

Primjena strojnog učenja u klasifikaciji slika kao tipičnih primjera nestrukturiranih tipova podataka

Ivandić, Tomislav

Master's thesis / Diplomski rad

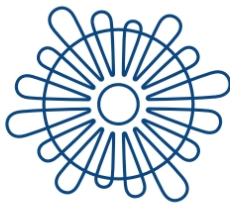
2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zadar / Sveučilište u Zadru**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:162:298962>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-02-23**



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Repository / Repozitorij:

[University of Zadar Institutional Repository](#)



Sveučilište u Zadru

Odjel za informacijske znanosti

Dvogodišnji diplomski sveučilišni studij informacijske znanosti (jednopedmetni izvanredni)

Tomislav Ivandić

**Primjena strojnog učenja u klasifikaciji slika kao tipičnih
primjera nestrukturiranih tipova podataka**

Diplomski rad

Zadar, 2023.

Sveučilište u Zadru

Odjel za informacijske znanosti

Dvogodišnji diplomski sveučilišni studij informacijske znanosti (jednopedmetni izvanredni)

**Primjena strojnog učenja u klasifikaciji slika kao tipičnih
primjera nestrukturiranih tipova podataka**

Diplomski rad

Student:

Tomislav Ivandić

Mentor:

Doc. dr. sc. Ante Panjkota

Zadar, 2023.



Izjava o akademskoj čestitosti

Ja, **Tomislav Ivandić**, ovime izjavljujem da je moj **diplomski** rad pod naslovom **Primjena strojnog učenja u klasifikaciji slika kao tipičnih primjera nestrukturiranih tipova podataka** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 18. ožujak 2023.

Sadržaj

| | |
|---|----|
| Uvod..... | 1 |
| 1 Veza između strojnog učenja i nestrukturiranih podataka | 4 |
| 1.1 Organizacija informacija | 4 |
| 1.2 Pretraživanje i dohvaćanje informacija | 6 |
| 1.3 Reprerentacija informacija..... | 7 |
| 1.4 Diseminacija informacija | 9 |
| 1.5 Kreiranje obogaćenih podatkovnih zbirki..... | 11 |
| 2 Slike kao tipičan primjer nestrukturiranih podataka | 13 |
| 2.1 Digitalna slika..... | 13 |
| 2.2 Model slikovnih podataka | 16 |
| 2.2.1 Fizička reprezentacija slike | 16 |
| 2.2.2 Logička reprezentacija slike..... | 16 |
| 2.3 Kontekstualne informacije | 18 |
| 2.4 Izdvajanje značajki slike..... | 19 |
| 2.4.1 Značajke boje..... | 19 |
| 2.4.2 Značajke oblika | 21 |
| 2.4.3 Značajke teksture | 22 |
| 2.5 Ostali izazovi vezani za klasifikaciju slika..... | 24 |
| 3 Strojno učenje | 26 |
| 3.1 Najvažniji tipovi zadataka u strojnom učenju | 27 |
| 3.1.1 Klasifikacija | 27 |
| 3.1.2 Regresija | 27 |
| 3.1.3 Grupiranje (eng. <i>clustering</i>)..... | 28 |
| 3.2 Osnovni problemi kod strojnog učenja | 28 |
| 3.3 Duboko učenje..... | 31 |
| 3.3.1 Umjetne neuronske mreže..... | 31 |
| 3.3.2 Perceptron..... | 32 |
| 3.3.3 Višeslojni perceptron (MLP) | 33 |

| | | |
|-------|---|----|
| 3.3.4 | Konvolucijskeneuronske mreže..... | 35 |
| 4 | Klasifikacija slika metodama strojnog i dubokog učenja | 38 |
| 4.1 | Klasični modeli strojnog učenja u klasifikaciji slika..... | 38 |
| 4.2 | Duboko učenje u klasifikaciji slika | 39 |
| 4.3 | Evaluacija uspješnosti klasifikacijskih modela | 40 |
| 5 | Istraživački dio (eksperimenti s klasifikacijom slika)..... | 43 |
| 5.1 | Eksperimenti s tradicionalnim modelima strojnog učenja..... | 46 |
| 5.1.1 | Podešavanje hiperparametara | 47 |
| 5.1.2 | Izdvajanje značajki | 47 |
| 5.1.3 | Evaluacija modela..... | 49 |
| 5.2 | Eksperimenti s modelima dubokog učenja | 50 |
| 1.5.1 | Klasifikacija temeljena na AlexNet modelu | 50 |
| 1.5.2 | Klasifikacija temeljena na VGG16 modelu..... | 53 |
| 5.2.1 | Klasifikacija temeljena na ResNet50 modelu | 56 |
| 5.2.2 | Klasifikacija temeljena na Visiontransformer modelu..... | 58 |
| 5.3 | Dodatna razmatranja uz podešavanjehiperparametara, evaluaciju i izbor modela | 60 |
| 5.3.1 | Regularizacija..... | 61 |
| 5.3.2 | Stopa učenja (eng. <i>learning rate</i>) | 62 |
| 5.3.3 | Normalizacija mini-grupe (eng. <i>batch normalization</i>)..... | 63 |
| 5.3.4 | Validacija, evaluacija i izbor modela..... | 63 |
| 6 | Rezultati i rasprava | 65 |
| 6.1 | Usporedba tradicionalnih modela | 65 |
| 6.1.1 | Rezultati klasifikacije "Vegetable image dataset" skupa podataka..... | 65 |
| 6.1.2 | Rezultati klasifikacije "Traffis Sign Classification and Recognition" skupa podataka..... | 67 |
| 6.1.3 | Rezultati klasifikacije "Grapevine Image" skupa podataka | 68 |
| 6.2 | Rezultati DL modela..... | 69 |
| 6.2.1 | Rezultati klasifikacije temeljene na AlexNet modelu | 69 |
| 6.2.2 | Rezultati klasifikacije temeljene na VGG16 modelu..... | 73 |
| 6.2.3 | Rezultati klasifikacije temeljene na ResNet50 modelu | 76 |
| 6.2.4 | Rezultati klasifikacije temeljene na Vision Transformer modelu | 80 |
| 6.3 | Usporedba rezultata tradicionalnih i DL modela..... | 84 |
| | Zaključak..... | 87 |

| | |
|------------------|----|
| Literatura | 89 |
|------------------|----|

SAŽETAK:

