

UPORABA NEVRONSKIH MREŽ PRI ISKANJU PAROV HIGGSOVIH BOZONOV

ANDRAŽ TOMŠIČ

Fakulteta za matematiko in fiziko
Univerza v Ljubljani

Ekperimentalna potrditev načina nastanka parov Higgsovih bozonov je ključnega pomena pri potrditvi teorije Standardnega modela oziroma preverjanju hipotez izven te teorije. Ta članek predstavi fizikalne osnove nastanka, razpada in detekcije takšnih parov. Prikazana je tudi uporaba nevronske mreže pri ločevanju detektiranih delcev, ki so nastali iz para Higgsovih bozonov in detektiranih delcev, nastalih iz drugih procesov.

THE USE OF NEURAL NETWORKS IN THE SEARCH FOR HIGGS BOSON PAIRS

Experimental confirmation of the formation of Higgs boson pairs is crucial in confirming the theory of the Standard Model or checking hypotheses outside of this theory. This article presents the physical basis of the formation, decay and detection of such pairs. Additionally it introduces the application of neural networks in the separation of detected particles created from a pair of Higgs bosons and detected particles created from other processes.

1. Uvod

Higgsov bozon (H) je osnovni delec znotraj Standardnega modela (SM) [1] in je nosilec Higgsovega polja, podobno kot je foton nosilec elektromagnetnega polja. Delec je prvič predstavil Peter Higgs in pet drugih znanstvenikov leta 1964 [2]. Teorija Higgsovega polja je predstavila rešitev problema nastanka mase delcev iz elektrostatskih interakcij z izjemno natančnimi napovedmi. Obstoj delca je bil potrjen leta 2012 s pomočjo trkalnika LHC in eksperimenta ATLAS [1]. Do današnjega dne je odkritje Higgsovega bozona največji uspeh eksperimentov CERN-a, ki je utrdil teorijo Standardnega modela.

Poleg enega Higgsovega bozona Standardni model napoveduje tudi pare, ki nastanejo preko sklopitve Higgsovega bozona s samim sabo, kar je bistvena napoved SM. Nastanek teh parov je mnogo bolj redek od nastanka H , zato jih še nismo uspeli potrditi. Poleg standardnega modela tudi druge teorije napovedujejo pare Higgsovih bozonov (HH), ki imajo drugačen način nastanka kot napoveduje SM.

Detekcija HH in določanje načina nastanka je ključnega pomena pri preverjanju temeljnih fizikalnih teorij. Ker iz detektorja ATLAS dobimo ogromno količino podatkov, potrebujemo napredne algoritme za njihovo obdelavo. Modeli strojnega učenja nam pomagajo določiti zapletene odvisnosti med fizikalnimi opazljivkami, saj nekaterih relacij ne moremo določiti s fizikalnim razmislekom. Izkaže se, da pri problemu iskanja parov HH odlično delujejo nevronske mreže, ki si jih bomo podrobneje ogledali kasneje. Poleg tega zelo dobro deluje tudi algoritem *Boosted decision trees*, ki spada v družino odločitvenih dreves.

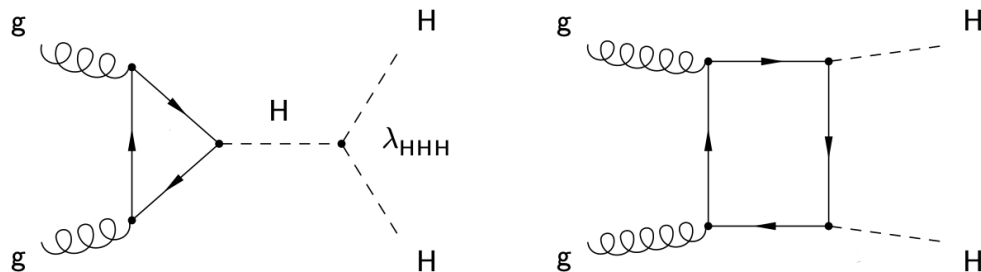
2. Nastanek parov Higgsovih bozonov

Pari Higgsovih bozonov lahko nastanejo pri trku dveh protonov v trkalniku LHC. Kvarki, ki sestavljajo proton, pri dovolj visokih energijah interagirajo med seboj preko barvne interakcije. Nosilec te interakcije je gluon, ki trči z drugim gluonom.

Predstavili bomo ne-resonančno in resonančno produkcijo parov HH [3] iz dveh gluonov. To sta najbolj pogosta načina nastanka HH , ki ju lahko ločimo po različnih intenzivnostih razpadnih kanalov zaradi drugačne razpoložljive energije HH glede na mehanizem produkcije. Posledično so tudi porazdelitve kinematičnih opazljivk razpadlih produktov različne.

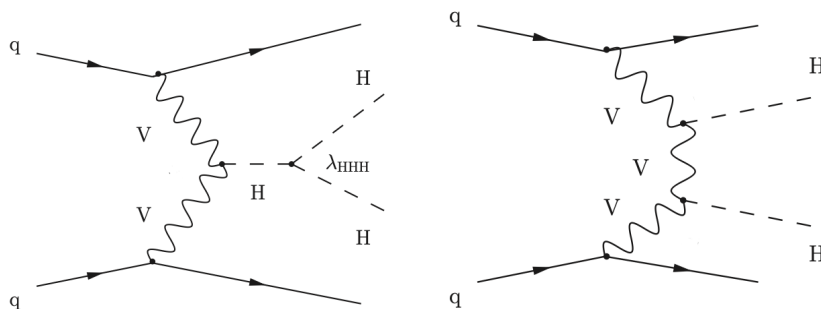
2.1 Produkcija HH po napovedih SM

Obravnavali bomo nastanek H iz dveh gluonov (gg). Standardni model napoveduje samo-sklopitev H , pri čemer nastane par HH . Presek te interakcije znaša $\sigma_{SM}(gg \rightarrow HH) = 31.1 \text{ fb}$ [4]. Ta proces si lahko ogledamo na sliki 1. Poleg resonančne produkcije preko samo-sklopitve H je prikazan tudi ne-resonančni nastanek preko sklopitve virtualnih kvarkov. Obstajajo tudi drugi načini nastanka



Slika 1. Diagram procesov nastanka HH iz dveh gluonov. Levi prikazuje nastanek preko samo-sklopitve H , desni pa ne-resonančni nastanek. Prirejeno po [4].

HH , ki imajo manjši presek in posledično še manjše verjetnosti nastanka. En izmed načinov, ki se poleg nastanka iz gg uporablja pri detekciji parov HH , je nastanek HH iz dveh vektorskih bozonov po različnih procesih. Dva izmed njih sta prikazana na sliki 2. Presek te interakcije znaša $\sigma_{SM}(VV \rightarrow HH) = 1.73 \text{ fb}$.



