

Optimizacija algoritma za identifikaciju hadronskih mlazova nastalih raspadom ultrarelativističkih Higgsovih bozona u dva b kvarka

MATEJ ROGULJIĆ

20. siječnja 2017.

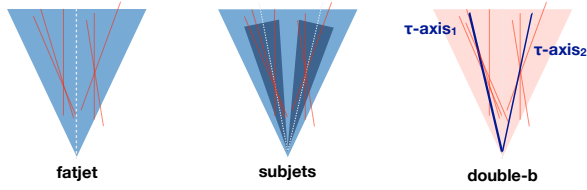
Sažetak

Proučavane su performanse algoritma za identifikaciju hadronskih b mlazova sa i bez primjene PUPPI algoritma za uklanjanje doprinosa dodatnih neelastičnih proton-proton sudara. Treniranje algoritma izvedeno je pomoću paketa TMVA koristeći kao signal simulirane procese raspada ultrarelativističkih Higgsovih bozona i kao pozadinu simulaciju hadronskih mlazova produciranih procesima kvantne kromodinamike. Motivacija za odabir ovog kanala raspada je činjenica da vezanje Higgsovog bozona na fermione još uvijek nije potvrđeno, a ultrarelativistički režim omogućava dodatno povećanje omjera signala i pozadine. Utvrđeno je da primjena PUPPI algoritma prilikom određivanja okusa hadronskih mlazova neznatno poboljšava diskriminatorsku moć algoritma.

1 Potraga za Higgsovim bozonom

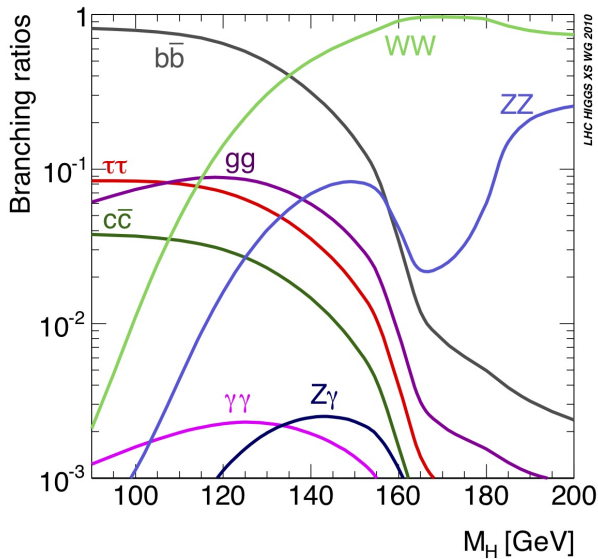
Higgsov bozon elementarna je čestica Standardnog modela, pobuđenje Higgsovog polja, čije se postojanje predviđalo još od 60-ih godina 20. stoljeća [1, 2, 3, 4, 5]. Do 2013. godine to je bila jedina nepotvrđena čestica Standardnog modela. Tada je na CERN-u potvrđen pronalazak nove čestice mase 125 – 127 GeV u kanalima raspada u dva fotona i kanalima raspada u dva Z bozona koji se naknadno raspadaju u elektrone ili mione. Daljnje su analize potrebne da bi se potvrdilo s kojim se modelom Higgsovog bozona mjerenja najbolje slažu. Interakcije Higgsovog bozona s fermionima opisuju se Yukawinom interakcijom i mogu se izmjeriti proučavanjem raspada Higgsovog bozona na fermione. Pogledom na omjere grananja raspada Higgsovog bozona (Slika 2) moglo bi se pomisliti da je to, budući da najveći omjer grananja ima raspad u $b\bar{b}$ kvarkove, relativno "lako" izmjeriti, ali do sada nije opažen Higgsov bozon u tom kanalu s pouzdanošću od