Porast količine nestrukturiranih tipova podataka postavio je nove izazove u raznim domenama informacijskih znanosti, čime se istaknula potreba za razvojem inovativnih, učinkovitih i isplativih metoda njihovog procesiranja među kojima istaknuto mjesto zauzimaju različite metode strojnog učenja. Ovaj diplomski rad istražio je razlike između tradicionalnog strojnog učenja i suvremenih metoda dubokog učenja, prvenstveno vezane za njihovu primjenu na zadaće klasifikacije slika, kao tipičnih primjera nestrukturiranog tipa podataka. Kako bismo usporedili izvedbu tradicionalnih algoritama strojnog učenja i modela dubokog učenja, proveli smo niz eksperimenata na tri različita skupa podataka ("Vegetable Image Dataset", "Traffic Signs Classification and Recognition" i "Grapevine Image Dataset") od kojih svaki svojim karakteristikama predstavlja poseban primjer tipičnih izazova za zadaću klasifikacije slika. Modeli dubokog učenja proizveli su značajno bolje rezultate klasifikacije na sva tri korištena skupa podataka.

Ključne riječi: strojno učenje; duboko učenje; klasifikacija slika; nestrukturirani podaci

Uvod

Tradicionalni informacijski sustavi rade s podacima organiziranim prema strogo definiranim podatkovnim modelima, najčešće tablicama čiji stupci i redovi jasno definiraju njihove atribute. Strukturiranjem podataka sukladno unaprijed poznatom modelu znatno olakšavamo njihovu organizaciju, skladištenje, indeksiranje, pretraživanje te općenito ekstrakciju korisnog znanja, neovisno o tome da li se njima služi osoba ili računalo. Primjerice, osobni podaci korisnika knjižnice mogu biti organizirani u tablici u kojoj svaki član zauzima po jedan redak, a svaki stupac sadrži točno određeni podatak o svakom pojedinom članu. Pronalaženje i izvlačenje informacija iz takvog izvora trivijalan je postupak, ali otvorimo li neku od knjiga s polica iste knjižnice suočit ćemo se s tekstem znatno slobodnije strukture. Treća riječ u petom retku označava nešto sasvim drugo nego treća riječ u petnaestom retku. Svaka od tih riječi povezana je s okolnim riječima u nebrojene kombinacije čije točno značenje može biti doslovno, preneseno, ovisiti o kontekstu ili na neki drugi način komunicirati znatno složenije informacije nego neka strogo strukturirana tablica ili baza podataka. Izvlačenje korisnog znanja iz takve strukture nije nimalo trivijalan zadatak za računalo, a sa mnogim će se tekstovima namučiti i ljudski subjekti.

Razvoj informacijsko – komunikacijskih tehnologija s posljedično raširenom upotrebom interneta, društvenih medija i različitih oblika elektroničkog poslovanja suočili su nas s ubrzanim gomilanjem prethodno neviđenih količina podataka od čega samo mali dio čine strukturirani podaci, dok sve ostalo (prema različitim procjenama između 80 i 95%) čine izrazito heterogeni skupovi nestrukturiranih podataka. Nestrukturirani podaci potencijalno su bogat izvori informacija kojima je znatno otežano pristupiti upravo zbog nepostojanja unaprijed definirane strukture, a u svrhu njihovog iskorištavanja neophodan je razvoj odgovarajućih podatkovnih modela te inovativnih, učinkovitih i isplativih metoda procesiranja među kojima istaknuto mjesto zauzimaju različite metode strojnog učenja (Gandomi i Haider 2015).

Tipičan primjer nestrukturiranog tipa podataka svakako su slikovni podaci koji, ponajviše zbog raširene upotrebe pametnih mobilnih telefona, samo 2017. godine bilježe porast od oko 80% (Rai 2017). Struktura zapisa digitalne slike u memoriji računala strogo je definirana u obliku matrica brojeva čiji elementi sadrže informacije potrebne za njihovu reprodukciju, ali nam ne govore ništa o njejoj semantičkoj dimenziji. Prikažemo li sliku na ekranu nekog uređaja ljudski subjekti će s relativnom lakoćom interpretirati njen sadržaj, ali zbog puke količine podataka s kojima smo suočeni neophodno je ovakve procese automatizirati.

Klasifikacija slika područje je primjene metoda strojnog učenja od neprocjenjive praktične vrijednosti u raznim domenama, primjerice sustavima za analizu satelitskih snimaka, kontrole kvalitete, medicinske dijagnostike, prepoznavanja teksta ili računalnog vida. Također nam olakšava izvršavanje zadaća koje se mogu svesti na uspoređivanje različitih mjera sličnosti među slikama, kao što su pretraživanje, pronalaženje ili ekstrakcija korisnih informacija. Razvoj učinkovitijih metoda klasifikacije slika između ostalog omogućava znatno smanjenje radnog opterećenja ljudskih operatera te olakšava donošenje kvalitetnih odluka. Radi se o interdisciplinarnom području koje objedinjuje doprinose različitih grana znanosti i inženjerstva, poput algoritama za obradu signala, senzora, slikovnih sustava, strojnog učenja, itd. (Javidi 2002; Labsky, Vacura, i Praks 2005; Andreopoulos i Tsotsos 2013; Xie i ostali 2015).

Problemu klasifikacije slika do relativno nedavno pristupalo se isključivo tradicionalnim metodama strojnog učenja koje još uvijek nalaze ograničenu primjenu na specijaliziranim zadaćama, ali porastom dostupnosti moćnih računalnih resursa, posebice grafičkih procesorskih jedinica, teorijske prednosti umjetnih neuronskih mreža sastavljenih od velikog broja slojeva konačno su se manifestirale u praksi pa u današnje vrijeme ove zahtjevne zadaće preuzimaju razne metode dubokog učenja, kao što su konvolucijske neuronske mreže ili transformer – enkoder arhitekture (Mahony i ostali 2020).

Cilj ovog diplomskog rada je pokazati osnovne principe i uobičajene izazove u klasifikaciji slika, kao tipičnog primjera nestrukturiranih podataka, primjenom metoda strojnog učenja, te istaknuti prednosti metoda dubokog učenja naspram tradicionalnih metoda.

