

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 298

**KLASIFIKACIJA BILJNIH VRSTA IZ SLIKA BILJAKA
TEMELJENA NA DUBOKOM UČENJU**

Marija Tišljar

Zagreb, veljača 2024.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 298

**KLASIFIKACIJA BILJNIH VRSTA IZ SLIKA BILJAKA
TEMELJENA NA DUBOKOM UČENJU**

Marija Tišljar

Zagreb, veljača 2024.

DIPLOMSKI ZADATAK br. 298

Pristupnica: **Marija Tišljar (0130309559)**

Studij: Računarstvo

Profil: Računarska znanost

Mentor: izv. prof. dr. sc. Alan Jović

Zadatak: **Klasifikacija biljnih vrsta iz slika biljaka temeljena na dubokom učenju**

Opis zadatka:

Cilj ovog diplomskog rada je istraživanje i primjena metoda dubokog učenja za klasifikaciju biljnih vrsta temeljem slika biljaka. Izazovi koji se pritom javljaju uključuju vizualnu sličnost između mnogih biljnih vrsta i neuravnoteženost broja primjeraka slika biljaka između biljnih vrsta. Ovi izazovi čine klasifikaciju izazovnom čak i za stručnjake. U ovom radu potrebno je koristiti veliki skup slikovnih podataka PlantNet (<https://github.com/plantnet/PlantNet-300K>). Ovaj skup podataka uključuje 306.146 slika koje predstavljaju 1081 različitu biljnu vrstu. Kako bi se riješio problem neuravnoteženosti, potrebno je upotrijebiti različite tehnike augmentacije podataka, kao što su rotacija, izrezivanje nasumičnih dijelova slike, translacija i druge. U okviru rada, potrebno je istražiti i usporediti različite arhitekture dubokog učenja (npr. ResNet50, DenseNet16 i druge) i utvrditi koja bi arhitektura pružala najbolje mogućnosti primjene na ovaj specifičan zadatak. Nakon odabira arhitekture, potrebno je provesti optimizaciju hiperparametara modela kako bi se postigla visoka kvaliteta klasifikacije. Dobivene rezultate modela potrebno je usporediti s postojećom literaturom kako bi se ocijenila njegova uspješnost. Model je potrebno ugraditi u jednostavnu web aplikaciju koja će omogućiti funkcionalnost klasifikacije biljne vrste na temelju snimljene slike.

Rok za predaju rada: 9. veljače 2024.

Hvala mojoj obitelji što su mi sve omogućili, mom mentoru na podršci i pomoći, i svima drugima koji su mi pomogli tijekom studija.

Sadržaj

1. Uvod	3
1.1. Opis skupa podataka	4
2. Duboko učenje	7
2.1. Duboko učenje	7
2.2. Motivacija za duboko učenje	7
2.3. Veliki skupovi podataka	8
2.4. Teorem "Nema besplatnog ručka"	8
3. Neuronske mreže	10
3.1. Učenje u neuronskim mrežama	11
3.2. Vrste neuronskih mreža	13
3.2.1. Potpuno povezane unaprijedne neuronska mreže	13
3.2.2. Konvolucijske neuronske mreže	13
3.2.3. Povratne neuronske mreže	14
4. Metodologija	16
4.1. Validacija	16
4.2. Usporedba arhitektura	17
4.3. Tehnike augmentacije	17
4.3.1. Vertikalno i horizontalno zrcaljenje	18
4.3.2. Varijacija boja	20
4.3.3. Promjena skale i omjera	20
4.3.4. Slučajno brisanje	21
4.3.5. Rotacija	22
4.3.6. Translacija	23

4.4. Pristupi u rješavanju problema neuravnoteženosti klasa	23
4.5. Stopa učenja	24
5. Eksperimenti i rezultati	26
5.1. Utjecaj augmentacije na kvalitetu učenja	26
5.2. Utjecaj naduzorkovanja na kvalitetu učenja	28
5.3. Utjecaj stope učenja na kvalitetu učenja	30
5.4. Usporedba s literaturom	30
6. Razvoj web aplikacije	32
6.1. Korištene tehnologije	32
6.2. Integracija modela u aplikaciju	32
7. Zaključak	35
Literatura	36
Sažetak	39
Abstract	40

1. Uvod

Na svijetu postoji oko 400 tisuća priznatih biljnih vrsta, i svake godine se otkrivaju nove [1]. Precizna identifikacija biljnih vrsta važna je za praćenje i očuvanje bioraznolikosti, ekološka istraživanja, obrazovanje i poljoprivredu. Duboko učenje i klasifikacijski modeli predstavljaju moćan alat s ogromnim potencijalom za rješavanje ovog izazovnog zadatka.

Cilj ovog diplomskog rada je istražiti problem klasifikacije biljnih vrsta, i testiranjem metoda dubokog učenja utvrditi koja daje najbolje rezultate na skupu podataka PlantNet-300K.

Klasificiranje biljnih vrsta težak je problem iz više razloga. Jedan je taj što je broj vrsta ogroman, a morfološke karakteristike koje ih razlikuju su vrlo raznolike, ali mogu biti i vrlo suptilne.

U procesu klasificiranja biljnih vrsta javljaju se dvije vrste nesigurnosti: aleatorička (statistička) nesigurnost i epistemička nesigurnost [2]. Aleatorička nesigurnost (engl. *aleatoric uncertainty*) tj. statistička nesigurnost, proizlazi iz prirodne inherentne slučajnosti i varijabilnosti u osobinama biljaka, i zbog toga ju je teško umanjiti.

Druga vrsta nesigurnosti je epistemička nesigurnost, koja se odnosi na nedostatak znanja ili informacija. Ona proizlazi iz činjenice da klasificiramo slike biljnih vrsta, a slika biljke sadrži samo djelomične informacije o biljci. Često nije moguće odrediti biljnu vrstu samo na temelju jednog dijela biljke (cvijeta, lista, sjemena, ploda itd.), a i ne nalaze se svi dijelovi biljke na njoj u isto vrijeme. Npr. prvo će se razviti cvijet te nakon što on otpadne razvit će se plod. Moguće je da dvije biljne vrste imaju isto lišće ali različito cvijeće, te ako pokušavamo klasificirati neku od tih vrsta samo na temelju slike lišća nikad nećemo sa sigurnošću moći reći o kojoj se vrsti radi jer nam nedostaje dio informacija.

1.1. Opis skupa podataka

Skup podataka Pl@ntNet-300K sadrži 306146 slika koje obuhvaćaju 1081 različitu biljnu vrstu [3]. Primjer slika iz tog skupa prikazan je na slici 1.1. Ovaj skup je odabran zbog svoje veličine i raznolikosti, što ga čini pogodnim za testiranje raznih modela dubokog učenja. Također, karakterizira ga neuravnoteženost; neke vrste su zastupljenije od drugih. Ova osobina čini ga pogodnim za ispitivanje tehnika poboljšavanja (augmentacije) podataka, kao i za analizu njihovog utjecaja na proces učenja.



Slika 1.1. Primjer slika iz skupa podataka Pl@ntNet-300K [3]

Skup podataka Pl@ntNet-300K napravljen je iz baze podataka Pl@ntNet. Pl@ntNet je građanska znanstvena platforma koja pomoću slika koje šalju korisnici identificira i popisuje biljne vrste. Aplikacija pomoću strojnog učenja i velike baze slika omogućuje korisnicima identificiranje biljnih vrsta, a korisnici učitavanjem slika povećavaju bazu podataka i pomažu u praćenju biljnih vrsta širom svijeta. PL@ntNet je jedna od najvećih svjetskih centara za praćenje bioraznolikosti s milijunima suradnika u preko 200 zemalja [2].

Čitava baza podataka Pl@ntNet zbog svoje veličine nije praktična za široku upotrebu u zajednici strojnog učenja, stoga je iz nje izdvojen manji reprezentativni skup podataka Pl@ntNet-300K, pazeći da se u njemu sačuva inherentna nesigurnost prisutna u cijeloj bazi podataka. Skup podataka predstavljen je u radu "Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution" [3]. Skup podataka sadrži samo podskup nasumično odabranih rodova iz čitave baze. Svaka vrsta unutar odabranog podskupa rodova uključena je u skup podataka. Ovaj pristup je očuvao prirodnu složenost i vizualnu sličnost vrsta unutar istog roda, što osigurava da je u podskupu podataka očuvana izvorna nesigurnost i raznolikost pune baze.

Dvije zanimljive i izazovne značajke Pl@ntNet-300K skupa podataka su: neuravnoteže-

nost klasa i vizualna sličnost različitih vrsta biljaka (slika 1.2.) [3]. Te karakteristike su inherentne načinu na kojem su slike prikupljene kao i urođenoj raznolikosti morfologije biljaka.

Neuravnoteženost klasa karakterizirana je dugorepom distribucijom (engl. *Long-Tailed Distribution*) [3]. Dugorepe distribucije karakterizira velik broj događaja ili vrijednosti koji su relativno rijetki, ali značajni. To znači da u takvoj distribuciji postoji veliki broj rijetkih događaja ili vrijednosti, koji se značajno razlikuju od većine drugih događaja. U kontekstu skupa podataka Pl@ntNet-300K to znači da je manji broj vrlo čestih vrsta predstavljen s jako puno slika, dok je većina vrsta koje su rijetke zastupljene s mnogo manjim brojem slika po vrsti.

Dugorepa distribucija može dovesti do pristranosti modela, gdje model bolje prepoznaje vrste s mnogo primjera, dok slabo prepoznaje one s manje primjera. Rješavanje ovog problema zahtijeva posebne tehnike u učenju modela, poput augmentacije podataka.

Plant identification: a difficult problem

Exercise: link the pictures to the right plant name

Slika 1.2. Prikaz vizualne sličnosti različitih biljnih vrsta [4]

Vizualna sličnost različitih vrsta biljaka izravno doprinosi visokoj neodređenosti oznaka (engl. *High Label Ambiguity*). Visoka neodređenost oznaka odnosi se na situacije gdje je teško jednoznačno odrediti pravu oznaku (engl. *label*) za određenu sliku biljke. To može biti zbog više faktora, uključujući sličnost između različitih vrsta biljaka, kvalitete slike, ili varijabilnosti u izgledu biljke (kao što su različite faze rasta, sezonske promjene ili jednostavne prirodne varijacije unutar iste vrste).

