

Potraga za raspadom stopa u kanal s 4 izlazne čestice korištenjem multivarijatnog pristupa i multiklasne neuronske mreže

Lucija Jukić

Prirodoslovno-matematički fakultet
Fizički odsjek

Sažetak

Predstavljen je prvi korak u potrazi za raspadom supersimetričnog partnera top kvarka (stopa) u kanal s 4 izlazne čestice za kojeg se, zbog pretpostavki teorije o gornjoj granici razlike mase stopa i neutralina, pretpostavlja da ima omjer grananja jednak jedinici. Selektiraju se događaji s točno jednim leptonom, barem jednim mlazom (eng. jet) i velikom vrijednošću nedostajuće energije. Opisan je postupak preselekcije i odabira ulaznih varijabli za multiklasnu neuronsku mrežu koja se koristi za konačnu selekciju - odabir signalnog područja. Radi se o početnom stadiju potrage u kojem se algoritmi strojnog učenja (u ovom slučaju neuronske mreže) nastoje optimizirati kako bi se maksimizirala signifikantnost za nalaženje signala u podacima dobivenim iz simulacija. Diskutira se opravdanost upotrebe multiklasne neuronske mreže za tu namjenu, s obzirom da su dosadašnje potrage uključivale binarne klasifikatore. Pokazuje se da bi se najbolji rezultat za signifikantnost dobio uz upotrebu multiklasne neuronske kao binarnog klasifikatora – u smislu da se u konačnoj analizi izostavljaju informacije o izlazima koji ne odgovaraju signalnom i odustaje se od definiranja kontrolnog područja unutar istog faznog prostora gdje se obavlja potraga za signalnim područjem.

I. UVOD

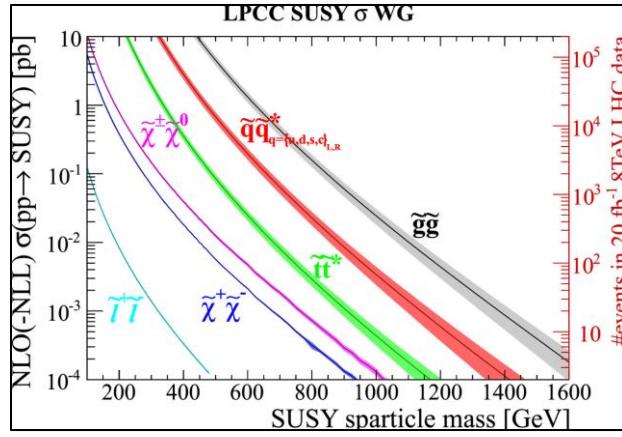
Standardni model za sada jako dobro reproducira podatke koji se mijere u eksperimentima na sudarivačima čestica, ali i dalje postoje otvorena pitanja koja se u sklopu standardnog modela ne mogu objasniti. Najpoznatiji primjeri vjerojatno su vezani uz podrijetlo tamne materije i mogućnost ujedinjenja svih interakcija na nekoj skali. Supersimetrije i dalje predstavljaju najperspektivniju ekstenziju standardnog modela koja razriješava ta i mnoga druga pitanja pa je potraga za supersimetričnim česticama od velike važnosti ne samo za fiziku elementarnih čestica, nego i za konceptualno razumjevanje (većeg dijela) svemira. Čestica koja je u centru razmatranja ovog rada je top skvark (stop) - supersimetrični partner top karka. Potraga se odvija u podacima iz eksperimenta na

detektoru CMS (eng. Compact Muon Solenoid) u sklopu Velikog hadronskog sudarivača (eng. Large Hadron Collider) na CERN-u. Treningom neuronskih mreža ideja je na podacima iz simulacija, koje predviđaju strukturu stvarnih podataka iz eksperimenta i gdje se simuliraju raspadi stopa i relevantnih pozadinskih procesa (iz standardnog modela), izvesti generalne zaključke o strukturi faznog prostora kinematičkih varijabli izlaznih čestica i dobiti informaciju o tome u kojim djelovima faznog prostora je udio signala (procesa koji odgovaraju raspadu stopa) značajan (signifikantan). Rad je zapravo ekstenzija prethodnog istraživanja [I] koje ima jednaki pristup problemu, ali se umjesto neuronskih mreža koristi BDT (eng. Boosted Decision Trees) kao alternativna metoda strojnog učenja. Glavna konceptualna razlika je u tome što se BDT koristi kao binarni klasifikator, dok se

Names	spin 0	spin 1/2	Names	spin 1/2	spin 1
squarks, quarks (3 x families)	$(\tilde{u}_L, \tilde{d}_L)$ \tilde{u}_R^*	(u_L, d_L) \tilde{u}_R	gluino, gluon	\tilde{g}	g
sleptons, leptons (3 x families)	d_R^* $(\tilde{\nu}_L, \tilde{e}_L)$ \tilde{e}_R^*	\tilde{d}_R (ν_L, e_L) \tilde{e}_R	winos, W bosons	\tilde{W}^\pm \tilde{W}^0	W^\pm W^0
Higgs, Higgsinos	(H_u^+, H_u^0) (H_d^0, H_d^-)	$(\tilde{H}_u^+, \tilde{H}_u^0)$ $(\tilde{H}_d^0, \tilde{H}_d^-)$	bino, B boson	\tilde{B}^0	B^0

Slika 1 Pregled čestica supersimetričnih modela.

multiklasna neuronska mreža koristi za odvajanje većeg broja različitih klasa – signala (raspada stopa) i odabranih pozadinskih procesa. U prvom djelu opisuje se konstrukcija samog faznog prostora i podataka iz simulacija, a u drugom mehanizam optimizacije treninga neuralnih mreža kako bi se iz danog seta podataka dobio maksimalni broj korisnih informacija.

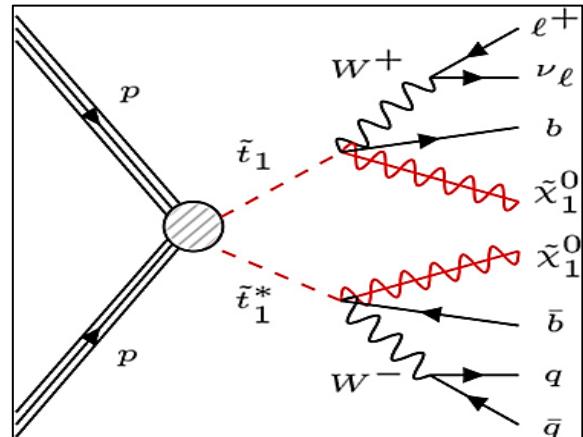


Slika 2 Izračunati udarni presjeci (u okviru MSSM-a) za proizvodnju parova različitih supersimetričnih čestica (pri energiji centra mase od 8TeV) u ovisnosti o masi supersimetrične čestice. Preuzeto iz [II].

II. SUPERSIMETRIJE

Glavna značajka svih supersimetričnih modela (SUSY) je pretpostavka postojanja dodatnih čestica koje se grupiraju u supersimetrične dublete s poznatim česticama iz standardnog modela (SM). Za svaki fermion iz standardnog modela postoji supersimetrični partner bozonskog karaktera i za svaki bozon postoji supersimetrični fermion. Na *Slici 1.* prikazan je pregled čestica supersimetričnih modela. Dvije čestice, SM čestica i njezin SUSY partner, dvije su različite realizacije istog objekta. Supersimetrije su u osnovi simetrija prostorvremena – ekstenzija Poincareove algebre koja uz standardne generatore prostornovremenskih translacija P^μ i rotacija $M^{\mu\nu}$ sadrži dodatne spin-1/2 generatore Q^α koji bozonsko polje pretvaraju u fermionsko i obratno (tzv. gradirana Poincareova algebra). Do loma simetrije mora doći jer dosada nije uočena niti jedna čestica koja ima masu jednaku masi neke od SM čestice i po karakteru (bozonskom ili fermionskom) se razlikuje od nje. Ako je teorija točna, u eksperimentima na sudarivačima čestica trebale bi nastajati supersimetrične čestice i uočiti se na detektoru kroz izlazne produkte kasnijeg raspada. Uz pretpostavku da je R-paritet (novi kvantni broj koji SM česticama pridaje vrijednost -1, a supersimetričnim partnerima 1) u rekacijama očuvan, supersimetrične

čestice nastaju u parovima (čestica-antičestica) i raspadaju se na najlakšu od njih koja bi unutar ovog modela trebala biti neutralino. Fizikalne čestice zapravo su svojstvena stanja operatora mase koja su mješavina čestica navedenih na *Slici 1.*, pa je tako neutralino linearna kombinacija fotina, zina i higgsina, a fizikalno mjerljiv stop \tilde{t}_1 mješavina je lijeve i desne kiralne komponente stopa.