Rad je koncipiran u pet glavnih cjelina. U prvoj cjelini obuhvaćena su važna područja primjene metoda strojnog učenja u domeni informacijskih znanosti, odnosno njihov utjecaj na unapređenje organizacije, pretraživanja, dohvaćanja, reprezentacije i diseminacije informacija, te izgradnju različitih vrsta obogaćenih podatkovnih zbirki.

Druga cjelina bavi se osnovama digitalne slike koje uključuju njihovu fizičku i logičku reprezentaciju, te izdvajanje značajki. Također kratkim osvrtom ističe važnost kontekstualnih informacija sadržanih u slici.

Tema treće cjeline su strojno i duboko učenje sa kratkim pregledom osnovnih zadaća i uobičajenih problema kod primjene ovih metoda, a u okviru četvrte cjeline obrađena je tema primjene metoda strojnog i dubokog učenja na klasifikaciju slika.

Peta cjelina sadrži prikaz provedenog istraživanja čija je svrha primijeniti, usporediti i evaluirati učinkovitost tradicionalnih naspram suvremenih metoda dubokog učenja na

problem klasifikacije slika, kao tipičnog primjera nestrukturiranih podataka. Posljednje dvije cjeline donose prikaz rezultata s pripadnom raspravom te zaključak.

1 Veza između strojnog učenja i nestrukturiranih podataka

Metode strojnog učenja omogućavaju otkrivanje obrazaca i odnosa unutar velike količine nestrukturiranih podataka, što ih čini posebno pogodnima za rješavanje brojnih problema iz domene informacijskih znanosti. Neke od najvažnijih područja primjene uključuju unapređenje organizacije, pretraživanja, dohvaćanja, reprezentacije i diseminacije informacija, te izgradnju različitih vrsta obogaćenih podatkovnih zbirki.

1.1 Organizacija informacija

Rastuća količina podataka koju svakodnevno generiraju pojedinci, organizacije i strojevi predstavlja izazov za njihovo učinkovito pretraživanje i iskorištavanje, stoga je presudni aspekt upravljanja informacijama u suvremenom dobu njihova organizacija. Jedna od ključnih strategija za svladavanje ovog izazova korištenje je metapodataka, taksonomija i ontologija.

Metapodaci nam pružaju informacije o drugim podacima, odnosno opisuju njihove najvažnije karakteristike, a mogu uključivati informacije poput datuma i vremena kada su stvoreni, ime autora, naslov, opis, format datoteke i slično. Ručno izdvajanje metapodataka dugotrajan je i mukotrpan proces podložan pogreškama, osobito kada se radi o velikim skupovima podataka, stoga se nameće potreba za njegovom učinkovitom automatizacijom. Primjena različitih metoda strojnog i dubokog učenja na automatsko i poluautomatsko izdvajanje metapodataka aktivno je područje istraživanja koje nam omogućava razvoj efikasnijih sustava organizacije velikih količina nestrukturiranih podataka kao što su tekstualni, slikovni, audio ili video materijali. Poluautomatsko izdvajanje metapodataka uključuje sudjelovanje ljudskih subjekata u njihovoj provjeri, korekciji i usmjeravanju procesa izdvajanja, čime se može dodatno povećati kvaliteta izdvojenih informacija (C. Liu i ostali 2019; Pal i ostali 2019; Boukhers i ostali 2021; Hasan Choudhury i ostali 2021).

Metapodaci opisuju pojedine instance podataka, ali nam ne govore mnogo o njihovim međuosobnim odnosima pa je u svrhu učinkovite organizacije informacija često potrebno primijeniti nešto složenije koncepte kao što su taksonomije i ontologije. Taksonomije su hijerarhijske strukture pomoću kojih se organiziraju koncepti na način da svaka kategorija čini podskup šire kategorije. Primjerice, knjižnični fond možemo podijeliti na više kategorija kao što su publicistika, beletristika ili referentna djela, pri čemu svaka od navedenih može sadržavati određeni broj podkategorija. Informacije organizirane na ovakav način omogućavaju

korisnicima da na relativno jednostavan i intuitivan način dođu do željenog izvora. Osim za izravno pretraživanje informacija, taksonomije se koriste u brojne druge svrhe, poput indeksiranja ili sustava za personaliziranu preporuku sadržaja (Doan i ostali 2004; Khoudja, Fareh, i Bouarfa 2018)(C. Zhang i ostali 2018).

Međutim, odnosi među konceptima često su nešto složeniji nego što ih mogu opisati taksonomije. Primjerice, biografije možemo kategorizirati kao dio publicistike, ali također postoje biografije izmišljenih likova koje bismo mogli istovremeno kategorizirati kao beletristiku. Opis ovakvih i složenijih međudnosa omogućavaju nam ontologije, koje osim hijerarhijski organiziranih koncepata uključuju njihove atribute i međusobne veze koje ne moraju nužno biti hijerarhijske. Na primjer, ontologija za medicinsku bazu podataka može uključivati kategorije kao što su bolesti, simptomi, tretmani i lijekovi, pri čemu svaka od ovih kategorija može biti međusobno povezana na različite načine. Korištenjem ontologije za organizaciju podataka unutar određene domene omogućava se uspostavljanje složenih veza između različitih kategorija podataka, te identifikacija međudnosa koji u protivnom možda ne bi bili vidljivi. Ontologije se primjenjuju u raznim granama kao što su ekstrakcija informacija, razvoj softvera, pretraživanje, upravljanje znanjem ili sustavi za personaliziranu preporuku sadržaja. Također čine temelj semantičkog weba čija se organizacija i pretraživanje zasniva na ontološkom znanju (Somodevilla García i ostali 2018; Lakzaei i Shamsfard 2021). Izrada taksonomija i ontologija na tradicionalan način dugotrajan je proces u kojem često sudjeluju čitavi timovi domenskih stručnjaka, ali jednako kao u slučaju izdvajanja metapodataka, eksplozivni rast količine nestrukturiranih oblika podataka koji je obilježio protekla desetljeća presudnom čini automatizaciju ovih zahtjevnih procesa. Primjerice, automatsko kreiranje taksonomija iz velikog korpusa tekstualnih podataka kao što su zbirke dokumenata temelj je za njihovu semantičku analizu, kojom se otvaraju različite mogućnosti praktične primjene. Primjena metoda strojnog i dubokog učenja na ove zadaće u literaturi se obično naziva učenje taksonomije, odnosno učenje ontologije. (C. Zhang i ostali 2018).