5 sigma. Razlog tomu je što postoji relativno mnogo drugih procesa čiji je čestični sastav konačnog stanja isti ("pozadina"). Količina pozadine smanjuje se na nekoliko načina. Ukoliko se ograničimo na ultrarelativistički režim Higgsovog bozona (npr. $p_T > 300$ GeV), odbacit ćemo mnogo raspada Higgsovih bozona, ali i relativno puno više pozadine [6]. Ultrarelativistički režim također dovodi i do zanimljivog ponašanja hadronskih mlazova. Iz sustava mirovanja Higgsovog bozona, b kvarkovi nastali raspadom imat će impulse jednakog iznosa i suprotnog smjera, ali to ne mora biti slučaj u laboratorijskom sustavu. Naime, zbog očuvanja impulsa, b kvarkovi (a zatim i hadronski mlazovi) do neke će mjere biti kolimirani. U ultrarelativističkom režimu to dovodi do djelomičnog preklapanja hadronskih mlazova (prikazano na Slici 1) što zahtijeva promjenu načina rekonstrukcije hadronskih mlazova. Tada se umjesto rekonstruiranja dva mlaza može konstruirati jedan veliki (tzv. "Fat Jet") koji u sebi sadrži i podatke o podstrukturi. U ovoj su se analizi upravo koristili mlazovi rekonstruirani na taj način.



Slika 1: Prikaz stvaranja "Fat" hadronskog mlaza. Preuzeto iz [7]

Za smanjivanje pozadine, osim odlaska u ultrarelativistički režim, postoje i algoritmi koji imaju sposobnost (do neke mjere) razlučiti radi li se o događaju koji je nama zanimljiv (npr. raspad Higgsovog bozona u $b\bar{b}$) ili ne. Razvoj i optimizacija takvih algoritama bitni su da bi se dobile što bolje diskriminatorne sposobnosti. U ovom se radu testira hoće li, i u kojoj mjeri, uključivanje PUPPI algoritma, opisanog u Poglavlju 2, u algoritam za prepoznavanje mlazova nastalih iz b kvakova poboljšati spomenuti algoritam.



Slika 2: Teorijska predviđanja omjera grananja Higgsovog bozona za različite mase. Vidimo da je na masi od 125 GeV najveći omjer grananja za raspad u par b kvarka i antikvarka

2 PUPPI

Unutar LHC-a (Large Hadron Collider) ubrzavaju se dva snopa protona te se, nakon što se postigne ciljana energija snopova, ukrštavaju na četiri lokacije gdje se nalaze detektori (CMS, Atlas, Alice, LHCb). Svakih 25 ns dođe do ukrštavanja tih dviju zraka. Prilikom svakog prolaska paketa protona doći će do mnogo sudara protona i nama su "zanimljivi" čeonu sudari što čini manjinu događaja. Da bi se samo takvi događaji bilježili, postoji detektorski sustav za okidanje ("trigger") koji daje signal za skupljanje podataka samo ukoliko su ispunjeni uvjeti koji vrijede za "zanimljive događaje". U svakom se događaju dogodi mnogo sudara protona pa čak i ako samo jedan od njih zadovoljava uvjete za "trigger", skupit će se i podatci o neželjenim dodatnim sudarima - "pileup". PUPPI algoritam [8] (Pileup Per Particle Identification) razvijen je upravo radi uklanjanja doprinosa neželjenih dodatnih sudara iz podataka čiji ćemo princip rada ukratko objasniti.

Podatci o putanji čestice dostupni su samo za nabijene čestice jer jedino one ostavljaju trag u detektorskom sustavu za praćenje putanja čestica. Iz putanja nabijenih čestica tada se rekonstruiraju točke iz kojih su došle čestice i tako se identificira primarna točka sudara (Primary vertex). Nabijene čestice koje ne dolaze iz te točke mogu se odbaciti kao "pileup", ali to se ne može napraviti za neutralne čestice jer za njih nisu dostupni podatci o putanji pa se zato koristi PUPPI algoritam. Za sve se čestice unutar jednog događaja (neutralne i nabijene) računa vrijednost parametra α koji ovisi o blizini drugih čestica i njihovim transverznim momentima.

$$\alpha_i = \log \sum_{j \in \text{event}} \xi_{ij} \times \Theta(R_{\min} \leq \Delta R_{ij} \leq R_0), \quad (2.1)$$

gdje je $\xi_{ij} = \frac{p_{Tj}}{\Delta R_{ij}}$. R_0 i R_{\min} parametri su algoritma koji su uobičajeno postavljeni na 0.3 i 0.02, respektivno. p_{Tj} je transverzalni impuls čestice j mjereno u GeV-ima. $\Theta(R_{\min} \leq \Delta R_{ij} \leq R_0)$ je skraćeni zapis za $\Theta(\Delta R_{ij} - R_{\min}) \times \Theta(R_0 - \Delta R_{ij})$, gdje je Θ Heaviside