Visoka neodređenost oznaka čini klasifikaciju biljnih vrsta posebno izazovnom, jer modeli dubokog učenja mogu imati poteškoća u učenju točnih karakteristika koje razlikuju slične vrste.

2. Duboko učenje

2.1. Duboko učenje

Duboko učenje je potpodručje strojnog učenja, koje koristi duboke umjetne neuronske mreže za učenje i interpretiranje velikih količina podataka. Neuronske mreže se nazivaju "duboke" jer imaju nekoliko (više od dva) slojeva neurona. Algoritmi dubokog učenja koriste slojeve neurona kako bi izlučili značajke više razine iz sirovih podataka. Primjerice u obradi slika, niži slojevi mogu identificirati jednostavne značajke poput rubova, dok viši slojevi mogu identificirati složenije koncepte poput lica ili znakova.

2.2. Motivacija za duboko učenje

Strojno učenje je jedno od područja umjetne inteligencije, koje se bavi razvojem algoritama koji se sami mogu poboljšavati kroz iskustvo i učenje.

U ranim fazama strojnog učenja dominantni pristup bili su plitki modeli koji su se oslanjali na ručno konstruirane značajke [5]. Značajan pomak donosi reprezentacijski pristup, gdje je cilj sirove podatke transformirati u oblik koji je pogodniji za algoritme strojnog učenja, tj. naučiti značajke i onda naučiti model. Ovaj pristup je omogućio modelima da otkriju složenije i suptilnije obrasce u podacima [5].

Duboko učenje predstavlja još jedan korak naprijed s pristupom učenja svega u isto vrijeme (engl. *end-to-end*) tj. učenje izravno od sirovih podataka do izlaznih rezultata pomoću niza nelinearnih transformacija [5]. Ovo omogućuje modelima da samostalno nauče i apstrahiraju značajke na puno dubljoj razini nego prije te je ovaj pristup otvorio mnoge nove mogućnosti u području umjetne inteligencije.

2.3. Veliki skupovi podataka

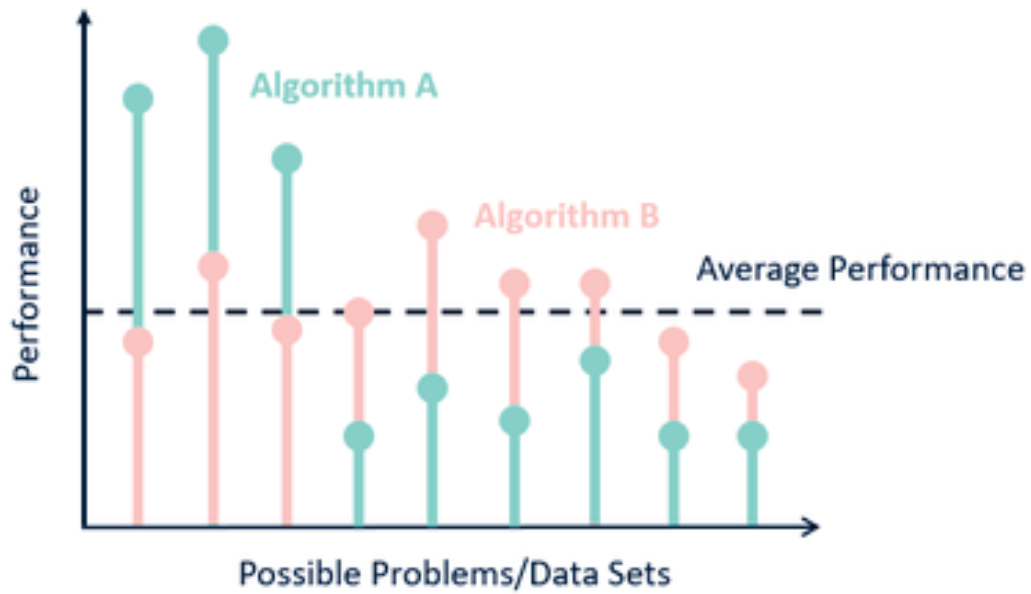
Pojava velikih skupova podataka igrala je ključnu ulogu u razvoju i popularnosti dubokog učenja. Veliki skupovi podataka su nužni za efikasno učenje dubokih modela s jako velikim brojem parametara. Količina i raznolikost podataka pomaže modelima da nauče složene obrasce i veze unutar podataka, što je ključno za dobro generaliziranje na novim, neviđenim podacima. Zbog svoje kompleksnosti, duboki modeli su skloni prenaučivosti ako su učeni na premalom skupu podataka jer mogu jednostavno zapamtiti skup za učenje umjesto da uče obrasce unutar podataka.

Uspjeh dubokih modela na velikim skupovima podataka također je populariziralo prijenosno učenje (engl. *transfer learning*), koncept gdje se model učen na jednom zadatku koristi kao početna točka za drugi zadatak. Modeli prednaučeni na velikim skupovima podataka mogu se fino podesiti na manjim skupovima podataka specifičnim za novi zadatak. Time se naučena znanja efikasno prenose na novi zadatak, što smanjuje resurse potrebne za učenje. Ovakav pristup često vodi do znatno boljih rezultata nego učenje sa slučajno inicijaliziranim težinama.

2.4. Teorem "Nema besplatnog ručka"

Teorem "Nema besplatnog ručka" (engl. *No Free Lunch Theorem*) govori da nema jednog općenitog algoritma koji je univerzalno najbolji. Teorem tvrdi da svi optimizacijski algoritmi rade jednako dobro kad se njihove performanse usrednje preko svih mogućih problema (slika 2.1.). Ne postoji jedan najbolji model za sve već model treba prilagoditi specifičnom problemu izborom najbolje arhitekture mreže, funkcije gubitka, veličine grupe za učenje ((engl. *batch*)), stope učenja i drugih hiperparametara [6].

Cilj prilagodbe modela specifičnom zadatku je postizanje bolje generalizacije na novim, neviđenim podacima. Ta performansa uvelike ovisi o pristranosti modela [5]. Ako pristranost modela odgovara podacima, model će moći naučiti pravu strukturu podataka, te će vjerojatno moći dobro generalizirati. Model može imati pristranost koja ne odgovara podacima, npr. ako model ima malu pristranost, može biti previše složen i pokušat se savršeno prilagoditi podacima za učenje (engl. *overfitting*). S druge strane, ako model ima visoku pristranost, može biti prejednostavan i neće moći naučiti složene obrasce u podacima. I jedan i drugi slučaj mogu dovesti do toga da model dobro radi na podacima



Slika 2.1. Teorem "Nema besplatnog ručka" [6]

za učenje, ali ima loše performanse na novim, neviđenim podacima.

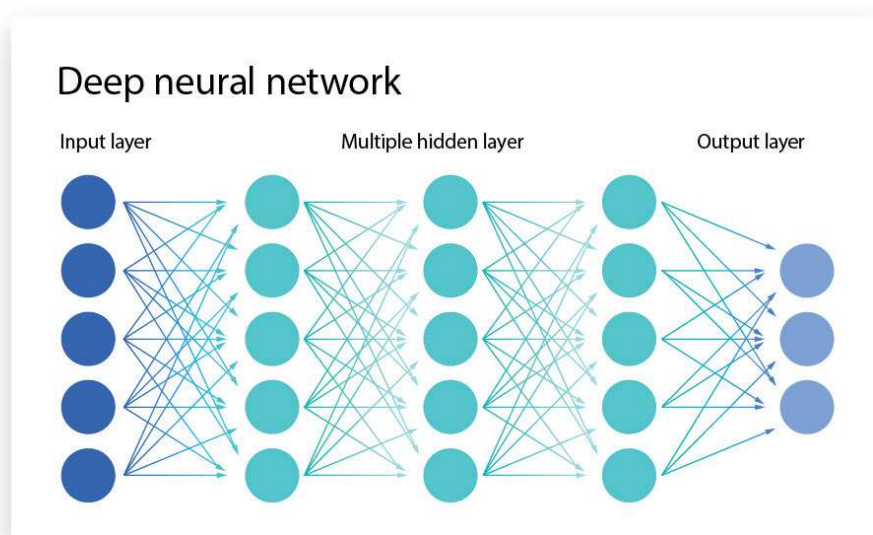
Zbog ovoga postupak rješavanja svakog pojedinačnog problema često uključuje eksperimentiranje sa različitim arhitekturama modela, tehnikama učenja i hiperparametrima.

3. Neuronske mreže

Neuronske mreže inspirirane su strukturom i funkcijom ljudskog mozga te su oblikovane kako bi replicirale njegovu sposobnost učenja i predstavljanja podataka.

Biološke neuronske mreže sastoje se od neurona koji su međusobno povezani komunikacijskim vezama, sinapsama. Svaki neuron može imati tisuće komunikacijskih veza što takvu mrežu čini nevjerovatno složenom. Također, mreže neurona imaju sposobnost mijenjati svoju strukturu i veze kroz proces učenja, skupljanja iskustva ili ozljeda. Komunikacijske veze koje se više koriste postaju jače, dok one koje se rijetko koriste slabe ili nestaju.

Neuronska mreža sastoji se od čvorova (sličnih neuronima) raspoređenih u slojeve (slika 3.1.). Ti čvorovi su međusobno povezani vezama koje imaju svoje težine. Mreža se najčešće sastoji od ulaznog sloja koji prima sirove podatke za obradu, više skrivenih slojeva koji obrađuju podatke putem niza matematičkih operacija i izlaznog sloja koji proizvodi konačan rezultat, poput klasifikacije ili regresije.