Organizacija informacija često uključuje istovremeno korištenje nekoliko ontologija koje mogu biti iz različitih domena sa svojim karakterističnim terminologijama pa uobičajen problem čini njihova heterogenost na semantičkoj, terminološkoj, sintaktičkoj ili semiotičkoj razini, stoga je nužno omogućiti njihovu interoperabilnost pronalženjem semantičkih odnosa ili podudaranja između entiteta različitih ontologija (eng. *ontology matching*). Metode strojnog i dubokog učenja mogu omogućiti učinkovitu automatizaciju ovog zahtjevnog procesa koji se također tradicionalno obavljao ručno (Doan i ostali 2004; Khoudja, Fareh, i Bouarfa 2018).

1.2 Pretraživanje i dohvaćanje informacija

Pretraživanje i dohvaćanje informacija važan su aspekt svakodnevnog života suvremenog čovjeka, temelj su znanstvenoistraživačkog rada, a od posebnog su interesa za mnoga znanstvena područja, poput računarstva, informacijskih znanosti ili knjižničarstva. Možemo ga definirati kao proces dobivanja relevantnih informacija iz velikog korpusa podataka, kao odgovor na korisnikov upit. Upit je formalni izraz korisnikovih informacijskih potreba na temelju kojeg sustav za pretraživanje pronalazi odgovarajuće predmete pretraživanja i rangira ih po relevantnosti, odnosno sličnosti sa sadržajem upita. Dakle, problem pretraživanja i dohvaćanja u suštini se može svesti na uspoređivanje različitih mjera sličnosti i rangiranje rezultata. Važno je istaknuti razliku u odnosu na pretraživanje u okvirima tradicionalnih sustava za upravljanje bazama podataka koji sadrže strogo strukturirane podatke u obliku povezanih tablica. Pretraga takvih baza u pravilu omogućava dobivanje vrlo specifičnih rezultata temeljenih na točnom podudaranju s upitom, odnosno radi se o procesu koji je po prirodi deterministički. Pretraživanje baza nestrukturiranih podataka znatno je složeniji proces čiji se rezultati u različitoj mjeri podudaraju s upitom pa u ovom slučaju govorimo o probabilističkom pretraživanju. Primjena strojnog učenja na organizaciju informacija u velikoj mjeri doprinosi kvaliteti sustava za pretraživanje i dohvaćanje, ali njihov doprinos nije ograničen samo na automatsko generiranje metapodataka, taksonomija i ontologija. Naime, suvremeno dohvaćanje informacija temelji se na složenim sustavima koji organiziraju i indeksiraju podatke, procesiraju korisničke upite, te koriste algoritme rangiranja kako bi omogućili preglednu prezentaciju relevantnih rezultata (Bajaj, Sharma, i Sangwan 2020; Guo i ostali 2020).

Temelje sustava za pretraživanje tradicionalno čini neki oblik indeksa, odnosno strukture koja povezuje važne attribute predmeta pretraživanja s njihovom lokacijom. Jednako kao kazalo pojmova na kraju knjige, koje povezuje ključne riječi sa stranicama na kojima se mogu pronaći, indeksi ubrzavaju proces pretraživanja jer, umjesto učitavanja i pregledavanja svakog pojedinog dokumenta, omogućavaju pretraživanje znatno jednostavnije podatkovne strukture. Metode dubokog i strojnog učenja omogućavaju automatizaciju procesa izrade indeksa, kojeg zbog ogromne količine podataka često nije moguće obaviti na tradicionalne načine (Gani i ostali 2015). Kvaliteta rezultata pretraživanja uvelike ovisi o kvaliteti korisničkih upita, stoga neki autori nude modele temeljene na strojnom učenju koji proširuju upite dodatnim pojmovima sukladno naučenim obrascima jezika karakterističnog za područje pretraživanja (Lucchese i ostali 2018).

Strojno učenje također može znatno unaprijediti personalizaciju sustava za pretraživanje koji uzimaju u obzir interese korisnika kod rangiranja rezultata. Umjesto fiksnog skupa pravila, tipičnih za tradicionalne modele personaliziranih tražilica, ovakvi sustavi uzimaju u obzir dinamičku prirodu procesa traženja informacija pa omogućavaju kontinuirano prilagođavanje pravila sukladno promjenama korisničkog informacijskog ponašanja (Yao i ostali 2020; 2021).

Metode strojnog učenja također mogu olakšati pronalaženje informacija omogućavanjem razvoja učinkovitijih sustava preporuka. Primjerice portal s vijestima može koristiti algoritme klasifikacije za preporuku članaka korisnicima na temelju njihovih interesa, odnosno analizom korisnikove povijesti čitanja i klasificiranjem članaka u relevantne kategorije, sustav može dati personalizirane preporuke za koje postoji veća vjerojatnost da će biti od interesa za pojedinog korisnika (S. Zhang i ostali 2019).

Jednako kao sveprisutnost nebrojenih izvora informacija, digitalno doba odlikuje raširena dostupnost različitih načina za njihovu diseminaciju, stoga pitanje kvalitete informacija koje pronalazimo predstavlja jedan od ključnih problema suvremenog društva. Donedavno je pristup masama bio ograničen samo više ili manje strogo kontroliranim medijima poput tiskovina, radija ili televizije, ali u današnje doba gotovo svatko ima mogućnost da najrazličitije informacije učini dostupnima širokoj publici. Informacijska sloboda svakako može biti pokretač brojnih pozitivnih društvenih procesa, ali donosi i obilje negativnih posljedica poput širenja štetnih, lažnih ili zavaravajućih informacija koje mogu naštetiti korisniku ili društvu cjelini. Analiza sadržaja metodama strojnog učenja omogućava klasifikaciju dostupnih izvora prema vjerodostojnosti pa su tako razvijeni brojni modeli namijenjeni otkrivanju lažnih vijesti, zavaravajućeg ili neprimjerenog sadržaja, omogućivši tako povećanje kvalitete rezultata pretraživanja (Shahrivari, Darabi, i Izadi 2020; Meesad 2021).

