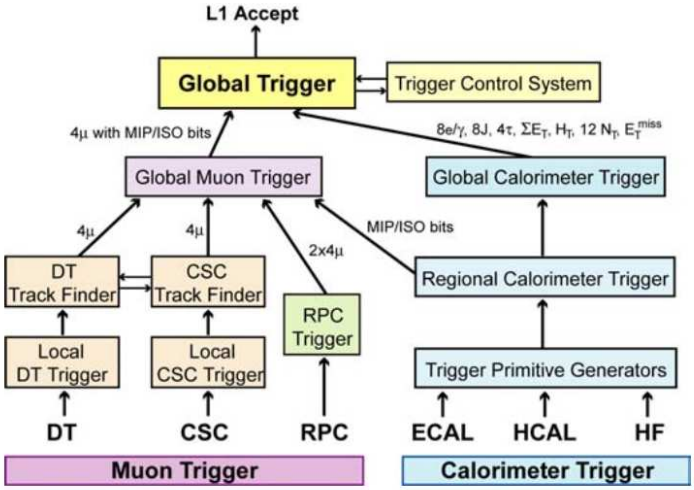

Figura 1.3: Schema del L1 trigger attuale. Crediti immagine [3]

dalle collisioni. Vi è perciò la necessità di un sistema di selezione in tempo reale (detto *trigger*) che diminuisca il rate di dati acquisiti al secondo e che permetta di raccogliere informazioni sugli eventi più significativi.

L'esperimento CMS implementa attualmente un sistema di trigger a due livelli composto dal Level-1 (L1), formato da schede e processori hardware, e dall'High Level Trigger (HLT) basato su una farm di server commerciali che svolgono la selezione basandosi sulla ricostruzione software delle collisioni. Il L1 riceve informazioni parziali dai sistemi calorimetrici e dalle camere a muoni, generando una selezione iniziale con una frequenza massima di uscita di 100 kHz. Quando viene inviato un segnale di accettazione da parte di L1, il rivelatore viene letto completamente e l'evento selezionato viene ricostruito nel HLT. La selezione HLT si basa su informazioni più dettagliate e permette di ridurre la frequenza di uscita a circa 1 kHz in media, salvando gli eventi più interessanti. [5]

Più nel dettaglio, il L1 trigger presenta componenti locali, regionali e globali. Alla base della gerarchia, i trigger locali, chiamati anche Trigger Primitive Generators (TPG), si basano sui depositi di energia nelle torri dei calorimetri e sui segmenti di traccia o le collisioni nelle camere muoniche. I trigger regionali combinano queste informazioni e ordinano gli oggetti assegnandogli un rango come candidati ad una certa tipologia di particella. Il rango di un evento viene determinato principalmente in base alla sua energia, al suo momento e alla qualità dei dati raccolti, influenzata dalla precisione dei rivelatori e dell'elettronica dei trigger utilizzati, così come dalla quantità di informazioni disponibili. I trigger globali dei calorimetro e i trigger globali muonici determinano gli oggetti calorimetrici e muonici con il rango più alto in tutto l'esperimento e li trasferiscono al trigger globale, l'entità superiore della gerarchia di L1. Quest'ultimo prende la decisione di respingere un evento o accettarlo per una valutazione ulteriore da parte dell'HLT. L'architettura del trigger L1 è illustrata nella figura 1.3. La latenza attualmente consentita per il trigger L1 tra un dato attraversamento del fascio e la distribuzione della decisione di trigger all'elettronica front-end del rivelatore è di 3.2 μs, cioè il L1 trigger ha a disposizione quel tempo per decidere se scartare o meno un evento rivelato. A causa della latenza del tracciatore a silicio, il trigger di L1 non è in grado di attendere il tempo necessario per l'invio e il processamento dei suoi dati, e quindi attualmente il L1 non contiene alcuna informazione sulle tracce rilasciate nel detector a silicio. [3]

1.2 Upgrade del trigger di primo livello a CMS

L'aggiornamento di LHC, il cui inizio è in programma per il 2027, porterà alla possibilità di raggiungere una luminosità integrata fino a 4000 fb^{-1}, da cui il nome del progetto *High Luminosity Large Hadron Collider* (HL-LHC). Questo permetterà l'accesso ad una più alta quantità di segnali rila-

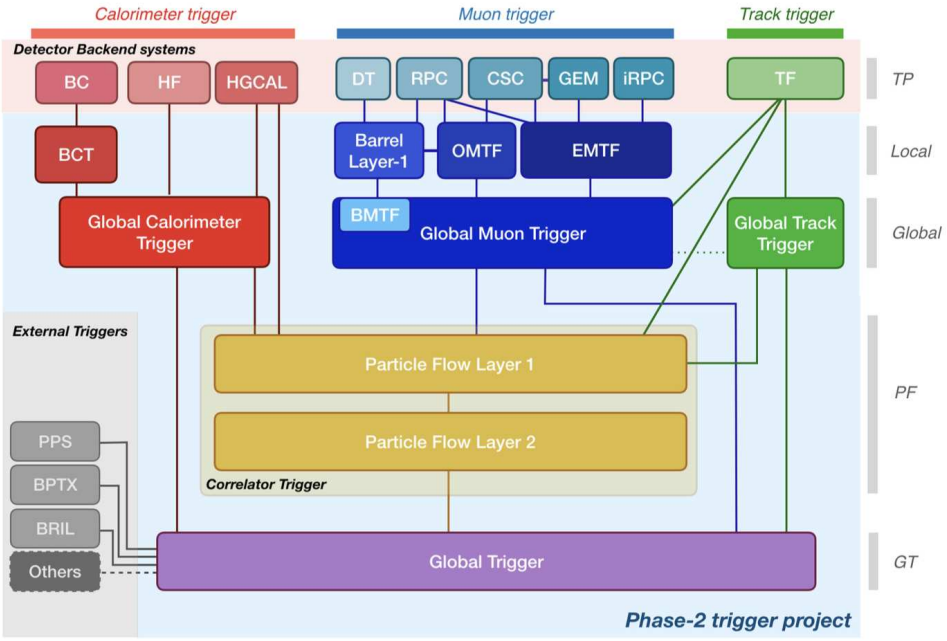

Figura 1.4: Schema proposto per l'upgrade del trigger L1. Crediti immagine [5]

sciati nei detector per collisione che necessiterà perciò di un adattamento dei sistemi di trigger degli esperimenti, tra cui CMS, per essere processata. L'aumento della luminosità di LHC comporterà un aumento del numero medio di interazioni simultanee nello stesso *bunch-crossing*, fattore che prende il nome di Pile-Up.
La fase 2 dell'aggiornamento del L1 trigger di CMS consisterà in un miglioramento dei già presenti trigger muonico e calorimetrico e nell'aggiunta di due nuove componenti: il Global Track Trigger (GTT) e il Correlator Trigger (immagine 1.4).
Il trigger calorimetrico avrà accesso a nuove informazioni sulla posizione delle particelle, tra cui informazioni ad alta granularità grazie anche alle migliorie apportate ai detector di posizione (non più perciò solo a quelle sui depositi energetici), mentre il trigger muonico agirà su un range più ampio ($|\eta| < 2.8$) e potrà unire informazioni del tracciatore al silicio a quelle delle camere a muoni. Il Global Track Trigger (GTT) riceverà informazioni dai tracciatori al silicio, i quali ricostruiscono le tracce di particelle cariche. Questo sistema di trigger utilizzerà poi queste tracce per costruire oggetti di trigger di alto livello che possono contribuire significativamente a compiti come la mitigazione del PileUp. Il Correlator Trigger aggregherà invece tutte le informazioni elaborate dai tre sistemi di trigger a monte (calorimetro, muonico e GTT) per ottenere le migliori prestazioni di trigger su topologie fisiche complesse. Ci sono due algoritmi essenziali nel correlator trigger: il Particle Flow identifica e ricostruisce tutte le particelle utilizzando le informazioni di tutti i sotto-rivelatori, mentre il Pileup Per Particle Identification (PUPPI) mitiga gli effetti del PileUp, permettendo di trascurare tutte le particelle non associate alla collisione principale dell'evento. [6]

Al termine dell'aggiornamento, il sistema L1 potrà processare fino a 750,000 eventi al secondo, cioè avere una frequenza di 750 kHz, e una latenza di 12.5 μs. [5]
L'aumento di frequenza e latenza e la possibilità di accedere a nuove informazioni da parte del L1 trigger, aprono la strada all'implementazione di algoritmi di machine learning per le selezione di eventi significativi. Lo scopo di questo progetto di tesi è sviluppare un possibile algoritmo per effettuare il riconoscimento dei jet di particelle provenienti dall'adronizzazione del quark b (anche detto b-tagging) da integrare nel L1 trigger dell'esperimento CMS. Da alcuni anni è infatti possibile implementare semplici modelli software di machine learning tradotti in hardware per mezzo di schede di elettronica programmabili FPGA.

Nel capitolo successivo verrà presentata una panoramica sulla fisica del quark b e dei jet da esso generati, sul flavour tagging e in particolare sul b-tagging al L1 a CMS.

Capitolo 2

Identificazione di jet

2.1 Bottom quark

Il Modello Standard (MS) è la teoria fisica che descrive le interazioni fondamentali e definisce una classificazione di tutte le particelle elementari note.
Le particelle possono essere suddivise a seconda della statistica a cui obbediscono: possono essere particelle a spin intero (bosoni) o semintero (fermioni). I fermioni possono essere a loro volta classificati sulla base di un altro numero quantico, la carica di colore: in tabella 2.1 sono distinti i leptoni dai quark, i primi con carica di colore nulla, gli altri con carica di colore non nulla.
Una evidenza empirica dell'esistenza della carica di colore è data dall'esistenza di stati legati di quark del tipo qqq. Questo tipo di stati, per il principio di esclusione di Pauli, possono essere ammessi in natura solo a patto che i tre quark non siano indistinguibili, ovvero differiscano per almeno un numero quantico. A questo scopo è stata introdotta la carica di colore (che può essere rosso, giallo o blu), che permette perciò l'esistenza di stati della forma $q_b q_r q_g$ e di carica di colore globalmente nulla (o bianco).

Fermioni	**1ª generazione**	**2ª generazione**	**3ª generazione**
Quark	Up u	Charm c	Top t
	Down d	Strange s	Bottom b
Leptoni	Neutrino elettronico ν_e	Neutrino muonico ν_μ	Neutrino tauonico ν_τ
	Elettrone e	Muone μ	Tau τ

Tabella 2.1: Organizzazione dei Fermioni per generazione

Il quark *bottom* o *beauty*, noto come quark b, è una delle sei varietà di quark previste dal Modello Standard. È il secondo quark più pesante, con una massa di circa 4.18 GeV/c^2 e un tempo di vita molto lungo rispetto ad altri quark pesanti, proprietà cruciale per la sua identificazione nei rivelatori. I decadimenti del quark b possono avvenire solo attraverso processi con salto di una o due famiglie:

$$b \to c \quad (\text{esempio: } B^- \to D^0\pi^-) \qquad b \to u \quad (\text{esempio: } B^0 \to \pi^+\pi^-) \tag{2.1}$$

Queste caratteristiche portano a due conseguenze importanti:

- La vita media degli adroni beauty è molto lunga: $\tau_{B^{0,\pm}} > 450\mu s$;
- Nei decadimenti possono essere osservati molti fenomeni rari e interessanti: decadimenti che avvengono tramite diagrammi a scatola, oscillazioni di sapore, grandi asimmetrie di CP.

Inoltre, il decadimento del bosone di Higgs in coppie di quark b è uno dei canali di decadimento più importanti e studiati nel contesto del Modello Standard della fisica delle particelle. Nel Modello Standard, il bosone di Higgs ha diverse modalità di decadimento, ma il decadimento in quark *bottom*

($H \to b\bar{b}$) è uno dei più probabili, con una *branching ratio* (probabilità di decadimento) di circa il 58%. Questo lo rende il canale di decadimento più comune per il bosone di Higgs.