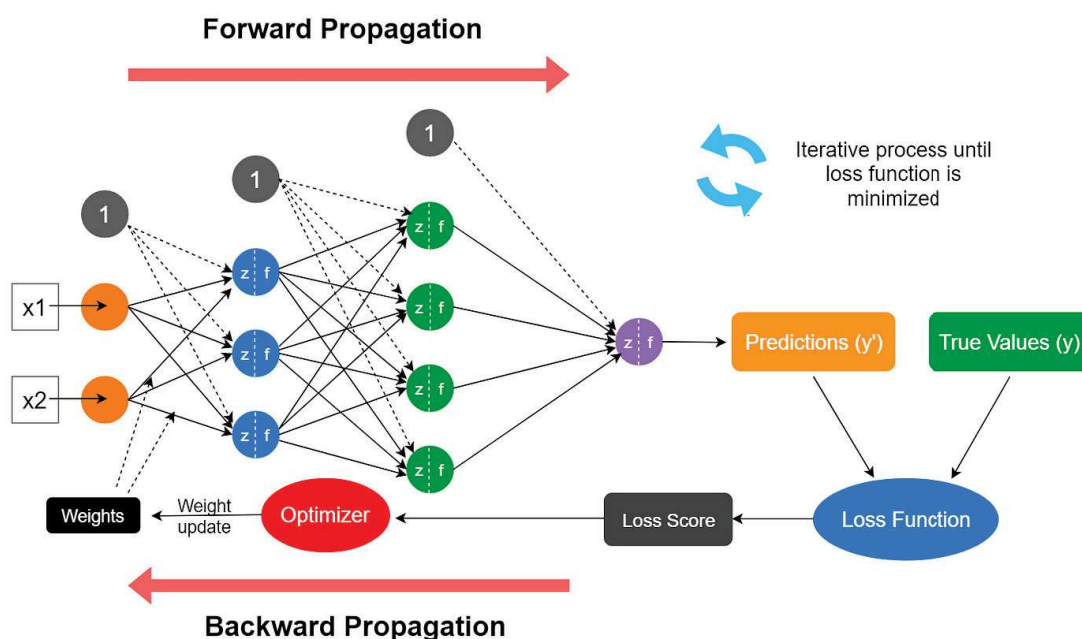


Slika 3.1. Slojevi neuronske mreže [7]

3.1. Učenje u neuronskim mrežama

Neuronske mreže uče kroz prilagođavanje svojih težina s ciljem minimiziranja razlike između dobivenih predikcija i željenih izlaza.

Prije početka učenja, inicijaliziraju se težine mreže. Te početne težine mogu biti nasumično odabrane, ili ako se koristi prijenosno učenje (engl. *transfer learning*), mogu biti preuzete iz već prednaučenog modela. Proces učenja započinje dovođenjem podataka za učenje na ulazni sloj mreže. Nakon toga podatci se propagiraju kroz skrivene slojeve mreže. Slojevi se sastoji od čvorova koji računaju izlaz koji se šalje u sljedeći sloj na temelju pojedinačnih ulaza i njihovih težina, primjenjujući na njih aktivacijsku funkciju. Aktivacijska funkcija omogućuje modelu da modelira složene, nelinearne odnose u podacima. Svaki sloj primi podatke, obradi ih, i proslijedi sljedećem sloju, sve dok podatci ne dođu do izlaznog sloja koji generira konačnu predikciju (slika 3.2.) [8].

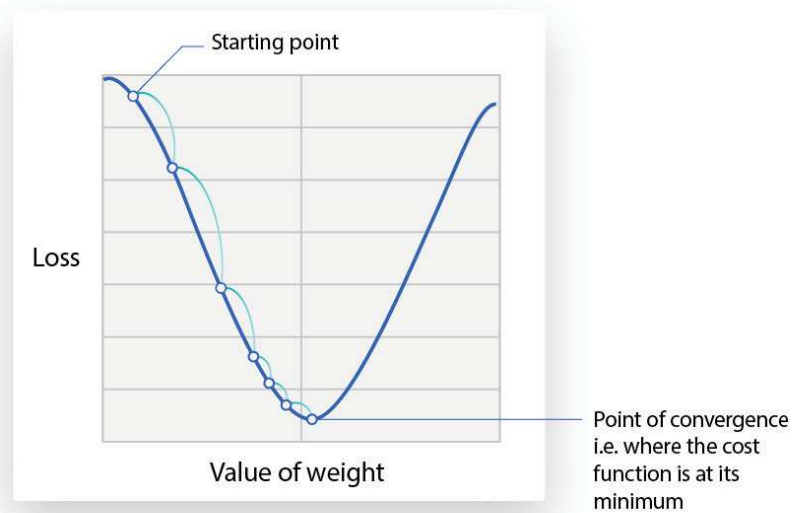


Slika 3.2. Učenje u neuronskoj mreži [9]

Uspješnost modela se procjenjuje na temelju usporedbe dobivenih predikcija i željenih izlaza. To se određuje pomoću funkcije gubitka, koja računa gubitak, odnosno razliku između dobivenih i predviđenih izlaza. Gubitak je mjera točnosti predikcija modela.

Gubitak se propagira unatrag (engl. *backpropagation*) kroz mrežu, kako bi mogli izračunati gradijent funkcije gubitka u odnosu na sve težine. Taj gradijent se koristi za ažuriranje težina pomoću algoritma gradijentnog spusta ili stohastičkog gradijentnog spusta.

Oba algoritma pomoću gradijenta računaju koji smjer pomaka težina najviše smanjuje gubitak. Razlika je u tome što gradijentni spust koristi cijeli skup podataka za ažuriranje težina u svakoj iteraciji, dok stohastički gradijentni spust ažurira težine za svaki uzorak ili za svaku mini grupu (engl. *mini-batch*), što ubrzava učenje ali ga čini potencijalno nestabilnijim.



Slika 3.3. Put prema konvergenciji [7]

Proces ažuriranja težina se iterativno ponavlja dok se ne postigne konvergencija ili se ne zadovolji neki drugi uvjet zaustavljanja, npr. jako mala promjena u gubitku između iteracija. Slika 3.3. prikazuje put kojim algoritam gradijentnog spusta putuje kroz prostor težina kako bi našao minimum funkcije gubitka. Svaki korak algoritma prilagođava težinu u smjeru koji smanjuje gubitak, sve dok ne dođe do točke konvergencije gdje funkcija gubitka postiže svoj minimum (slika 3.3.). Tijekom učenja, model će proći kroz sve podatka više puta, kako bi naučio složene obrasce unutar podataka. Jedan prolazak kroz čitav skup podataka naziva se epoha. Broj epoha ovisi o stopi učenja, arhitekturi modela i ostalim hiperparametrima. Ako se odabere prevelik broj epoha može doći do prenaučivosti modela (engl. *overfitting*), jer se model previše prilagodi podacima za učenje i izgubi moć generalizacije na novim, neviđenim podacima.

3.2. Vrste neuronskih mreža

Arhitektura neuronskih mreža može biti specifično prilagođena određenom tipu zadatka ili vrsti podataka, što znači da nema jedne najbolje arhitekture za sve zadatke. U ovom poglavlju razmotrit će se osnovne vrste neuronskih mreža i njihova prikladnost za rješavanje problema klasifikacije biljnih vrsta na temelju slika biljaka.

3.2.1. Potpuno povezane unaprijedne neuronska mreže

Potpuno povezane unaprijedne neuronske mreže (engl. *Feedforward Neural Networks (FNNs)*), poznate još kao i višeslojni perceptron, su najosnovnija vrsta neuronskih mreža. Sastoje se od minimalno tri sloja, ulaznog sloja, jednog ili više skrivenih slojeva i jednog izlaznog sloja. Svaki čvor u sloju je povezan sa svakim čvorom u susjednom sloju. Informacije kroz mrežu struje od ulaza prema izlazu i mreža ne sadrži petlje [8].

Ovakve mreže su pogodne za razne zadatke, uključujući klasifikaciju i regresiju, no nisu najbolji izbor za podatke koji uključuju prostornu ili vremensku komponentu kao što su slike ili vremenski slijedovi. Razlog je što mreža razmatra sve ulaze neovisno i nema sposobnost očuvanja vremenske ili prostorne komponente podataka.

3.2.2. Konvolucijske neuronske mreže

Konvolucijske neuronske mreže (engl. *convolutional neural network (CNN)*) su posebno osmišljene za podatke koji imaju topologiju rešetke, poput slika i vremenskih slijedova. Mreža se tipično sastoji od niza konvolucijskih slojeva i slojeva sažimanja [10].

Konvolucijski slojevi koriste operaciju konvolucije. Konvolucija je skalarni produkt jedne funkcije s obzirom na posmaknutu i reflektiranu drugu funkciju. Matematički, konvolucija se može zapisati kao:

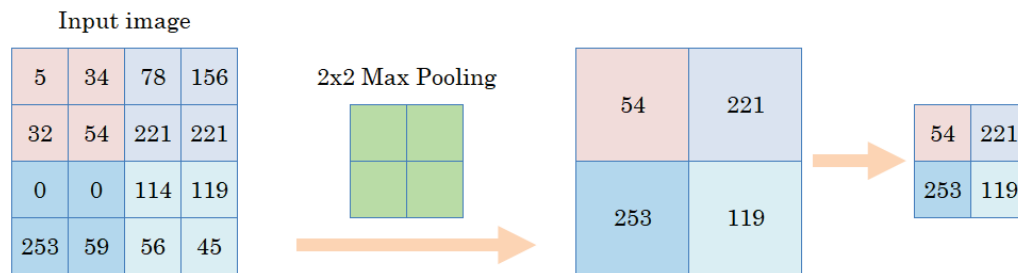
$$h(t) = (w * x)(t) = \int_{-\infty}^{\infty} w(\tau) \cdot x(t + \tau) d\tau,$$

gdje w predstavlja funkciju jezgre, a x ulazni signal [10].

Konvolucijski sloj sastoji se od skupa filtara malih prostornih dimenzija (npr. 3x3, 5x5). Filtri računaju mape značajki klizeći preko ulazne slike. Mapa značajki predstavlja rezultat konvolucije određenog filtra i ulazne slike. Elementi mape značajki predstavljaju prisutnost određene značajke (npr. ruba, teksture, oblika) u odgovarajućem dijelu slike

[11]. Iako je konvolucija slična potpuno povezanom sloju, neuroni nisu potpuno povezani s neuronima iz susjednog sloja već svaki neuron povezan samo s malim brojem lokalnih neurona. To omogućuje svakom neuronu da se specijalizira za učenje značajki iz malog prostornog područja ulaza.

Slojevi sažimanja (engl. *pooling layers*), poput sažimanja najvećom (slika 3.4.) ili srednjom vrijednošću smanjuju prostornu dimenziju ulaznih mapa značajki čime se smanjuje i računski složenost modela. Slojevi sažimanja povećavaju invarijantnost na pomak, što je korisno kada je za raspoznavanje bitnije detektirati prisutnost koncepta nego lokaciju.