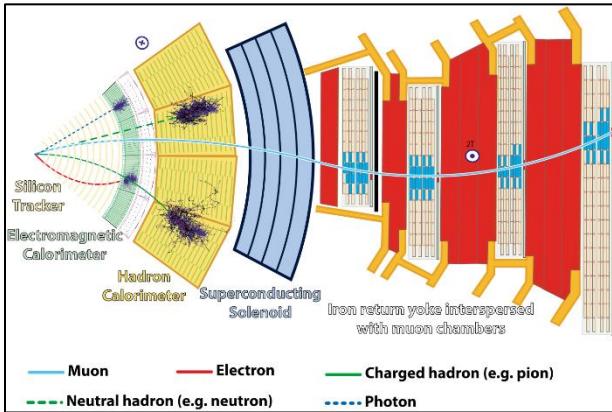


Slika 3 Shematski prikaz nastanka stopa i antistopa u proton-proton sudaru i kasnijeg raspada na četiri izlazne čestice. Na dijagramu je zbog jednostavnosti prikaza izostavljen nastanak virtualne čestice (b skvark najvjerojatniji) koja se raspada na neutralino $\tilde{\chi}_0$ (neutralino) i b kvark (ili antikvark). W je zbog pretpostavki modela također virtualan. Stop i antistop mogu se raspasti kroz leptonski ili hadronski kanal. Slika je preuzeta iz [III].

A. Potraga za supersimetrijama

S obzirom na relativno veliki broj supersimetričnih čestica može se postaviti pitanje zašto se potraga za supersimetrijama obavlja upravo potragom za proizvodnjom parova stop-antiststop. Na *Slici 2* prikazan je udarni presjek za proizvodnju parova niza različitih supersimetričnih čestica u ovisnosti o pretpostavci za vrijednost njihove mase. Za danu vrijednost mase, proizvodnja parova sgluina te skvarkova prve i druge generacije je vjerojatnija pa zbog toga možda i bolji kandidat za obavljanje potrage za supersimetrijama. Pretpostavka je međutim da je masa skvarkova nižih generacija veća od mase skvarkova treće generacije zbog čega je produkcija parova stop-antastop vjerojatnija. I za sgluine se također prepostavlja da imaju masu veću od stopa, ali zbog velike vrijednosti konstante vezanja jake interakcije i dalje predstavljaju perspektivan kanal za obavljanje potrage pa se istraživanja često fokusiraju na njega. Pretpostavke o masama supersimetričnih čestica uglavnom proizlaze iz želje da se određeni problemi standardnog modela u okviru SUSY modela razriješe

„prirodnim putem“, tj. bez uvođenja dodatnih parametara u teoriju. Važan je primjer problem divergencije koji se javlja pri računu mase Higgsovog bozona – u višim korekcijama (interakcija Higgsovog polja s virtualnim česticama) javljaju se divergencije (proporcionalne kvadratu Planckove skale) koje su naravno u suprotnosti s očekujimima jer je izmjerena masa konačna (125GeV). U okviru supersimetričnih modela se zbog uvođenja novih čestica ti divergentni doprinosi pokrate. Pri razriješavanju tog problema stop zapravo ima značajnu ulogu – zbog velike mase top kvarka stop uklanja najveći doprinos.



Slika 4 Shematski prikaz CMS detektora.

B. Kanali raspada stopa

Zbog sačuvanja R-pariteta stop \tilde{t}_1 i antistop \tilde{t}_1^* moraju se u eksperimentima stvarati u parovima te raspadi na najlakšu supersimetričnu česticu – neutralino $\tilde{\chi}_0$. Razlika između mase stopa i mase neutralina predstavlja energiju dostupnu ostalim produktima raspada i utječe na dostupnost pojedinih kanala. Pretpostavka je da se masa stopa ne razlikuje puno od mase neutralina – utemeljena je na kozmološkim razmatranjima svojstava tamne materije i želi da se problem konačne mase Higgsovog bozona u sklopu supersimetričnih modela razriješi prirodnim putem zbog kojih masa stopa ne bi trebala biti prevelika. U ovom projektu se gornja granica za razliku mase stopa i neutralina postavlja na $\Delta m(\tilde{t}, \tilde{\chi}_0) < 80\text{GeV}$ (manje od mase W bozona) i zasebna ispitivanja provode za vrijednosti unutar tog raspona u koracima od 10GeV . Ovdje konkretno razmatranja se ograničavaju na slučaj $\Delta m(\tilde{t}, \tilde{\chi}_0) = 30\text{GeV}$. Uz pretpostavku da su procesi s promjenama generacija kvarkova značajno potisnuti isključuju se raspadi oblika $\tilde{t} \rightarrow c\tilde{\chi}_0$. Zbog navedene pretpostavke o gornjoj granici razlike masa stopa i neutralina procesi $\tilde{t} \rightarrow t\tilde{\chi}_0$ (zbog $\Delta m(\tilde{t}, \tilde{\chi}_0) < m(t)$) i raspad na tri izlazne čestice oblika $\tilde{t} \rightarrow W^+ b\tilde{\chi}_0$ (zbog $\Delta m(\tilde{t}, \tilde{\chi}_0) < m(W)$) su kinematički zabranjeni. Kao

najbolja sljedeća opcija preostaju raspadi na četiri izlazne čestice $\tilde{t}_1 \rightarrow bff'\tilde{\chi}_0$, gdje su f i f' oznake za fermione (mogu biti kvarkovi ili leptoni) – omjer granjanja za taj kanal u ovakvim uvjetima je približno jednak jedinici. Shematski prikaz raspada stopa u kanal s 4 izlazne čestice prikazan je na *Slici 3* – stop i antistop mogu se raspasti hadronski (izlazni fermioni su kvarkovi) ili leptonski (izlazni fermioni su lepton i neutrino).

III. EKSPERIMENT NA CMS-u

U eksperimentima koji se provode na CMS-u analiziraju se proton-proton sudari na energiji u sustavu centra impulsa od 13TeV – odnosi se na mjerena obavljena 2016. godine kada je ukupni integrirani luminozitet bio 35.9fb^{-1} . Iako se mjereni podaci koriste tek u kasnijim fazama istraživanja, podatak o godini mjerena bitan je i za razumijevanje podataka iz simulacija pošto se pri generiranju podataka trebaju uzeti u obzir stanja različitih poddetektora na CMS-u koja se mijenjaju ovisno o godini izvođenja eksperimenta.