Slika 2. Diagram procesov nastanka HH iz dveh vektorskih bozonov. Vektorski bozon je označen z V , kvark pa s q . Prirejeno po [4].

Ključna napoved, na kateri sloni veljavnost standardnega modela, je Higgsov potencial [5]

$$\Phi(\psi_H) = \frac{1}{2}m_H^2 \psi_H^2 + \lambda_{HHH} \psi_H^3 + \frac{1}{4}\lambda_{HHH} \psi_H^4 + \mathcal{O}(\psi_H^5) \quad (\lambda_{HHH}^{SM} \simeq 0.13),$$

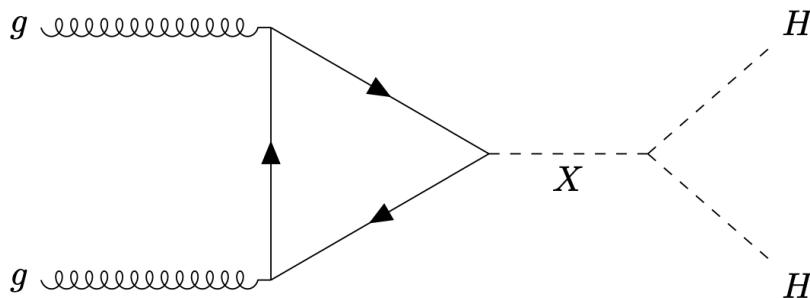
kjer je ψ_H valovna funkcija Higgsovega bozona, $\nu = 246 \text{ GeV}$ [5] in $m_H = 125 \text{ GeV}$ masa Higgsovega bozona. Obravnava izvora potenciala Φ ni pomembna za razumevanje našega problema.

Potrditev nastanka parov oziroma relativno natančna meritev konstante λ_{HHH} , ki predstavlja verjetnost oziroma kanal za samo-sklopitev $H \rightarrow HH$, je ključna za potrditev napovedi Standardnega modela. Odstopanje od te vrednosti bi lahko nakazovalo na fiziko onkraj Standardnega modela (angl. *Beyond Standard model*, *BSM*).

2.2 Resonančna produkcija HH onkraj SM

Po nekaterih teorijah onkraj Standardnega modela obstaja možnost nastanka HH preko drugačne resonančne produkcije. Tu je posrednik pri nastanku HH težka (spin-0) resonanca X [3]. Pri tem izračunu se tipično upošteva samo najbolj dominanten način nastanka težke resonance t.j. iz dveh

gluonov. V aktualnih teoretičnih modelih BSM se uporablja resonance z maso $m_X \in [251 \text{ GeV}, 1600 \text{ GeV}]$ [6]. Proces je prikazan na sliki 3.



Slika 3. Diagram procesa nastanka parov HH preko težke resonance X . Vzeto iz [7].

3. Razpad in detekcija parov HH

Detekcija parov Higgsovih bozonov je izjemno težavna naloga, saj je verjetnost za nastanek nizka zaradi majhnega preseka interakcije ($gg \rightarrow HH$), poleg tega pa so ti delci nestabilni in prehitro razpadejo, da bi jih lahko neposredno zaznali. Zaradi tega poizkušamo detektirati delce nastale iz razpadov HH , ki jih imenujemo “signal”. Z detekcijo signala namreč potrdimo obstoj parov Higgsovih bozonov. Če obravnavamo problem v težiščnem sistemu reakcije, ima signal (relativistično) Breit-Wignerjevo (BW) verjetnostno porazdelitev po energijah centrirano pri invariantni masi Higgsovega bozona

$$f(E) = \frac{2E}{\pi} \frac{m_H \Gamma}{(E^2 - m_H^2)^2 + m_H^2 \Gamma^2} \theta(E) \quad [8] \quad (1)$$

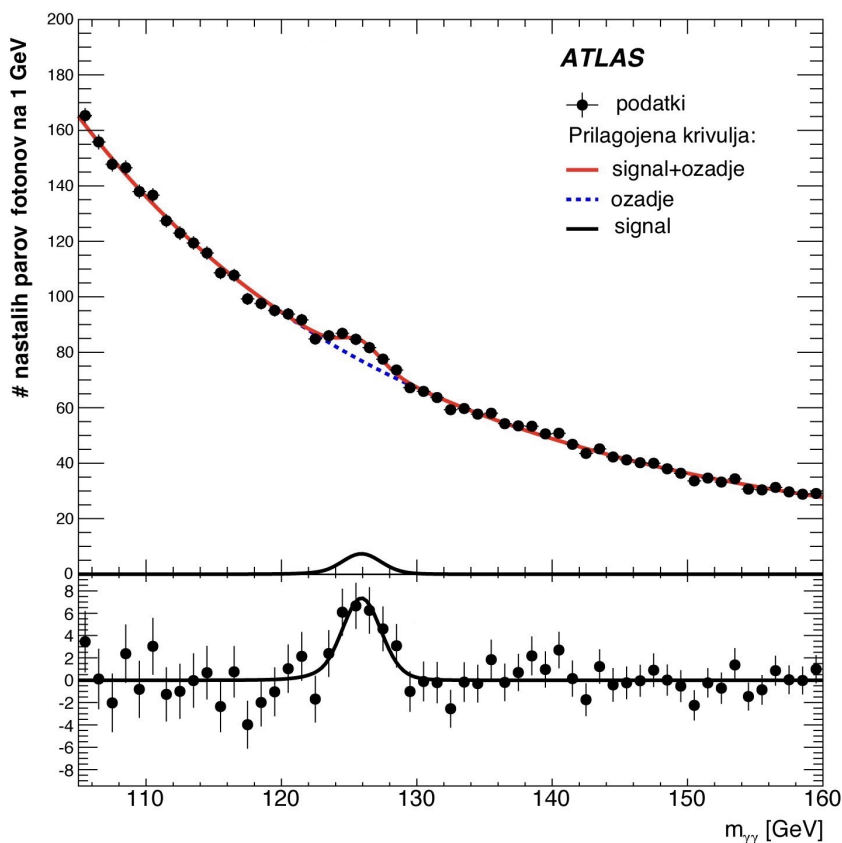
pri čemer je razpoložljiva energija v sistemu označena z E , naravna širina Higgsovega bozona z $\Gamma = 4.1 \text{ GeV}$ [9], njegova mirovna masa z $m_H = 125 \text{ GeV}$ [6] in Heavisidova funkcija s θ .