1.3 Rerezentacija informacija

Osim nepostojanja strogo definirane strukture, nestrukturirane tipove podataka odlikuje velika količina šuma, odnosno nevažnih ili redundantnih informacija koje dodatno otežavaju pristup korisnom znanju, stoga je sirove podatke potrebno transformirati u smisleniji i kompaktniji prikaz koji obuhvaća njihove temeljne obrasce. Rerezentaciju nestrukturiranih podataka obično nazivamo značajkama ili atributima, a jedan od jednostavnijih primjera kada govorimo

o tekstualnim podacima je model vreće riječi (eng. *bag-of-words*) koji dokumente predstavlja kroz učestalost ponavljanja pojedinih riječi u tekstu. Na sličan način slike možemo predstaviti pomoću histograma boja koji pruža informacije o broju piksela svake boje sadržane u slici. Nedostaci obje metode proizlaze iz činjenice da su informacije koje nose dokumenti i slike znatno kompleksnije od samog broja gradivnih elemenata pa je za zahtjevnije zadaće potrebno primijeniti složenije reprezentacije koje obuhvaćaju širi raspon informacija sadržanih u nestrukturiranim oblicima podataka (Murala, Gonde, i Maheshwari 2009; Qader, M. Ameen, i Ahmed 2019; Y. Wang i Jordan 2022).

Samo izdvajanje značajki automatiziran je proces, ali njihov se odabir tradicionalno radi ručno, odnosno odgovarajući stručnjak ili tim stručnjaka mora prema osobnom znanju i iskustvu odabrati značajke koje smatra najinformativnijima za obavljanje potrebne zadaće. Metode strojnog i dubokog učenja mogu značajno unaprijediti ovaj korak kroz tzv. učenje reprezentacija ili učenje značajki. Učenje značajki temeljni je aspekt strojnog učenja s brojnim praktičnim primjenama, a posljednjih je godina doživio revoluciju razvojem metoda dubokog učenja. Klasičan primjer ove metode su duboke konvolucijske neuronske mreže (eng. *convolutional neural networks* - *CNN*) koje kroz proces učenja prilagođavaju vrijednosti filtara za detekciju značajki slike, poput različitih vrsta rubova ili tekstura, ovisno o njihovoj relevantnosti za odabranu zadaću. Rezultat ovog procesa hijerarhijska je reprezentacija slikovnih podataka u obliku mapa značajki. Jednako kao što ljudi kompleksne koncepte prikazuju kao kompoziciju jednostavnijih, hijerarhija mapa značajki opisuje podatke kroz nekoliko razina apstrakcije pa tako značajke niske razine kao što su jednostavni kutovi, rubovi, teksture i boje međusobnim kombinacijama čine složenije značajke više razine kao što su oblici ili konkretni predmeti. Metode strojnog učenja sposobne su razotkriti obrasce unutar skupa podataka koje su ljudskom subjektu teško vidljive, stoga učenje značajki osim automatizacije procesa može značajno unaprijediti njihovu kvalitetu (Yoshua Bengio, Courville, i Vincent 2013; Alzubaidi i ostali 2021; Gupta, Pathak, i Kumar 2022).

Dodatna prednost reprezentacija kreiranih na ovakav način je da mogu prenijeti znanje naučeno na jednom skupu podataka (eng. *transfer learning*), odnosno mogu se primijeniti na druge skupove koji zbog svoje veličine ne omogućavaju izdvajanje dovoljnog broja kvalitetnih značajki pa tako CNN modeli prethodno trenirani na ImageNet skupu podataka, koji sadrži preko 14 milijuna slika u 1000 kategorija, mogu učinkovito klasificirati višestruko manje skupove podataka. Značajke dobivene na velikom skupu heterogenih podataka mogu biti dovoljno apstraktne da omogućavaju primjene na skupove iz potpuno nepovezanih domena. Primjerice, modeli prethodno trenirani na ImageNet skupu uspješno se mogu

primijeniti na klasifikaciju slika iz domene medicinske dijagnostike, iako nijedna klasa ovog skupa ne sadrži slike takvog sadržaja (Kim i ostali 2022).

Duboko učenje znatno je unaprijedilo i reprezentaciju informacija na području obrade prirodnog jezika, ponajviše razvojem transformer arhitektura dubokih neuronskih mreža. Transformer arhitekture sastoje se od dva ključna segmenta – enkodera i dekodera. Enkoder je zadužen za transformaciju ulaznih podataka u njihovu nižedimenzionalnu reprezentaciju, a dekodeer je segment arhitekture koji temeljem spomenute reprezentacije rekonstruira izvorne podatke. U usporedbi sa jednostavnim *bag-of-words* modelom, reprezentacije koje stvara enkoder uključuju informacije o međusobnim odnosima riječi unutar teksta, te na taj način preciznije enkodiraju značenje koje tvore svojim kombinacijama. Navedena osobina čini ih iznimno uspješnima na području strojnog prevođenja, a njene različite varijante također se mogu primijeniti na rekonstrukciju slika ili videa. Upotreba reprezentacija slikovnih podataka dobivenih ovim metodama na zadaće klasifikacije pokazale su obećavajuće rezultate koji u nekim slučajevima nadilaze konvolucijske neuronske mreže, stoga posljednjih godina primjena transformera na probleme računalnog vida predstavlja rastuće područje istraživanja (Wu i ostali 2020).

1.4 Diseminacija informacija

Metode dubokog i strojnog učenja omogućavaju nam razotkrivanje znanja skrivenog u velikim količinama nestrukturiranih podataka, te njegovu učinkovitu prezentaciju u korisniku razumljivim oblicima koji omogućavaju dublje uvide i kvalitetnije donošenje odluka. Njihov učinak na diseminaciju informacija vjerojatno je najočitiiji kroz sustave koji omogućavaju pronalaženje relevantnih izvora, kao što su sustavi za pretraživanje i dohvaćanje, kreiranje preporuka ili personaliziranih sadržaja, ali teško je zamisliti područje ljudskog djelovanja koje ne bi moglo imati koristi od diseminacije informacija koje bi u protivnom ostale skrivene u gomilama nestrukturiranih podataka (Serrano-Guerrero i ostali 2011).

Primjerice, Chen, Gryllias, i Li (2019) donose unaprijeđenu metodu detekcije defekata u proizvodnji dijelova rotacijskih strojeva koja uz pomoć konvolucijske neuronske mreže izdvaja značajke visoke razine iz sirovih vibracijskih signala i prosljeđuje ih klasifikatoru koji prepoznaje defekte.