Le leggi della Cromodinamica Quantistica (QCD) impongono l'impossibilità di trovare un quark libero in natura: i quark posseggono infatti carica di colore non nulla, ma gli unici stati osservabili in natura sono gli stati che possiedono globalmente carica di colore nulla. Questa proprietà dei quark è nota come confinamento: la forza di legame tra due quark cresce all'aumentare della distanza.
Quando due quark vengono separati, o quando viene irraggiato un gluone isolato, come accade nelle collisioni negli acceleratori di particelle, l'energia necessaria cresce fino al punto in cui diventa energeticamente favorevole la creazione di una coppia quark-antiquark dal vuoto piuttosto che permettere ai quark di allontanarsi ulteriormente. Perciò, anziché rilevare i quark singoli nei rivelatori, vengono osservate particelle neutre rispetto alla carica di colore, come mesoni e barioni, che appaiono raggruppate. Questo fenomeno, noto come adronizzazione, è uno dei processi fondamentali della fisica delle particelle odierna.
Poiché il contenuto degli adroni porta una carica di colore, ciascuno di questi frammenti di decadimento porterà una parte della carica di colore. Per far sì che anche questi frammenti rispettino il confinamento, devono necessariamente essere creati altri oggetti colorati (come gluoni o altri quark) per formare oggetti privi di colore.
Questo insieme di prodotti di decadimento e frammenti deve conservare il momento del loro adrone originale e, di conseguenza, tenderà a viaggiare in direzioni simili, spesso formando un gruppo a forma di cono con diverse traiettorie. Questa cascata risultante di oggetti colorati è nota come un jet.[2]

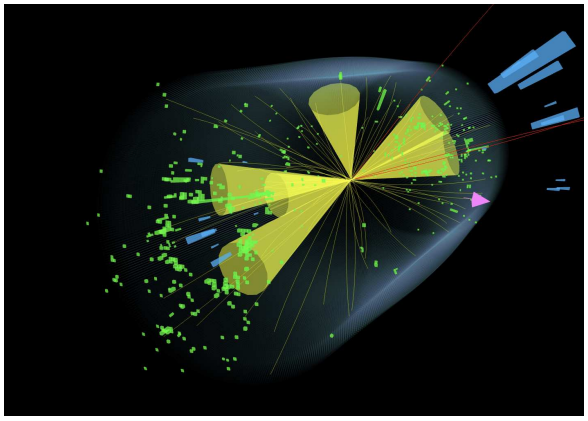

Figura 2.1: Collisione protone-protone a 13 TeV nel rivelatore CMS nel 2016. La figura mostra le particelle che emergono dalla produzione di una coppia di quark top, che a loro volta decadono in due bosoni W e due quark b. I prodotti di decadimento di uno dei quark b includono un mesone J/ψ che decade in una coppia di muoni con carica opposta. Ciascuno dei coni gialli in figura rappresenta la direzione di un jet associato all'adronizzazione di un quark o gluone. Crediti immagine: CMS Experiment

Dallo studio delle proprietà dei jet è possibile risalire al quark da cui il jet ha avuto origine. Scopo di questa tesi sarà proprio studiare e proporre algoritmi per l'identificazione dei jet originati dall'adronizzazione di un quark b o un quark pesante in prodotti di collisione protone-protone da poter applicare al L1 trigger dell'esperimento CMS dopo il suo upgrade di Fase 2. Questo compito è detto b-tagging, o heavy flavour tagging.

2.2 Flavour tagging a CMS

Effettuare flavour tagging è uno dei compiti principali a CMS. Ciò consente di risalire a posteriori alle particelle elementari prodotte nelle collisioni e di studiare decadimenti o fenomeni fisici rari avvenuti nel rivelatore.
Nel corso degli anni a CMS sono stati sviluppati e perfezionati algoritmi sempre più precisi nell'individuare il sapore adronico dei jet rivelati.

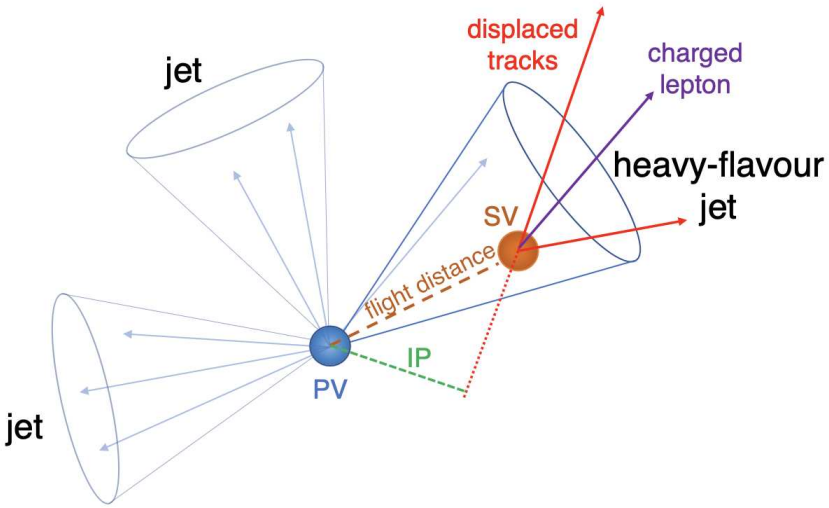

Figura 2.2: Illustrazione di un jet pesante e del vertice secondario dovuto al decadimento di un quark b o c. Crediti immagine [7]

Gli algoritmi di b-tagging sfruttano informazioni provenienti da variabili che hanno un forte potere di discriminazione, come le informazioni sul secondo vertice di decadimento e sul parametro di impatto dei jet.
Una caratteristica dei jet b e c, provenienti da quark con una lunga vita media, è infatti quella di avere un secondo vertice di decadimento oltre a quello primario in prossimità della collisione. Ciò è illustrato in figura 2.2, dove si vede che è possibile ricostruire un vertice secondario (SV) di decadimento a partire dalle informazioni sulla direzione delle tracce dei componenti del jet ed individuare di conseguenza un elevato parametro di impatto (IP), dell'ordine dei millimetri.
Un'altra variabile che può fornire importanti informazioni riguardo la natura dei jet, è la presenza o meno di leptoni, assieme alle grandezze cinematiche ad essi associate e relative al jet di appartenenza. Questa tipologia di informazioni viene però sfruttata in algoritmi di b-tagging che possono usare la presenza di leptoni all'interno del jet per la discriminazione dei quark (tipicamente questi algoritmi sono detti soft-lepton taggers), nonostante la percentuale di jet che presentano leptoni sia generalmente bassa, inferiore al 20%.
Il flavour tagging, ad oggi, prima della Fase 2 dell'aggiornamento di CMS, avviene con precisione a livello di offline, cioè solo dopo aver acquisito tutte le informazioni a disposizione del rivelatore e dopo che i due trigger hanno già effettuato tagli significativi ai dati a disposizione. Infatti, sino ad ora non è stato possibile effettuare flavour tagging al livello del L1 trigger a causa dell'assenza di informazioni a sufficienza a quello stadio e del brevissimo tempo di latenza a disposizione. In maniera approssimata esso può avvenire solo a livello di HLT, dopo che la riduzione della rate da 40MHz a 100KHz (quindi dopo che è stato scartato il 99.75% degli eventi) è già avvenuta.
L'upgrade del L1 trigger consentirà invece di avere a disposizione a quel livello nuove informazioni su tracce ed energia e un tempo di latenza maggiore oltre ad un nuovo livello di correlator, come già discusso in sezione 1.2: si apre perciò la possibilità di selezionare eventi significativi sin da subito tramite tecniche di machine learning.
A differenza di quanto avviene offline, al livello del L1 trigger, anche dopo l'aggiornamento, non avremo a disposizione tutti i dati più significativi per la discriminazione: in particolare non saranno a disposizione le informazioni sulla ricostruzione dei vertici secondari e dei parametri di impatto a causa dell'assenza dei dati provenienti dai pixel del tracciatore più interno. Sarà invece possibile usufruire di informazioni energetiche sui jet e di dati sui costituenti, oltre a dati parziali sulle tracce.

Nel capitolo seguente saranno discussi ed analizzati i dati a disposizione al L1 trigger dopo l'upgrade, al fine di selezionare le variabili più significative per l'implementazione di un algoritmo di machine learning che effettui il b-tagging.

Capitolo 3

Analisi dei dati

3.1 Sistema di coordinate di CMS

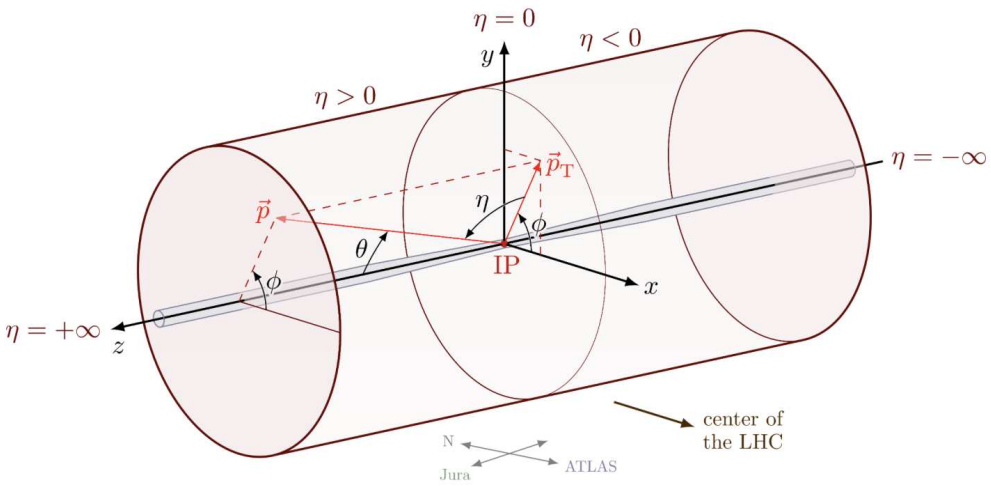

Figura 3.1: Sistema di coordinate di CMS. Crediti immagine: [8]

Prima di entrare nel dettaglio con l'analisi dei dati, è importante comprendere il sistema di coordinate dell'esperimento CMS (figura 3.1), nel quale sono descritti i jet e i loro costituenti. Le particelle sono individuate dal seguente sistema di coordinate:

- p_T: momento trasverso, ovvero la componente del momento della particella in direzione radiale rispetto alla sezione del cilindro, cioè la sezione lungo la quale avviene la curvatura delle traiettorie delle particelle cariche per effetto del campo magnetico solenoidale;
- $\eta \equiv -\ln\tan(\frac{\theta}{2})$: pseudorapidità, una grandezza legata all'angolo polare θ rispetto all'asse del cilindro, ovvero la direttrice lungo la quale si muovono i fasci di particelle nel detector;
- ϕ: angolo azimutale rispetto all'asse x, dove x è positivo verso il centro dell'anello di LHC;
- m : massa della particella.

La descrizione di una particella in questo sistema di coordinate, è equivalente alla descrizione del quadrimomento, che viene generalmente utilizzata nella descrizione relativistica del moto delle particelle.

$$\text{Sistema di coordinate CMS:} \quad (p_T, \ \eta, \ \phi, \ m) \quad \rightarrow \quad \text{Quadrimomento:} \quad \left(\frac{E}{c}, \ p_x, \ p_y, \ p_z\right)$$

3.2 Dati dalle simulazioni

Le prese dati di HL-LHC non avranno inizio prima del 2030, risulta perciò necessario effettuare i nostri studi su dati simulati. Le simulazioni che sono state adoperate nello sviluppo di questo lavoro di tesi ricostruiscono collisioni protone-protone nelle quali avviene un processo di produzione di una coppia di quark top ($t\bar{t}$). Questo è un processo ad elevata sezione d'urto, che prevede una elevata produzione di quark b, motivo per cui viene spesso utilizzato negli esperimenti per lo sviluppo e la misura delle performance degli algoritmi di b-tagging. Inoltre, ad ogni processo di decadimento del quark top è associata la presenza di un quark b, perciò nel campione di dati sono presenti sia jet da quark b che da quark leggeri (in cui si intendono tutti i jet che sono generati dall'adronizzazione di quark u, d, s e da gluoni). Frequente è anche la presenza di quark charm (quark c), un quark più leggero del quark b, ma che ha caratteristiche che lo rendono in parte simile a un quark pesante. Per questo motivo il quark c verrà trattato separatamente nel corso dello sviluppo degli algoritmi di b-tagging. Il PileUp delle particelle (cioè il numero medio di collisioni contemporanee) nelle simulazioni è di 200, che è ciò che ci si aspetta a HL-LHC.
I dati a disposizione provengono da una simulazione dei segnali a disposizione al livello di L1 trigger dopo l'upgrade, in particolare di ciò che è disponibile a livello del Correlator del futuro L1 trigger. Poichè si tratta di dati simulati, si è in possesso delle informazioni del generatore, ovvero la verità Monte Carlo e le condizioni iniziali a partire dalle quali sono state effettuate le simulazioni.
I jet sono stati ricostruiti tramite un algoritmo veloce di ParticleFlow che mira a ricostruisce l'intero evento sulla base di tutta l'informazione disponibile identificando ciascuna particella prodotta, e quindi ad associare a ciascun jet le particelle candidate identificate (e.g. elettroni, pioni, fotoni, ...). I costituenti sono poi passati attraverso un altro algoritmo per la riduzione delle particelle non provenienti dalla collisione primaria, noto come PUPPI (PileUp Per Particle Identification).

I dati ottenuti dalle simulazioni sono strutturati nel seguente modo:

n^o collisione	n^o jet per collisione	$p_{T,jet}$	η_{jet}	ϕ_{jet}	$p_{T,cost,i}$	$\eta_{cost,i}$	$\phi_{cost,i}$	$PDGID_{cost,i}$	...	hadron flavour
...	...	...	...	...	...	...	...	...	...	...

dati dei jet — dati dei costituenti dei jet

Si può vedere che sono note le variabili cinematiche associate ai jet e ai loro costituenti, il PDGID dei costituenti che è un codice identificativo (sulla base della codifica del Particle Data Group [9]) che viene assegnato a ciascuna particella costituente dall'algoritmo di Particle Flow per indicarne la tipologia, ed infine il sapore adronico del jet, che corrisponde alla verità Monte Carlo associata al jet. In particolare un sapore adronico pari a 5 implica che il jet in questione proviene da adronizzazione di un quark b, un sapore adronico uguale a 4 corrisponde ad una adronizzazione di un quark c, ed infine se pari a 0 si ha un jet da quark leggero o gluone (udsg).
Di seguito si riporta un'immagine che permette di visualizzare nello spazio delle coordinate angolari (η, ϕ) alcuni jet e i loro costituenti (figura 3.2).