Ker je presek nastanka HH izredno majhen, detektiramo poleg trkov oziroma dogodkov signala tudi ozadje. To so dogodki, ki so zelo podobni signalu, vendar niso nastali preko tvorbe para HH . Razlikujejo se po verjetnostni porazdelitvi, ki je odvisna od porabljene energije pri trku dveh p . Ta ne vsebuje ekstrema pri invariantni masi Higgsovega bozona. Na sliki 4 je prikazan masni spekter števila nastalih parov fotonov pri iskanju enega H . Nekateri fotoni so nastali iz Higgsovega bozona in za njih velja porazdelitev iz enačbe 1. Zaradi tega ima spekter lokalni vrh pri m_H . Podobno porazdelitev pričakujemo za HH le z veliko več šuma in šibkejšim signalom, saj je nastanek para še mnogo bolj redek. Pri iskanju s strojnim učenjem je potrebno upoštevati tudi ostale spremenljivke, ki vplivajo na verjetnost nastanka parov, ne le energijske odvisnosti.

3.1 Razpad v druge delce

Par Higgsovih bozonov lahko razpade v več različnih delcev. Izbira razpadnega kanala za obravnavo je težavna naloga, saj je potrebno najti dober kompromis. Po eni strani si želimo močen signal (veliko razpadov), vendar imajo taki razpadi običajno močno ozadje, ki nas ovira. Pri izbiri kanala je potrebno upoštevati tudi ločljivost detektorja za posamezne delce, enostavnost proženja in več ostalih eksperimentalnih podrobnosti [1]. Vsak razpadni kanal ima več prednosti in slabosti zato v praksi znanstveniki uporabljajo več različnih razpadnih kanalov in na različne načine poizkušajo potrditi obstoj para HH .

Tabela 1 prikazuje verjetnost razpadov HH v različna končna stanja. Poleg vseh ostalih lastnosti je potrebno pri izbiri razpadnega kanala upoštevati tudi energijo nastalih delcev, ki sledi porazdelitvi iz enačbe 1. Ta vpliva na verjetnosti nastanka drugih delcev iz HH . Na sliki 5 so prikazane



Slika 4. Histogram detektiranih fotonov v odvisnosti od energije pri iskanju enega Higgsovega bozona s trki p^+p^+ pri težišču energije 7 oz. 8 TeV. Prirejeno po [10].

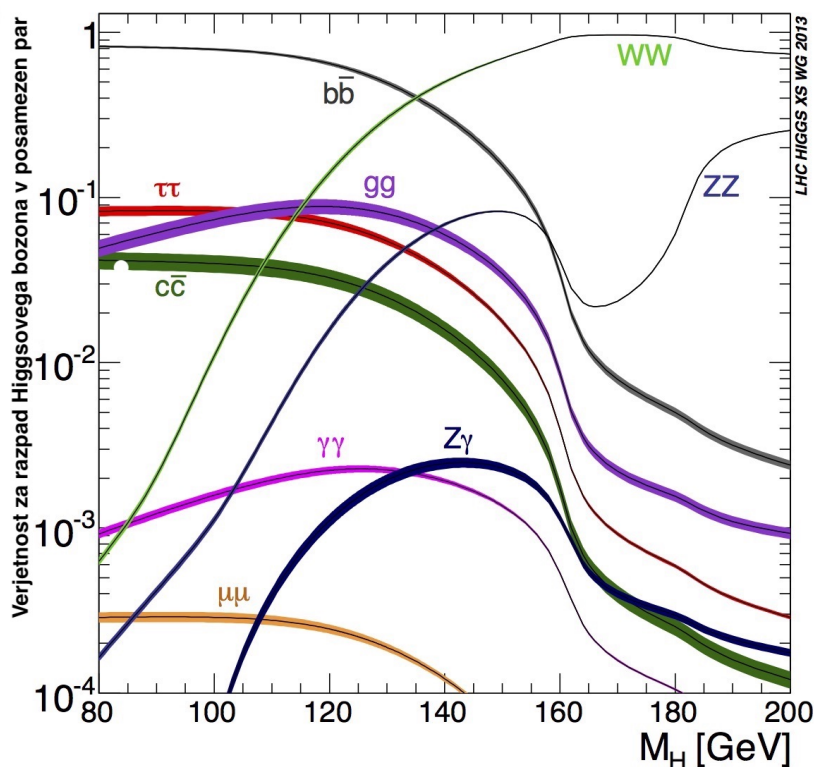
	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	33%				
WW	25%	4.6%			
$\tau\tau$	7.3%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

Tabela 1. Verjetnosti razpadov para Higgsovih bozonov v različne možnosti dveh parov delcev. Vzeto iz [3].

verjetnosti razpada H z negotovostmi v odvisnosti od energije nastalih delcev. To odvisnost je potrebno upoštevati pri simulacijah trkov in učenju modela za klasifikacijo trkov t.j. ugotavljanju ali so delci nastali iz HH ali po kakšnem drugem procesu.

3.2 Primer izbire razpadnih kanalov

Naš cilj je izbrati statistično neodvisno kombinacijo razpadnih kanalov, ki imajo čim večjo občutljivost t.j. čim več signala v primerjavi s šumom. V praksi se uporablja veliko različnih možnosti za učenje več vrst modelov. Eden izmed načinov je kombinacija kanala $b\bar{b}\tau^+\tau^-$ in $b\bar{b}\gamma\gamma$ za neresonančno produkcijo. Za resonančno produkcijo porabimo poleg teh dveh tudi razpad v $b\bar{b}b\bar{b}$ [3].