A. Izgled izmjerениh podataka

Eksperimenti koji se provode na sudsarivačima čestica na CERN-u mogu se konceptualno rastaviti na određeni broj koraka. Prvi od njih je generacija proton-proton sudara. Protoni su duž staze LHC-a grupirani u nakupine (*eng. bunches*) i ubrzani do brzina bliskih brzini svjetlosti. Dvije nakupine fokusiraju se i sudsaraju na pozicijama na kojima su smješteni detektori. Pod jednim događajem podrazumijeva se zapravo prolazak (sudar) dviju nakupina, a unutar njega može se dogoditi veći broj proton-proton sudara (najčešće između 30 i 60 njih). Sljedeći je korak detekcija čestica. Shematski prikaz CMS detektora prikazan je na *Slici 4*. Sastoje se od niza poddetektora koji mijere energiju (kalorimetri) i detektora tragova koji registriraju putanju čestica. Direktnom detekcijom se može uočiti samo mali broj čestica: elektroni, mioni, protoni, neutroni, kaoni i nabijeni pioni. Kriterij je da čestice imaju vrijeme života dovoljno dugo da uspiju proći kroz detektor. Ako prilikom proton-proton sudara nastaju kvarkovi u detektoru se uočavaju mlazovi (*eng. jets*) – kolimirani snopovi čestica koji nastaju kao rezultat hadronizacije kvarkova i kasnijeg raspada nastalih hadrona na „stabilne“ (opazive) čestice. Prisutnost neutrina (i neutralina) prepoznaje se po nedostajućoj energiji i impulsu. S obzirom da se svake sekunde unutar CMS detektora odvija otprilike 40 milijuna događaja (prolazaka nakupina), potreban je sustav okidača koji će selektirati događaje od interesa i spremati ih za daljnju obradu. Potrage koje se fokusiraju na kandidate za tamnu materiju i novu fiziku uglavnom selektiraju događaje s

visokoenergetskim mlazovima i velikom vrijednošću nedostajuće energije. U konačnici se spremi tek 1000 događaja po jednoj sekundi. Zadnji korak obrade podataka je rekonstrukcija. Tijekom tog koraka se kombiniraju informacije dobivene sa svih poddetektora kako bi se utvrdile vrste nastalih čestica i iz kojeg vrha (*eng. vertex*) potječe - u jednom događaju može se odviti veći broj proton-proton sudara pa je potrebno utvrditi od kojeg od njih dolazi koja čestica. Od svih sudara koji se odviju tijekom prolaska dviju nakupina (događaja) u pravilu je samo jedan visokoenergetski – slučaj kada pojedinačni partoni koji sudjeluju u sudaru preuzimaju na sebe veći udio ukupne kinetičke energije protona. Samo ti visokoenergetski procesi se u konačnici analiziraju, a ostali se podrazumijevaju pod nazivom nagomilavanje (*eng. pileup*). Finalni podaci spremaju se za daljnju analizu i sadrže i informacije prikupljene sa detektora i one dobivene tijekom rekonstrukcije – vrijednosti kinematičkih varijabli izlaznih čestica, vrstama čestica, kojim vrhovima čestice pripadaju...

B. Podaci iz simulacija

Simulacije zapravo imaju slične korake kao i eksperiment. Najprije se simuliraju proton-proton sudari, u smislu da se određuju Feynmanovi dijagrami relevantnih procesa i računaju kvadrati amplituda do željenog reda. Fazni prostor izlaznih čestica probira se na način da se nasumično odabiru vrijednosti njihovih impulsa i varijabli u partonskim distribucijskim funkcijama te računaju diferencijalni udarni presjeci danog ishoda (Monte Carlo simulacije – program MadGraph). Izlazne čestice sa tako odabranim vrijednostima impulsa raspadaju se na „stabilne“ čestice (npr. program PYTHIA) te šalju kroz detektor – program koji simulira prolaz čestica kroz materijale od kojih su sastavljeni poddetektori CMS-a (npr. GEANT4). Rezultat toga su podaci koji izgledaju jednakim kao podaci mjereni na detektorima u pravom eksperimentu. Takvi podaci se na kraju šalju u algoritam za rekonstrukciju (*Particle Flow Algorithm*). Time se u konačnici dobiva skup podataka nalik iz eksperimenta, s iznimkom da je sada poznato koji podatak (događaj) pripada kojoj klasi procesa (je li signal ili jedna od pozadina iz standardnog modela). Svakom tako dobivenom događaju pridružen je težinski koeficijent određen relacijom (za klasu/proces C)

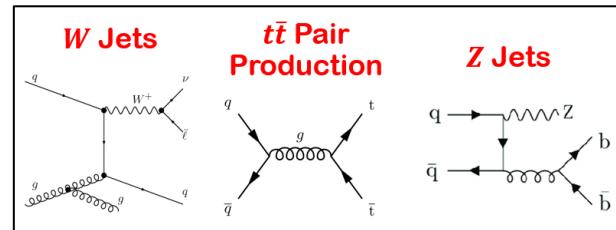
$$w_{event,C}^{(k)} = \frac{1}{N_C} \mathcal{L} \sigma_k P_k$$

gdje je sa N_C označen ukupni broj simuliranih događaja klase C (taj faktor omogućuje da se iz podataka isključi informacija o ukupnom broju simuliranih događaja s

obzirom da je to nešto što je proizvoljno odabранo), \mathcal{L} je luminozitet, σ_k oznaka za diferencijalni udarni presjek k -og događaja, a P_k koeficijent unutar kojeg su uzeti u obzir efekti prisutni u ostalim koracima simulacija – npr. efikasnosti detektora i procesa rekonstrukcije, efekti nagomilavanja i sl. U konačnici svaka težina $w_{event,C}^{(k)}$ zapravo govori o vjerojatnosti da se k -ti događaj dogodi u pravom eksperimentu (zapravo broj događaja k -og tipa koji se očekuje vidjeti u eksperimentu).

C. Odabir događaja

Od svih događaja koji su izmjereni na detektoru cilj je odabrati one za koje je najizglednije da odgovaraju signalu. Zbog toga se postavljaju uvjeti selekcije koji uzimaju u obzir što veći udio signalnih događaja i izbacuju procese standardnog modela (pozadine). Pošto raspadom stopa nastaju neutralni i neutrini, kroz selekciju se uzimaju u obzir samo događaji koji imaju veliku vrijednost nedostajuće energije i impulsa. Kako bi se umanjio broj pozadina koje generiraju veliki broj mlazova (*eng. multijet*) postavlja se zahtjev da u događaju mora biti detektiran točno jedan lepton i barem jedan mlaz. Time se zapravo odabiru samo događaji u kojima se jedan od para stop-antistop raspada leptonski, a drugi hadronski. Zahtjev da se oba stopa raspadaju leptonski isključio bi prevelik broj signalnih događaja. Preselekcija je daljnji zahtjev na vrijednosti izmjerenih varijabli kako bi se dodatno umanjio broj pozadinskih procesa. Zove se preselekcija zato što se odvija prije konačne selekcije – odabira signalnog područja – koja se obavlja upotrebom algoritama strojnog učenja (u ovom slučaju multiklasnih neuronskih mreža).

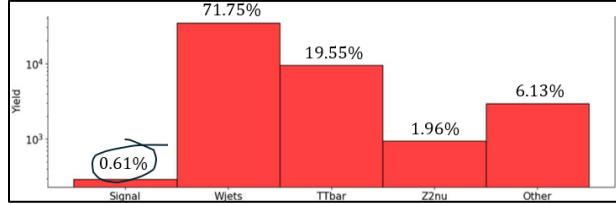


Slika 5 Feynmanovi dijagrami tri najdominantnija pozadinska procesa iz standardnog modela.

D. Pozadine iz standardnog modela

Pozadine iz standardnog modela koje ulaze u zahtjeve preselekcije su W +mlazovi, proizvodnja para $t\bar{t}$, Z +mlazovi, te ostale manje pozadine (Drell-Yan, dvobožonski procesi ...). Feynmanovi dijagrami triju procesa koji doprinose s najvećim brojem događaja prikazani su na *Slici 4*, dok su njihovi relativni odnosi

prikazani na *Slici 5*. I nakon preselekcije, raspodjelja stopa predstavlja tek oko 0.61% ukupnog broja događaja, a s navećim udjelom ulaze $W+$ mlazovi zbog velikog udarnog presjeka i sličnosti sa signalnim događajima. $Z+$ mlazovi, iako su događaji s velikim udarnim presjekom ulaze s manjim udjelom jer u tom slučaju mlaz mora biti pogrešno rekonstruiran kao lepton. Kako bi proces prošao odabir događaja Z bozon se mora raspasti na par $v\bar{v}$ zbog zahtjeva na prisutnost nedostajuće energije, a onda lepton može nastati jedino ako je mlaz interpretiran kao lepton.