Si può da subito notare il maggior numero di costituenti che caratterizza i jet generati per adronizzazione del quark b: vedremo che questa sarà una delle principali caratteristiche discriminanti per la classificazione dei jet.

3.3 Analisi delle distribuzioni delle principali grandezze fisiche associate ai jet

Per poter studiare un algoritmo finalizzato alla classificazione dei jet, è necessario analizzare i dati a disposizione per effettuare una scelta riguardo a quali variabili risultano più informative in relazione alla natura dei jet e sono quindi indicate come input per una rete neurale.
In primo luogo è stato effettuato uno studio preliminare dei dataset completi e delle variabili cinematiche più semplici a disposizione, lasciando momentaneamente da parte il contributo delle grandezze associate ai costituenti.

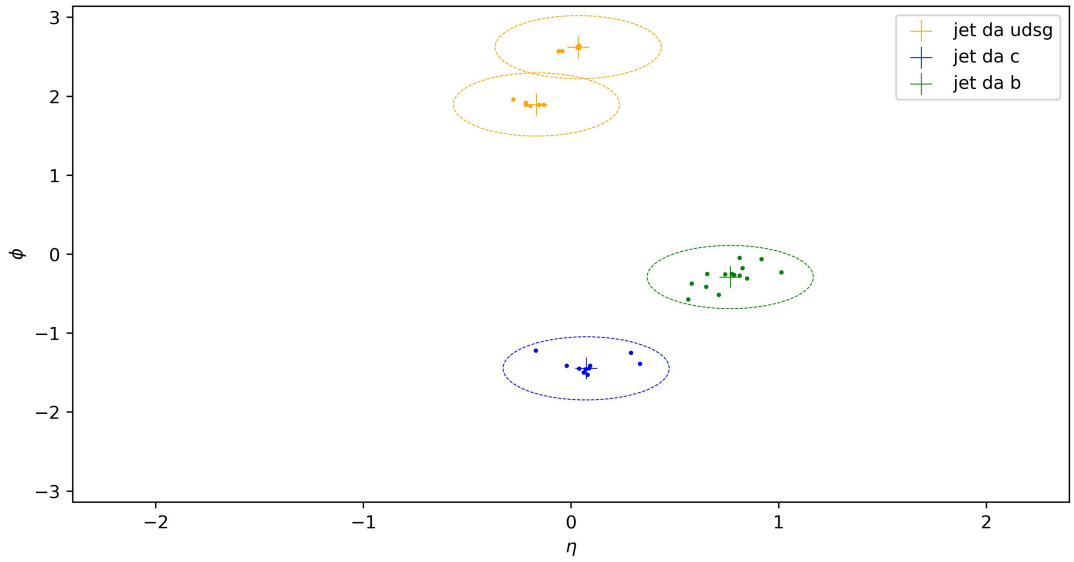

Figura 3.2: Visualizzazione di una collisione con produzione di quattro jet, tra cui uno da b (in verde) ed uno da c (in blu). Il centro del jet è rappresentato dal marker a croce, con rispettivi costituenti rappresentati dai marker circolari. Le linee tratteggiate delineano un raggio di 0.4 nello spazio delle coordinate (η, ϕ), che è il valore del parametro per la ricostruzione di un jet da parte dell'algoritmo.

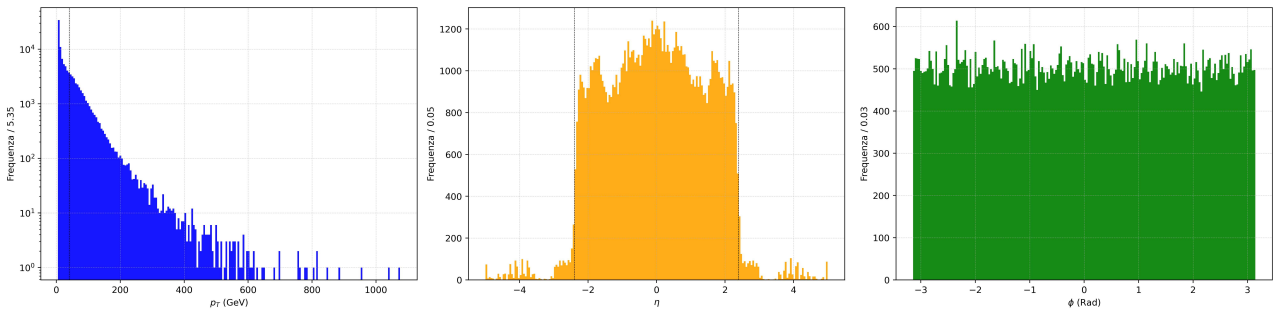

Figura 3.3: Istogrammi delle frequenze delle variabili cinematiche associate ai jet prima di tagliare i dati. Le rette verticali tratteggiate indicano i filtri che verranno implementati successivamente.

Innanzitutto è stato necessario effettuare un taglio del dataset, mantenendo solo i dati negli intervalli $p_T > 40\,\text{GeV}$, e $|\eta| < 2.4$, sappiamo infatti che i tracciatori non sono attivi in un intervallo angolare maggiore. A livello di energia invece il taglio si è rivelato utile in quanto a CMS si è interessati a eventi ad alta energia, per questo vengono scartati dati di jet con momento trasverso basso. Questo ha diminuito significativamente la statistica del dataset: si vede infatti dall'istogramma del momento trasverso (in scala logaritmica) che la maggior parte dei dati possiede bassa energia.

Nei grafici in figura (3.3) non è stata fatta una distinzione tra jet provenienti da adronizzazione di quark b, c o $udsg$: una prima analisi è stata proprio cercare differenze tra gli andamenti delle variabili cinematiche al variare della tipologia di jet al quale sono associate. In figura 3.4 si può vedere che dall'andamento delle principali variabili cinematiche non è possibile riconoscere a primo impatto differenze sostanziali al variare della tipologia di jet a cui sono associate.

Si è proceduto allora studiando i costituenti di ogni jet, a partire dal numero di costituenti per jet: ci aspettiamo che i jet provenienti da adronizzazione del quark b presentino in media un maggior numero di costituenti, poiché conosciamo i processi di decadimento a catena che caratterizzano il quark b. In effetti, come possiamo vedere dall'istogramma in figura 3.5, questo è esattamente ciò che avviene.

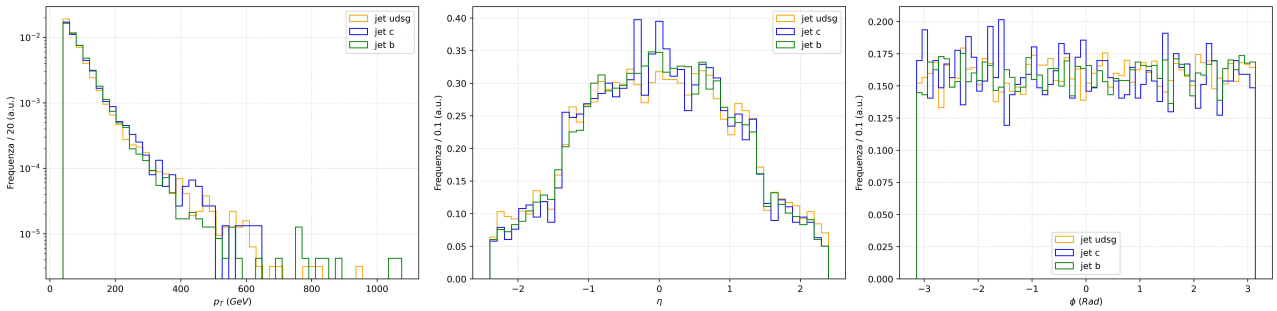

Figura 3.4: Istogrammi sovrapposti delle frequenze normalizzate alla stessa area (unitaria). Gli istogrammi dei jet provenienti dalle 3 diverse componenti sono separati nei diversi colori: verde per i jet da quark-b; blu per i jet da quark-c; giallo per i jet da quark leggeri (usdg).

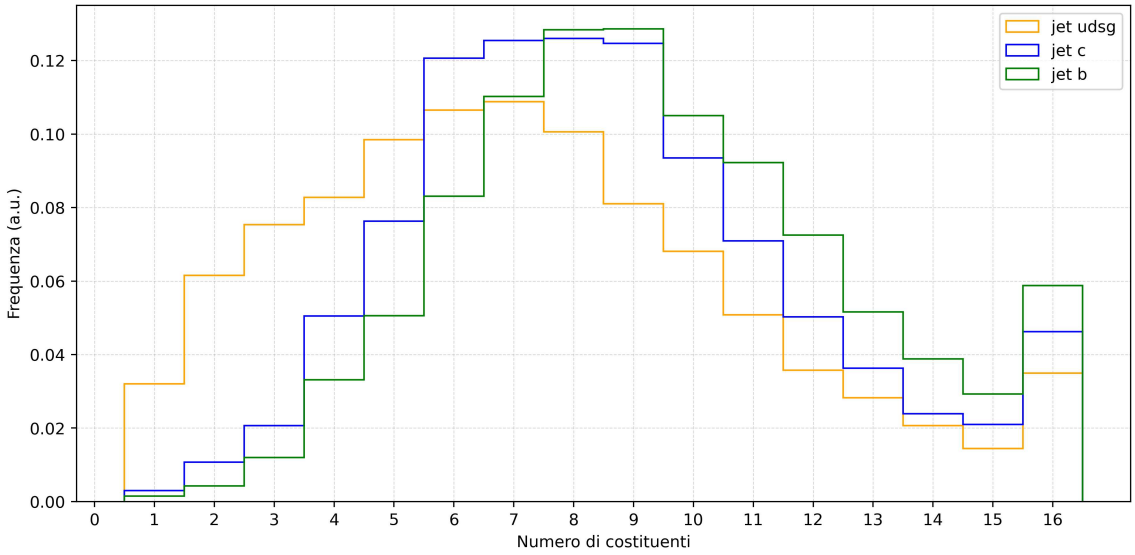

Figura 3.5: Istogrammi sovrapposti per la distribuzione del numero dei costituenti per jet. Le aree di ogni istogramma sono normalizzate all'unità. L'ultimo bin contiene i valori di overflow.

Come è possibile vedere dal grafico, sono disponibili informazioni sui primi 16 costituenti per ogni jet. Risulta essere un'informazione rilevante la tendenza dei jet da quark b ad avere mediamente un maggior numero di costituenti, e il fatto che i jet da c dimostrino un comportamento intermedio tra i quark leggeri e i b .

Prima di procedere ulteriormente con l'analisi è stato effettuato un ulteriore selezione al dataset. Sono stati eliminati tutti i jet con un singolo costituente, poiché perderebbe senso la nozione stessa di jet, e tutti i jet costituiti da due particelle delle quali una delle due avente il 90% del momento a disposizione. Per comprendere le ragioni di questo filtro, è opportuno pensare a come vengono ricostruiti i jet a partire dall'algoritmo di Particle Flow. Esso ricostruisce infatti i "candidati particella" a partire dai rilasci energetici nei rivelatori, particelle che vengono poi raggruppate in jet dall'algoritmo anti k_t con un parametro di clustering di 0.4.
Con questo tipo di ricostruzione, nonostante la complessità degli algoritmi, è possibile che possa essere ricostruito un jet costituito da una singola particella molto energetica accompagnata da una poco energetica, potenzialmente irradiata dalla particella energetica stessa. Il taglio dei dati in questione ha lo scopo di limitare il numero di jet ricostruiti a fronte di casi limite degli algoritmi di clustering.

Dopo l'ultima selezione operata sul dataset, si è andati ad analizzare le tipologie di costituenti dei jet, informazione alla quale abbiamo accesso tramite il PDGID dei costituenti fornito nel dataset.
In figura 3.6 è riportata la distribuzione delle particelle presenti nel dataset e come si distribuiscono in percentuale all'interno delle tre tipologie di jet analizzate. Come atteso tutti i jet contengono una

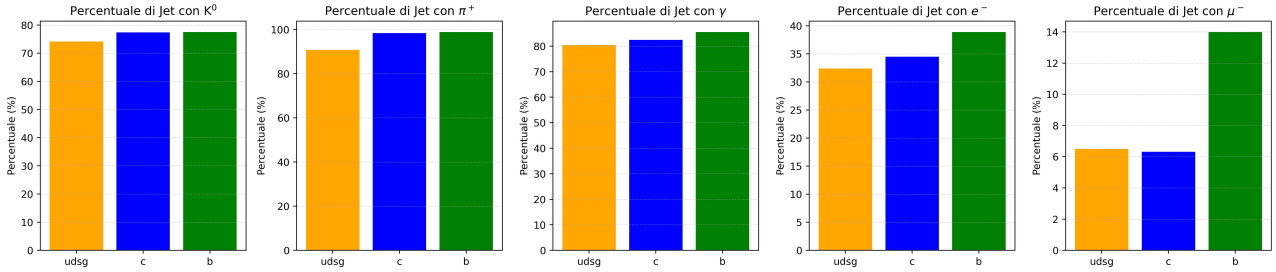

Figura 3.6: Distribuzione percentuale delle tipologie di costituenti per classe di jet

frazione estremamente alta di pioni e kaoni. È poi utile osservare come i jet provenienti dall'adronizzazione di quark b tendano ad essere maggiormente caratterizzati dalla presenza di un leptone carico, come in effetti previsto e descritto in sezione 2.2.