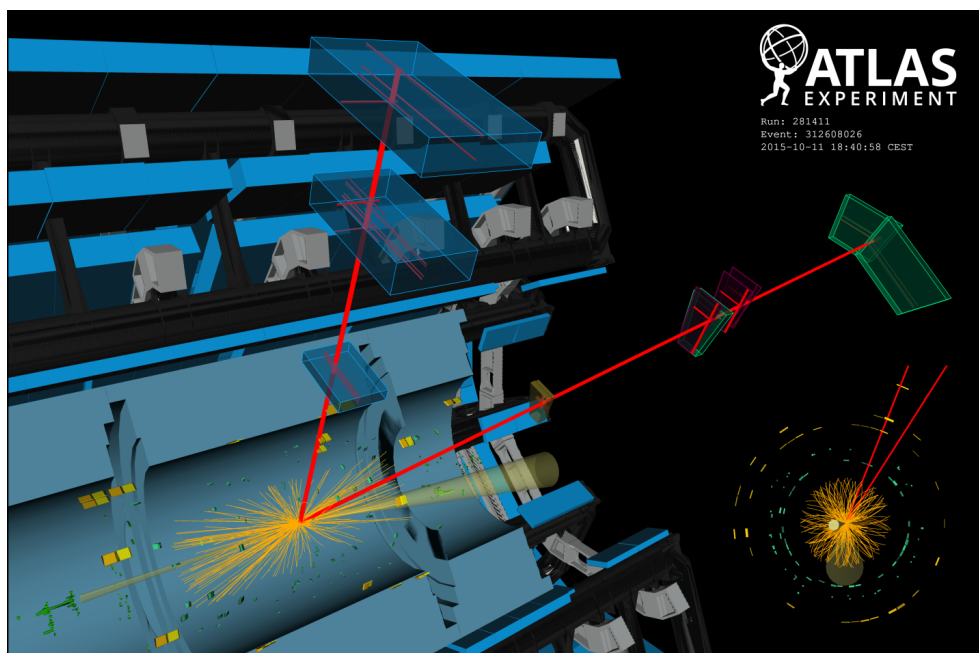
Slika 5. Verjetnosti razpada H v različne pare delcev v odvisnosti energije H . Prirejeno po [11].

4. Opis trkalnika LHC in delovanja detektorja ATLAS

LHC (large hadron collider) [12] je največji in najmočnejši pospeševalnik delcev na svetu. 27 km obroč superprevodnih magnetov hlajenih na -271.3°C s pomočjo tekočega helija pospeši protone na izredno visoke energije z Lorentzovim faktorjem γ reda 10^4 . Tisoče različnih magnetov upogiba tok protonov. Pred trkom poseben magnet "stisne" tok protonov, s čimer povečamo luminoznost (večja gostota trkov na enoto časa) in tako na bolj učinkovit način dobimo več fizikalno zanimivih procesov [12].

ATLAS je en izmed dveh večnamenskih detektorjev. Detektor ima obliko 46 m dolgega in 25 m širokega cilindra. Detektor je osno simetričen. To pomeni, da je polarni kot ortogonalna spremenljivka glede na verjetnost nastanka HH , kar bomo uporabili pri validaciji nevronske mreže. Tehta približno 7000 ton, kar je podobno masi Eifflovega stolpa. Šest različnih podsistemov, zloženih po plasteh, se uporablja za merjenje različnih fizikalnih opazljivk kot so gibalna količina, naboj, trajektorija in energija delca [13, 14]. Shema detektorja si lahko ogledamo na sliki 6.

V ATLAS-u se zgodi ogromno trkov z podatkovnim tokom več kot 60 terabajtov na sekundo. Le nekateri od teh trkov vsebujejo zanimive karakteristike, ki jih lahko uporabimo za odkrivanje nove fizike. Podatke se zato filtrira s pomočjo naprednega prožilnega sistema [16], ki pove detektorju, katere dogodke naj zapiše in katere zavrže. Prožilni sistem je sestavljen iz dveh delov. Prvi del prožilnega sistema je vgrajen v strojni opremi in prepusti okrog 100000 dogodkov na sekundo. Drugi del je vgrajen v programski opremi, ki s pomočjo 40000 procesorskih jeder procesira podatke in prepusti okrog 1000 dogodkov v sekundi, kar lahko shranimo in kasneje uporabimo za obdelavo [16].



Slika 6. Prikaz potencialnega H , ki razpada v dva miona (rdeči črti) in dva curka hadronov znotraj detektorja ATLAS. Izvzeto iz [15].

5. Nadzorovano strojno učenje in njegova uporaba pri iskanju procesa HH

Strojno učenje uporabljamo v fiziki, kadar želimo uporabiti kombinacijo vhodnih spremenljivk oziroma značilk, da napovemo ciljno spremenljivko, pri tem pa ne poznamo relacij med vsemi značilkami in ciljno spremenljivko [17]. Dodaten razlog, da si pri iskanju HH pomagamo s strojnimi učenjem, je izredno šibek signal v primerjavi z ozadjem, saj je nastanek HH še mnogo bolj redek proces kot proizvodnja posameznih H . Slabost nekaterih metod strojnega učenja je, da nam ne olajšajo razumevanja povezav med spremenljivkami, saj so to lahko izredno zapletene funkcije. Ena izmed teh metod je (globoka) nevronska mreža, ki si jo bomo podrobneje ogledali kasneje.

Fizikalno razumevanje vsaj nekaterih relacij med spremenljivkami ali simetrij je izjemno pomembno, saj algoritmu strojnega učenja poenostavi nalogo, ki posledično proizvede boljše rezultate. Analogno si lahko predstavljamo, da vsaj delen analitičen rezultat omogoča nižjo časovno zahtevnost in boljše natančnost klasični numerični metodi.

Nabori podatkov iz raznih področij, v katerih določamo ciljno spremenljivko, imajo različno naravo oziroma lastnosti. Opis teh značilnosti so meta-podatki. Zaradi različnih karakteristik, v katere se ne bomo spuščali, delujejo nekatere metode odlično na enem naboru podatkov, na drugem pa popolnoma odpovejo. V primeru določanja HH se izkaže, da v splošnem dobro delujejo (globoke) nevronske mreže in odločitvena drevesa, bolj specifično metoda z imenom *boosted decision trees* [6].

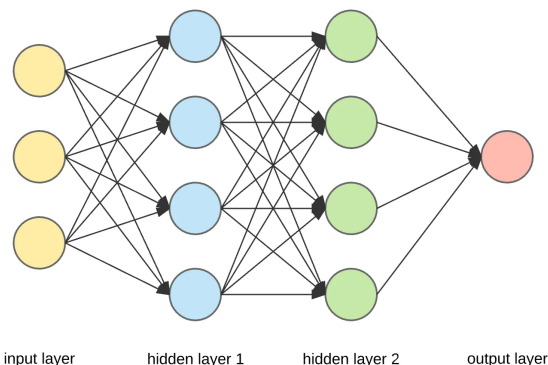
Ker se področje strojnega učenja v zadnjem času izredno hitro razvija, obstajajo tudi naprednejše in boljše metode npr. različne variacije nevronske mreže, ki še niso bile uporabljene pri nalogi iskanja HH , čeprav kažejo obetavne rezultate pri primerjalni analizi algoritmov na javno dostopnem naboru podatkov, kot je nabor HIGGS [18].