Strojno učenje može se koristiti i za predviđanje prometnih obrazaca, prepoznavanje potencijalnih nesreća i optimiziranje planiranja rute kao dijela pametnih prometnih rješenja,

odnosno sustava koji se oslanjaju na korištenje naprednih tehnologija za poboljšanje učinkovitosti i sigurnosti prometa. U kombinaciji s tehnologijom Interneta stvari (eng. *Internet of Things*), navedene metode omogućavaju učinkovitu diseminaciju ovih informacija između različitih uređaja, vozila, putnika i ostalih sudionika u prometu (Zantalis i ostali 2019).

Duboko učenje može unaprijediti diseminaciju informacija u kontekstu tehnologije pametnih gradova. Primjerice, iMARS sustav temelji se na modelu dubokog učenja treniranog na velikom skupu povijesnih podataka kako bi pružio točna predviđanja budućih događaja i uvjeta. Dobiveni podaci dostavljaju se korisnicima putem mobilne aplikacije koja u realnom vremenu pruža informacije poput situacija u prometu, vremenskih uvjeta ili događanja. Sustav također omogućava prikaz različitih informacija u obliku nadzorne ploče namijenjene predstavnicima gradskih vlasti zaduženih za nadzor i upravljanje sustavom (Basori i ostali 2019).

Na području medicinske dijagnostike strojno i duboko učenje omogućava prepoznavanje i klasifikaciju različitih patoloških promjena pa tako Yadav i Jadhav (2019) predlažu model za klasifikaciju slika raka kože temeljen na konvolucijskoj neuronskoj mreži. Metode strojnog učenja također omogućavaju analizu i razotkrivanje skrivenih obrazaca u podacima prikupljenima nosivim senzorima koji bilježe fiziološke parametre ljudskog tijela (WBAN) u svrhu pružanja zdravstvenih informacija u realnom vremenu, te predviđanja različitih zdravstvenih ishoda (Punj i Kumar 2021).

Doprinos ovih metoda nije zaobišao ni geografske informacijske sustave te sustave za upravljanje prirodnim nepogodama. Primjerice, analizom podataka prikupljenih sustavima za daljinsko istraživanje, kao što satelitske snimke, LIDAR-i ili SAR radari moguće je otkriti različite geološke fenomene koji nisu vidljivi golim okom, što može omogućiti i olakšati ranu detekciju lokacija na kojima postoji visoka vjerojatnost pojave različitih nepogoda poput odrona, te na taj način doprinijeti prevenciji materijalnih šteta i gubitka ljudskih života (Mohan i ostali 2021).

Alati temeljeni na strojnom i dubokom učenju mogu značajno povećati točnost i brzinu obavljanja različitih poslova vezanih za znanstveno istraživački rad. Na primjer, automatskim generiranjem sažetaka, pojašnjenjem dijelova složenih znanstvenih publikacija ili mapiranjem literature, navedeni alati mogu ne samo ubrzati i olakšati istraživački rad, nego omogućiti vrijedne uvide koji bi u protivnom mogli proći nezapaženo. Također mogu omogućiti automatizaciju napornih, repetitivnih poslova, te na taj način omogućiti istraživačima da svoje

vrijeme i energiju usmjere na složenije ili kreativnije aspekte znanstvenog rada (Chandha, R, i Ghosal 2023).

Sve u svemu, metode strojnog i dubokog učenja značajno su unaprijedile načine na koje izvlačimo korisno znanje iz velikih količina nestrukturiranih podataka, te omogućile njihovo prezentiranje korisnicima i drugim informacijskim sustavima u oblicima koji omogućavaju stvaranje korisnih uvida ili olakšavanje donošenja važnih odluka u okvirima širokog spektra ljudskih aktivnosti.

1.5 Kreiranje obogaćenih podatkovnih zbirki

Ekstrakcijom metapodataka iz nestrukturiranih podataka moguće je semantički obogatiti podatkovne zbirke iz različitih područja ljudskog djelovanja. Primjerice, sustavi za automatsku anotaciju slika temeljem analize njihovog sadržaja generiraju odgovarajuće ključne riječi koje ih dodatno opisuju. Navedenom se problemu tradicionalno pristupalo ručno ili metodama temeljenima na vjerojatnosti i statistici, a proteklih nekoliko godina porasla je popularnost primjene metoda proizašlih iz teorije grafova, te metoda strojnog i dubokog učenja (Ke, Zou, i Niu 2019).

Semantičko obogaćivanje zbirki slika nije ograničeno samo na automatsko generiranje ključnih riječi pa tako Abgaz i ostali (2021) predstavljaju metodologiju koja kombiniranom primjenom konvolucijske neuronske mreže i tehnologija semantičkog weba ekstrahira znanje skriveno unutar kolekcije slika kulturne baštine, te na taj način omogućava bolji uvid u političke, društvene, ekonomske i kulturne vrijednosti vremena u kojem su nastale.

Ali i ostali (2022) nude nešto složeniji sustav namijenjen obogaćivanju zbirki slika iz digitaliziranih novinskih isječaka, koji uz pomoć metapodataka izdvojenih iz članaka i klasifikacije slika metodama strojnog učenja otkriva geografske lokacije motiva s fotografija, te na ovaj način omogućava izradu refotografija, odnosno slika koje kombiniraju povijesne fotografije iste lokacije s fotografijama iz današnjeg vremena.

Primjenom metoda dubokog učenja moguće je obogatiti i digitalne zbirke geografskih karata. Primjerice, Hu i ostali (2022) razvili su model temeljen na kombinaciji konvolucijskih i povratnih neuronskih mreža koji pronalazi i izdvaja metapodatke iz geografskih karata, kao što su lokacija, mjerilo i orijentacija. Dobiveni podaci mogu se upotrijebiti u svrhu unaprijeđenja sustava za pretraživanje i dohvaćanje geografskih karata.

Karnani i ostali (2022) ističu kako digitalne zbirke slika bioloških uzoraka često sadrže manjkave ili pogrešne metapodatke, a navedene izazove nastoje riješiti primjenom metoda računalne obrade slike i strojnog učenja. Rješenje koje predlažu omogućava automatsko izdvajanje metapodataka koje brzinom i točnošću znatno nadmašuje ljudske subjekte. Sustav je za sada ograničen na slike uzoraka riba, ali sličnim pristupom moguće je obogatiti i zbirke slika ostalih oblika bioloških uzoraka. Metapodaci izdvojeni na ovaj način mogu se upotrijebiti za unaprjeđenje usluga digitalnih knjižnica koje omogućavaju pristup ogromnom broju slika sličnog sadržaja.