Slika 3.4. Sažimanje najvećom vrijednošću [12]

Konvolucijske mreže rade vrlo dobro na problemima klasifikacije slika, zbog svoje mogućnosti detekcije i hijerarhijske reprezentacije značajki unutar slika. Raniji slojevi detektiraju jednostavnije značajke poput rubova, boja i tekstura, što omogućuje kasnijim slojevima kombiniranje ovih značajki u složenije strukture [10].

3.2.3. Povratne neuronske mreže

Povratne neuronske mreže (engl. *recurrent neural network, RNN*) posebno su dizajnirane za obradu sekvencijskih podataka, kao što su vremenski nizovi. Tok podataka u povratnoj mreži omogućuje mreži pamćenje informacija iz prethodnog koraka. To je posebno korisno za obrađivanje podataka u kojima je bitan redoslijed.

U svakom koraku čvor prima ulaz iz sekvence (npr. riječ u rečenici) i skriveno stanje prethodnog sloja, zatim kombinirajući to dvoje računa izlaz. Nakon toga ažurira se skriveno stanje te ono postaje ulaz za sljedeći korak [13].

Jednadžbe koje definiraju povratnu neuronsku mrežu su:

1. Ažuriranje skrivenog stanja:

$$h^{(t)} = \tanh(W_{hh}h^{(t-1)} + W_{xh}x^{(t)} + b_h), \quad (3.1)$$

2. Projekcija u izlazni sloj:

$$o^{(t)} = W_{hy}h^{(t)} + b_o. \quad (3.2)$$

Skriveno stanje h omogućuje mreži da kroz pamćenje informacija iz prethodnih koraka razumije kontekst i ovisnosti unutar sekvence. Na kraju se dodaje linearan izlazni sloj koji obavlja projekciju iz skrivenog u izlazni prostor [13].

Iako su povratne neuronske mreže moćan alat za učenje sekvencijskih podataka, imaju problem s učenjem dugoročnih ovisnosti unutar podataka zbog problema eksplodirajućeg ili nestajućeg gradijenta. Do problema eksplodirajućeg gradijenta dolazi kada vrijednosti gradijenta postanu nekontrolirano velike, što dovodi do nestabilnosti u učenju zbog ažuriranja težina u prevelikim koracima. Do problema nestajućeg gradijenta dolazi kada vrijednosti gradijenta postaju sve manje i manje prilikom propagiranja kroz mrežu zbog čega se težine ažuriraju vrlo malo i dolazi do jako sporog učenja ili čak potpunog prestajanja učenja.

4. Metodologija

4.1. Validacija

Model na izlazu daje vektor vjerojatnosti. Svaki element tog vektora predstavlja vjerojatnost da ulazna slika pripada određenoj klasi. Klasa s najvećom vjerojatnošću obično se smatra predikcijom modela.

Neka model klasifikacije razlikuje N klasa. Vektor vjerojatnosti \mathbf{p} , gdje svaki element p_i predstavlja vjerojatnost da ulazna slika pripada određenoj klasi i tada se zapisuje kao:

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix},$$

gdje je:

$$\sum_{i=1}^N p_i = 1.$$

Metrika top-1 točnost (engl. *top-1 accuracy*) govori u koliko je posto slučajeva klasa za koju model daje najveću vjerojatnost točna. Ovo je vrlo stroga metrika i koristi se kada je bitno da je model vrlo precizan.

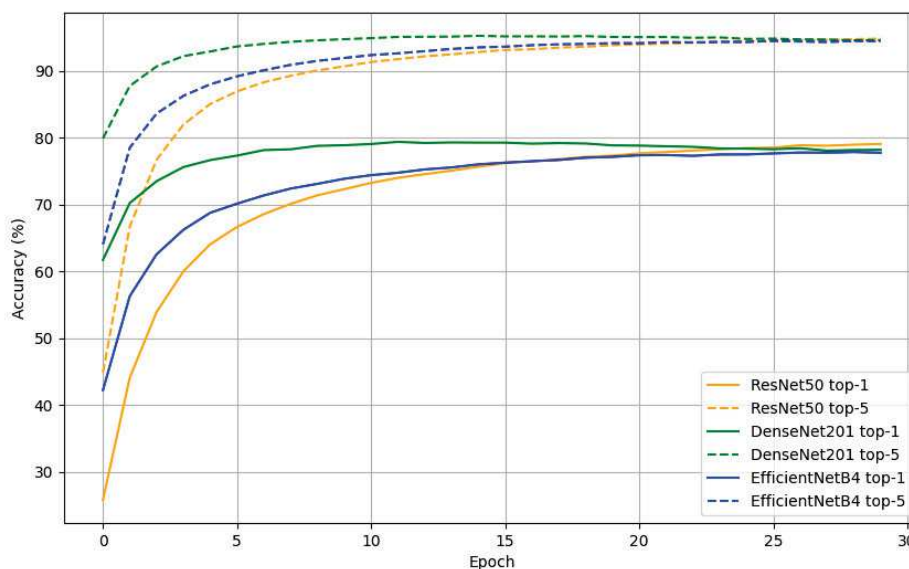
Top-5 točnost (engl. *top-5 accuracy*) govori u koliko je posto slučajeva 5 predikcija s najvećom vjerojatnosti sadrži točnu klasu. Ova metrika može biti korisna u situacijama gdje su razlike između klasa suptilne ili kada je više klasa podjednako relevantno za određeni ulaz.

4.2. Usporedba arhitektura

Za ovaj rad odabrane su za usporedbu tri arhitekture konvolucijskih mreža: EfficientNetB4 [14], ResNet50 [15] i DenseNet201 [16].

U provedenim testiranjima, ResNet50 je postigao bolje performanse u odnosu na EfficientNetB4. Također, u usporedbi s DenseNet201, ResNet50 je pokazao slične rezultate, ali s prednošću kraćeg vremena potrebnog za učenje, slika 4.1.

Modeli su učeni tijekom 30 epoha s veličinom grupe 96 i stopom učenja 0,00001. Sva tri modela su inicijalizirana s prednaučenim težinama s ImageNeta. EfficientNetB4 je ostvario top-1 točnost od 77,83 %, DenseNet201 79,36 % a ResNet50 79,1 %. Zbog kraćeg vremena učenja ResNet50 i slične točnosti top-1 kod ResNet50 i DenseNet201, daljnji eksperimenti provodit će se samo s arhitekturom ResNet50.



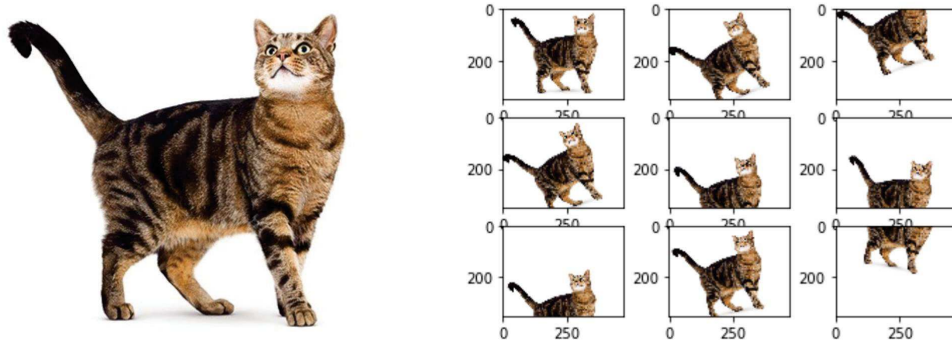
Slika 4.1. Usporedba modela

4.3. Tehnike augmentacije

Augmentacija podataka je tehnika povećavanja skupa podataka za učenje bez potrebe za prikupljanjem novih podataka. Skup podataka se povećava stvaranjem varijacija iz postojećih podataka primjenom različitih transformacija kao npr. rotacije, translacije, skaliranja, dodavanje šuma, mijenjanje boje i drugih [17]. Tim postupcima od jedne

slike možemo dobiti veliki broj slika (vidi sliku 4.2.).

Primjena augmentacije podataka može poboljšati generalizacijsku moć modela jer povećavanje skupa podataka stvaranjem varijacija uvodi šum u podatke, što pomaže modelu postati otporniji na različite uvjete i bolje generalizirati na nevidenim primjerima. Primjena različitih transformacija pomaže modelu i postati invarijantan na te transformacije.



Slika 4.2. Kreiranje velikog broja slika iz jedne slike [18]

Augmentacija podataka posebno je važna kada postoji neuravnotežen skup podataka, gdje model daje bolje predikcije za klase s puno primjeraka jer je njih uspio naučiti, dok klase s manje podataka ostaju podnaučene. Ipak, potrebno je paziti da ne dođe do prenaučavanja manje čestih klasa. To se može dogoditi ako prilikom povećanja skupa podataka ne uvedemo dovoljno raznolikosti, pa model nauči prepoznavati proširene podatke umjesto generalizirajućih značajki.

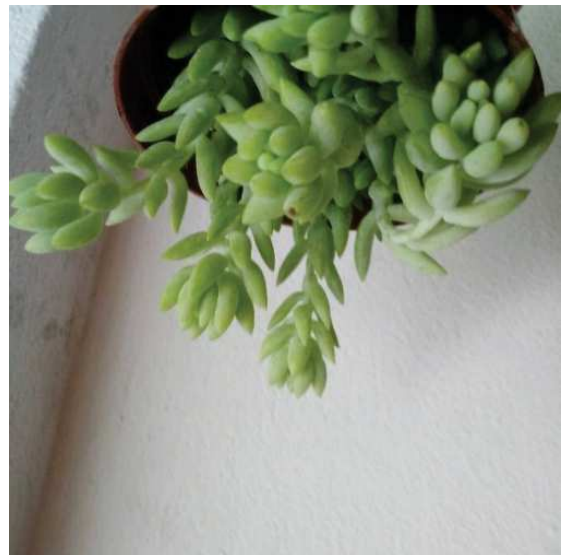
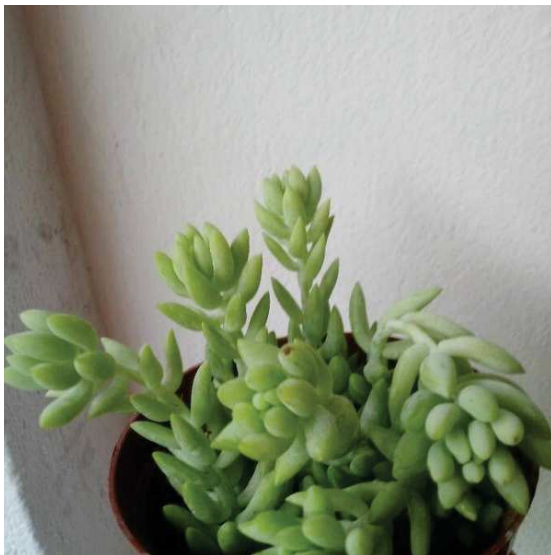
4.3.1. Vertikalno i horizontalno zrcaljenje

Horizontalno zrcaljenje slike (engl. *Horizontal Flip (HFlip)*) je zrcaljenje slike oko njene vertikalne osi, što rezultira time da lijevo postane desno i obrnuto, slika 4.3. Ova transformacija se koristi kada su objekti na slikama invarijantni na promjenu orijentacije duž vertikalne osi (npr. ljudska lica).