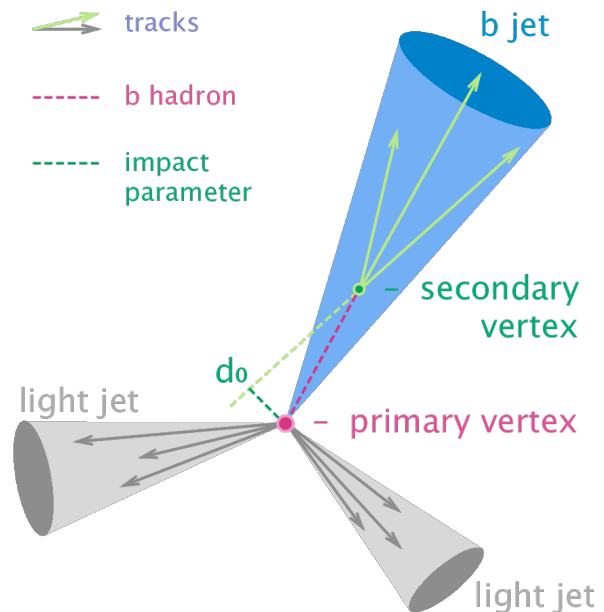
step funkcija, a ΔR_{ij} udaljenost čestica i i j u prostoru pseudorapiditeta i azimutalnog kuta. Sumirati se može po svim česticama u događaju ili pak samo po nabijenim česticama koje dolaze iz primarne točke sudara. Podatci o putanjama nabijenih čestica dostupni su samo u centralnom dijelu detektora pa za čestice u tom dijelu postoji mogućnost sumiranja samo po nabijenim česticama iz primarne točke sudara dok se u prednjem dijelu detektora, gdje nemamo informacije o putanjama, mora sumirati po svim česticama. Ako smo zbrajali doprinose svih čestica, tada se parametar označava sa α^F ("forward"), a ako smo samo izvršili sumaciju po nabijenim česticama iz primarne točke sudara, parametar označavamo sa α^C ("central").

Promatranjem jednadžbe 2.1 vidimo da će čestica imati veću vrijednost α što se više čestica nalazi u njenoj blizini. Čeoni sudari dovode do procesa hadronizacije prilikom kojeg nastaje mnoštvo čestica unutar uskog stošca nazvanog hadronskim mlazom. Iz tog će razloga čestice iz primarne točke sudara uglavnom imati u blizini i druge čestice došle iz te točke što dovodi do većih vrijednosti α . Čestice došle iz dodatnih sudara naći će se u blizini drugih čestica samo igrom slučaja što će dovesti do manjih vrijednosti α .

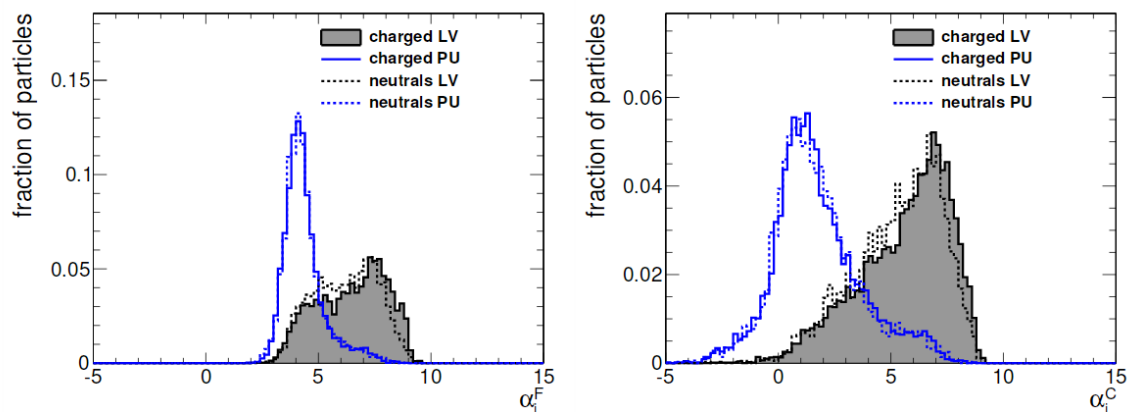
Za nabijene čestice iz centralnog dijela detektora izračunamo sljedeće raspodjele: $\alpha_{PU}^C, \alpha_{LV}^C, \alpha_{PU}^F$ i α_{LV}^F (LV označava da je parametar α izračunat za čestice koje su došle iz primarne točke sudara, a PU za one koje nisu). Zatim se pretpostavi da su distribucije iste za nabijene i neutralne čestice i pogledom na Sliku 3 vidimo da je pretpostavka razumna. Na taj se način dobiju distribucije parametra α za neutralne čestice. Vidimo da veličina α ima neku moć razlikovanja čestica iz vodećeg vrha i onih iz "pileup-a". Karakteristike distribucije (medijan i RMS) koriste se za pridruživanje težina česticama na način da sve čestice čiji je α_i manji od medijana dobiju težinu 0, a što je veća vrijednost α_i to će pridružena težina čestici biti bliže jedinici. Dakako, nabijene čestice uvijek imaju pridruženu težinu jednaku nuli ili jedinici jer su nam dostupne njihove putanje dok neutralne čestice imaju težine iz intervala $[0, 1]$.