Metode dubokog i strojnog učenja vezane za područje obrade prirodnog jezika uspješno se primjenjuju za obogaćivanje zbirki dokumenata. Safder i ostali (2020) nude rješenje izazova vezanih za pretraživanje i dohvaćanje dokumenata specifičnih sadržaja poput znanstvenih članaka koji uključuju informacije vezane za računalne algoritme kao što su uzorci pseudokoda, informacije o korištenim skupovima podataka, performansama ili kompleksnosti algoritama.

Podaci generirani u procesu industrijske proizvodnje također mogu predstavljati opsežne podatkovne zbirke koje je moguće obogatiti primjenom sličnih metoda. Costa i ostali (2017) predlažu sustav temeljen na kombinaciji tehnika strojnog učenja i semantičke analize sadržaja uz pomoć ontologija u svrhu otkrivanja implicitnih informacija unutar zbirki podataka o proizvodima iz građevinske domene. Ovakav pristup omogućava učinkovitije pretraživanje i dohvaćanje informacija o proizvodima, posebice u usporedbi s najčešće korištenim sustavima temeljenima na indeksiranju i ekstrakciji ključnih riječi, koje ne mogu u odgovarajućoj mjeri obuhvatiti informacijsko bogatstvo sadržano u ovakvim zbirkama podataka.

Navedeni primjeri predstavljaju samo neke od mogućih primjena metoda strojnog učenja na izdvajanje metapodataka u svrhu obogaćivanja velikih podatkovnih zbirki, koje postaju sveprisutne zbog ubrzanog rasta količine nestrukturiranih vrsta podataka.

2 Slike kao tipičan primjer nestrukturiranih podataka

Slikovni podaci zapisani u digitalnom obliku slijede strogo definiranu strukturu koja određuje njihov fizički prikaz, ali svojim sadržajem predstavljaju tipičan primjer nestrukturirane vrste podataka. Sadržaj i značenje slike uvelike ovise o subjektivnoj interpretaciji promatrača, stoga je uz strukturu njihovog digitalnog zapisa potrebno definirati podatkovne modele koji na odgovarajući način obuhvaćaju elemente njihovog sadržaja.

2.1 Digitalna slika

Glavni preduvjet za računalnu obradu slike je da ona bude zapisana u digitalnom obliku. Slika može biti u potpunosti generirana pomoću računala, ali većina digitalnog slikovnog materijala na neki je način prenesena iz fizičkog svijeta u računalni oblik kroz proces uzorkovanja i kvantizacije. Slika se u fizičkom svijetu najprije snima pomoću polja senzora koji na razne načine pretvaraju energiju svjetlosnog zračenja u električni signal. Uzorkovanje je proces diskretizacije ovog kontinuiranog električnog signala, čime se snimljena slika raščlanjuje na diskretne prostorne koordinate. Vrijednosti amplituda ovog kontinuiranog električnog signala nose informacije o boji i svjetlini pojedinog dijelica slike, a njihovu diskretizaciju nazivamo kvantizacijom (Acharya i Ray 2005).

Kvantizacija se provodi u dvije faze, a to su odabir ograničenog skupa boja koje će diskretnim veličinama predstavljati boje iz kontinuiranog raspona s izvorne slike i dodjeljivanje odgovarajuće vrijednosti pojedinom pikselu. Većina suvremenih metoda kvantizacije temelji se na nekom od algoritama grupiranja (Celebi 2011).

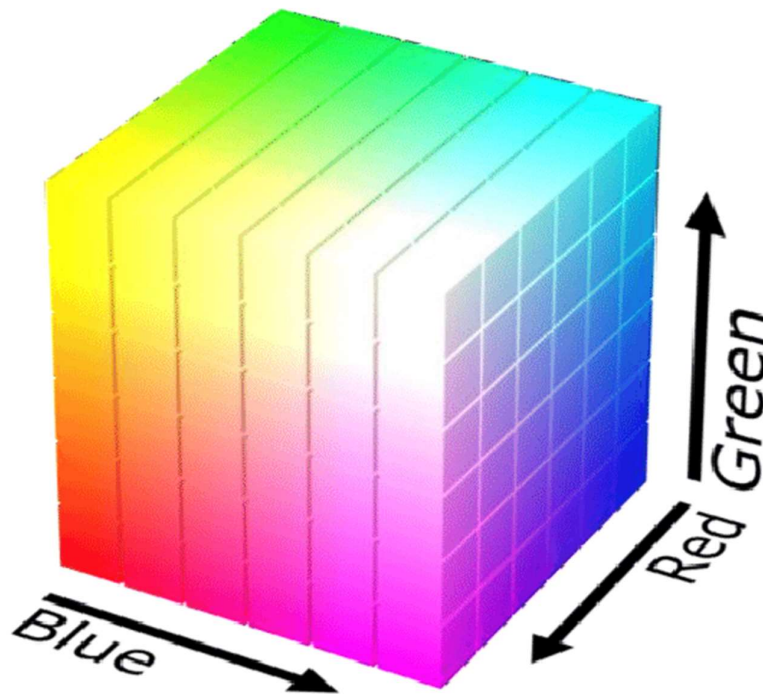
Riječ piksel (eng. *pixel*) izvedenica je od engleskog „picture element“, a označava sićušne dijelove koji zajedno poput mozaika čine čitavu sliku. Informacije o prostornom smještaju i boji pojedinog elementa slike spremaju se u matrice čije dimenzije odgovaraju dimenziji slike u pikselima. Ukupan broj piksela nazivamo rezolucijom koju najčešće izražavamo kao dimenzije na ovaj način digitalizirane slike, primjerice 32x32, 800x600 ili 1024x768.

Svaki element matrice sadrži cijeli broj koji opisuje boju pojedinog piksela numeričkim vrijednostima iz određenog modela boja, odnosno matematičkog modela sastavljenog od reprezentacije svake boje iz odabranog raspona u obliku n-torki cijelih brojeva. Model boja za crno-bijelu sliku najjednostavniji je primjer jer sadrži samo različite vrijednosti intenziteta, koje percipiramo kao različite nijanse sive boje, smještene u jednoj matrici. Model za prikaz

slika u boji nešto je složeniji jer opis njihovih piksela zahtijeva više cjelobrojnih vrijednosti koje označavaju zasićenost pojedine osnovne boje, nijanse, svjetline ili nekog drugog podatka ovisno o odabranom modelu (Acharya i Ray 2005).