Uno studio dedicato è stato quindi riservato ai jet contenenti leptoni: è noto infatti l'elevato potere discriminante delle variabili legate ai leptoni nei jet, proprietà ampiamente sfruttata nel flavour tagging offline a CMS.
Ad esempio, è stata notata una differenza nel numero di muoni nei costituenti dei jet b rispetto ai jet c e leggeri. Mediamente, circa il 6% dei jet b che presentano almeno un muone tra i loro costituenti, sono composti in realtà da più di un muone, contro il 3% dei jet c e meno dell'1% dei jet leggeri. Per i b è inoltre giustificato il fatto di poter avere più leptoni a causa del possibile decadimento a cascata $b \to c \to light$ in cui entrambi il b e il c decadono in un leptone. Un'analisi più approfondita delle variabili cinematiche associate ai jet contenenti leptoni è trattata nella sezione seguente.

3.4 Analisi delle variabili cinematiche associate ai costituenti dei jet

Oltre al numero dei costituenti e al PDGID di ogni costituente, è stato opportuno analizzare il comportamento delle variabili cinematiche di ogni costituente relativamente al jet di appartenenza. In particolare è stata definita la variabile

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

Dove

$$\Delta\eta = |\eta_{jet} - \eta_{cost,i}| \qquad \Delta\phi = |\phi_{jet} - \phi_{cost,i}|$$

Il ΔR rappresenta lo spostamento relativo nello spazio dei parametri (η, ϕ) di ogni costituente rispetto al jet di appartenenza. Allo stesso modo è stato preso in considerazione il momento trasverso di ogni costituente normalizzato al momento del jet, $p_{T,cost,i}/p_{T,jet}$, per cercare le relazioni tra la tipologia di jet e il comportamento dei rispettivi costituenti rispetto al jet cui appartengono.

Dal plot in figura 3.7 delle distribuzioni delle variabili appena definite non è possibile notare differenze significative nell'andamento delle curve dei momenti normalizzati e dei ΔR.
L'analisi è stata allora approfondita andando a studiare le stesse variabili per ogni costituente dei jet in ordine di $p_{T,i}$ decrescente.

In questo caso è invece possibile notare una chiara differenza soprattutto nel comportamento del $p_{T,cost,i}/p_{T,jet}$ dei jet leggeri rispetto ai jet da b.
Per il primo costituente dei jet derivati da adronizzazione di jet leggeri si rileva infatti un'accumulazione significativa di eventi a valori di $p_{T,cost,i}/p_{T,jet} > 0.7$: mediamente ci sono molti più jet leggeri il cui primo costituente possiede più del 70% del momento a disposizione.
Conseguentemente al fatto che, come avevamo visto, i jet leggeri presentano mediamente un minor numero di costituenti e alla condizione di conservazione del momento per cui

$$\sum_{i=1}^{n} p_{T,cost,i} = p_{T,jet}$$

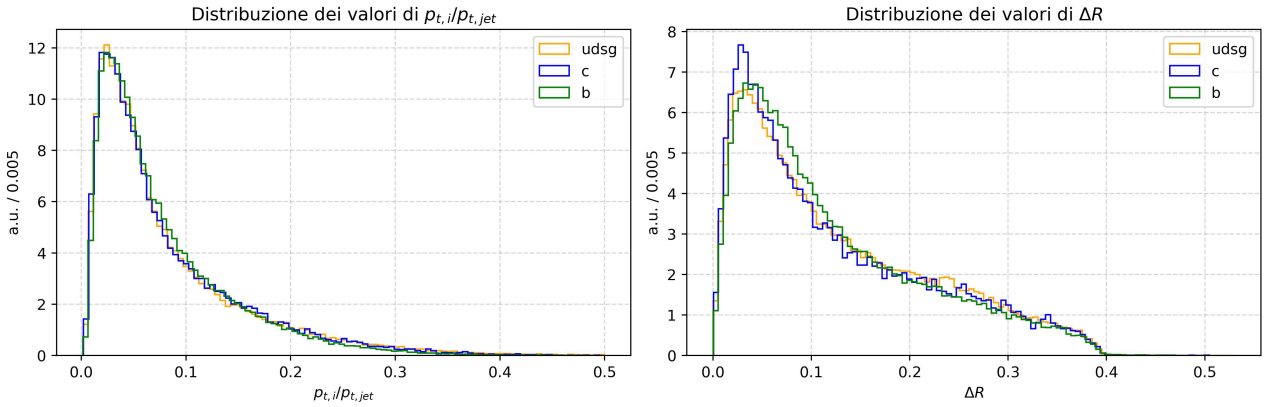

Figura 3.7: Distribuzione delle variabili $p_{T,cost,i}/p_{T,jet}$ e ΔR per tutti i costituenti dei jet. Le aree degli istogrammi sono normalizzate all'unità.

è attesa allora una accumulazione di eventi a basso $p_{T,cost,i}/p_{T,jet}$ per i costituenti successivi dei jet leggeri, ed effettivamente è proprio ciò che si ottiene, come mostrato negli istogrammi in figura 3.9.

Un'analisi più approfondita è stata invece dedicata ai jet contenenti leptoni, distinguendo ciò che avviene per i jet con muoni e i jet con elettroni. Nel flavour tagging offline risulta avere un forte potere discriminante la variabile p_T^{rel}, definita come

$$p_T^{rel} = \frac{|\vec{p}_{lep} \times (\vec{p}_{jet} - \vec{p}_{lep})|}{|\vec{p}_{jet} - \vec{p}_{lep}|}$$

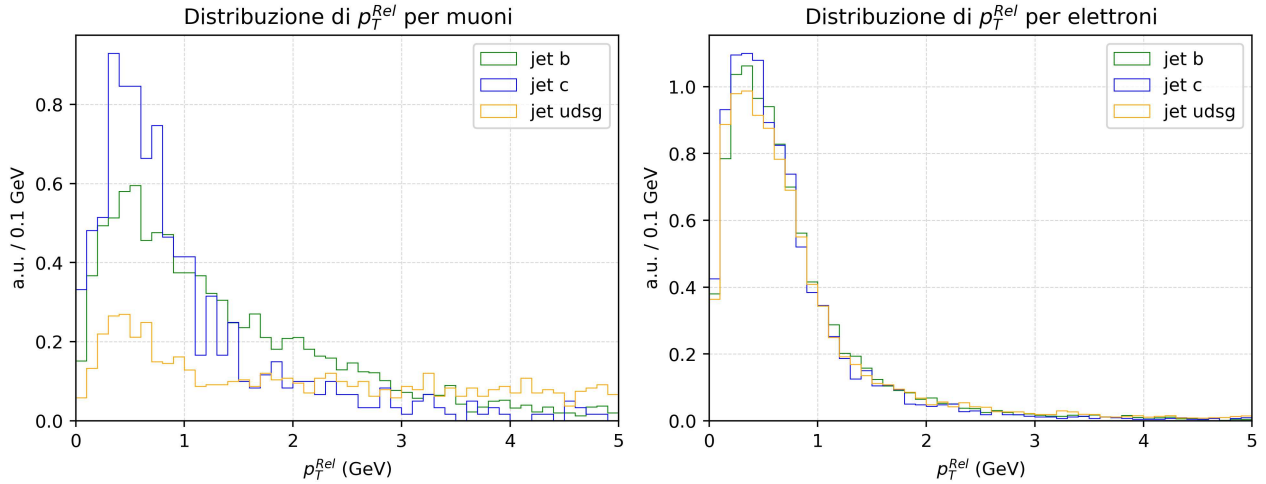

Figura 3.8: Distribuzione della variabile p_T^{rel} per jet contenenti muoni ed elettroni. Le aree degli istogrammi sono normalizzate all'unità.

Dal plot delle distribuzioni in figura 3.8 è possibile notare una chiara differenza tra il comportamento dei jet contenenti muoni e i jet contenenti elettroni. Nel caso dei muoni il p_T^{rel} è mediamente più basso per i costituenti appartenenti a jet da quark pesante rispetto a quelli da jet leggeri. Per gli elettroni le differenze sono invece minime: a questo stadio non si hanno abbastanza informazioni per poter affermare che anche gli elettroni appartenenti a jet pesanti tendono ad avere un p_T^{rel} mediamente minore, come ci aspetteremmo. Per quanto la variabile p_T^{rel} abbia sicuramente del potere discriminante, non è sufficiente per poter descrivere un classificatore che miri ad avere una buona efficienza, a causa della percentuale dei jet da b con leptoni bassa, e purezza, a causa del non perfetto potere discriminante, soprattutto per gli elettroni.

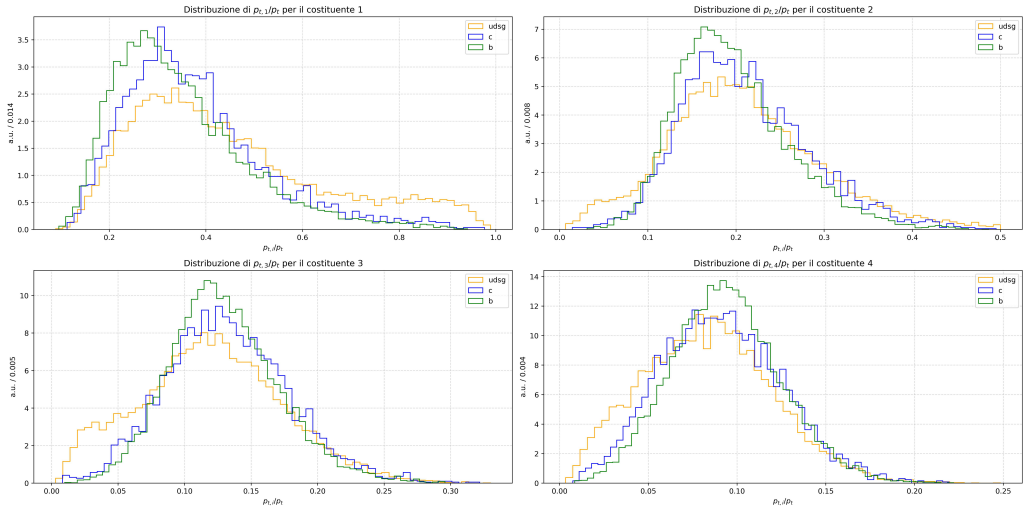

Figura 3.9: Distribuzione del $p_{T,cost,i}/p_{T,jet}$ per i primi 4 costituenti di ogni jet. Le aree degli istogrammi sono normalizzate all'unità. L'ordine dei costituenti è definito dall'ordine decrescente del momento trasverso del costituente.

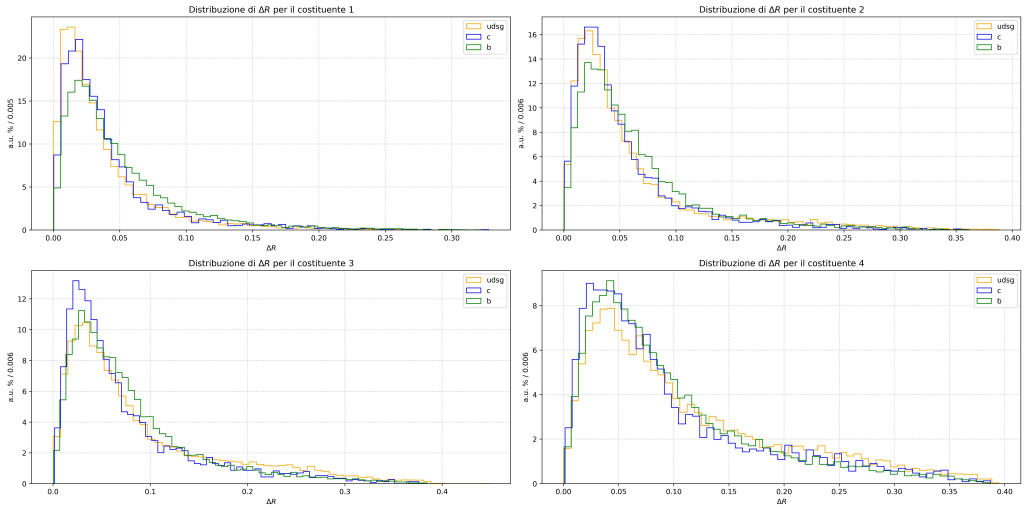

Figura 3.10: Distribuzione del $\Delta R_{cost,i}$ per i primi 4 costituenti di ogni jet. Le aree degli istogrammi sono normalizzate all'unità. L'ordine dei costituenti è definito dall'ordine decrescente del momento trasverso del costituente.