6. Delovanje nevronske mreže

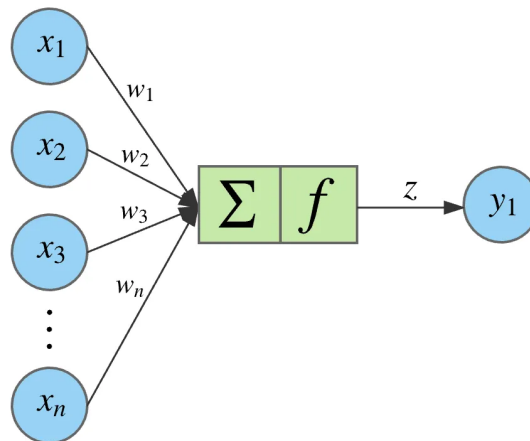
Nevronska mreža (NN) [17] je nelinearna funkcija, sestavljena iz preprostih elementov (neuronov), ki jo uporabimo za določanje ciljne spremenljivke v odvisnosti od vhodnih spremenljivk. V našem primeru je ciljna spremenljivka vrednost med 0 in 1, ki predstavlja verjetnost, da so detektirani

delci nastali iz HH . Preprost model NN je prikazan na sliki 7.

Na sliki 8 si lahko podrobneje ogledamo posamezen nevron. Posamezno vozlišče iz n vhodnih spremenljivk, zloženih v vektor $\vec{x} \in \mathbb{R}^n$, proizvede uteženo vsoto, ki je nato uporabljena kot argument t.i. aktivacijske funkcije f . Posledično je celoten sloj nevronske mreže $\vec{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ definiran kot $\vec{F}(\vec{x}) = \vec{f}(\mathbf{W}\vec{x} + \vec{b})$ [19]. Aktivacijska funkcija pri tem deluje na vsako komponento argumenta dolžine m posebej in vrne vektor dolžine m , zato je označena z \vec{f} . Cilj učenja NN na podatkih je določitev matrike uteži $\mathbf{W} \in \mathbb{R}^{m \times n}$ in vektorja pragov $\vec{b} \in \mathbb{R}^m$.



Slika 7. Preprost model NN sestavljen iz dveh skritih plasti, ciljne plasti in vhodne plasti iz treh spremenljivk. Vzeto iz [19].



Slika 8. Delovanje posameznega vozlišča, ki izhod izračuna preko aktivacijske funkcije z argumentom utežene vsote vhodnih spremenljivk. Vzeto iz [19].

Najbolj pogosto uporabljena je aktivacijska funkcija ReLU (*rectified linear unit*), definirana kot

$$\text{ReLU}(x) = \max(0, x),$$

saj je njen izračun izredno hiter, hkrati pa je v mnogo primerih izredno uspešna. Obstaja mnogo drugih aktivacijskih funkcij npr. hiperbolični tangens, sigmoid... [20]

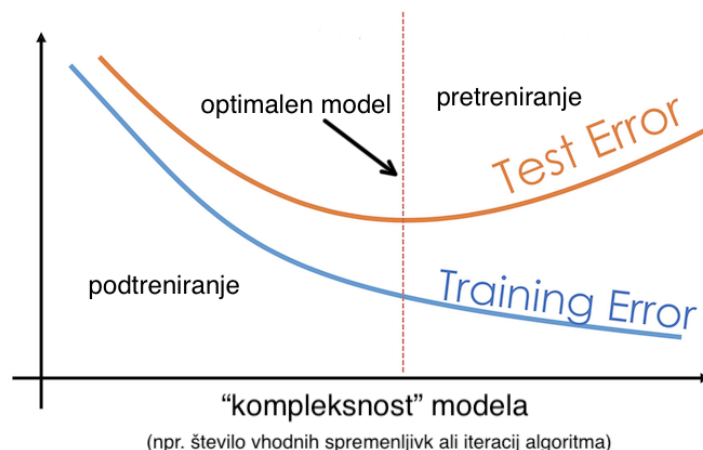
Aktivacijska funkcija, algoritem učenja, število skritih plasti itd. so prosti parametri modela, ki so določeni pred učenjem. Običajno je nepredvidljivo, katera kombinacija parametrov bo najboljša, zato naučimo več modelov in nato izberemo tistega, ki proizvede najboljše rezultate znotraj določene evaluacijske metrike (npr. lažen pozitiven delež) oziroma znotraj kombinacije teh metrik.

6.1 Učenje nevronske mreže

Podatki za učenje in validacijo algoritma vsebujejo nabor lastnosti oziroma karakteristik (npr. gibalna količina, invariantna masa, energija, ...) in pravilen odgovor za posamezen vnos. Te podatke dobimo iz simulacij trkov [21], ki so ustvarjeni iz teoretične napovedi, saj imajo tako podatki vključeno vrednost izhodne spremenljivke, torej verjetnost zaznanega HH . Večji del podatkov uporabimo za učenje nevronske mreže, nekaj pa jih odvezamemo in kasneje uporabimo za validacijo oziroma testiranje nevronske mreže.

Pomembno je, da algoritem posploši napoved na podatke, ki jih še ni videl. Premajhna napaka na podatkih za učenje (angl. *training error*) glede na napako na testnih podatkih nakazuje na pretreniran model na že videnih podatkih. V takem primeru se je potrebno vrniti nekaj korakov nazaj, saj tako dobimo manjšo testno napako (angl. *test error*). Ta proces prikazuje slika 9.

Nato poizkušamo izbrati dober začetni približek za nabor prostih parametrov. V primeru učenja metod na podatkih iz detektorja je poleg poznanih relacij med posameznimi spremenljivkami vredno upoštevati osnovne fizikalne zakone, kot je ohranitev energije.



Slika 9. Napaka na podatkih za učenje in testna napaka v odvisnosti od kompleksnosti algoritma. Prirejeno po [22].

Algoritem učenja oziroma optimizator, kot sta *back-propagation* [23] in *Adam* [24] iščeta minimum funkcije izgube L . Recimo, da so prave izhodne spremenljivke vrednosti \vec{y} , napovedi modela pa $\vec{\hat{y}}$. Enostaven primer je funkcije izgube povprečne kvadratne napake (MSE)

$$L_{MSE} = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2. \quad (2)$$

N v enačbi 2 predstavlja *batch size*. To je število učnih primerov znotraj ene iteracije učnega algoritma, saj se izkaže, da pristop delitve podatkov proizvede boljše rezultate. Običajno je *batch size* precej manjši od števila vnosov podatkov. Cilj optimizatorja je poiskati \mathbf{W} in \vec{b} za (lokalni) minimum L . Po več iteracijah poizkušamo priti čim bližje globalnemu minimumu, pri čemer pazimo na pretreniranost modela.