Slika 4.3. Primjer primjene horizontalnog zrcaljenja slike

Vertikalno zrcaljenje slike (engl. *Vertical Flip (VFlip)*) je zrcaljenje slike oko njene horizontalne osi, slika 4.3. To znači da vrh slike postane dno i obrnuto. Transformacija je korisna u slučajevima kada su objekti invarijantni na promjenu orijentacije duž horizontalne osi (npr. životinje, biljke, automobili).



Slika 4.4. Primjer primjene vertikalnog zrcaljenja slike

Treba biti oprezan prilikom primjene ovih transformacija jer one mogu dovesti do promjene konteksta ili klase slike. Npr. ako horizontalno zrcalimo sliku sata koji pokazuje 3:00 dobit ćemo sat koji pokazuje 9:00, što mijenja značenje slike.

U kontekstu odabranog skupa podataka *Pl@ntNet-300k*, koji je namijenjen za klasifikaciju biljaka, obje transformacije su prigodne i korisne jer biljke često imaju strukture koje

su invarijantne na promjenu orijentacije (npr. oblik lista ili cvijeta).

4.3.2. Varijacija boja

Transformacija nasumične varijacije boja (engl. *color jitter*) uključuje nasumične promjene u boji, svjetlini, zasićenosti, kontrastu ili nijansi slike, slika 4.5. Ova tehnika može biti korisna u kontekstu klasifikacije slika biljaka, jer uči modelu da bude otporniji na promjene osvjetljenja i boja, koje su česte u prirodnim scenarijima. Npr. biljke slikane u drugo doba dana ili u različitim vremenskim uvjetima.



Slika 4.5. Primjer primjene varijacije boja nad slikom

4.3.3. Promjena skale i omjera

Promjena skale (engl. *scale*) transformira ulaznu sliku smanjivanjem ili povećavanjem, slika 4.6. Promjena omjera mijenja širinu ili visinu slike, slika 4.7. U kontekstu skupa podataka, ove tehnike mogu pomoći modelu da nauči prepoznavati biljke slikane iz različitih perspektiva i udaljenosti. Naprimjer, biljka slikana iz daleka je vrlo malena dok ista biljka slikana iz blizine zauzima cijelu sliku.



Slika 4.6. Primjer povećanja slike



Slika 4.7. Primjer promjene omjera širine i visine slike

4.3.4. Slučajno brisanje

Slučajno brisanje (engl. *random erase*) je transformacija koja zacrni ili popuni nekom odabranom bojom slučajno odabrani pravokutnik i time ga efektivno izbriše iz slike, slika 4.8. Ova transformacija čini model robustnijim jer ga uči da bude otporniji na odsutnost određenih dijelova slike. U kontekstu raspoznavanja biljaka, ova transformacija pomaže modelu da prepozna biljku i ako su neke njezine strukture odsutne, npr. prisutan je samo list ali ne i cvijet. Prilikom primjene ove transformacije treba pripaziti da se ne koristimo prevelike pravokutnike ili prečesto brisanje kako ne bi uk-

lonili previše informacija i time smanjili mogućnost modela da nauči određene značajke.



Slika 4.8. Primjer primjene slučajnog brisanja dijelova slike

4.3.5. Rotacija

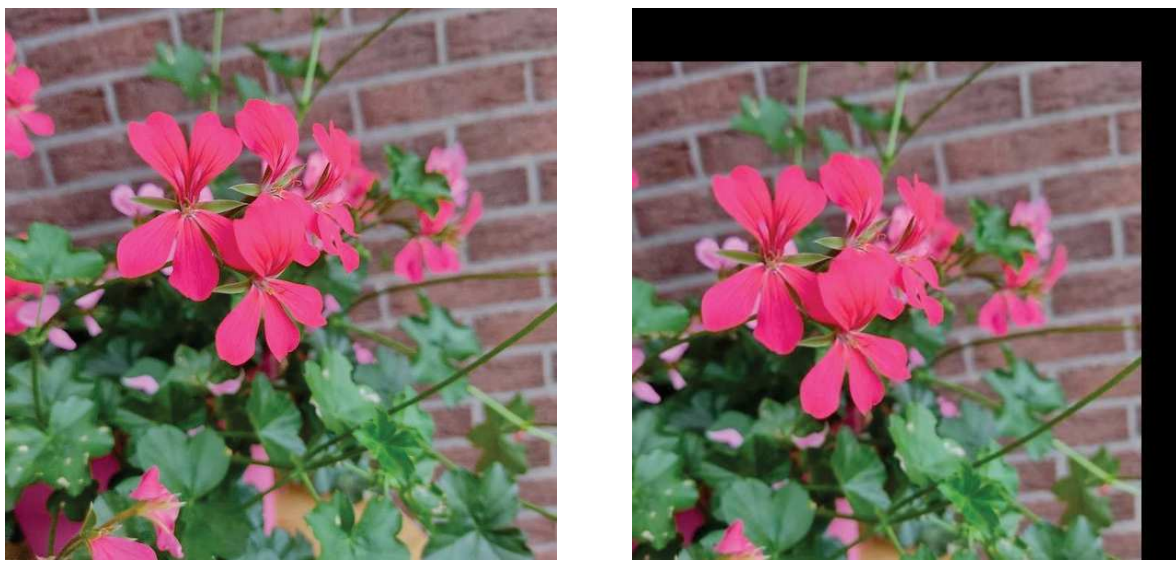
Rotacija (engl. *rotation*) transformira ulaznu sliku okrećući je pod određenim kutom, slika 4.9. Ova transformacija pomaže modelu da prepozna objekte bez obzira na njihovu orijentaciju. U kontekstu skupa podataka *Pl@ntNet-300k* to je korisno, jer biljke mogu biti slikane iz različitih kutova.



Slika 4.9. Primjer primjene rotacije na sliku

4.3.6. Translacija

Translacija (engl. *translation*) pomiče cijelu sliku u horizontalnom ili vertikalnom smjeru, slika 4.10. Ovo može pomoći modelu da nauči prepoznavati objekte i kada oni nisu centrirani. Biljka se može pojaviti u različitim dijelovima slike i ova tehnika nam pomaže simulirati različite uvjete slikanja.



Slika 4.10. Primjer primjene translacije na sliku

4.4. pristupi u rješavanju problema neuravnoteženosti klasa

U slučaju velike neuravnoteženosti klasa model može postati pristran prema klasama koje sadrže puno primjeraka i davati loše rezultate na drugim klasama. Postoji mnoštvo tehnika koje mogu pomoći u ublažavanju tog problema.

Podešavanja težina klasa (engl. *class weights*) je tehnika u kojoj klasama koje imaju manje primjeraka dodaju se veće težine u funkciji gubitka kako bi model postao osjetljiviji prema njima i smanjio svoju pristranost prema većinskim klasama. Podešavanje težina klasa omogućava modificiranje funkcije gubitka tako da greške na primjercima iz manjinskih klasa pridonose ukupnom gubitku s većom težinom. Težine koje se dodjeljuju klasama se mogu odrediti kao, naprimjer, obrnuti omjer učestalosti klasa, čime bi manje zastupljene klase dobile veće težine.

Ponovno uzorkovanje podataka uključuje metode naduzorkovanja ili poduzorkovanja

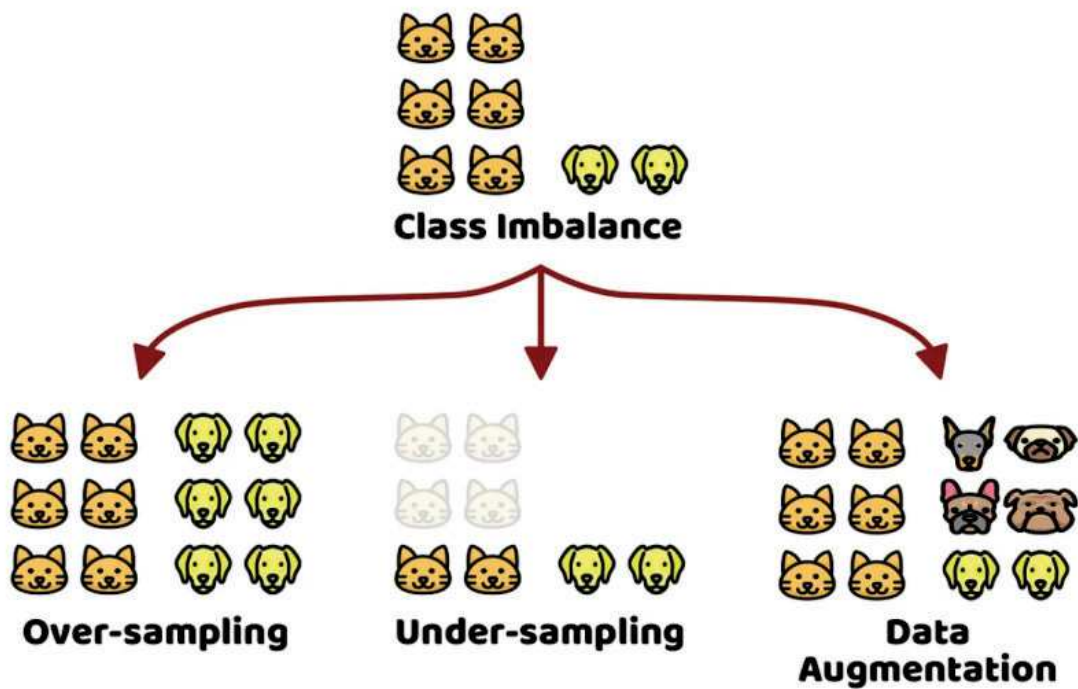
kako bi se postigla veća ravnoteža klasa, slika 4.11. Naduzorkovanje (engl. *oversampling*) je metoda koja povećaje broj primjeraka manjinskih klasa. To se može postići npr. dupliciranjem postojećih primjeraka, što može pomoći, ali ne dodaje nove informacije modelu. Drugi način na koji se to može postići je sintetičkim generiranjem novih primjeraka iz postojećih primjeraka (npr. pomoću metode SMOTE (engl. *Synthetic Minority Oversampling Technique*) [19]).