3 b označavanje

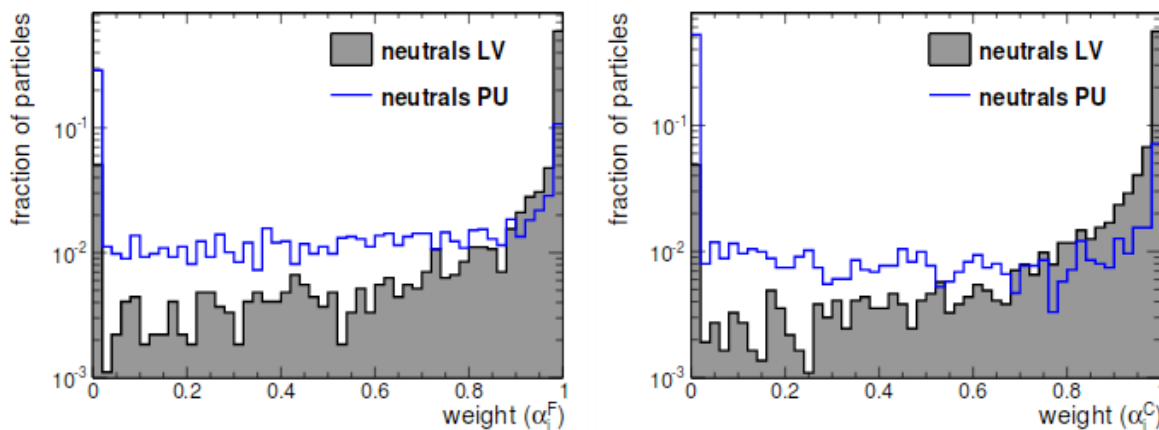
Hadronizacija je proces formiranja hadrona iz kvarkova i gluona koji se odvija nakon visoko energetskog sudara u kojem se stvore kvarkovi i gluoni. Zbog "zatočenja boje", kvarkovi i gluoni ne mogu biti izolirani pa se iz vakuuma spontano stvaraju kvarkovi i antikvarkovi koji tada čine hadrone. Skup čestica nastalih hadronizacijom naziva se hadronskim mlazom i ima karakterističan oblik stošca. Raspad Higgsovog bozona u dva b hadrona dovodi do stvaranja dva hadronska mlaza i to je bitna karakteristika toga raspada te je stoga bitno moći prepoznati od kojeg je kvarka došao pojedini mlaz. b označavanje ("b tagging") je rekonstrukcijska tehnika koja svakom mlazu pridruži vjerojatnost da je došao iz b kvarka. To se postiže koristeći poznata svojstva b kvarka i izmjerena svojstva hadronskog mlaza poput parametra sudara, duljine preleta i mase verteksa. Karakteristično svojstvo b hadrona nastalih sudarom protona u LHC-u je to da stignu prijeći put od nekoliko milimetara prije nego li se raspadnu. To svojstvo utječe na neke od parametara koji se koriste u b označavanju kao što je prikazano na Slici 5.