Capitolo 4

Reti neurali per il b-tagging

4.1 Introduzione alle reti neurali per la classificazione binaria

Il Machine Learning è una branca dell'intelligenza artificiale che permette l'apprendimento basato direttamente sui dati. Questo apprendimento può avvenire in tre modalità principali: supervisionato, dove il modello viene addestrato su un set di dati etichettati per fare previsioni; non supervisionato, in cui il modello cerca di trovare strutture o pattern nascosti nei dati non etichettati; e reinforcement learning, dove l'algoritmo impara attraverso interazioni con l'ambiente ricevendo ricompense o penalità. Questi approcci sono utilizzati per risolvere svariate tipologie di problemi tra cui i principali sono la classificazione, dove l'obiettivo è assegnare categorie ai dati, e di regressione, in cui si prevede un valore continuo dati degli input.

Le reti neurali artificiali (ANN) sono uno dei modelli di machine learning più generali, ispirati alla struttura e al funzionamento del cervello umano, in cui una serie di unità funzionali (neuroni) sono interconnesse tra loro in livelli (o layers). Queste reti sono in grado di apprendere rappresentazioni complesse dei dati, rendendole particolarmente efficaci per compiti come il riconoscimento delle immagini, la traduzione automatica e molto altro.

L'unità di calcolo fondamentale di una rete neurale artificiale è il neurone, che processa tramite una trasformazione, in generale non lineare, un vettore x di valori in ingresso restituendo uno scalare *y(x)* in output. I vettori in ingresso sono moltiplicati per un vettore di pesi w e la somma algebrica pesata degli ingressi, a seconda dell'architettura della rete, serve come input per uno strato (layer) successivo di neuroni, fino a giungere ad un ultimo stadio in cui viene confrontata con un valore di soglia θ, per mezzo del quale viene definito l'output binario $\{1; 0\}$[10]. In termini matematici ciò significa

$$y(x) = g\left(\sum_{i=1}^{n} w_i x_i - \theta\right) \equiv g(w^T x - \theta)$$

dove g è detta funzione di attivazione, e può assumere diverse forme a seconda del problema che si sta affrontando. In questo lavoro di tesi verranno spesso utilizzate le funzioni di attivazione Rectified Linear Unit (ReLU), Sigmoid e Tanh, definite come segue

$$\text{Tanh:} \quad \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad \text{ReLU:} \quad g(x) = \begin{cases} x & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases} \qquad \text{Sigmoid:} \quad \sigma(x) = \frac{1}{1 + e^{-x}}$$

L'aspetto interessante è che i valori dei pesi e del valore di soglia possono essere determinati tramite un processo di apprendimento automatico a partire da un insieme di addestramento composto da coppie input-output note. Una struttura di questo tipo è detta Perceptron e geometricamente corrisponde alla calibrazione di parametri per la ricerca di un iperpiano che separi i vettori di input nei due spazi corrispondenti all'output binario $\{0;1\}$. Questo procedimento è attuabile se e soltanto se gli

spazi in questione sono linearmente separabili: ciò rappresenta una delle maggiori limitazioni del Perceptron.[10]

Per superare le limitazioni del Perceptron è necessario ricorrere a reti multistrato, o reti multilayer feed-forward, che consentono in linea teorica, sotto ipotesi opportune sulle funzioni di attivazione dei neuroni, di approssimare con la precisione voluta una qualsiasi funzione continua su un insieme compatto e quindi anche, in particolare, di risolvere problemi di classificazione di insiemi non separabili linearmente.
La struttura di una rete multistrato, schematizzata in figura 4.1, può essere riassunta come segue:

- n nodi di ingresso corrispondenti alle n features di ingresso alla rete;
- un insieme di neuroni organizzati in L strati, di cui *L-1* strati nascosti (*hidden layers*) e uno strato di output, corrispondente all'output della rete;
- un insieme di archi pesati e orientati che consentono la connessione tra strati successivi, ma non tra neuroni nello stesso strato.

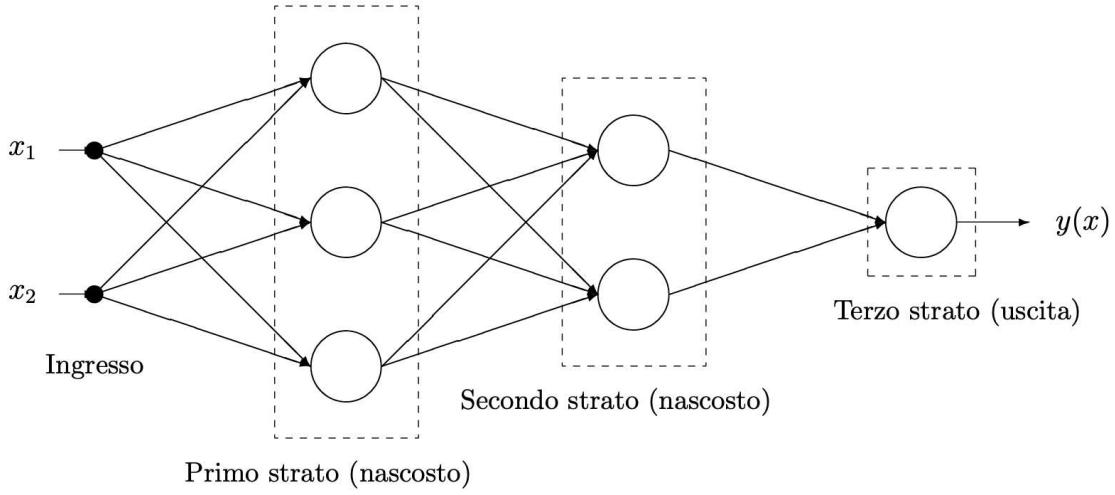

Figura 4.1: Schema di una rete multilayer a 3 strati, con 2 strati nascosti, 2 ingressi e un'uscita. Crediti immagine: [10]

Per rendere una rete neurale in grado di classificare correttamente dati, è necessario effettuare prima un addestramento su un campione di dati la cui classificazione è nota.
Costruire una rete multistrato per la classificazione binaria con n ingressi significa scegliere l'architettura (numero di strati e numero di neuroni di ogni strato), e addestrare la rete, ossia determinare il vettore $w \in R^m$ le cui componenti corrispondono ai pesi e soglie dei neuroni nei vari strati.
Deve perciò essere risolto un problema di minimizzazione di una certa funzione di distanza dal valor vero che può assumere varie forme e prende il nome di funzione di costo o perdita (Loss function). In questo lavoro verrà utilizzata la funzione *Binary Cross Entropy*, definita come segue

$$\text{BCELoss}(x, y) = -\frac{1}{N}\sum_{i=1}^{N}[x_i \log(y_i) + (1 - x_i)\log(1 - y_i)]$$

dove x_i è il valore target per l'esempio i, y_i è la previsione del modello per l'esempio i, N è il numero di esempi.
Per addestrare una rete viene perciò effettuato il calcolo di gradienti per individuare la direzione di maggiore decrescita della funzione in questione. In particolare in questo lavoro è stato utilizzato il metodo della *backward propagation*, che consiste nel calcolo dell'errore tra l'uscita prevista e quella reale che viene poi propagato all'indietro per aggiornare i pesi della rete e ridurre l'errore stesso in un

processo ricorsivo che termina dopo aver fatto osservare alla rete un certo numero di volte il dataset di allenamento, cioè dopo n epoche.

4.2 Studio delle variabili di input

4.2.1 Scelta delle variabili di input per la rete

A seguito dello studio delle variabili discusse nel capitolo precedente, è stata fatta una selezione per determinare quelle più informative, che potessero essere significative per permettere alla rete neurale di apprendere e creare correlazioni tra la tipologia di jet e le variabili di input.

Innanzitutto, è stato ritenuto opportuno scegliere in input le principali variabili cinematiche di ogni jet, cioè p_T, η, ϕ. Questo perché, nonostante nelle distribuzioni non siano stati rilevati comportamenti significativamente distintivi da parte delle diverse tipologie di jet, queste risultano essere le principali variabili descrittive del jet. L'obiettivo è mirare alla creazione di correlazioni tra le variabili più descrittive e l'energia e gli angoli di emissione dei jet. Fondamentale invece è, come già commentato, l'inserimento della variabile corrispondente al *numero di costituenti* per jet.

Come visto in precedenza, inoltre, sicuramente risultano avere un forte potere discriminante le variabili associate ai primi 4 costituenti di ogni jet. In particolare sono state scelte il $\Delta R_{cost,i}$ e il $p_{T,cost,i}/p_{T,jet}$, grandezze che inglobano al loro interno le informazioni più importanti sulla cinematica dei costituenti. I costituenti con un $p_{T,cost,i}$ minore dei primi 4 sono invece stati esclusi. Questo perché, da un'analisi del tutto analoga a quella avvenuta in sezione 3.4, è stato possibile vedere che, per tutte le tipologie di jet, le curve che descrivono gli andamenti delle variabili in questione tendono a schiacciarsi verso valori prossimi allo 0, come in realtà ci aspetteremmo che accada, e quindi che il loro potere discriminante tra le classi in esame tenda rapidamente a diventare nullo.

Un'altra informazione importante in input per la rete neurale è la composizione del jet, perciò è stato ritenuto necessario inserire il PDGID dei costituenti tra le variabili selezionate.
Inserire il PDGID, che è una variabile categorica, come semplice variabile all'interno della rete avrebbe portato ad una confusione del modello. Se rappresentiamo infatti il PDGID semplicemente come numeri interi, la rete interpreterebbe questi numeri come aventi un ordine o una scala, cosa che non ha senso nel contesto delle particelle. Ad esempio, confrontando il PDGID = 11 (elettrone) e il PDGID = 13 (muone) il modello potrebbe erroneamente individuare un muone come se fosse "più particella" di un elettrone.
Per evitare questa ambiguità, è stato necessario effettuare *one-hot encoding* sul PDGID delle particelle presenti nei jet. Questo approccio garantisce che il modello tratti ogni tipo di particella come una categoria distinta senza assumere relazioni ordinali inesistenti. Fare *one-hot encoding* significa rappresentare le n tipologie di particelle come un vettore binario n-dimensionale le cui componenti sono tutte nulle, fatta eccezione per quelle in corrispondenza della particella che è presente nel jet, a cui è assegnato il valore 1. Questo processo è stato ripetuto per ogni costituente di ogni jet, costituente che è perciò identificato da un vettore pentadimensionale (130 = Kaoni, 211 = pioni, 22 = fotoni, 11 = elettroni, 13 = muoni) con una sola entrata non nulla, corrispondente al tipo di particella a cui è associata. Questo implica un totale di $16 \times 5 = 80$ features aggiuntive in input al modello.

Infine commentiamo l'analisi relativa ai jet contenenti leptoni. Una variabile ritenuta significativa che è stata inserita tra le features di input è il numero di muoni presenti nel jet. Infatti, come già analizzato in sezione 3.3, risulta più frequente la presenza di più di un muone in jet da b piuttosto che in jet leggeri o c.
È stata invece fatta la scelta di escludere dagli input le informazioni relative al p_T^{rel} sia per gli elettroni che per i muoni. Per gli elettroni era già stato commentato che le distribuzioni del p_T^{rel} non fossero sufficientemente caratteristiche per poter rappresentare una variabile significativa. Per i muoni invece il p_T^{rel} avrebbe potuto svolgere un ruolo importante nella classificazione: il problema che sorge è però la bassissima percentuale di jet nei quali è stata riscontrata la presenza di muoni, motivo per cui non ci aspettiamo che l'aggiunta di una variabile legata ai muoni possa influenzare in maniera significativa

il risultato dell'addestramento della rete, anche perchè facilmente molto correlata con ΔR e $p_T/p_{T,jet}$ del muone in questione, se tra i primi 4 costituenti del jet.

4.2.2 Definizione del target per la classificazione

L'obiettivo di questo lavoro di tesi è, come già chiarito in precedenza, la scrittura e il test di un algoritmo di machine learning di classificazione binaria che possa effettuare il l'identificazione dei jet provenienti da quark b e la loro separazione dalla categoria di jet più facilmente presenti nelle collisioni, e cioè i jet prodotti da quark leggeri. In particolare il focus del lavoro verte sul riuscire a classificare i jet provenienti dall'adronizzazione del quark b rispetto a quelli generati da adronizzazione di jet leggeri o gluoni.
Nonostante ciò, l'analisi dei dati è stata svolta su un dataset contenente anche i jet da quark c, che, a fronte della loro massa e vita media, hanno caratteristiche intermedie rispetto alle altre due tipologie di jet, in alcuni casi con una leggera tendenza a somigliare ai jet da b. Possiamo perciò supporre che i jet da c costituiscano una vera e propria categoria a sé stante, che sarebbe forzato inglobare nei jet leggeri definendo un classificatore binario che distingua i jet b dai *non* b.
Per questo motivo è stato deciso in ultima analisi di non inserire i jet da c all'interno del dataset di training della rete neurale: non solo sarebbe stato forzato includerli all'interno dei jet leggeri, ma avrebbero potuto potenzialmente causare confusione in fase di addestramento nell'individuazione delle features proprie dei jet b, a causa della loro somiglianza sotto certi aspetti con i jet c.
Il target della classificazione dei jet è stato quindi scelto come binario: la classe 0 rappresenta i jet da quark leggeri, e la classe 1 rappresenta i jet da quark b. I jet da quark c sono esclusi dall'addestramento, ma sono poi stati inclusi a posteriori negli studi di valutazione delle performance del tagger su dati mai osservati dalla rete nelle fasi di addestramento.