Algoritem *back-propagation* je eden izmed prvih in najenostavnejših optimizatorjev. Osnova tega algoritma je izračun gradienta L na posameznem koraku. Definirajmo $\mathbf{W}_b = \mathbf{W} \oplus \vec{b} \in \mathbb{R}^{(m+1) \times n}$, kjer je \oplus operator konkatencije oz. spajanja. Glavni korak postopka povzema enačba

$$\mathbf{W}_b \leftarrow \mathbf{W}_b - \alpha \nabla_{\mathbf{W}_b} L,$$

pri čemer je α prost parameter, in sicer stopnja učenja. Gradient funkcije izgube $\nabla_{\mathbf{W}_b} L$ izračunamo s posplošitvijo verižnega pravila na matrike. Večina ostalih optimizatorjev je evolucija algoritma *back-propagation*.

6.2 Simulacije in njihova validacija

Potrebne podatke za učenje algoritmov strojnega učenja simuliramo s kombinacijo metode *Monte-Carlo* in drugih metod, ki delujejo na podlagi podatkov. Znotraj teorije SM se modelira ozadje in ne-resonančna produkcija HH . Modelira se tudi resonančna produkcija HH iz teorij onkraj Standardnega modela [6].

Simulacije so izredno zapletene, saj imajo mnogo parametrov. Pomembno je preveriti ujemanje simuliranih podatkov in podatkov iz detektorja pri nižjih energijah in drugih fizikalnih procesih ter s tem potrditi korektnost simulacije.

6.3 Validacija nevronske mreže

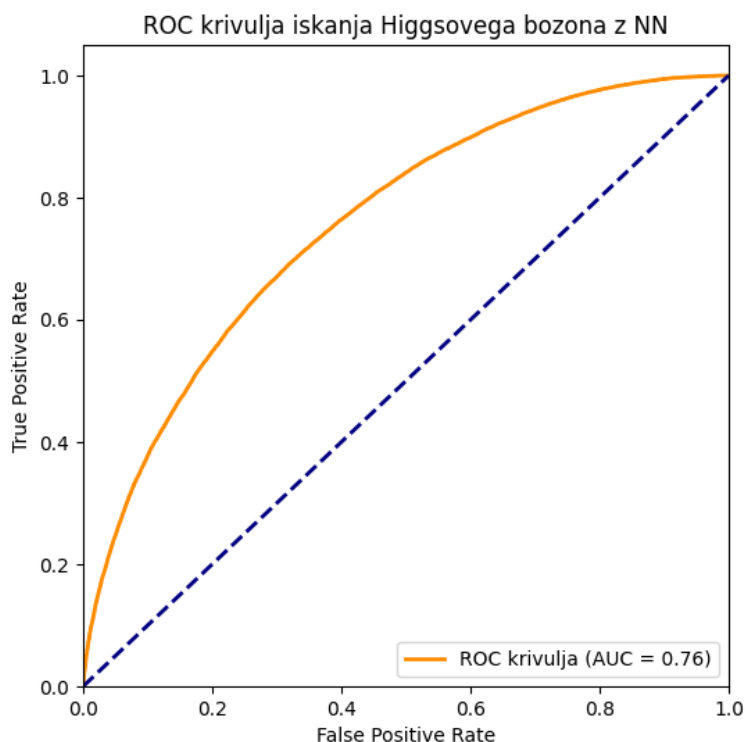
Definicija uspešnosti binarne klasifikacijske metode, torej ločevanja iskanega signala ali ozadja, je običajno prikazana s krivuljo ROC (*receiver operating characteristics*). Funkcija izgube, ki jo uporablja optimizator, se običajno ne uporablja kot metrika uspešnosti preiskovane metode. Uspešnost

modela rajši opišemo z eno samo številko. Ploščina pod krivuljo ROC (angl. *area under ROC curve*, AUC) je odlična evaluacijska metrika za primer binarnega klasifikatorja, saj predstavlja dober kompromis med deležem pravilno pozitivnih primerov TPR (angl. *true positive rate*) in deležem napačno pozitivnih primerov FPR (angl. *false positive rate*), ki sta definirana kot

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}},$$

pri čemer so pravilni pozitivni primeri TP (angl. *true positives*), napačno pozitivni primeri FP (angl. *false positives*), pravilni negativni primeri TN (angl. *true negatives*) in napačni negativni primeri FN (angl. *false negatives*).

Krivulja ROC je parametrična krivulja in prikazuje soodvisnost med TPR in FPR pri različnih pragovih. Primer take krivulje prikazuje slika 10. Prag (angl. *threshold*) je parameter krivulje. Krivuljo ROC konstruiramo tako, da pri posamezni vrednosti pragu vse napovedi večje od te vrednosti klasificiramo kot pravilne, napovedi manjše od pragu pa kot napačne. Nato narišemo točko, ki jo enolično določata TPR in FPR pri izbranem pragu. Za naključno ugibanje ciljne spremenljivke je značilna krivulja ROC identiteta. Idealna krivulja ROC je stopničasta (Heavisidova) funkcija, kateri se poizkušamo čim bolj približati.



Slika 10. Primer krivulje ROC pri klasifikaciji verjetnosti nastanka delcev iz H z NN.

Poznavanje relacij med spremenljivkami ima poleg poenostavitve problema za NN več dodatnih prednosti. Ob upoštevanju, da je poljubna fizikalna opazljivka "ortogonalna", torej ne vpliva na verjetnost za nastanek HH , si pri iskanju HH posledično lahko pomagamo z validacijo na podatkih razdeljenih glede na neodvisno spremenljivko.

Pri opisu detektorja ATLAS smo omenili osno simetričnost detektorja. Verjetnost, da so zaznani delci nastali iz HH je zaradi tega neodvisna od polarnega kota. Posledično lahko podatke razdelimo na dva ali več delov. Na enem učimo model, drugi del podatkov pa uporabimo za validacijo.

7. Opis postopka uporabe nevronske mreže pri iskanju HH preko razpada $b\bar{b}\tau^+\tau^-$

Oglejmo si primer iskanja HH preko razpadnega kanala $b\bar{b}\tau^+\tau^-$. Verjetnost za ta način razpada je 7.3 %, kar podaja tabela 1. Slaba stran nizke verjetnosti je, da imamo manj podatkov za učenje modela, vendar ima ta kanal eno izmed manj intenzivnih ozadij v primerjavi z intenziteto signala.

Enako pomembno kot izbira prostih parametrov modela je preprocesiranje podatkov pred učenjem modela. V našem primeru so podatki zbrani pri težišču energije protonov 13 TeV med letoma 2015 in 2018 [6]. LHC je bil zasnovan za trke do 14 TeV, vendar je prilagajanje superprevodnih magnetov na višjo energijo dolgotrajen proces [25] zato uporabljamo energijo centrirano pri 13 TeV.