Slika 6 Odnosi broja događaja pojedine klase koji se očekuje vidjeti u eksperimentu.

IV. PRONALAZAK ČESTICA I SIGNIFIKANTNOST

Veličina koja služi kao mjeri uspješnosti izdvajanja signala od pozadinskih procesa zove se signifikantnost (*eng. figure of merit*) i računa se prema sljedećoj formuli

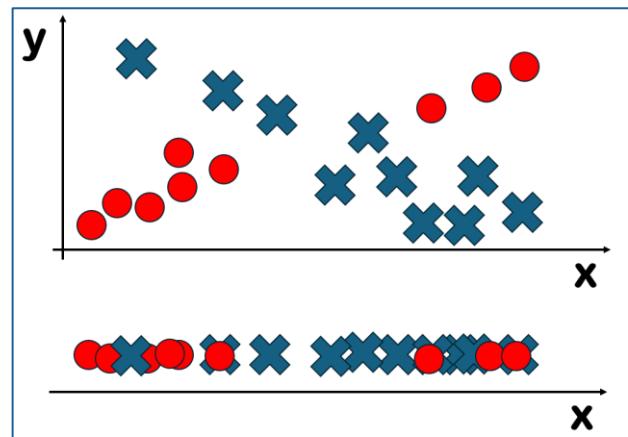
$$FOM = \frac{S}{\sigma_{tot}} = \frac{S}{\sqrt{\sigma_{sig}^2 + \sigma_{Bkg}^2 + \sigma_{sys}^2}}$$

gdje je S (B) očekivani broj signalnih (pozadinskih) događaja predviđen na temelju simulacija, a σ_{tot} predviđena pogreška izmjerjenih podataka – tu ulaze u obzir i pogreške mjerjenja pozadinskih ($\sigma_{Bkg} = \sqrt{B}$) i signalnih ($\sigma_{sig} = \sqrt{S}$) događaja, te sistematske pogreške za koje se procjenjuje da iznose $\sigma_{sys} = \sqrt{0.2B}$. Signifikantnost zapravo govori o tome koliko je pouzdano da je dani rezultat mjerjenja nagovještaj nekog novog efekta, a ne posljedica statističke fluktuacije.

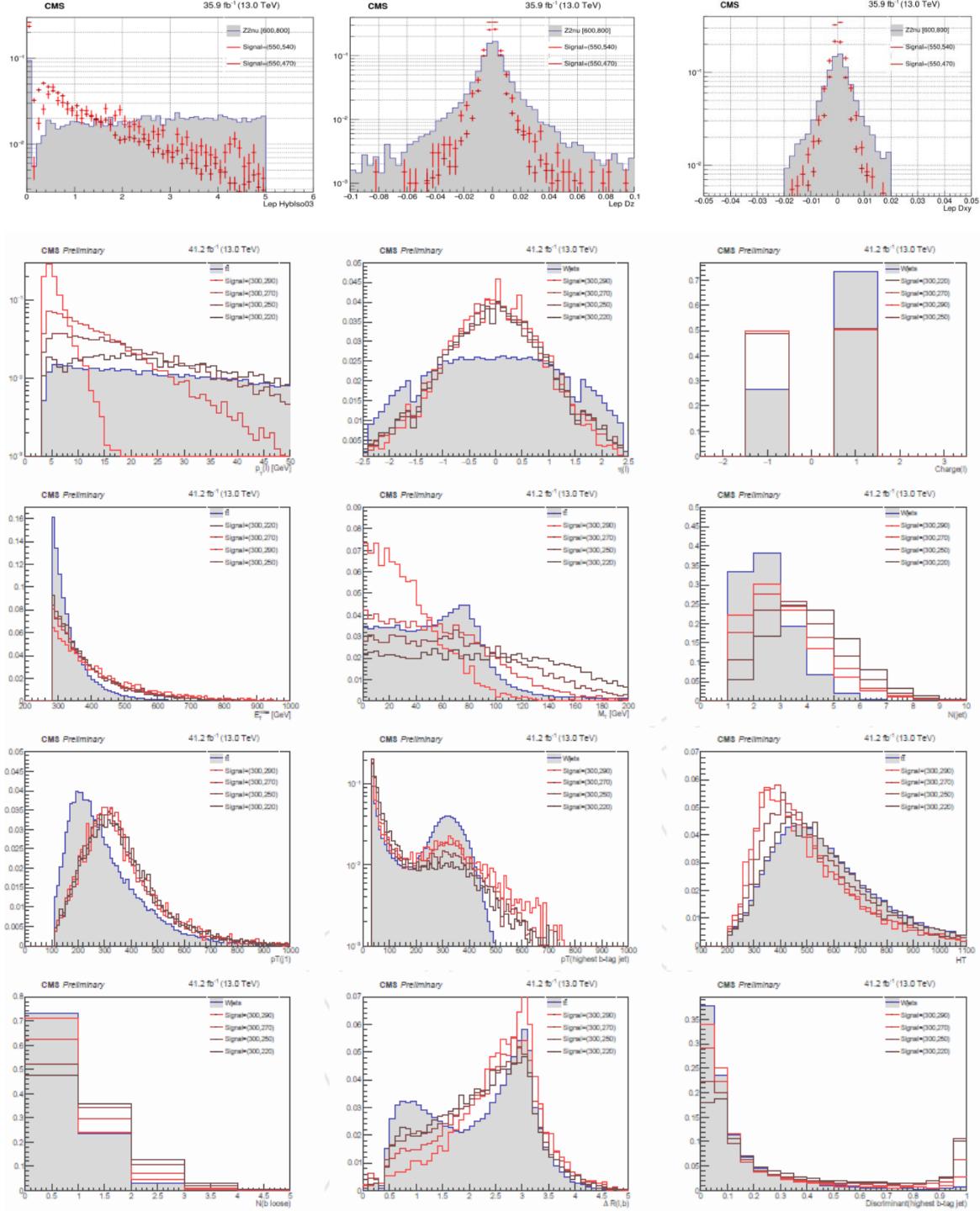
V. MULTIVARIJATNI PRISTUP

Kako bi se dodatno odvojio signal od pozadinskih procesa mogu se iskoristiti razlike u distribuciji broja događaja po kinematičkim varijablama izlaznih čestica. Za procese čije raspodjеле preferiraju drukčije vrijednosti dane varijable (veći udio događaja je raspodjeljen na tim vrijednostima) u odnosu na signalne događaje kaže se da pokazuju diskriminaciju u odnosu na njih. Set kinematičkih varijabli koji se koristi u ovom radu za diskriminaciju signala od ciljanih pozadinskih procesa i ovisnost

normaliziranih distribucija broja događaja o tim varijablama prikazan je na *Slici 8*. Želja je iskoristiti raspodjelu kako bi maksimalno odvojili signal od pozadinskih procesa – zadržali što veći udio signalnih događaja, a odbacili pozadine. Prva ideja bi bila postaviti linearne rezove na vrijednosti diskriminirajućih varijabli. Time se postavlja uvjet da događaji koji ulaze u konačnu selekciju (signalno područje) moraju imati vrijednosti tih varijabli unutar odabranih intervala. Signifikantnost koja se dobiva takvom procedurom iznosi 0.027, dok se uzimanjem u obzir svih događaja iz preselekcije dobiva 0.018. Primjer je naveden radi ilustracije koliko se bolji rezultat za signifikantnost dobiva jednom kada se selekcije rade u višedimenzionalnom prostoru razapetom zadanim setom diskriminirajućih varijabli što se obavlja u daljnjoj analizi ovog rada uz pomoć neuronskih mreža. To je zapravo glavna ideja multivarijatnog pristupa – set varijabli koje pokazuju određenu diskriminaciju između dviju različitih klasa (procesa) mogu se dodatno odvojiti u višedimenzionalnom prostoru ako se varijable različito koreliraju za dvije klase. Ilustracija takvog slijeda razmišljanja prikazana je na *Slici 7*. Odabir seta diskriminirajućih varijabli koji će biti ulazne varijable neuronske mreže temelji se na poznavanju kinematike raspada stopa (signala) i procesa standardnog modela koji ulaze kao pozadine. Npr. za diskriminaciju $Z+$ mlazova koriste se varijable vezane uz prepoznavanje mlazova pošto je u tom slučaju mlaz pogrešno interpretiran kao lepton. Za diskriminaciju $W+$ mlazova koriste se varijable za prepoznavanje b-kvarkova jer je u tim procesima vrlo malo vjerojatno da nastaje b-kvark, dok je u signalu za onaj stop koji se raspada hadronski to