4.3 Struttura della rete neurale

4.3.1 Rete feed-forward: ottimizzazione di iperparametri

Dopo aver concluso lo studio delle variabili più informative riguardo la natura dei jet, è stato definito uno scheletro di rete neurale feed-forward della quale si è scelto di ottimizzare alcuni dei parametri, in maniera tale da avere un algoritmo il più efficiente possibile.
L'algoritmo è stato scritto in linguaggio Python utilizzando la libreria PyTorch.[11]
Gli iperparametri che sono stati ottimizzati sono i seguenti:

- *grandezza del batch*: indica il numero di campioni di addestramento che vengono passati al modello prima che i gradienti vengano calcolati e i pesi del modello vengano aggiornati;
- *numero di layer nascosti*: indica il numero di strati presenti tra lo strato di input e quello di output;
- *numero di nodi per layer nascosti*: è il numero di neuroni presenti in ogni strato nascosto;
- *learning rate*: è un parametro di regolazione in un algoritmo di ottimizzazione che determina la dimensione del passo ad ogni iterazione mentre ci si dirige verso un minimo della funzione di perdita;
- *percentuale di dropout*: è la frazione di neuroni che vengono disattivati casualmente durante l'addestramento per prevenire l'overfitting e migliorare la generalizzazione del modello.

L'ottimizzazione di questi cinque iperparametri è stata effettuata mediante la libreria *Optuna* [12], addestrando e testando la rete al variare dei parametri in determinati intervalli, per un totale di 150 iterazioni. In figura 4.3 è possibile vedere le combinazioni di parametri testate e gli intervalli nei quali essi sono stati scelti.
La metrica di valutazione scelta per determinare i parametri migliori è l'*Area Under the Curve* (AUC), metrica che verrà discussa più in dettaglio in seguito, ma che è tanto migliore quanto più il suo valore è prossimo ad 1.

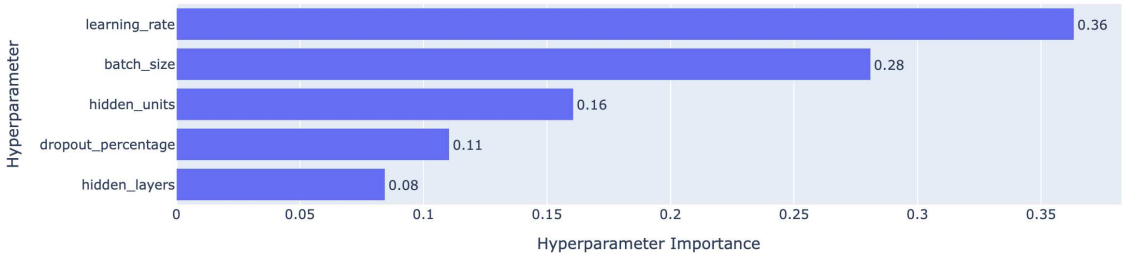

Figura 4.2: Importanza degli iperparametri nel processo di ottimizzazione

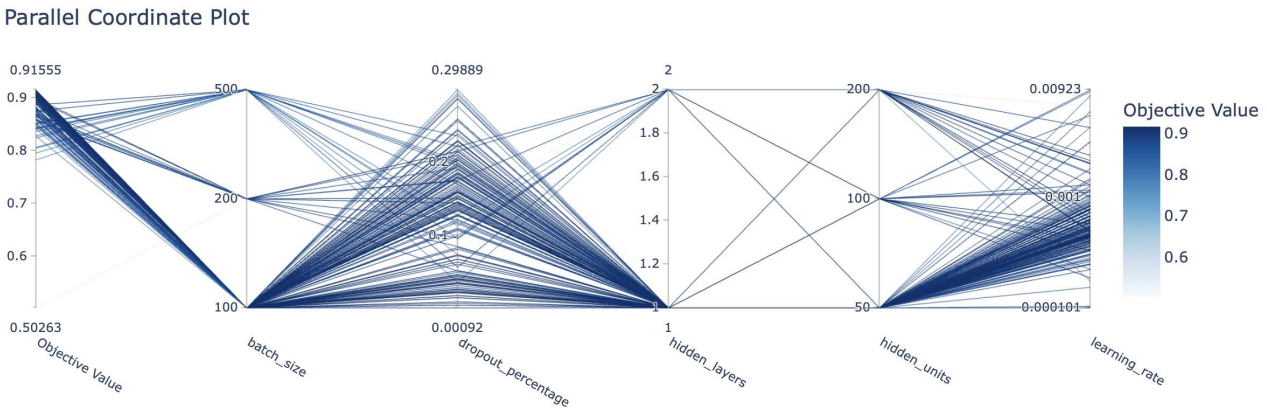

Figura 4.3: Grafico a coordinate parallele. Ogni percorso dato da una linea spezzata corrisponde a un tentativo di ottimizzazione. I parametri *Learning Rate* e *Dropout Percentage* sono stati campionati in intervalli continui di valori, mentre tutti gli altri parametri in insiemi discreti.

Dal grafico in figura 4.3 è possibile trarre informazioni anche sul peso che ogni parametro ha nell'ottimizzazione. Si vede ad esempio che la percentuale di dropout non ha grande rilevanza, dato che vengono provati più o meno tutti valori nell'intervallo senza prediligere un intervallo più stretto. Ciò invece non vale per il learning rate, per il quale dopo alcuni tentativi l'ottimizzatore ha ritenuto opportuno concentrarsi su un intervallo di valori ben preciso e più ristretto. Ciò è confermato anche dal grafico che il software ci offre riguardo all'importanza degli iperparametri riportato in figura 4.2.

L'ottimizzazione ha prodotto i risultati in tabella 4.1, corrispondenti al modello che massimizza la figura di merito scelta per lo studio.

Parametro	**Valore**
AUC	0.9156
Dropout Percentage	0.0399
Batch Size	100
Learning Rate	0.00037
Hidden Layers	1
Hidden Units	50

Tabella 4.1: Migliori iperparametri stimati

La rete definitiva presenta perciò un singolo layer nascosto con 50 nodi con attivazione ReLU e un layer di output attivato da una sigmoide[1]. L'ottimizzatore scelto è stato *Adam*, mentre la Loss Function una *Binary Cross Entropy with logits*, che risultano essere le scelte più indicate per una rete per la classificazione binaria.

[1]L'attivazione sigmoide del layer di output è in realtà inclusa nella funzione di perdita *BCEwithLogits* di PyTorch.

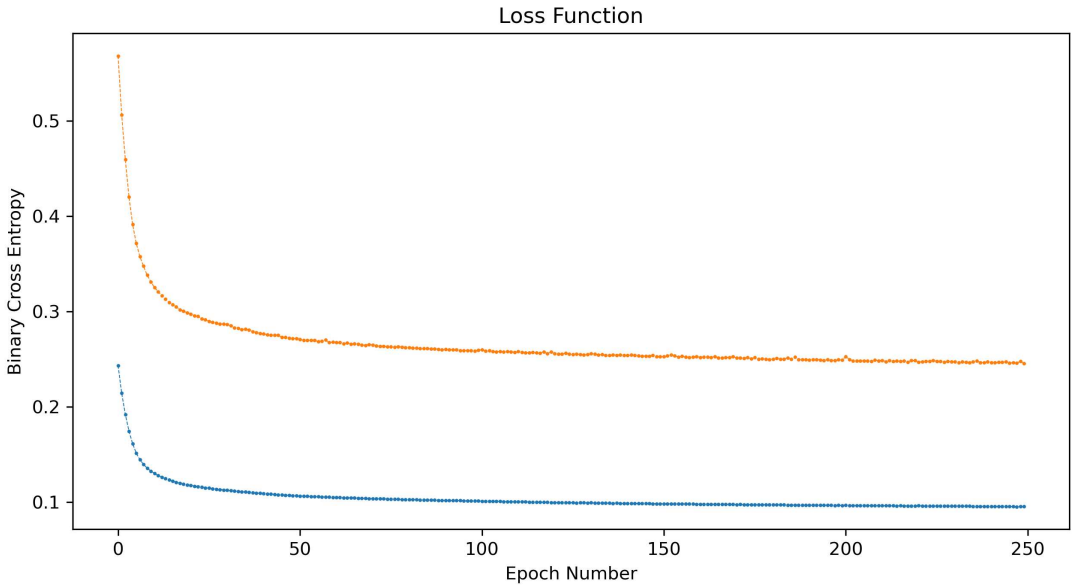

Figura 4.4: Plot della funzione di perdita dell'addestramento (azzurro) e della validazione (arancione)

Il dataset è stato inoltre pesato per permettere alla rete di avere una buona rappresentazione della classe b che è statisticamente sottorappresentata nel campione a disposizione.

I risultati di questa rete sono descritti nella sezione 5.1.1.

4.3.2 Addestramento e validazione della rete

Una volta definiti i parametri principali della rete e le variabili di input si è proceduto alla fase di addestramento della rete neurale. L'addestramento è stato effettuato su un totale di 250 epoche, fornendo i dati di addestramento in batch di dimensione 100. Parallelamente all'addestramento è stata portata avanti anche la validazione della rete. Parte del dataset di test è stata perciò riservata a validare l'addestramento della rete epoca per epoca, cioè effettuare una vera e propria fase di test alla fine di ogni epoca, per monitorare eventuali situazioni di overfitting dei dati.

È possibile infatti che una rete possa apprendere troppo bene i dati che vengono forniti in addestramento e non si riesca poi ad ottenere un modello generale che possa classificare un dataset nuovo. Il plot delle funzioni di perdita in figura 4.4 ci dà due importanti informazioni:

- La loss function di addestramento tende molto rapidamente a zero, perciò la rete sta effettivamente "imparando" qualcosa dal training;
- Anche la funzione di perdita della validazione tende a diminuire verso un certo valore di minimo, ovviamente maggiore di quello della funzione di perdita di training. Se ci fossimo ritrovati in una situazione in cui la funzione di perdita di validazione avesse subito una risalita verso l'alto dopo un certo numero di epoche, avremmo potuto concludere di essere in una situazione di overfitting, in cui il modello presenta dei buoni risultati in addestramento, ma non riesce a fare alrettanto in fase di test.

4.3.3 Studio preliminare per l'implementazione di una Convolutional Neural Network (CNN)

Un'altra direzione in cui si è pensato di muoversi è quella delle reti neurali convoluzionali (CNN), tipicamente utilizzate per il processamento di matrici di dati, come ad esempio i pixel di un'immagine. Nel caso in questione, i jet sono stati riorganizzati in forma matriciale: ogni jet è stato rappresentato con un matrice 16×11 dove 11 sono le features in input (che vediamo in tabella 4.2) e 16 sono i

costituenti dei jet, con un valore di padding impostato a -10 per tutte le entrate vuote delle matrici associate a jet con un numero di costituenti inferiore a 16.

γ	K	μ	e	π	p_T	η	ϕ	ΔR	$\Delta p_{T,i}/p_T$	carica elettrica	jet 1
0	0	0	0	1	0.355	-0.001	0.5	0.1	0.02	-1	1º costituente
1	0	0	0	0	0.338	0.029	1.2	0.2	0.05	0	2º costituente
0	1	0	0	0	0.079	-0.049	2.5	0.15	0.1	1	3º costituente
0	0	0	0	1	0.065	0.020	0.8	0.05	0.02	-1	4º costituente
0	0	1	0	0	0.032	-0.027	1.0	0.3	0.08	-1	5º costituente
0	0	0	1	0	0.029	0.064	1.5	0.2	0.04	-1	6º costituente
0	0	0	0	1	0.019	0.208	0.7	0.1	0.03	1	7º costituente

Tabella 4.2: Esempio di funzionamento di un kernel convoluzionale quadrato

Il funzionamento dei layer convoluzionali in una rete può essere compreso attraverso gli schemi riportati nelle tabelle 4.2 e 4.3. In una rete convoluzionale, l'intera stringa di dati associata al jet non viene processata simultaneamente; piuttosto, la rete riceve in input singoli pacchetti di informazioni, la cui dimensione è determinata dalle dimensioni del kernel convoluzionale. Questo approccio consente all'algoritmo di creare correlazioni locali tra i costituenti del jet, potenzialmente migliorando il potere discriminante della rete.