Uporabljali bomo izključno dogodke, ki vključujejo vsaj en zaznan lepton, ostale pa izločimo. Posamezen dogodek v povprečju vsebuje 34 istočasnih neelastičnih trkov pp [6]. Po filtriranju podatkov je celotna integrirana luminoznost L_{trkov} (gostota trkov) $139.0 \pm 2.4 fb^{-1}$ [6]. To pomeni, da je pričakovana vrednost števila razpadlih parov HH v leptone za naš nabor podatkov, pri čemer predpostavimo nastanek $gg \rightarrow HH$ po teoriji SM, enaka

$$\langle N_{HH} \rangle = \sigma_{SM}(gg \rightarrow HH) \cdot \langle L_{trkov} \rangle = 4320$$

S pomočjo statističnih testov je mogoče potrditi "ortogonalnost" nekaterih spremenljivk, ki ne vplivajo na verjetnost za nastanek HH in jih izločiti ter s tem omogočiti boljšo kategorizacijo NN. Pomembna naloga podatkovnih znanstvenikov, ki poznajo delovanje metode do potankosti, je ostalo preprocesiranje podatkov. V to kategorijo spada normiranje podatkov, prilagajanje izstopajočih meritev itd.

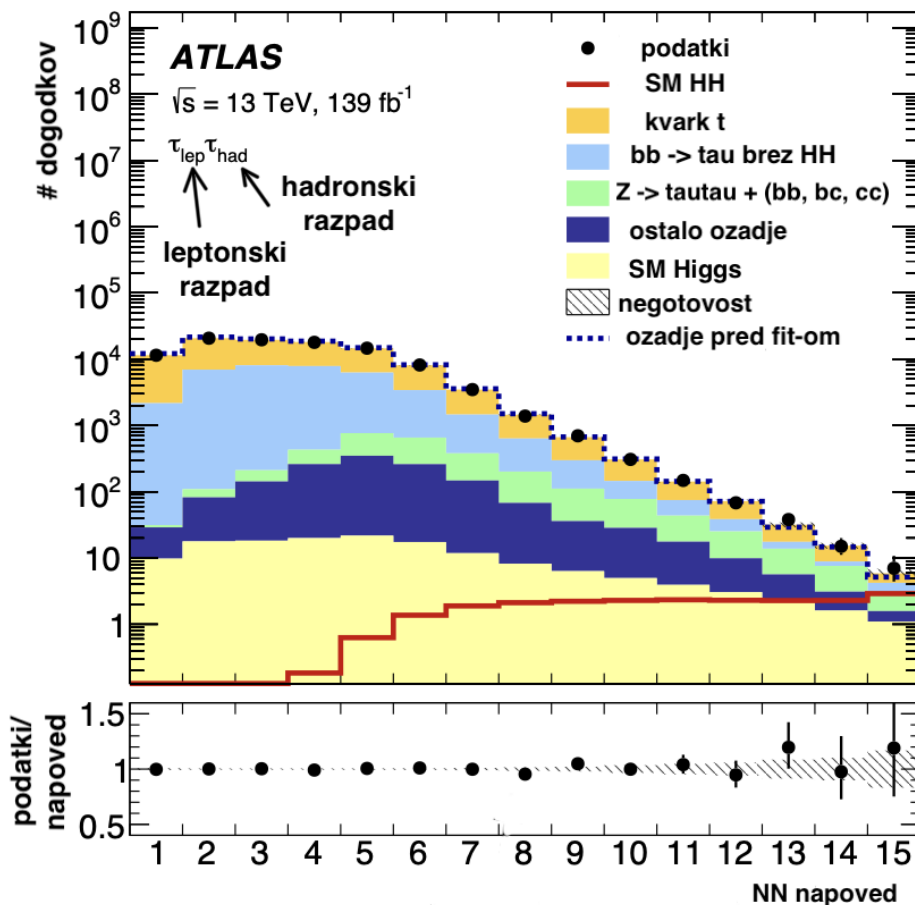
Po obdelavi podatkov jih lahko podamo NN. Podrobno poznavanje delovanja metode nam omogoča dober začetni približek za nekaj množic parametrov. Po validaciji NN izberemo tistega z najboljšo kombinacijo željenih evaluacijskih metrik (AUC, TPR, FPR, L, \dots). V našem primeru je glede na redkost nastanka HH gotovo pomembno, da imamo izredno nizek FPR. Dodatno se lahko ukvarjamo z optimizacijo parametrov za majhne izboljšave rezultatov.

Ko smo zadovoljni z rezultati modela na validacijskih podatkih, so koeficienti NN \mathbf{W}_b določeni. Primer napovedi NN, pri čemer en lepton τ razpade leptonsko, drug pa hadronske, je prikazan na sliki 11. Zaradi namena vizualizacije je širina predalov na histogramu prikazana kot enotna, čeprav imajo robovi teh predalov pri prilagajanju napovedi NN različen razmik. Dodatno je pomembno opaziti, da na histogramu ni prikazano celotno območje vrednosti napovedi NN. Pri nižjih vrednostih prevladuje ozadje, pri višjih pa signal. Število napovedi NN z visoko vrednostjo je zelo majhno, saj je signal zelo šibek. Posledično je napaka napovedi z večjimi vrednostmi znotraj posameznega predala prevelika, da bi lahko iz njih prebrali kakršnokoli uporabno informacijo. Ocenimo lahko, da so na histogramu prikazane napovedi NN do okvirne vrednosti $\frac{1}{3}$, saj je v skrajno desnem predalu histograma ozadje približno dvakrat močnejše od pričakovanega signala.

Ker se podatki iz detektorja odlično ujema s simuliranim ozadjem, NN ne pokaže dovolj izrazitega signala. To je vidno na spodnjem delu slike 11, ki prikazuje razmerje števila napovedanih dogodkov in izmerjenih dogodkov znotraj posameznega predala. Zaradi premajhnega odstopanja signala je napoved v okviru napake ozadja. Pri napovedih z višjimi vrednostmi bi zaradi pomanjkanja zadostnega števila dogodkov imeli še mnogo višjo negotovost ozadja in napovedi.