Poduzorkovanje (engl. *undersampling*) je tehnika uklanjanja primjeraka iz većinskih klasa kako bi se smanjila njihova zastupljenost i povećala ravnoteža klasa.

U ovom radu u eksperimentima je korištena tehnika naduzorkovanja i više različitih tehnika augmentacije.

4.5. Stopa učenja

Stopa učenja je hiperparametar koji određuje veličinu koraka u kojem se ažuriraju težine modela tijekom učenja kako bi se minimizirao gubitak [20]. Odabir odgovarajuće stope učenja je bitan, jer premala stopa učenja može dovesti do toga da model presporo konvergira prema optimalnom rješenju, a prevelika stopa učenja može dovesti do toga da model radi prevelike korake koji mogu uzrokovati divergenciju.



Slika 4.11. Slika prikazuje tri metode rješavanja problema neravnoteže klasa u skupu podataka: povećavanjem broja primjera manje zastupljenih klasa (engl. *over-sampling*), smanjivanjem broja primjera prekomjerno zastupljene klase (engl. *under-sampling*) i stvaranjem novih, raznovrsnih primjera (engl. *data augmentation*) [21]

5. Eksperimenti i rezultati

Nakon inicijalnog odabira arhitekture ResNet50 provedeni su eksperimenti kako bi se utvrdilo kakav utjecaj na performanse ovog klasifikacijskog modela imaju različite metode augmentacije podataka, naduzorkovanje, stopa učenja, broj epoha i ostali hiperparametri.

5.1. Utjecaj augmentacije na kvalitetu učenja

U svim eksperimentima vezanim za augmentaciju podatka održavani su fiksni parametri kako bi se osigurala dosljednost rezultata. Ti fiksni parametri uključuju korištenje optimizatora Adam, konstantnu veličinu grupe 96, upotrebu prethodno naučenih težina s ImageNeta, fiksnu stopu učenja postavljenu na $1,000e-05$ i 30 epoha učenja. Odabrana je malo niža stopa učenja kako bi se na početku učenja izbjegla prenapla promjena prednaučenih težina, što bi moglo poremetiti proces prilagodbe modela na novi skup podataka. Za metriku procjene uspješnosti koristi se top-1 točnost i srednji Jaccardov indeks za usporedbu utjecaja naduzorkovanja.

Model s tim hiperparametrima i bez dodatnih augmentacija podataka ostvario je top-1 točnost od 76,29 %.

U tablici 5.1. prikazane su performanse modela s različitim vrstama augmentacija, omogućujući brzi pregled i bolje razumijevanje specifičnog doprinosa svake tehnike. U prvom eksperimentu ispitivao se utjecaj vertikalnog i horizontalnog zrcaljenja. Vjerojatnost da se dogodi transformacija je postavljena na 50 %. Takav model je dostigao točnost od 77,19 %.

U drugom eksperimentu razmatrao se utjecaj primjene slučajnog skaliranja veličine i omjera slika. Definiran je raspon unutar kojeg je veličina svake slike slučajno skalirana, između 8 % (0.08) i 100 % (1.0) originalne veličine. Definiran je i raspon unutar kojeg se omjer stranica slike slučajno prilagođava na [0.75, 1.33], što znači da se omjer stranica

svake slike može se slučajno prilagoditi tako da bude između 75 % i 133 % izvornog omjera stranica. Cilj je bio poboljšati sposobnost modela da prepozna biljke i strukture različitih veličina. S primjenom ove augmentacije model je dosegao točnost od 77,49 %.

U trećem eksperimentu ispitivao se utjecaj primjene nasumične varijacije boja (engl. *color jitter*) s vjerojatnošću od 40 %, uključujući izmjene u svjetlini, kontrastu i zasićenju boja. Primjena ove tehnike za cilj imala povećati robusnost modela simulirajući različite prirodne varijacije u svjetlosti na slikama biljaka. Model naučen na tako augmentiranim podacima je dosegao točnost 77,29 %.

U četvrtom eksperimentu ispitivao se utjecaj slučajnog brisanja (engl. *random erase*) dijelova slike na kvalitetu učenja modela. Transformacija je uključena s vjerojatnošću od 30 %. Slučajno odabrani pravokutnik ispunjavao se nasumičnom bojom, pri čemu svaki piksel može biti različite boje. S primjenom ove augmentacije model je dosegao točnost od 77,59 %.

U petom eksperimentu ispitivao se utjecaj rotacije slike. Rotacija je primijenjena s vjerojatnošću od 50 %. Kao kut rotacije mogao je biti odabran kut između 0 i 360 stupnjeva. Cilj ovog eksperimenta bio je povećati sposobnost modela da prepozna objekte bez obzira na njihovu orijentaciju u prostoru. Model naučen s ovom vrstom augmentacije postigao je točnost od 77,21 %.

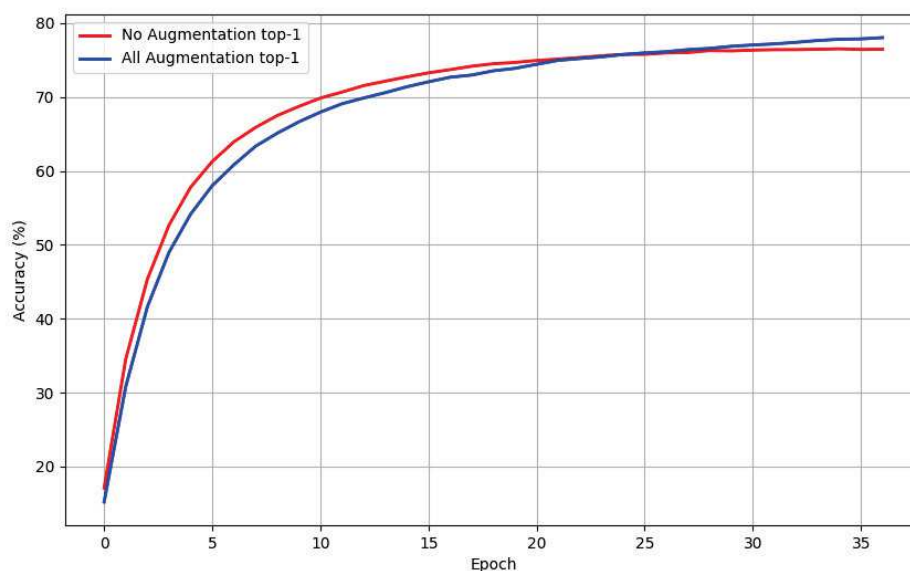
Šesti eksperiment se fokusirao na ispitivanje utjecaja translacije na učenje modela. Cilj je bio naučiti model da dobro prepozna strukture i biljke čak i kada nisu centrirani u kadru, simulirajući situacije gdje su objekti djelomično izvan kadra. S primjenom ove augmentacije, model je postigao točnost od 77,25 %.

Model s uključenim svim dosadašnjim augmentacija postigao je točnost od 76,88 %.

U tablici 5.1. može se vidjeti da svaka opisana augmentacija podataka doprinosi robusnosti modela. Na grafu 5.1. može se vidjeti da model koji uči na augmentiranim podacima u početku uči sporije od modela koji uči na neaugmentiranim podacima, zato što augmentacija efektivno povećava skup podataka. Međutim, nakon dovoljnog broja epoha, model naučen na augmentiranim podacima postiže bolju top-1 točnost. Time se može zaključiti da augmentacija skupa podataka doprinosi robusnosti i točnosti modela.

Augmentacija	Top-1 točnost
Vertikalno i horizontalno zrcaljenje	77,19 %
Varijacija boja	77,29 %
Promjena skale i omjera	77,49 %
Slučajno brisanje	77,59 %
Rotacija	77,21 %
Translacija	77,25 %
Sve augmentacije	76,88 %
Bez augmentacija	76,29 %

Tablica 5.1. Rezultati modela sa i bez primjene augmentacija



Slika 5.1. Rezultati modela sa i bez primjene augmentacija

5.2. Utjecaj naduzorkovanja na kvalitetu učenja

Naduzorkovanje je korišteno kako bi se povećanjem broja primjeraka manjinskih klasa smanjila neuravnoteženost klasa. U ovom eksperimentu korišteni su parametri koji uključuju optimizator Adam, konstantnu veličinu mini grupe 96, 60 epoha učenja, upotrebu prethodno naučenih težina s ImageNeta i stopu učenja koja se stupnjevito smanjuje

od $1,000e-05$ uz faktor smanjenja 0.3.

U eksperimentu je cilj bio usporediti rezultate modela naučenog uz korištenje svih gore opisanih augmentacija i modela koji uz te augmentacije koristi i naduzorkovanje. Pokazalo se da model s augmentacijama ostvaruje top-1 točnost od 81,07 % a model sa naduzorkovanjem dosiže samo 73,73 %. S druge strane metrika top-1 točnost ne uzima u obzir neuravnoteženost klasa. Kako bi se točnije odredilo kvalitetu modela s obzirom na sve klase promatrat će se metrika srednjeg Jaccardovog indeksa.