È stata fatta la scelta di utilizzare un kernel rettangolare, di dimensione $2 \times n_{features}$, come mostrato in tabella 4.3, con un passo di 2, cioè il kernel si muove di due righe alla volta verso il basso. L'idea è quella di dare in input alla rete tutte le informazioni su due costituenti alla volta, con l'obiettivo di calcolare correlazioni tra due costituenti successivi[2].

γ	K	μ	e	π	p_T	η	ϕ	ΔR	$\Delta p_{T,i}/p_T$	carica elettrica	jet 1
0	0	0	0	1	0.355	-0.001	0.5	0.1	0.02	-1	1º costituente
1	0	0	0	0	0.338	0.029	1.2	0.2	0.05	0	2º costituente
0	1	0	0	0	0.079	-0.049	2.5	0.15	0.1	1	3º costituente
0	0	0	0	1	0.065	0.020	0.8	0.05	0.02	-1	4º costituente
0	0	1	0	0	0.032	-0.027	1.0	0.3	0.08	-1	5º costituente
0	0	0	1	0	0.029	0.064	1.5	0.2	0.04	-1	6º costituente
0	0	0	0	1	0.019	0.208	0.7	0.1	0.03	1	7º costituente

Tabella 4.3: Esempio di funzionamento di un kernel convoluzionale rettangolare

La struttura scelta per la rete è la seguente:

- Un layer convoluzionale con kernel 2×11 e stride 2, con 1 canale di input e 4 di output. La funzione di attivazione per il primo layer è la $Tanh$, per evitare di perdere informazioni sui valori negativi.
- tre layer fully connected con, rispettivamente, 200, 10, e 1 nodo di output, attivati da una funzione $ReLU$ e una sigmoide in uscita. È stata scelta una percentuale di dropout del 20%, e gli stessi pesi per regolare lo sbilanciamento delle classi utilizzati per la rete feed forward.

La funzione di perdita e l'ottimizzatore sono gli stessi della rete feed-forward: *BCEwithLogits* e *Adam*.

Lo studio dell'architettura di questa seconda rete non è stato approfondito nel dettaglio, ed è fornita come un primo tentativo di implementazione di una tecnica alternativa a quella delle reti feed-forward descritte nel paragrafo 4.3.1. I risultati, anche se preliminari, di questa implementazione di CNN per il problema della classificazione di b-tagging sono presentati nel capitolo 5.1.2.

[2]Si ricorda che i costituenti sono ordinati per p_T decrescente

Capitolo 5

Discussione dei risultati

5.1 Metriche di valutazione delle prestazioni della rete

5.1.1 Rete feed-forward

Una volta terminato l'addestramento della rete, è necessario misurarne le prestazioni testandola su un dataset diverso da quelli usati in addestramento e validazione per verificarne il vero potere di classificazione confrontando l'ouput della rete con la verità Monte Carlo associata agli input.
Quando viene dato in input ad una rete di classificazione binaria un vettore da classificare, la rete restituisce in output un numero compreso tra 0 ed 1. Questo coefficiente è interpretabile come una probabilità: più esso è vicino ad 1, maggiore è la probabilità calcolata dall'algoritmo che l'oggetto in questione appartenga alla classe positiva, nel nostro caso la classe dei jet b. Ad ogni elemento in fase di test viene perciò assegnata una diversa probabilità di appartenere alla classe dei jet b: sta a noi scegliere una soglia di probabilità per la quale un jet venga classificato come jet b o leggero. La scelta dei punti di lavoro svolge un ruolo fondamentale nell'ambito del b-tagging e verrà approfondita nella sezione 5.2. Di seguito analizziamo le metriche principali per valutare le prestazioni della rete.

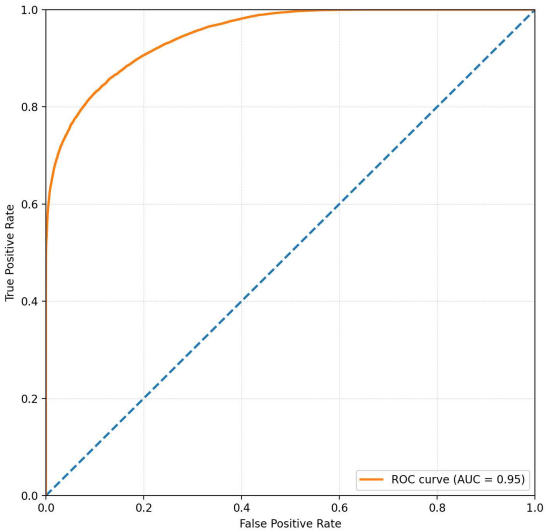

Figura 5.1: ROC curve della rete feed-forward ottenuta in fase di test. L'area sotto la curva (AUC) è pari a 0.95.

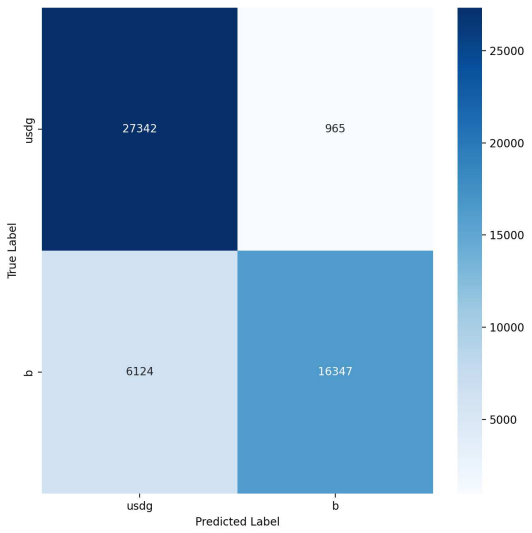

Figura 5.2: Matrice di confusione della rete feed-forward in fase di test. L'intensità del colore corrisponde a una maggiore popolazione dell'entrata.

Innanzitutto, una volta fissata una soglia per la classificazione e nota la verità Monte Carlo, è necessario definire quattro tipologie di classi:

- Veri positivi (TP): l'insieme dei jet b che viene correttamente individuato come tale;
- Veri negativi (TN): l'insieme dei jet leggeri che viene correttamente individuato come *non* b;
- Falsi positivi (FP): l'insieme dei jet leggeri che viene erroneamente individuato come jet b;
- Falsi negativi (FN): l'insieme dei jet b che vengono erroneamente individuati come *non* b;

Sulla base di questo tipo di classificazione viene costruita la matrice di confusione (confusion matrix) che troviamo in figura 5.2. Tramite questa metrica è possibile visualizzare l'operato del classificatore una volta fissata la soglia di probabilità per il tagging. Nel caso in figura la soglia è stata arbitrariamente impostata a 0.5. Ogni colonna della matrice rappresenta i valori predetti, mentre ogni riga rappresenta i valori reali. L'elemento sulla riga i e sulla colonna j è il numero di casi in cui il classificatore ha classificato la classe "vera" i come classe j. Il caso ideale sarebbe rappresentato da una matrice perfettamente diagonale, con entrate nulle in corrispondenza dell'antidiagonale, ovvero dei falsi positivi e falsi negativi. Nel caso reale, l'obiettivo è quello di minimizzare il numero di falsi positivi e falsi negativi, cioè diminuire al massimo la confusione del modello. La matrice di confusione è una delle prime conferme del corretto funzionamento della rete, mostrando chiaramente che i casi in cui il classificatore commette errori sono molto minori di quelli in cui esso restituisce la risposta corretta.

La Receiver Operating Characteristic curve (ROC curve) è una metrica di valutazione fondamentale per i classificatori binari. In figura 5.1 è riportata la ROC curve per la rete feed-forward discussa in precedenza. Sull'asse delle ordinate è riportato il tasso di veri positivi, mentre sull'asse delle ascisse il tasso di falsi positivi.

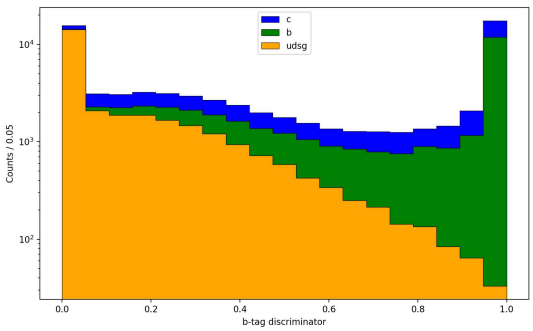

Figura 5.3: Istogrammi impilati delle probabilità predette dal classificatore per jet generati da adronizzazione di quark leggeri (giallo), b (verde) o c (blu). L'asse delle ordinate è in scala logaritmica.

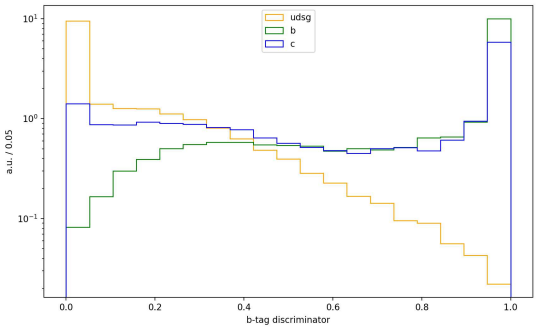

Figura 5.4: Istogrammi sovrapposti delle probabilità predette dal classificatore per jet generati da adronizzazione di quark leggeri (giallo), b (verde) o c (blu). Le aree degli istogrammi sono normalizzate all'unità, l'asse delle ordinate è in scala logaritmica.

La costruzione di una ROC curve avviene variando la soglia del discriminatore in un insieme molto denso di punti nell'intervallo [0;1] e riportando nel grafico il tasso di falsi positivi e veri positivi. A posteriori perciò la curva tracciata rappresenta il numero dei veri positivi che ci aspettiamo di ottenere fissando un valore di falsi positivi che riteniamo accettabile. La riga diagonale rappresenta invece il comportamento di un algoritmo che, indipendentemente dalla soglia applicata sul discriminatore, presenta sempre una rate identica di falsi positivi e di veri positivi, rappresentando quindi l'incapacità del classificatore a distinguere le due classi. Al contrario, quanto più la ROC curve è rigonfiata verso l'angolo in alto a sinistra, allontanandosi dalla diagonale, maggiore sarà l'affidabilità delle rete nella classificazione.

Il parametro più importante per valutare la prestazione di una rete binaria è infatti quello dell'Area Under the Curve (AUC) che ci permette di quantificare il rigonfiamento della curva verso l'angolo superiore sinistro tramite il calcolo di un'area. Il caso ideale di classificatore perfetto corrisponderebbe all'avere un'area sotto la curva unitaria, cioè ad avere una ROC curve che ricalchi i lati superiore e sinistro del quadrato in cui è compresa: ciò significherebbe poter impostare una soglia per la rete neurale tale per cui si possano azzerare i falsi positivi e avere il 100% di veri positivi, una rete che non commette errori. Nel caso reale perciò una rete è tanto migliore quanto più il valore della AUC si avvicina ad 1. Nel caso in questione, la AUC è pari a 0.95, si ha perciò un classificatore molto efficiente.

La figura 5.3 mostra la distribuzione delle probabilità predette dal classificatore sul dataset di test in tre istogrammi impilati, uno per classe. Il grafico in figura 5.4 rappresenta invece i tre istogrammi sovrapposti, in scala logaritmica, con aree degli istogrammi normalizzati all'unità. In giallo sono rappresentati i jet leggeri, ai quali riconosciamo essere assegnata una probabilità generalmente bassa di appartenere alla classe dei b. Sottolineiamo la scala logaritmica dei grafici, per cui l'area nella parte superiore del grafico ha un peso molto maggiore rispetto a ciò che è nella regione inferiore. Per i jet da b la situazione è opposta: le probabilità loro assegnate tendono all'unità. Ciò conferma quanto visto in precedenza con la matrice di confusione e la ROC curve: il classificatore svolge con elevata accuratezza il suo ruolo di discriminare jet da quark leggeri e jet da b.
Una discussione a parte invece è dedicata alla classe dei jet da c (in blu nelle immagini). Il classificatore, infatti, non è stato addestrato per riconoscere questa classe di jet ed effettivamente il risultato è una distribuzione di probabilità abbastanza omogenea, ad eccezione di una leggera accumulazione di conteggi attorno alla probabilità più alta. Ciò conferma l'ipotesi fatta in fase di addestramento per la quale si è scelto di escludere i jet da c dal training: la classe dei jet da c è a tutti gli effetti una categoria che non è corretto inglobare in nessuna delle altre due classi, nonostante vi siano alcuni jet con caratteristiche molto simili a quelle dei jet da b.

5.1.2 Rete neurale convoluzionale

In questa sezione discuteremo i risultati ottenuti dal modello di rete convoluzionale, costruita ad uno stadio preliminare nell'ottica di futuri studi e miglioramenti.