8. Zaključek

Zaenkrat znanstveniki še niso uspeli določiti samo-sklopitvene konstante λ_{HHH} , da bi lahko ta rezultat uporabili kot kriterij za izločanje teorij onkraj Standardnega modela ali potrditev vrednosti, ki jo napoveduje SM. Področje je vsekakor perspektivno, saj se nove izboljšane metode strojnega učenja hitro razvijajo. Poleg tega se načrtuje nov trkalnik FCC (Future Circular Collider) za



Slika 11. Histogram reskalirane napovedi NN. S črnimi pikami so označeni podatki iz detektorja, s črtkano črto napovedi NN in z rdečo črto simuliran signal po napovedih SM. Ostale oznake na grafu prikazujejo simulirano ozadje iz različnih virov. Spodnji del slike prikazuje razmerje med podatki iz detektorja in napovedmi NN.

pospeševanje delcev na energije okrog 100 TeV, kar je skoraj za celoten velikostni red boljše od LHC. Zagon pospeševalnika je pričakovan okrog leta 2040. Novi prispevki k znanosti bi lahko omogočili zaznavo HH in s tem naredili korak k potrditvi oziroma zavrnitvi nekaterih teorij izven Standardnega modela. Trenutna naloga fizikov je odkriti nove relacije med verjetnostjo za nastanek para HH in ostalimi fizikalnimi opazljivkami in tako optimizirati analizo zajetih podatkov. Poleg tega je potrebno preizkusiti najnovejše metode npr. sodobne grafovske nevronske mreže na podatkih, ki jih že imamo. Na ta način bomo prišli do boljšega približka za λ_{HHH} in stopili korak bližje kakšni novi teoriji ali utrdili teorijo Standardnega modela.

9. Zahvala mentorju

Zahvaljujem se mentorju prof. dr. Borutu Paulu Kerševanu, ki mi je tekom pisanja tega članka namenil veliko nasvetov in pomoči.

LITERATURA

- [1] ATLAS Collaboration, *A detailed map of higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature **607** (2022), no. 7917, 52–59, <https://www.nature.com/articles/s41586-022-04893-w>.
- [2] CERN, *The higgs boson*, 2023, <https://home.cern/science/physics/higgs-boson>.
- [3] C. L. Cheng and Atlas Team, *Combination of searches for resonant and non-resonant higgs boson pair production in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}b\bar{b}$ decay channels using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detectors*, APS April Meeting Abstracts, vol. 2022, 2022, https://indico.cern.ch/event/949705/contributions/4577077/attachments/2368750/4048151/Poster_for_LP2021_Conference_Alkaid_Cheng.pdf, pp. L10–001.

- [4] The CMS collaboration, A.M. Sirunyan, A. Tumasyan, et al., *Search for nonresonant higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ tev*, J. High Energ. Phys. **2021** (2021), 257, [https://doi.org/10.1007/JHEP03\(2021\)257](https://doi.org/10.1007/JHEP03(2021)257).
- [5] F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, *Trilinear higgs coupling determination via single-higgs differential measurements at the LHC*, The European Physical Journal C **77** (2017), 1–24, <https://doi.org/10.1140/epjc/s10052-017-5410-8>.
- [6] ATLAS Collaboration, *Search for resonant and non-resonant higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 tev pp collision data from the ATLAS detector*, arXiv preprint arXiv:2209.10910 (2022), <https://arxiv.org/abs/2209.10910>.
- [7] ———, *Search for resonant pair production of higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ tev with the ATLAS detector*, Physical Review D **105** (2022), no. 9, 092002, <https://arxiv.org/pdf/2202.07288.pdf>.
- [8] F. Giacosa, A. Okopińska, and V. Shastry, *A simple alternative to the relativistic breit-wigner distribution*, The European Physical Journal A **57** (2021), no. 12, 336, <https://doi.org/10.1140/epja/s10050-021-00641-2>.
- [9] ATLAS Collaboration, *ATLAS measures higgs boson's decay width*, 2022, <https://home.cern/news/news/physics/atlas-measures-higgs-bosons-decay-width>.
- [10] ———, *Measurement of the higgs boson mass from the $h \rightarrow \gamma\gamma$ and $h \rightarrow zz^* \rightarrow 4l$ channels with the ATLAS detector using 25 fb⁻¹ of pp collision data*, Phys. Rev. D **90** (2014), 052004, <https://atlas.cern/updates/briefing/precise-measurement-higgs-boson-mass>.
- [11] ———, *The higgs boson*, 2021, http://opendata.atlas.cern/books/current/get-started/_book/the-higgs-boson.html.
- [12] CERN, *The large hadron collider*, 2023, <https://home.cern/science/accelerators/large-hadron-collider>.
- [13] ATLAS experiment, *The ATLAS detector*, 2023, <https://atlas.cern/Discover/Detector>.
- [14] CERN, *ATLAS*, 2023, <https://www.home.cern/science/experiments/atlas>.
- [15] ———, *ATLAS images gallery*, 2023, <https://home.cern/resources/image/experiments/atlas-images-gallery>.
- [16] ATLAS experiment, *Trigger and data acquisition system*, 2023, <https://atlas.cern/Discover/Detector/Trigger-DAQ>.
- [17] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, vol. 112, Springer, 2013.
- [18] Y. Chen, E.A. Huerta, J. Duarte, et al., *A FAIR and AI-ready higgs boson decay dataset*, Sci Data **9** (2022), 31, <https://doi.org/10.1038/s41597-021-01109-0>.
- [19] A. Dertat, *Applied deep learning - part 1: Artificial neural networks*, 2017, <https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6>.
- [20] S. Sharma, S. Sharma, and A. Athaiya, *Activation functions in neural networks*, Towards Data Sci **6** (2017), no. 12, 310–316, <https://www.ijeast.com/papers/310-316>, Tesma412, IJEAST.pdf.
- [21] The ATLAS Collaboration, G. Aad, B. Abbott, et al., *The ATLAS simulation infrastructure*, Eur. Phys. J. C **70** (2010), 823–874, <https://doi.org/10.1140/epjc/s10052-010-1429-9>.
- [22] S. Lau, J. Gonzalez, and D. Nolan, *Principles and techniques of data science*, 2020, https://www.samlau.me/test-textbook/ch/15/bias_cv.html.
- [23] Y. LeCun, D. Touresky, G. Hinton, and T. Sejnowski, *A theoretical framework for back-propagation*, Proceedings of the 1988 connectionist models summer school, vol. 1, 1988, https://www.researchgate.net/profile/Yann-Lecun/publication/2360531_A-Theoretical-Framework-for-Back-Propagation/links/0deec519dfa297eac1000000/A-Theoretical-Framework-for-Back-Propagation.pdf, pp. 21–28.
- [24] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980 (2014), <https://doi.org/10.48550/arXiv.1412.6980>.
- [25] CERN, *Restarting the LHC: Why 13 tev?*, 2023, <https://home.cern/science/engineering/restarting-lhc-why-13-tev>.