Slika 7 Ilustracija ideje multivarijatnog pristupa. Dvije klase (X i O) pokazuju veću diskriminaciju u dvodimenzionalnom prostoru razapetom varijablama x i y (gore) nego u jednodimenzionalnom (dole). Razlog je u tome što se varijable x i y različito koreliraju za dvije klase – za klasu O korelacija je ≈ 1 dok je za klasu X ona ≈ -1 .

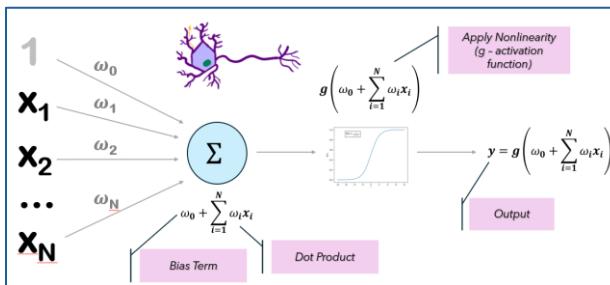


Slika 8 Usporedba normaliziranih distribucija broja događaja po diskriminirajućim varijablama koje se koriste kao ulazne varijable neuronske mreže za signal i za odabranu pozadinu koju bi ta varijabla trebala odvajati od signala. Za signal su na pojedinim grafovima prikazane distribucije za različite vrijednosti $\Delta m(\tilde{t}, \tilde{\chi}_0)$. U ovom radu se koriste samo podaci koji odgovaraju vrijednosti $\Delta m(\tilde{t}, \tilde{\chi}_0) = 30\text{GeV}$.

najčešći slučaj pošto su procesi s promjenama generacije kvarkova potisnuti. Varijable vezane uz nedostajuću energiju i impuls koriste se za odvajanje signala i od $W+mlazova$ i od procesa proizvodnje para $t\bar{t}$ s obzirom da imaju samo jedan neutrino kao izvor nedostajuće energije, dok signal ima dva dodatna neutralina. Transverzalni impulsi vodećeg mlaza (mlaz s najvećom vrijednošću transverzalnog impulsa) i mlaza za koji je najizglednije da dolazi kao rezultat hadronizacije b kvarka koristi se za odvajanje proizvodnje para $t\bar{t}$ jer su te varijable osjetljive na razliku Δm dostupnu ostalim produktima raspada – u slučaju stopa radi se o razlici mase stopa i neutralina, dok se u slučaju proizvodnje para $t\bar{t}$ radi o razlici mase t kvarka i W bozona. Konačna odluka hoće li varijabla ući u konačnu analizu ili ne, donosi se na temelju izlaznih varijabli algoritama strojnog učenja koje se koriste za linearno odvajanje signala od pozadine. Ako rezultiraju u povećanju signifikantnosti kinematičke varijable se uzimaju u obzir.

VI. MULTIKLASNA NEURONSKA MREŽA

Neuronska mreža dobila je ime po svojoj umreženoj strukturi izgrađenoj od osnovnih gradivnih jedinica, neurona, koji po obliku i principu funkciranja ugrubo podsjećaju na biološke neurone u mozgu. Jedan neuron na ulazu prima određeni set ulaznih varijabli koji skalarno pomnožen sa setom težina bude dalje stavljen u (nelinearnu) aktivacijsku funkciju (*Slika 9*).



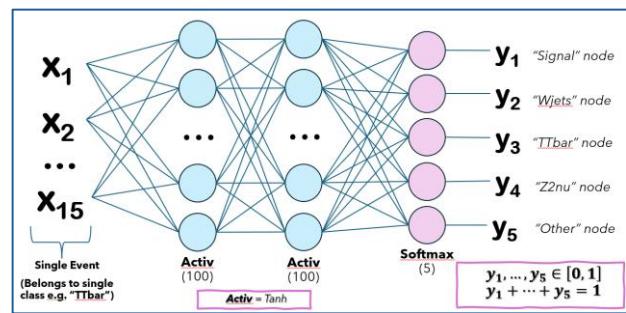
Slika 9 Prikaz neurona

Strukturu neuronske mreže je zapravo lako shvatiti ako se uzme u obzir analogija s biološkim neuronima – svaki neuron karakteriziran je vlastitim setom težina i izlaz jednog može poslužiti kao ulaz drugima. U shematskim prikazima neuronske mreže svaki je prikazan jednim čvorom. Neuronska mreža koja se ovdje koristi sastoji se od ulaza s 15 (kinematičkih) varijabli, dva unutarnja sloja od kojih svaki ima po 100 čvorova i gdje je na svakom odabran tanh za aktivacijsku funkciju, te izlaznog sloja s 5 čvorova (*Slika 10*). Svaki čvor pripada

jednoj od pet klasa u koje su se sortirali događaji iz simulacija – signalu, $W+mlazovima$, proizvodnji para $t\bar{t}$, $Z+mlazovima$, te ostalima (svi ostali manji pozadinski procesi iz standardnog modela koji ulaze u preselekciju). Ideja je da za dani set od 15 kinematičkih varijabli (događaj) neuronska mreža odgovori na pitanje: „Kojoj klasi prpada?“, odnosno da na svakom izlazu daje okvirnu informaciju o vjerojatnosti da dani događaj pripada klasi kojoj taj izlaz odgovara (na prvom za signal, drugom za $W+mlazove$ itd.). Za tu namjenu na izlazu je za aktivacijsku funkciju odabran softmax – ograničava vrijednost izlaznih varijabli na interval $[0,1]$ i postavlja dodatni uvjet da suma svih vrijednosti na izlazu mora biti jednaka jedinici:

$$\text{softmax}(y_j) = \frac{e^{y_j}}{\sum_{i=1}^5 e^{y_i}}$$

Glavna je želja da neuronska mreža tu funkciju obavlja na podacima izmjerenima u eksperimentu. Kako bi to uspjela, potrebno je da na podacima iz simulacija, gdje je poznato kojoj klasi koji događaj pripada, tijekom treninga nauči kako kako klasificirati događaje. Pritom zapravo na temelju podataka iz simulacija ugrubo određuje koji su udijeli 5 klasi u pojedinim dijelovima 15-dimenzionalnog prostora razapetog odabranim setom diskriminirajućih varijabli.