Jaccardov indeks je metrika koja se koristi za računanje sličnosti između skupova. Vrijednost može biti od 0 do 1 [22]. Jaccardov indeks se za pojedinu klasu računa pomoću formule:

$$J_i = \frac{|Y_i \cap P_i|}{|Y_i \cup P_i|}, \quad (5.1)$$

gdje je J_i Jaccardov indeks za klasu i , Y_i skup stvarnih oznaka klase i , a P_i skup dobivenih oznaka klase i . Prosječni Jaccardov indeks preko svih klasa računa se pomoću formule:

$$J_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N J_i. \quad (5.2)$$

Za razliku od top-1 točnosti koja može biti visoka u neuravnoteženim skupovima podataka ako model jednostavno uvijek predviđa najčešću klasu, Jaccardov indeks daje bolju mjeru performansi modela u prepoznavanju manje zastupljenih klasa.

Prosječni Jaccardov indeks osigurava da performansa modela u svakoj klasi doprinosi ukupnoj ocjeni performansi.

Za model s augmentacijom, prosječni Jaccardov indeks bio je 0.322, dok je za model s naduzorkovanjem i augmentacijama, prosječni Jaccardov indeks bio 0.335, ukazujući na poboljšanje u klasifikaciji manje zastupljenih klasa kada se koristi naduzorkovanje. Iako je top-1 točnost lošija kada se promatra cijeli test skup podataka koji je također nebalansiran, prema strožoj metrici - Jaccardovom indeksu, možemo zaključiti da se manje zastupljene klase bolje klasificiraju te da naduzorkovanje pomaže u rješavanju problema nebalansiranosti skupa podataka.

5.3. Utjecaj stope učenja na kvalitetu učenja

Eksperimenti su provedeni uz fiksne parametre: optimizator Adam, veličina mini grupe 96, upotrebu prethodno naučenih težina s ImageNeta, broj epocha 30.

Prvi eksperiment je bilo korištenje kosinusne stope učenja (engl. *cosine learning rate*). Kosinusna stopa učenja mijenja se u svakoj epohi slijedeći oblik kosinusa. Početna stopa učenja bila je 0,0001 a minimalna stopa 0,000005. Model naučen uz ovakvu stopu učenja postigao je top-1 točnost 77,34 %.

Drugi eksperiment je korištenje stope učenja koja se stupnjevito smanjuje (engl. *step decay*). Početna stopa učenja je bila 0.0001, faktor smanjenja 0.3, intervali smanjenja 10,15 i 20. Model je postigao točnost od 77,48 %.

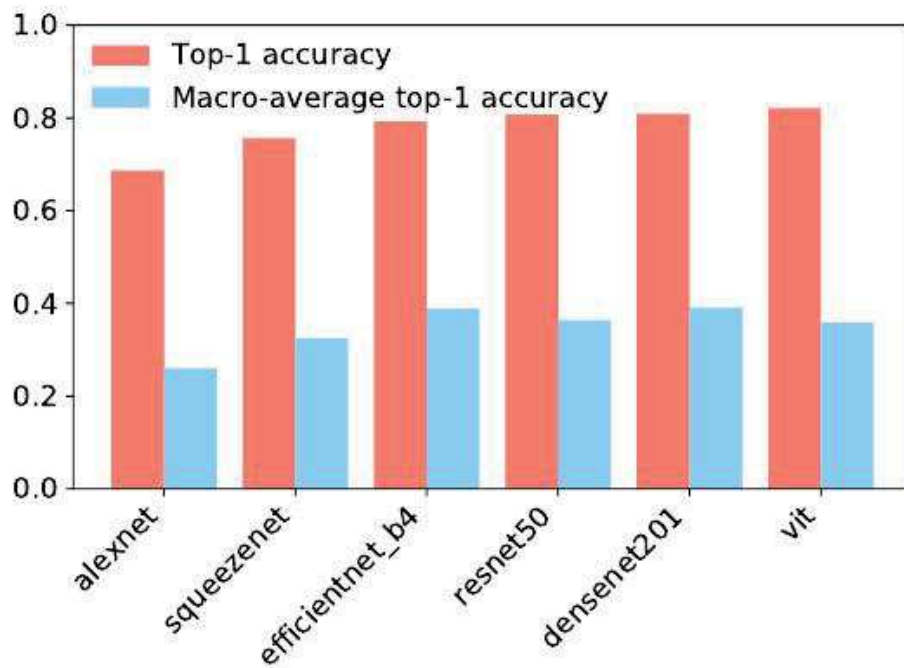
Iako su oba pristupa pokazala slične performanse, za daljnje eksperimente je korištena stopa učenja koja se stupnjevito smanjuje.

5.4. Usporedba s literaturom

Model koji je postigao najbolju top-1 točnost u ovom diplomskom radu bio je ResNet50 učen uz primjenu svih u radu opisanih tehnika augmentacije. On je ostvario točnost od 81,071 %.

Uspoređujući ovaj rezultat s podacima iz literature, s istraživanjem predstavljenom u radu pod nazivom "Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution" [3], možemo vidjeti da je to dobra performansa. Naime, u navedenom radu, model baziran na ResNet arhitekturi ostvario je top-1 točnost koja premašuje 80 % (slika 5.2.).

Visoka točnost modela ResNet50 u ovom diplomskom radu ukazuje na efikasnost primijenjenih tehnika augmentacije slika.



Slika 5.2. Rezultat metrike top-1-accuracy iz rada: Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution [3]

6. Razvoj web aplikacije

Za potrebe ovog diplomskog rada razvijena je jednostavna web aplikacija koja sadrži sučelje gdje se može učitati slika biljke, slika 6.1. Kao odgovor, aplikacija pruža popis pet vrsta biljaka za koje model s najvećom vjerojatnošću smatra da odgovaraju učitanj slici, slika 6.2.

6.1. Korištene tehnologije

Aplikacija je razvijena koristeći Django [23]. Django je web framework koji pruža arhitekturu za brz razvoj sigurnih i održivih web aplikacija. Za frontend je korišten Vue.js [24], JavaScript framework koji omogućava izgradnju interaktivnih korisničkih sučelja. Ove tehnologije odabrane su zbog njihove jednostavnosti i lakoće korištenja.

Jaccardov indeks izračunat je koristeći Pythonovu PyCM biblioteku [25] koja se koristi za analizu performansi klasifikacijskih modela.

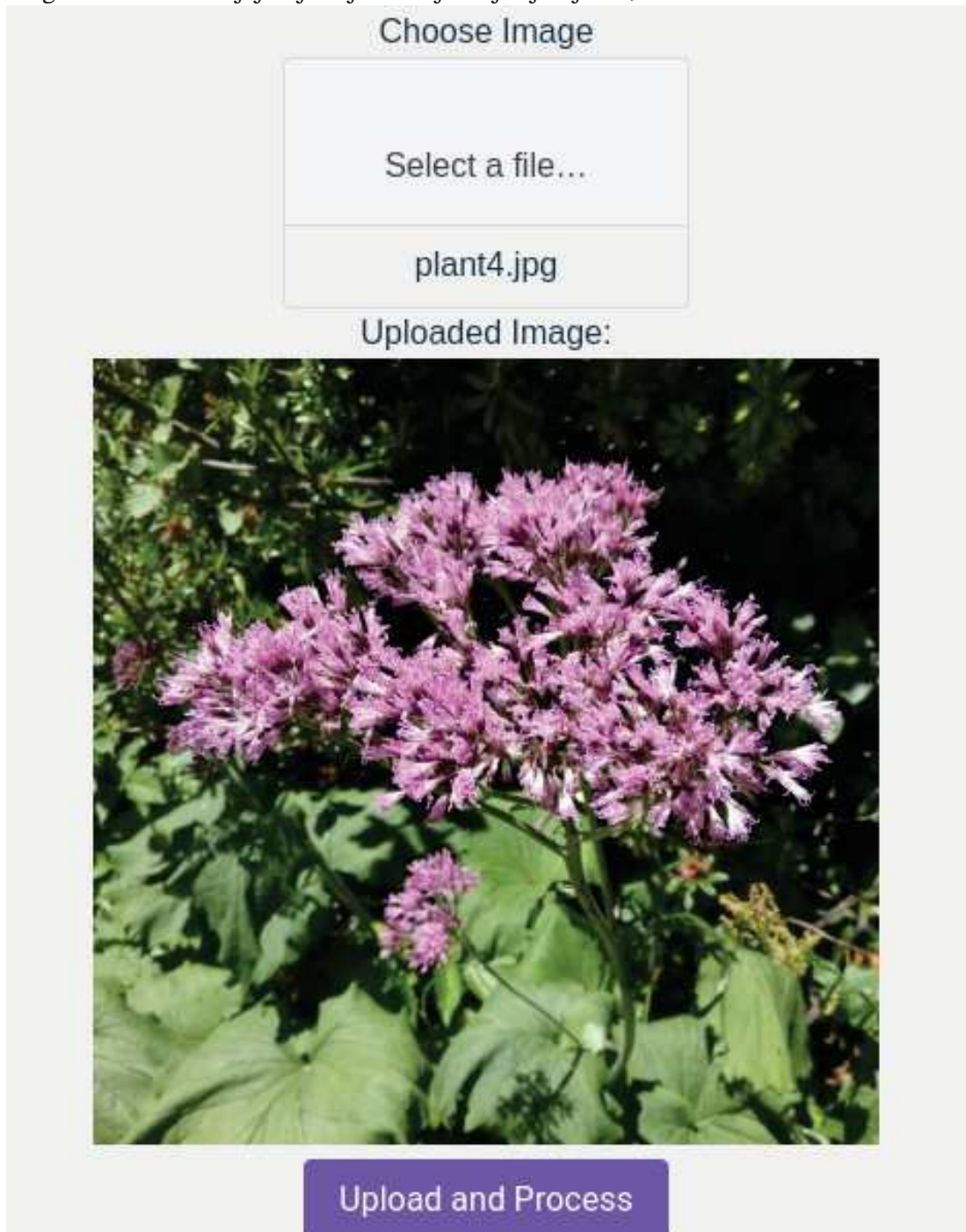
Za učenje i vrednovanje modela korištena je biblioteka otvorenog koda PyTorch Image Models [26].

6.2. Integracija modela u aplikaciju

Zaključivanje (engl. *inference*) je proces koji omogućuje aplikaciji da koristi naučeni model u stvarnom vremenu. Zaključivanje omogućuje aplikaciji predviđanja klasa na temelju učitane slike.