Slika 5: Shema b mlaza. Parametar sudara označen je sa d_0 , a duljina preleta je udaljenost između primarnog i sekundarnog verteksa. Preuzeto iz [9]



Slika 3: Distribucije parametara α^F i α^C za različite kategorije čestica. LV (Leading Vertex) označava čestice iz primarne točke sudara, a PU (pileup) čestice iz dodatnih sudara. Puna linija označava nabijene čestice, a isprekidana linija neutralne. Preuzeto iz [8]

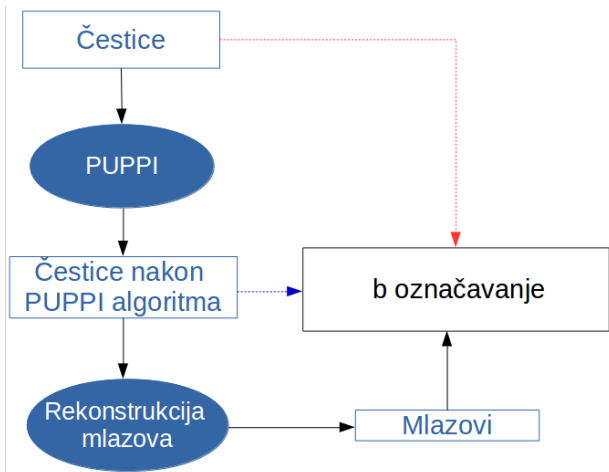


Slika 4: Histogram težinskih faktora koji dodjeljuje PUPPI algoritam za čestice iz prednjeg (lijevo) i centralnog (desno) dijela detektora. Preuzeto iz [8]

4 Algoritam za dvostruko b označavanje

PUPPI algoritam koristi se prilikom rekonstrukcije mlazova i tako se indirektno koristi i u b označavanju, a može se i direktno koristiti tako da se primijeni na čestice koje su ulaz b označavanju kako je prikazano na Slici 6. U ovom se radu uspoređuju performanse algoritma za b označavanje sa i bez direktne uporabe PUPPI algoritma. Cijeli proces može se razložiti u sljedeće korake: priprema Monte Carlo uzoraka raspada Higgsovog bozona u dva b kvarka i pripadajuće QCD pozadine, treniranje algoritma korištenjem spomenutih uzoraka, evaluacija diskriminatora dobivenog

algoritmom i prikazivanje performansi na grafu. Ulazni podatci koje se koriste prilikom treniranje algoritma su hadronski mlazovi koji zaista dolaze od raspada Higgsovog bozona u dva b kvarka (takvi se mlazovi nazivaju "signalni mlazovi") i pozadinski mlazovi. Hadronski su mlazovi rekonstruirani anti-kt algoritmom [10] s parametrom $R = 0.8$. Algoritam za dvostruko b označavanje trenira se pomoću TMVA paketa (Toolkit for Multivariate Analysis), konkretnije, korištenjem BDT metode ("Boosted Decision Tree") [11].

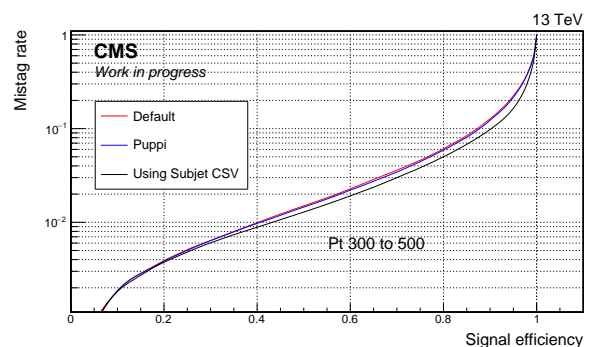


Slika 6: Shema koja prikazuje primjenu PUPPI algoritma u b označavanju. Uspoređujemo efikasnost algoritma sa (crvena strijelica) i bez primjene PUPPI algoritma (plava strijelica) na čestice proslijeđene algoritmu za b označavanje. U oba se slučaja PUPPI primjenjuje prije rekonstrukcije mlazova.