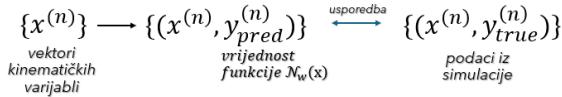


Slika 10 Multiklasna neuronska mreža

A. Trening

Kako bi naveli neuronsku mrežu da obavlja željenu funkciju treba pronaći za to odgovarajući set težina $\{\omega_i\}$. Neuronska mreža može se shvatiti kao funkcija oblika $N_\omega: \mathbb{R}^N \rightarrow [0,1]^M$ na prostoru ulaznih varijabli ($N=15$, $M=5$), a set težina kao set parametara te funkcije – jednako kao što nagib i odsječak na ordinati predstavljaju parametre linearne funkcije. Trening sada predstavlja neki oblik prilagodbe funkcije N_ω na dani set podataka – konceptualno je jednaka procedura kao u linearnoj regresiji kada se parametri pravca namještaju tako da suma svih kvadrata odstupanja pravca od mjerenih

podataka bude minimalna. U ovom slučaju radi se o prilagodbi na podatke iz simulacija:



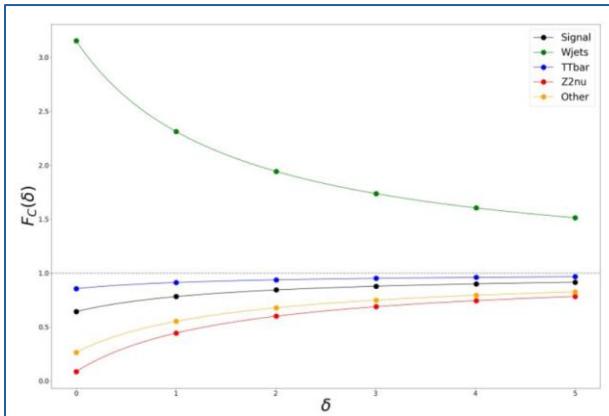
gdje su $y_{true}^{(n)}$ 5-komponentni vektori s jedinicom na jednom, a nulama na svim ostalim mjestima (na kojem mjestu se nalazi jedinica ovisi o klasi kojoj n -ti događaj u setu podataka iz simulacija pripada, eng. *one-hot encoded*), a $y_{pred}^{(n)}$ vektori koji predstavljaju vrijednost funkcije N_ω u danoj točki ulaznog prostora (n -tog događaja) i dani set težina $\{\omega_i\}$. Funkcija koja se minimizira zove se los funkcija i dana je izrazom

$$\mathcal{L}(\omega) = \frac{1}{K} \sum_k \ell(y_{true}^{(k)}, y_{pred}^{(k)}(\omega))$$

gdje suma ide po svim događajima u podacima iz simulacija, a K je ukupni broj događaja u setu podataka. U ovom slučaju za funkciju ℓ je odabrana kategorička logistička regresija

$$\ell(y_{true}, y_{pred}) = -y_{true}^T \ln(y_{pred})$$

koja se temelji na KL divergenciji, funkciji koja daje prirodnu mjeru „udaljenosti“ između dviju distribucija.



Slika 11 Konkretnе vrijednosti faktora F_C koje se koriste u radu.

Iako je ideja treninga neuronske mreže konceptualno slična linearnoj regresiji, proces minimizacije je bitno složeniji s obzirom da se radi o funkciji s velikim brojem parametara. Los funkcije \mathcal{L} , kao funkcija seta težina $\omega \equiv \{\omega_i\}$, numerički se postepeno minimizira tako da se najprije nasumično odabere jedan set ω^0 , odredi gradijent los funkcije $\Delta\omega_i^0 \equiv \frac{\partial\mathcal{L}}{\partial\omega_i}|_{\omega=\omega^0}$, te potom korigira na novu

vrijednost tako da se kreće u suprotnom smjeru od gradijenta. To se ponavlja za veći broj koraka, svaki se od njih zove epoha, a korekcija težina za epohu formalno je dana izrazom

$$\omega_i^{t+1} = \omega_i^t - R e^{-Dt} \Delta\omega_i^t$$

gdje su R (eng. *learning rate*) i D (eng. *decay rate*) parametri koji vode minimizaciju, a odabiru se vrijednosti koje rezultiraju najboljim performansom.

A. Balansiranje događaja

S obzirom na to kako su generirani podaci iz simulacija, raspodjela događaja u prostoru diskriminirajućih varijabli ne reflektira stvarnu raspodjelu podataka iz eksperimenta. Zapravo ovisi o tome koliko je događaja generirano za koju klasu procesa – nešto što je proizvoljno. Informacija o tome koliko je pojedini događaj vjerojatan zapisan je u težinama $\{w_{event}^{(k)}\}$ iz III.B. Način na koji se one uzimaju u obzir prilikom treninga neuronske mreže je kroz težine $\{\tilde{w}^{(k)}\}$ za los funkciju:

$$\mathcal{L}(\omega) = \frac{1}{K} \sum_k \tilde{w}^{(k)} \ell(y_{true}^{(k)}, y_{pred}^{(k)}(\omega))$$

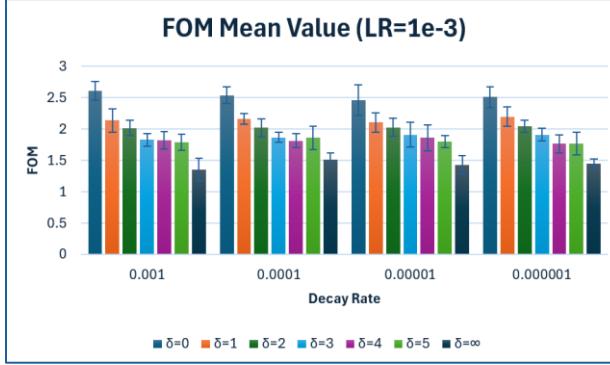
Standardna procedura za tretiranje težina los funkcije za multiklasnu neuronsku mrežu je da se tretira svaka klasa posebno. Za težine klase C se ovdje konkretno uzimaju izrazi:

$$\tilde{w}_C^{(k)} = \frac{w_{event,C}^{(k)}}{\sum_i w_{event,C}^{(i)}} N_S \times f_C$$

gdje suma ide po svim događajima koji odgovaraju klasi C , a N_S broj događaja u setu podataka iz simulacija koji odgovaraju signalu. Prvim dijelom izraza raspodjeli broja događaja za svaku klasu se normiraju i množe s konstantom normalizacije N_S koja je odabrana samo iz numeričkih razloga i nema utjecaj na performans. Konstanta skaliranja svake klase f_C trebala bi se odabrati tako da reflektira odnose između stvarnog broja događa između različitih procesa (već prije diskutiranog u III.D i prikazanog na Slici 6). Prvi dio izraza stoga govori o odnosima između pojedinih događaja unutar jedne klase, dok f_C govori o globalnim odnosima između klasa (npr. uzima se u obzir informacija da su događaji koji potječu od W+mlazova najbrojniji). S obzirom da se za konstante skaliranja f_C u principu može odabrati bilo koji pozitivni realni broj dobro bi došao određeni sustav vođenja evidencije – određivanja koji su odabiri konstanti skaliranja međusobno ekvivalentni. Zbog toga se uvodi nova veličina

$$F_C = \frac{f_C}{\bar{f}} ; \quad \bar{f} = \frac{1}{5} \sum_c f_c$$

koja ima praktična svojstva $F_C \in [0,5]$ i $\sum_c F_C = 5$. Sheme balansiranja događaja s istim izborom $\{F_C\}$ su međusobno ekvivalentna (u smislu da je minimum los funkcije na istoj kombinaciji $\{\omega_i\}$ – minimum se ne mijenja ako se los funkcija pomnoži s nekom konstantom). U izrazu za težine $\tilde{w}_C^{(k)}$ bez smanjenja općenitosti konstante skaliranja f_C mogu se zamijeniti s F_C .



Slika 12 Usporedba signifikantnosti dobivene za različite sheme balansiranja događaja karakterizirane s δ . Usporede su napravljene za treninge s 4 različite vrijednosti parametra D i jednom vrijednosti parametra L .