Za aplikaciju je odabran model ResNet50 jer je u eksperimentima davao najbolje performanse. Njegovi parametri su: optimizator Adam, veličina grupe 96, inicijalizacija s prethodno naučenim težinama s ImageNeta, stopa učenja koja se stupnjevito smanjuje od $1,000e-05$ uz faktor smanjenja 0.3, 60 epoha učenja, primjena augmentacije: vertikalno i horizontalno zrcaljenje, varijacija boja, promjena skale i omjera, slučajno brisanje, translacija i rotacija.

U aplikaciju je učitani naučeni model. Kako bi model korisniku mogao dati točnu predikciju potrebno je obaviti predobradu slike, tj. pripremiti tu sliku kako bi odgovarala zahtjevima modela. Predobrada uključuje normalizaciju piksela i promjenu veličine slike. Nakon što model da predikciju za sliku, rezultati se prikazuju korisniku kao lista of pet mogućih klasa od najvjerojatnije do najmanje vjerojatne, slike 6.1 i 6.2.



Slika 6.1. Učitavanje slike u aplikaciju

Uploaded Image:



Upload and Process

Processing Complete

Species Name	Percentage
Lactuca virosa L.	72.27 %
Lactuca serriola L.	14.44 %
Lactuca virosa Habl.	11.58 %
Helminthotheca echioides (L.) Holub	1.3 %
Daphne laureola L.	0.07 %

Slika 6.2. Prikaz rezultata

7. Zaključak

Iz provedenih eksperimenata i analize rezultata može se zaključiti da je augmentacija podataka igrala veliku ulogu u povećanju robusnosti i točnosti modela. Primjena različitih tehnika augmentacije, kao što su horizontalno i vertikalno zrcaljenje, promjena skale i omjera, nasumična varijacija boja, slučajno brisanje, rotacija i translacija, poboljšala je sposobnost modela da prepozna biljke iz različitih perspektiva i uvjeta.

Odabir metoda augmentacije vrlo je bitan kada se model uči na skupu podataka koji ima veliku neuravnoteženost klasa. Metode augmentacije potrebno je odabrati prema specifičnom skupu podataka tako da transformacije imaju smisla s obzirom na slike i objekte koji se na njima nalaze.

Tehnika naduzorkovanja se također pokazala korisnom u rješavanju problema neuravnoteženosti klasa, što se moglo vidjeti kroz poboljšanje srednjeg Jaccardovog indeksa. Rezultati ukazuju na to da je model postao osjetljiviji na manje zastupljene klase, čime je postignuta bolja balansiranost u prepoznavanju svih klasa.

Odabrana arhitektura modela, ResNet50, pokazala se kao efikasna u kombinaciji s primijenjenim tehnikama augmentacije i naduzorkovanjem, ostvarujući dobre rezultate.

Razvijanje jednostavne web aplikacije omogućuje korisniku interakciju s modelom u stvarnom vremenu.

Daljnji rad na ovom problemu mogao bi ići u smjeru istraživanja novih i složenijih arhitektura koje bi mogle pružiti dodatna poboljšanja u performansama modela. Također, uz više vremena, mogla bi se provesti opsežnija analiza i optimizacija hiperparametara. Iako se naduzorkovanje pokazalo korisnim, druge metode poput poduzorkovanja ili generiranja sintetičkih podataka pomoću GAN-ova (engl. *Generative Adversarial Networks*) [27] mogle bi pružiti dodatna poboljšanja. Web aplikacija bi se mogla nadograditi dodavanjem funkcionalnosti da korisnik može pružiti povratnu informaciju o kvaliteti klasifikacije određene slike.

Literatura

- [1] Wikipedia, “Biljke - wikipedia”, <https://hr.wikipedia.org/wiki/Biljke>, posjećeno 3. veljače 2024.
- [2] PL@ntNet, “Pl@ntnet: Plant identification guide”, <https://plantnet.org/>, posjećeno 2. veljače 2024.
- [3] C. Garcin, A. Joly, P. Bonnet, A. Affouard, J. Lombardo, M. Chouet, M. Servajean, T. Lorieul, i J. Salmon, “Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution”, u *NeurIPS Datasets and Benchmarks 2021*, 2021.
- [4] Pl@ntNet Team, “A pl@ntnet dataset for machine learning researchers”, <https://plantnet.org/en/2021/03/30/a-plntnet-dataset-for-machine-learning-researchers/>, posjećeno 8. veljače 2024.
- [5] S. Šegvić, “Uvodno predavanje i pregled osnova dubokog učenja”, PDF prezentacija, 2023., dostupno na: <https://www.zemris.fer.hr/~ssegvic/du/du0intro.pdf>.
- [6] A. Community, “There is no free lunch in data science”, Alteryx Community Blog, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: <https://community.alteryx.com/t5/Data-Science/There-is-No-Free-Lunch-in-Data-Science/ba-p/347402>
- [7] IBM, “What are neural networks?” IBM Topics, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: <https://www.ibm.com/topics/neural-networks>
- [8] S. Šegvić, “Duboke unaprijedne mreže”, PDF prezentacija, 2023., dostupno na: <https://www.zemris.fer.hr/~ssegvic/du/du1feedforward.pdf>.
- [9] M. D. S. 365, “Overview of a neural networks learning process”, Medium, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: <https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa>
- [10] S. Šegvić, “Konvolucijski modeli”, PDF prezentacija, 2023., dostupno na: <https://www.zemris.fer.hr/~ssegvic/du/du2convnet.pdf>.
- [11] Duboko učenje, FER, “Konvolucijski modeli, mnist, cifar:”, <https://dlunizg.github>.

- io/lab2/, posjećeno 1. veljače 2024.
- [12] M. A. Lab, “Do convolutional neural networks mimic the human visual system?” Michigan AI Lab Blog, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: https://msail.github.io/post/cnn_human_visual/
 - [13] S. Šegvić, “Povratne neuronske mreže”, PDF prezentacija, 2023., dostupno na: <https://www.zemris.fer.hr/~ssegvic/du/du5recurrent.pdf>.
 - [14] M. Tan i Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks”. PMLR, 2019. [Mrežno]. Adresa: <https://arxiv.org/abs/1905.11946>
 - [15] K. He, X. Zhang, S. Ren, i J. Sun, “Deep residual learning for image recognition”. IEEE, 2016. [Mrežno]. Adresa: <https://arxiv.org/abs/1512.03385>
 - [16] G. Huang, Z. Liu, L. Van Der Maaten, i K. Q. Weinberger, “Densely connected convolutional networks”. IEEE, 2017. [Mrežno]. Adresa: <https://arxiv.org/abs/1608.06993>
 - [17] PyTorch, “Transforming and augmenting images”, PyTorch Documentation, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: <https://pytorch.org/vision/stable/transforms.html>
 - [18] UbiAI, “What are the difficulties associated with data augmentation?” UbiAI Blog, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: <https://ubiai.tools/what-are-the-difficulties-associated-with-data-augmentation/>
 - [19] J. Brownlee, “Smote oversampling for imbalanced classification”, <https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/>, 2023., posjećeno 8. veljače 2024.
 - [20] W. contributors, “Learning rate — wikipedia,” https://en.wikipedia.org/wiki/Learning_rate, 2024., posjećeno 8. veljače 2024.
 - [21] T. D. Science, “Class imbalance, random sampling, and data augmentation with imbalanced-learn”, Towards Data Science Blog, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: <https://towardsdatascience.com/class-imbalance-random-sampling-and-data-augmentation-with-imbalanced-learn-63f3a92ef04a>
 - [22] Wikipedia contributors, “Jaccard index”, Wikipedia, The Free Encyclopedia, 2024., posjećeno 3. veljače 2024. [Mrežno]. Adresa: https://en.wikipedia.org/wiki/Jaccard_index
 - [23] Django Software Foundation, “Django”, <https://www.djangoproject.com/>, 2023.

- [24] Evan You and the Vue.js team, “Vue.js - the progressive javascript framework”, <https://vuejs.org/>, 2023.
- [25] PyCM, “Pycm: Multi-class confusion matrix library in python”, <https://www.pycm.io/>, 2023., posjećeno 3. veljače 2024.
- [26] R. Wightman, “PyTorch Image Models (timm)”, <https://github.com/huggingface/pytorch-image-models>, 2023., posjećeno 8. veljače 2024.
- [27] W. contributors, “Generative adversarial network — wikipedia,”, 2023., posjećeno 8. veljače 2024. [Mrežno]. Adresa: https://en.wikipedia.org/wiki/Generative_adversarial_network

Sažetak

Klasifikacija biljnih vrsta iz slika biljaka temeljena na dubokom učenju

Marija Tišljar

Učenje modela na skupu podataka sa velikom neuravnoteženosti klasa je izazovan problem. Ovaj rad kroz niz eksperimenata i analizu rezultata dolazi do zaključka da augmentacija podataka igra važnu ulogu u povećanju robusnosti i točnosti modela. Tehnika naduzorkovanja, također se pokazala efikasnom u rješavanju problema neuravnoteženosti klasa, doprinoseći boljem prepoznavanju manje zastupljenih klasa. Odabrana arhitektura modela, ResNet50, u kombinaciji s augmentacijom i optimizacijom hiperparametara, pokazala se uspješnom u postizanju visoke točnosti klasifikacije, dosegovši klasifikacijsku točnost od približno 81 %, što je usporedivo sa srodnim rezultatima iz literature. Dodatno, razvoj web aplikacije omogućio je interakciju korisnika s modelom u stvarnom vremenu.

Ključne riječi: augmentacija podataka; neuravnoteženost klasa; klasifikacija slika; duboko učenje; naduzorkovanje; neuronske mreže; PlantNet-300K; ResNet50

Abstract

Classification of Plant Species from Plant Images Based on Deep Learning

Marija Tišljar

Training models on a dataset with significant class imbalance is a challenging problem. This thesis, through a series of experiments and result analysis, concludes that data augmentation plays an important role in enhancing the robustness and accuracy of the model. The technique of oversampling has also proved effective in addressing the problem of class imbalance, contributing to better recognition of less represented classes. The chosen model architecture, ResNet50, combined with data augmentation and hyperparameter optimization, has been successful in achieving high classification accuracy, reaching accuracy of 81 %, which is comparable to the results of related work. Furthermore, the development of a web application has enabled real-time user interaction with the model.

Keywords: data augmentation; class imbalance; image classification; deep learning; oversampling; neural networks; PlantNet-300K; ResNet50