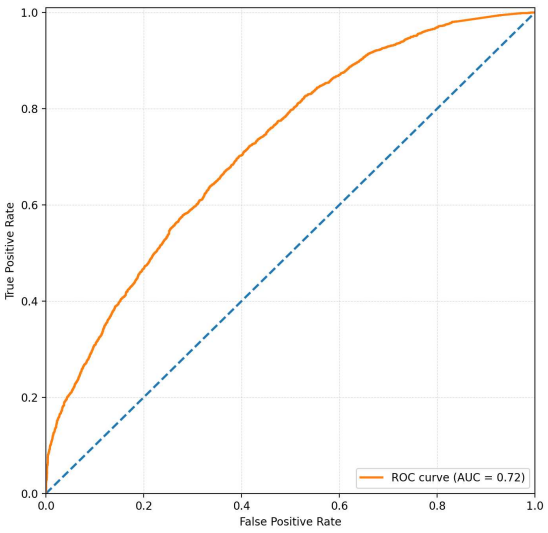

Figura 5.5: ROC curve associata alla rete convoluzionale in fase di test. L'area sotto la curva è pari a 0.72.

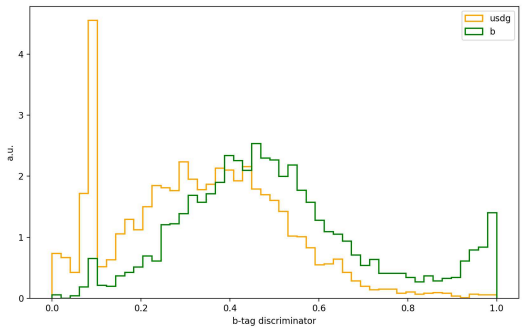

Figura 5.6: Istogrammi sovrapposti delle probabilità predette dalla rete neurale convoluzionale per jet leggeri (giallo) e b (verde). Le aree degli istogrammi sono normalizzate all'unità.

I risultati della CNN non sono stati altrettanto validi: seppure il modello dimostri una certa capacità di discriminazione delle due classi, dalla ROC curve e dalla AUC (in figura 5.5) si vede che le prestazioni della rete convoluzionale sono peggiori di quelle della rete feed-forward.
Da un'analisi delle distribuzioni di probabilità assegnate dal discriminatore ai jet in fase di test, riportate in figura 5.6, si può ottenere una conferma della non eccelsa capacità di classificazione della rete, se confrontata con la rete feed-forward.
Tuttavia, le distribuzioni delle probabilità delle due classi appaiono come un risultato promettente: è chiara la tendenza ad assegnare probabilità maggiori a jet da b ed inferiori ai jet leggeri. Questo lascia presupporre che uno studio futuro più approfondito del modello e una modifica nella struttura della rete possano aprire la strada a risultati alla pari o migliori di quelli ottenuti con la rete feed-forward.

5.2 Scelta dei punti di lavoro

La scelta dei punti di lavoro, cioè definire una soglia di probabilità per il b-tagging, svolge un ruolo cruciale nell'utilizzo di un modello di classificazione.

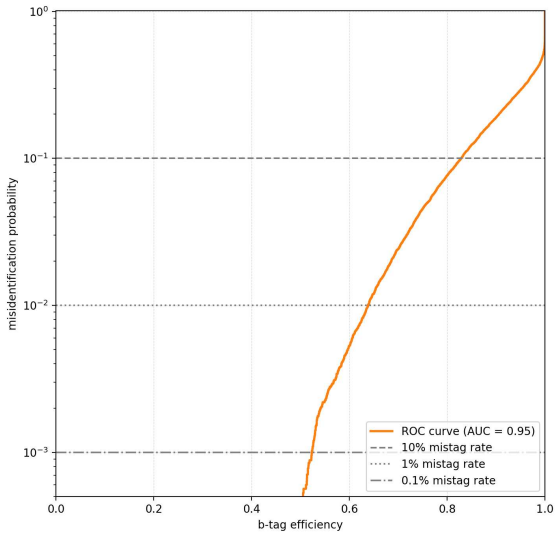

Figura 5.7: Ulteriore tipologia di ROC curve, tipicamente usata a CMS, associata alla rete feed-forward in fase di test. Sull'asse delle ordinate, in scala logaritmica, viene posta la probabilità di identificare erroneamente un jet leggero come jet da b; sull'asse delle ascisse troviamo l'efficienza nel b-tagging, ovvero la percentuale di jet da b che effettivamente vengono riconosciuti come tali.

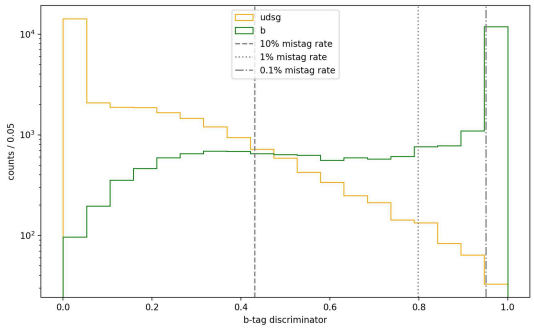

Figura 5.8: Scelta dei punti di lavoro a partire dagli istogrammi di distribuzione delle probabilità. La scala dell'asse delle ordinate è logaritmica. Le rette verticali rappresentano tre diversi punti di lavoro, rispettivamente, da sinistra a destra, Loose, Medium e Tight, che corrispondono al 10%, 1%, 0.1% di probabilità di avere un falso positivo.

Nella sezione 5.1.1 abbiamo visto una matrice di confusione realizzata fissando una soglia di probabilità del 50%, totalmente arbitraria, per l'assegnazione delle etichette da parte del classificatore, ma questa non è l'unica via percorribile, anzi, non è la maniera con la quale i discriminatori di questi algoritmi sono utilizzati nei contesti della fisica delle alte energie.

Ad esempio, la diminuzione di questa soglia permette di accrescere il numero di veri positivi, al costo di aumentare anche il tasso di falsi positivi.
Nell'ambito del b-tagging a CMS è importante definire dei punti di lavoro sulla base della probabilità di ottenere un falso positivo. Si distinguono solitamente 3 punti di lavoro, corrispondenti a 3 livelli ammessi di misidentificazione di jet leggeri come jet provenienti dall'adronizzazione di quark pesanti. I punti di lavoro sono solitamente nominati Loose, Medium e Tight, in corrispondenza di tassi di

misidentificazione del 10%, 1% e 0.1% rispettivamente. Queste tre classi sono poi utilizzate nei diversi contesti delle analisi e ricerche di fisica all'esperimento CMS.
Ad esempio, nell'idea di selezionare un campione esteremamente puro di jet da quark b, è necessario applicare una soglia sul discriminatore di b-tagging molto stringente, che produca una misidentificazione tipicamente a livello del permille (cioè il punto di lavoro "Tight"). D'altro canto, se fosse ad esempio necessario porre un veto sulla presenza di jet da b nei prodotti della collisione, sarebbe possibile applicare una selezione con una probabilità di misidentificazione piu' larga, a livello del 10% (cioè il punto di lavoro "Loose").

Nell'immagine 5.7 vediamo una diversa tipologia di ROC curve in cui sono evidenziate delle soglie corrispondenti alla percentuale di eventi identificati erroneamente come "da b" e dalla quale è stato possibile ottenere i risultati riportati in tabella 5.1.

Punto di Lavoro	**Tasso di misidentificazione**	**Soglia**	**Efficienza di b-tag**
Loose	10%	0.431	$\sim 83\%$
Medium	1%	0.799	$\sim 64\%$
Tight	0.1%	0.951	$\sim 52\%$

Tabella 5.1: Soglie per il tasso di misidentificazione, corrispondente efficienza di b-tag e punto di lavoro associato.

Ciò significa che, ad esempio, imponendo al modello di identificare come jet da b solo i jet alla quale viene associata una probabilità superiore alla soglia di 0.951 di appartenere alla classe positiva, si riuscirebbero a individuare e a salvare il 52% degli eventi provenienti da adronizzazione di quark b, al prezzo di una percentuale dello 0.1% di eventi falsi positivi.
Questo processo è fondamentale, perchè consente di scegliere a seconda dell'esigenza quale percentuale di eventi positivi salvare e la corrispondente percentuale di eventi misidentificati.

Una visualizzazione più immediata della scelta dei punti di lavoro è riportata in figura 5.8: tutto ciò che è a destra della retta di soglia scelta viene identificato come jet da b.

Capitolo 6

Conclusioni

La presente tesi ha esplorato in dettaglio il concetto di b-tagging all'interno dell'esperimento CMS, evidenziando il ruolo cruciale di questa tecnica nell'identificazione dei jet generati dall'adronizzazione dei quark pesanti. La fase 2 dell'aggiornamento di CMS consentirà l'implementazione di algoritmi di machine learning, come quelli discussi in questo progetto di tesi, a livello del trigger L1 tramite schede di elettronica programmabili.

È stato utilizzato un dataset simulato di collisioni protone-protone a HL-LHC nell'esperimento CMS, in cui sono stati studiati i prodotti di decadimento di quark top per lo sviluppo di algoritmi di b-tagging per il Level-1 trigger dell'esperimento CMS. Sono state poi studiate le variabili del dataset identificando quelle più promettenti per l'identificazione di jet provenienti dall'adronizzazione di quark pesanti. Successivamente è stato sviluppato un algoritmo di classificazione supervisionata, ed è stata usata la verità Monte Carlo delle simulazioni per addestrare il modello a separare jet da adronizzazione di quark b e jet da quark leggeri.

I risultati ottenuti dagli algoritmi sviluppati e testati nel presente progetto di tesi mostrano che la rete feed-forward, discussa dettagliatamente nel capitolo 4, si presenta come un valido candidato per l'implementazione in CMS. Le prestazioni della rete risultano infatti molto valide, come deducibile dalla curva ROC e dal valore della AUC di 0.95, estremamente prossimo a 1, oltre a garantire un'ampia libertà nella scelta dei punti di lavoro, che possono rendere l'efficienza nel b-tagging della rete regolabile in base alle esigenze dell'esperimento.
Questo risultato apre la strada a un ulteriore miglioramento delle prestazioni della rete, oltre a un'espansione delle capacità del sistema. In particolare, il classificatore, attualmente binario, potrebbe essere ulteriormente potenziato per diventare un classificatore a più classi. Questo miglioramento permetterebbe di distinguere non solo tra i jet b e quelli leggeri, ma anche di identificare specificamente altre categorie di jet, come quelli provenienti da quark charm, le cui proprietà sono state analizzate e discusse in questo stesso lavoro, aumentando così la precisione dell'analisi complessiva.

È stato poi prodotto uno studio preliminare di un secondo modello, basato su una rete neurale convoluzionale, descritto in sezione 4.3.3, usando un kernel rettangolare di dimensione $2 \times n_{features}$ e stride 2. Lo studio ha prodotto risultati preliminari di AUC di 0.72. Le prestazioni di questo modello sono quindi buone ma attualmente meno performanti di quelle della rete feed forward.
Tuttavia, è d'obbligo sottolineare che lo studio della CNN portato avanti in questa tesi fosse uno studio preliminare, con lo scopo di tracciare una possibile direzione per lo sviluppo degli algoritmi di machine learning per il b-tagging a CMS.
Sotto questo aspetto, i risultati si sono rivelati molto promettenti: dalle distribuzioni di probabilità assegnate dalla rete convoluzionale si evince un interessante potenziale nella capacità di discriminazione della rete, che lascia pensare che studi più approfonditi del modello e un eventuale riadattamento della struttura della rete possano permetterle di ottenere prestazioni migliori.

Bibliografia

[1] CMS Collaboration. *The Compact Muon Solenoid Technical Proposal*. Rapp. tecn. CERN-LHCC-94-38. CERN, 1994. URL: http://cdsweb.cern.ch/record/290969.

[2] Aidan D. Chambers. "Neural-Network Based b-Tagging in the CMS Level 1 Trigger System". Massachusetts Institute of Technology, giu. 2023.

[3] The CMS Collaboration et al. "JINST 3 S08004". In: *Journal of Instrumentation* 3 (2008). DOI: 10.1088/1748-0221/3/08/S08004.

[4] G.L. Bayatian et al. "CMS Physics Technical Design Report: Addendum on High Density QCD with Heavy Ions". In: (gen. 2007).

[5] *The Phase-2 Upgrade of the CMS Level-1 Trigger Technical Design Report*. Rapp. tecn. CERN-LHCC-2020-004. CERN, 2020. URL: https://cds.cern.ch/record/2714892.

[6] Claire Savard. "Overview of the HL-LHC Upgrade for the CMS Level-1 Trigger". In: *EPJ Web of Conferences*. CHEP 2023, on behalf of the CMS Collaboration. University of Colorado, Boulder. 2024. DOI: 10.1051/epjconf/202429502022.

[7] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011.

[8] Izaak Neutelings. *CMS Coordinate System*. https://tikz.net/axis3d_cms/. 2021.

[9] S. Navas et al. "Review of Particle Physics". In: *Phys. Rev. D* 110.3 (2024). 59 citations counted in INSPIRE as of 21 Aug 2024, p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[10] L. Grippo e M. Sciandrone. *Metodi di ottimizzazione per le reti neurali*. Dispense del corso, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza". URL: mailto:grippo@dis.uniroma1.it.

[11] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: 1912.01703 [cs.LG].

[12] Takuya Akiba et al. "Optuna: A Next-generation Hyperparameter Optimization Framework". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019.