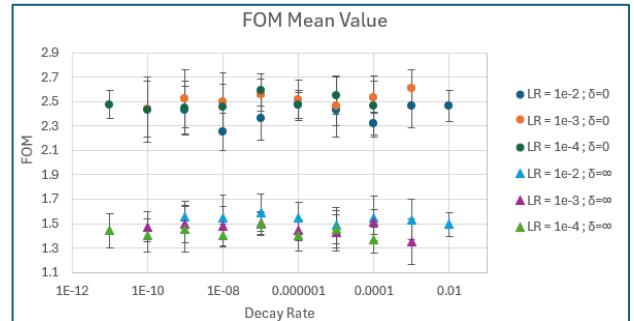
VII. REZULTATI I DISKUSIJA

Za odabir konstanti skaliranja ovdje se konkretno uzimaju vrijednosti dane izrazima:

$$f_C = \frac{N_C^{sample}}{N_S^{sample}} + \delta$$

gdje N_C^{sample} za pozadinske procese predstavlja ukupni (u eksperimentu) očekivani broj događaja te klase predviđen na temelju Monte Carlo simulacija, dok je N_S^{sample} uvećan s obzirom da se u setu podataka iz simulacija nalaze rezultati za različite kombinacije masa neutralina i stopa (ali istu razliku $\Delta m(\tilde{t}, \tilde{\chi}_0) = 30 GeV$). Pošto je stvarna statistika broja događaja iz eksperimenta za stop značajno potisnuta, a najbitnija klasa za izdvajanje, ovime se umjetno povećava za onaj iznos koliko je uvećana statistika uzorka iz simulacija. Istovremeno se ovim odabirom za konstante f_C čuva relativni odnos između stvarnog broja događaja ostalih klasa. Konstanta δ uvedena je kako bi se postepeno uvodio balans između klasa. Ovisnost konstanti F_C o parametru δ za ovaj izbor

konstanti skaliranja svake klase prikazan je na *Slici 12*. Slučaj $\delta = \infty$ odgovara situaciji kada su sve klase stavljene na istu razinu, tj. $F_C(\infty) = 1$. Treninzi neuronske mreže obavljeni su za šest vrijednosti parametra $\delta = 0, 1, 2, 3, 4, 5, \infty$ i uspoređeni rezultati za signifikantnost. Signifikantnost se računa na način da se obavlja rez na izlazu neuronske mreže koji odgovara signalu za raspodjelu događaja koje je neuronska mreža kategorizirala kao signalne događaje. Za jednu kombinaciju parametara L i D iz IV.A. obavljeno je 10 treninga i za svaki od njih izračunata signifikantnost – predstavljene su očekivane vrijednosti i standardne devijacije signifikantnosti kroz tih 10 treninga. Rezultati koji uspoređuju izračunatu signifikantnost za šest shema balansiranja događaja karakterizirane s $\delta = 0, 1, 2, 3, 4, 5, \infty$ i 4 različite vrijednosti parametra D te jednu vrijednost L prikazani su na *Slici 12*. Uočava se silazni trend signifikantnosti na svim predstavljenim vrijednostima parametara L i D kako problem postepeno postaje balansiran. Radi potpunosti, za dvije rubne vrijednosti $\delta = 0$ i $\delta = \infty$ obavljeni su treninzi za veći broj kombinacija L i D , a rezultati su prikazani na *Slici 13*.



Slika 13 Usporedba signifikantnosti dobivene za dvije granične sheme balansiranja ($\delta = 0$ i $\delta = \infty$) za veći broj kombinacija parametara L i D korištenih tijekom treninga neuronske mreže.

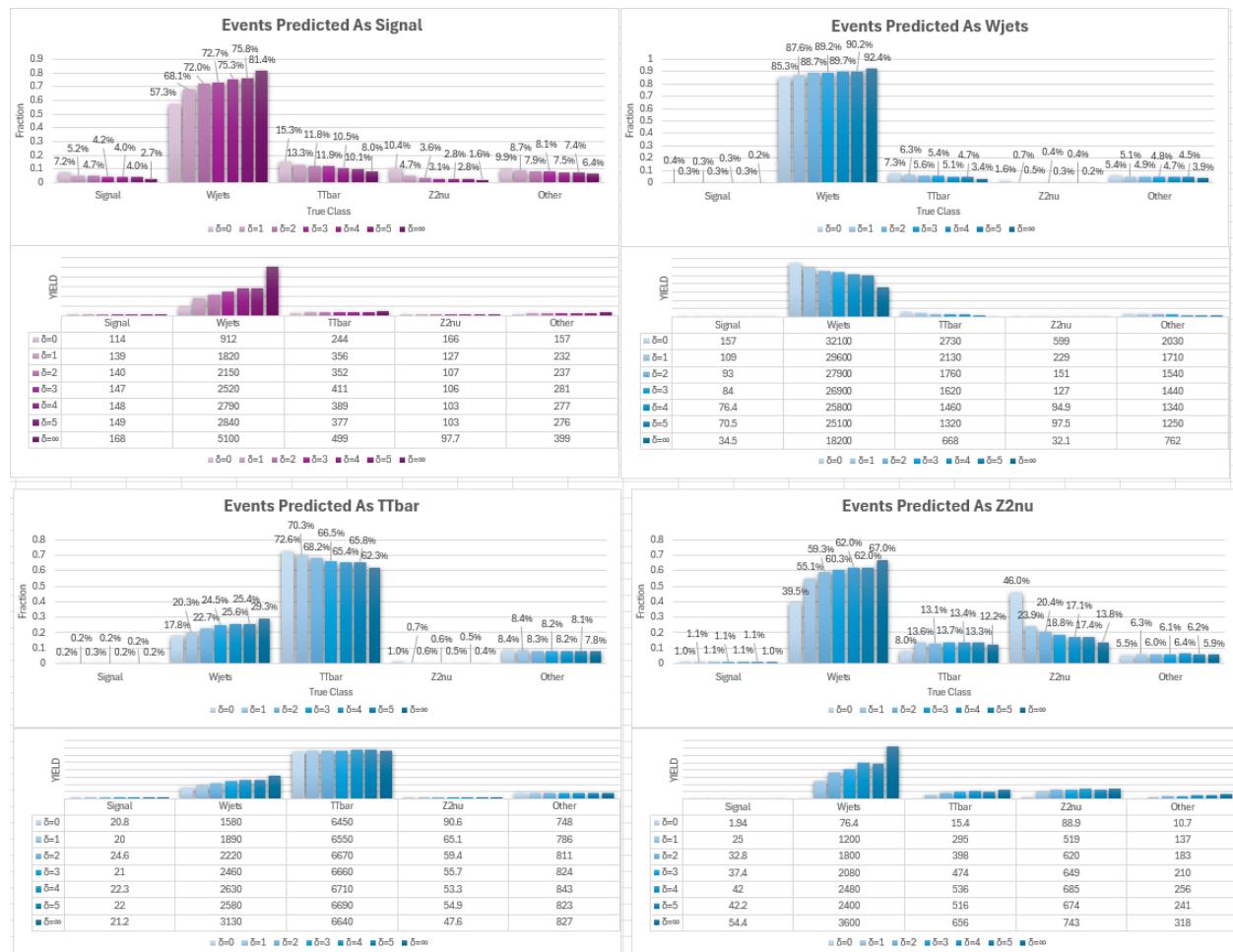
Potpuno balansiran problem $\delta = \infty$ kroz težine los funkcije $\tilde{w}_C^{(k)}$ u trening uvodi samo informaciju o raspodjeli vjerojatnosti nalaženja pojedine klase u danoj točki (k-tog događaja C-te klase) prostora diskriminirajući varojabli, ali relativno u odnosu na tu klasu. Može govoriti o određenom postotku događaja dane klase, ali ne i o stvarnom očekivanom broju događaja. Rezultat toga je vrlo loše razdvajanje signalnih događaja od W+mlazova jer se npr. 1% signalnih događaja tretira jednako važno (jednako brojno) kao 1% W+mlazova. Pogledaju li se stvarni odnosi broja događaja, taj je omjer zapravo $\approx 1:60$. Taj se efekt dodatno vidi ako se pogleda raspodjela događaja po klasama na signalnom izlazu – prelaskom sa

slučaja $\delta = \infty$ na $\delta = 0$ broj signalnih događaja se smanjio za faktor 0.3, a $W+mlazova$ za faktor 0.82.