Ukupno je korišteno 27 ulaznih varijabli tijekom treniranja, poput mase i energije verteksa, duljine preleta i parametra sudara. Neke se varijable odnose na svojstva cijelog hadronskog mlaza dok se neke varijable odnose na njegovu podstrukturu. Podsjetimo se, "Fat" hadronski mlaz sastoji se od 2 spojena hadronska mlaza te su nam dostupni podatci o samom "Fat" mlazu i podatci o mlazovima od kojih je građen. Ovaj je skup varijabli odabran jer se koristio u prethodnim algoritmima za b označavanje [7], ali može se i mijenjati. Npr. varijable koje opisuju podstrukturu "fat" hadronskog mlaza možemo zamijeniti sa "subjekt CSV" ("combined secondary vertex") [12] varijablom koja je kombinacija dostupnih podataka o podstrukturi.

Nakon učitavanja podataka poziva se funkcija iz paketa TMVA koja pokušava što bolje razdvojiti signalne i pozadinske mlazove. Rezultat te funkcije je .xml datoteka koju možemo koristiti za buduće analize. Naime, u TMVA paketu također postoji i funkcija koja učitava .xml datoteku i svakom mlazu pridruži vrijednost diskriminatora. Što je veća vrijednost, veća je i vjerojatnost da se radi o signalu, a ne pozadini. Rezultati se prikazuju crtanjem efikasnosti signala i efikasnosti pozadine na grafu. Efikasnost se definira kao $\epsilon = \frac{N_{\text{pass}}}{N_{\text{total}}}$,

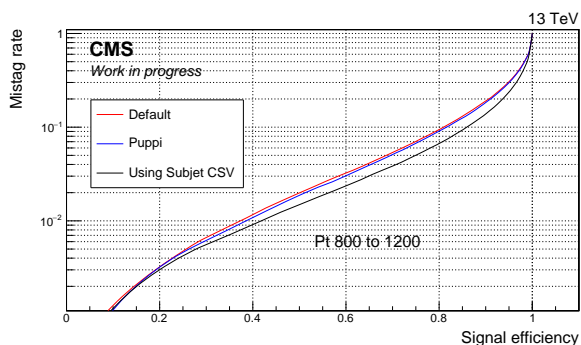
gdje je N_{pass} broj događaja čija je vrijednost diskriminatora iznad odabranog praga, a N_{total} ukupan broj događaja. U idealnom bi slučaju imali $\epsilon_{\text{signal}} = 1$ i $\epsilon_{\text{pozadina}} = 0$. Da bismo dobili jednu točku na grafu, odaberemo vrijednost diskriminatora koja će nam biti prag i za tu vrijednost izračunamo ϵ_{signal} i $\epsilon_{\text{pozadina}}$. Različitim odabirom praga diskriminatora dobijemo skup točaka koje prikazujemo na grafu. Dobivena krivulja naziva se ROC krivulja ("Receiver operating characteristic") i u ovom su radu sve ROC krivulje prikazane u logaritamskoj skali na y-osi.



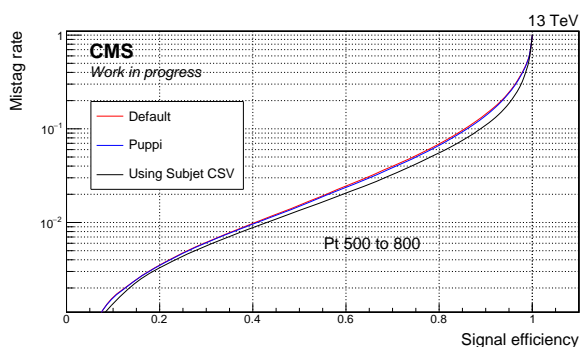
Slika 7: ROC krivulje u rasponu transverzalnih momenata mlazova [300, 500] GeV. Krivulje "PUPPI" i "Default" odnose se na slučajeve gdje se koristi i ne koristi PUPPI tijekom b označavanja, respektivno, a krivulja "Using subjekt CSV" odnosi se na slučaj gdje koristimo manji set varijabli, opisan u Poglavlju 4. Zanimljivo je da smo smanjivanjem ukupnog broja varijabli dobili nešto bolje performanse algoritma.