S obzirom da je najvažniji aspekt ovog istraživanja pronaći područja prostora razapetog odabranim diskriminirajućim varijablama koja su bogata signalom, nije odmah jasno zašto se u tu svrhu koristi multiklasna neuronska mreža – najbitnija informacija je da li događaj pripada signalu, a odgovor je binarne prirode, da ili ne, tj. klase su signal i pozadina. Izdvajanje drugih klasa, $W+mlazova$, proizvodnje $t\bar{t}$ i $Z+mlazova$, iako nije jedanko važno kao signal, bitno je za definiranje kontrolnih područja koja u konačnici omogućuju da se preciznije odrede očekivani doprinosi tih klasa u ukupnom broju događaja. Multiklasna neuronska mreža zapravo u tom pogledu nudi praktično rješenje jer istovremeno s potragom signala, izdvaja i kontrolna područja koja bi se mogla koristiti u kasnijoj analizi. Raspodjela događaja po klasama na izlazu $W+mlazova$, proizvodnje $t\bar{t}$ i $Z+mlazova$ prikazana je na *Slici 14*. Kako bi kontrolna područja bila upotrebljiva za daljnju analizu bitno je da kontaminacija signalom (udio signalnih događaja u

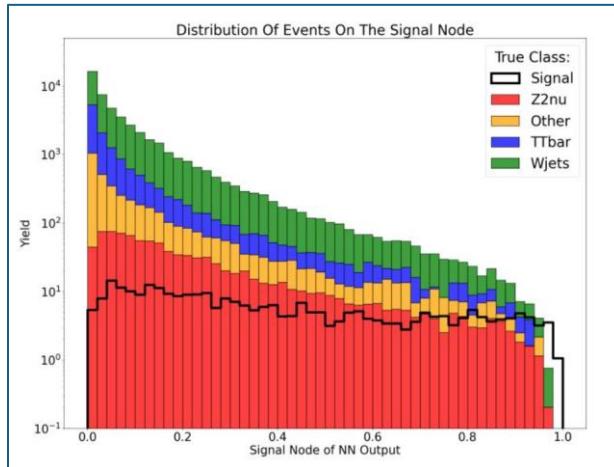
kontrolnom području) bude mala – idealno manja od 1%, križna kontaminacija (udio ostalih pozadinskih procesa u kontrolnom području dane klase) ne bude značajna – većinski udio bi trebao dolaziti od klase kojoj odgovara dano kontrolno područje, te na kraju, bitno je da ima dovoljno statistike (ukupni broj događaja). Za bilo koji izbor sheme balansiranja kontrolna područja za procese $W+mlazova$ i proizvodnje para $t\bar{t}$ mogu se upotrijebiti – u tom slučaju je možda najbolji odabir $\delta = 0$ uzimajući u obzir obe kontrolne zone. Problem se javlja kod definicije kontrolnog područja za $Z+mlazove$. Udeo tih procesa u ukupnom broju događaja u izlazu za $Z+mlazove$, iako je najbolji za slučaj $\delta = 0$, problematičan je zbog smanjene statistike. U ostalim slučajevima, problem je također nezanemariva križna kontaminacija, ali je ipak upotrebljiva za konačnu analizu.

Cijena koju se mora platiti ako se žele koristiti kontrolne zone određene klasifikacijom neuronske mreže je smanjena statistika (broj događaja) u zoni gdje se obavlja izdvajanje signalnog područja (izlaz neuronske mreže koji odgovara signalu) – uključuju samo događaje



Slika 14 Raspodjela broja događaja po klasama koje u koje ih je svrstala neuronska mreža.

koji su neuronskom mrežom klasificirani kao signal. Za velike vrijednosti parametra δ , statistika, iako je smanjena, upotrebljiva je za obavljanje prilagodbe. Za slučaj $\delta = 0$ (za koju se dobiva najveća signifikantnost), to postaje problematično. Međutim, u tim slučajevima multiklasna neuronska mreža može se shvatiti kao binarni klasifikator i analizirati raspodjelu broja svih događaja na izlazu neuronske mreže koji odgovara signalu (*Slika 15*). Upotreba aktivacijske funkcije *softmax* osigurava da takva procedura nema utjecaj na izračunate vrijednosti signifikantnosti – rezovi koji maksimiziraju signifikantnost u svim se slučajevima nalaze iznad vrijednosti signalnog izlaza od 0.5. Prebacivanjem događaja klasificiranih kao neki drugi proces na signalni izlaz time je osigurano da se nužno nalaze ispod te vrijednosti (ako je događaj klasificiran kao neki proces to znači da je vrijednost na pripadnom izlazu neuronske mreže veća od vrijednosti na svim ostalim izlazima – maksimalna vrijednost na izlazu koji ne odgovara tom procesu je stoga 0.5). Definiranje kontrolnih područja se u tom slučaju treba obavljati u novom prostoru koji isključuje događaje prostora u kojem se obavljala trenutna analiza – takva je procedura napravljena u prethodnim istraživanjima s BDT-om.



Slika 15 Raspodjela očekivanog broja događaja po vrijednosti izlazne varijable neuronske mreže koja odgovara signalu.

VIII. ZAKLJUČAK

Potraga za raspadom stopa u kanal s 4 izlazne čestice za koji se zbog prepostavki modela o razlici mase stopa i neutralina $\Delta m(\tilde{t}, \tilde{\chi}_0) = 30\text{GeV}$ smatra da ima omjer grananja jednak jedinici, obavljena je na podacima iz simulacija u prostoru razapetom odabranim setom kinematičkih varijabli i određenom uvjetima preselekcije. Multiklasna neuronska mreža u tom prostoru obavlja

klasifikaciju signala i pozadinskih procesa iz standardnog modela: $W+mlazova$, proizvodnje $t\bar{t}$, $Z+mlazova$ te ostalih manjih pozadina. Pokazano je da se multiklasna neuronska mreža općenito može koristiti poput binarnog klasifikatora u kojem se izlaz neuronske mreže koji odgovara signalu koristi za linearno odvajanje signala od pozadinskih procesa. Prednost multiklasne mreže je u tome što se kroz balasiranje događaja svaka klasa može tretirati posebno te odgovarajući parametri u shemama balansiranja namještati tako da daju željeni performansi. U ovom konkretnom slučaju upitno je da li je neuronsku mrežu najpovoljnije koristiti kao binarni klasifikator iako se u tom slučaju dobivaju najveće vrijednosti očekivane signifikantnosti (FOM=2.6). Razlog je u tome što se odustaje od definiranja kontrolnih područja u istom prostoru u kojem se obavlja potraga za signalnim čime se smanjuje preciznost konačne analize. Stoga je alternativa koristiti sheme balansiranja koje su između dvije krajnosti; potpuno balansiranog slučaja ($\delta = \infty$) i gore diskutiranog ($\delta = 0$). Ako se za vrijednost parametra koji postepeno uvodi balans između klase odabere vrijednost $\delta = 2$ dobivena vrijednost signifikantnosti je FOM=2. Taj rezultat i dalje je bolji od onih dobivenih u prethodnim istraživanjima sa binarnim klasifikatorima: binarnom neuronskom mrežom (FOM=1.8) i BDT-om (FOM=1.6).

LITERATURA

- [I] CMS Collaboration, „*Search for top squarks in the four-body decay mode with single lepton final states in proton-proton collisions at $\sqrt{s} = 16\text{TeV}$* “, Journal of High Energy Physics (2023)
- [II] Anadi Canepa, „*Searches for supersymmetry at the Large Hadron Collider*“, Reviews in Physics (2019)