5 Rezultati

ROC krivulje za raspon transverzalnih momenata mlazova od 300 do 500 GeV prikazane su na Slici 7. S grafa je vidljivo da dodavanje PUPPI algoritma tijekom b označavanja tek neznatno poboljšava performanse algoritma. Također, treba provjeriti mijenja li se ponašanje novog diskriminatora s promjenama transverznog momenta. Usporedbom Slika 7, 8 i 9 na kojima su prikazane ROC krivulje za transverzne momente mlazova u intervalima [300, 500], [500, 800] i [800, 1200] GeV, respektivno, vidi se da se performanse značajno ne mijenjaju.



Slika 8: ROC krivulje za transverzne momente u rasponu [500, 800] GeV



Slika 9: ROC krivulje za transverzne momente u rasponu [800, 1200] GeV

6 Zaključak

Diskriminatori za identifikaciju b mlazova nastalih raspadom ultrarelativističkih Higgsovih bozona uspješno su dobiveni sa i bez primjene PUPPI algoritma prilikom b označavanja. Evaluirane su performanse dobivenih diskriminatora i utvrđeno je da uključivanje PUPPI algoritma u b označavanje tek neznatno poboljšava diskriminatorske sposobnosti. Također, ispitana je i ovisnost diskriminatora o transverzalnemu momentu mlazova i pokazano je da se performanse ne mijenjaju značajno. Dodatno poboljšanje performansi algoritma dobiveno je zamjenom svih varijabli koje se odnose na podstrukturu mlazova sa njihovom kombinacijom, "subjet CSV".

Literatura

- [1] P.W. Higgs. "Broken symmetries, massless particles and gauge fields". *Physics Letters* 12.2 (rujan 1964.), str. 132–133. DOI: 10.1016/0031-9163(64)91136-9. URL: [https://doi.org/10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9).
- [2] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Physical Review Letters* 13.16 (listopad 1964.), str. 508–509. DOI: 10.1103/physrevlett.13.508. URL: <https://doi.org/10.1103/physrevlett.13.508>.
- [3] G. S. Guralnik, C. R. Hagen i T. W. B. Kibble. "Global Conservation Laws and Massless Particles". *Physical Review Letters* 13.20 (studeni 1964.), str. 585–587. DOI: 10.1103/physrevlett.13.585. URL: <https://doi.org/10.1103/physrevlett.13.585>.
- [4] Peter W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". *Physical Review* 145.4 (svibanj 1966.), str. 1156–1163. DOI: 10.1103/physrev.145.1156. URL: <https://doi.org/10.1103/physrev.145.1156>.
- [5] T. W. B. Kibble. "Symmetry Breaking in Non-Abelian Gauge Theories". *Physical Review* 155.5 (ožujak 1967.), str. 1554–1561. DOI: 10.1103/physrev.155.1554. URL: <https://doi.org/10.1103/physrev.155.1554>.
- [6] Jonathan M. Butterworth i dr. "Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider". *Physical Review Letters* 100.24 (lipanj 2008.). DOI: 10.1103/physrevlett.100.242001. URL: <https://doi.org/10.1103/physrevlett.100.242001>.
- [7] *Identification of double-b quark jets in boosted event topologies*. Teh. izv. CMS-PAS-BTV-15-002. Geneva: CERN, 2016. URL: <http://cds.cern.ch/record/2195743>.
- [8] Daniele Bertolini i dr. "Pileup per particle identification". *Journal of High Energy Physics* 2014.10 (listopad 2014.). DOI: 10.1007/jhep10(2014)

059. URL: [https://doi.org/10.1007/jhep10\(2014\)059](https://doi.org/10.1007/jhep10(2014)059).

- [9] *b tagging Wikipedia*. <https://en.wikipedia.org/wiki/B-tagging>. [Link aktivan 18.1.2017.]
- [10] Matteo Cacciari, Gavin P Salam i Gregory Soyez. “The anti- k_t jet clustering algorithm”. *Journal of High Energy Physics* 2008.04 (travanj 2008.), str. 063–063. DOI: 10.1088/1126-6708/2008/04/063. URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [11] *Toolkit for Multivariate Analysis (TMVA)*. <http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf>. [Link aktivan 18.1.2017.]
- [12] *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. Teh. izv. CMS-PAS-BTV-15-001. Geneva: CERN, 2016. URL: <http://cds.cern.ch/record/2138504>.