U kontekstu digitalne slike često se spominje pojam prostora boja, koji opisuje specifičan, mjerljiv i fiksni raspon mogućih boja i vrijednosti svjetline. Njegova najosnovnija praktična funkcija je opisati mogućnosti uređaja za snimanje ili reprodukciju informacija o boji. S druge strane, model boja temelj je podatkovne strukture na kojoj izvodimo operacije procesiranja slike, a predstavlja matematičku apstrakciju unutar koje organiziramo informacije o bojama određenog raspona. Neki od često korištenih modela boja su RGB i HSI.

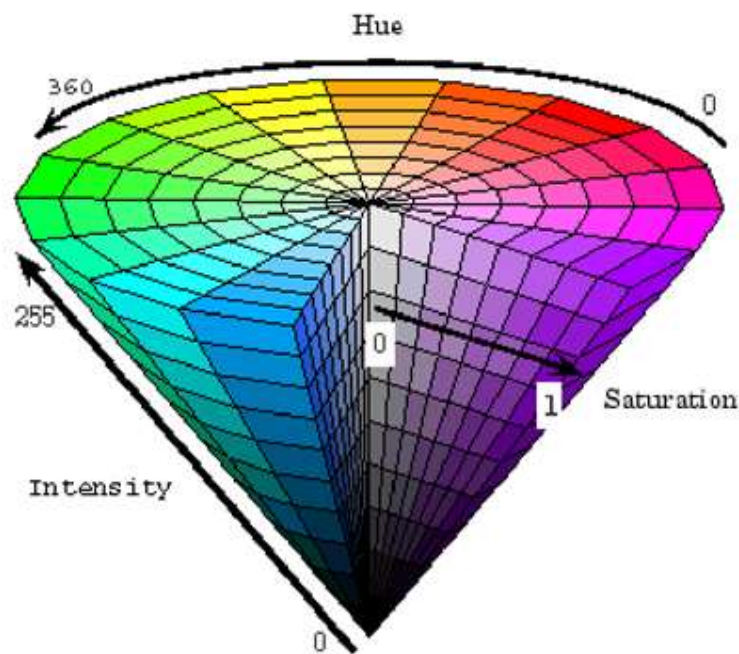
RGB (Red, Green, Blue) model najčešće je korišten model boja, a temelji se na aditivnoj kombinaciji tri osnovne boje prvog reda - crvene, zelene i plave. Pojedinačni pikseli opisani su vrijednostima intenziteta svake od ovih boja u unaprijed određenom rasponu, primjerice od 0 do 255. Koncept dobivanja čitavog raspona sekundarnih boja miješanjem nekolicine osnovnih boja proizlazi izravno iz teorije ljudske vizualne percepcije poznate davno prije pojave elektroničkih uređaja, stoga ne iznenađuje da se radi o prvom i još uvijek najčešće korištenom modelu. Projekcija RGB modela na kocki prikazana je na Slici 1.



Slika 1 Prikaz RGB modela mapiranog na kocku (Vieira 2021, 32)

RGB model idealan je za generiranje slike kod njenog prikaza na ekranu ili snimanja kamerom, ali nije najbolji izbor želimo li razvijati algoritme za obradu slike. Naime, ljudsko oko primjećuje tri osobine slike – ton (eng. *hue*), zasićenost (eng. *saturation*), odnosno mjeru kojom je ton boje razrijeđen određenim udjelom bijele boje, te svjetlinu (eng. *brightness*) koju nije moguće izravno kvantificirati na zadovoljavajući način jer je u potpunosti subjektivna. Učinkovito rješenje ovog problema korištenje je jednog od ključnih elemenata svjetline, a to je intenzitet (eng. *intensity*), vrijednost koja sadrži informaciju o razini nijanse sive boje.

HSI (Hue, Saturation, Intensity) predstavlja implementaciju ovog rješenja, a slikovito je prikazanona Slici 2.



Slika 2 Prikaz HSI modela boja mapiranog na stožac(Ismael i Kashmar 2019, 6)

Raspon kuta koji određuje ton je od 0 do 360 u kojem 0, 120 i 240 predstavljaju redom crvenu, zelenu i plavu boju. Suprotni vrhovi stožaca predstavljaju bijelu, odnosno crnu boju. Osim što su njegove komponente usko povezane s načinom na koji ljudi percipiraju boje, važna karakteristika HSI modela boja leži u tome da promjena vrijednosti komponente intenziteta nema nikakav učinak na informacije o boji slike. Navedene osobine čine ovaj model posebno pogodnim za primjenu različitih metoda obrade slike. Primjerice, za razliku od RGB modela koji može biti izrazito osjetljiv na promjene osvjetljenja, HSI i srodni modeli kao što su HSV i HSB vjernije zadržavaju informacije o boji, što ih čini pogodnijima za

izdvajanje značajki boje, a odvojenost intenziteta kao neovisne komponente omogućava lakše izdvajanje značajki koje ovise o promjenama svjetline, kao što je detekcija rubova (Gonzalez i Woods 2018).

2.2 Model slikovnih podataka

Preduvjet za semantičku obradu nestrukturiranih podataka kao što su slike izrada je njihovog strukturiranog prikaza, odnosno apstrakcije sastavljene od niza koncepata koji opisuju njihov sadržaj. Mnoštvo je takvih podatkovnih modela, ali su obično vezani za specifično područje primjene, stoga ćemo za potrebe ovog rada koristiti opći model kojeg su analizirajući različite domene, kao što su medicinske slike, planovi dizajna namještaja za kuće, katalogi elektroničkih shema i geografski informacijski sustavi, razvili Grosky i Stanchev (2002). Prema navedenom modelu, sliku možemo raščlaniti na njenu fizičku i logičku reprezentaciju.

2.2.1 Fizička reprezentacija slike

Neovisno da li je otisnuta na papiru, naslikana na platnu ili nacrtana na zidu, analogna slika fizički je neodvojiva od svog zapisa. Digitalna slika nešto je složenija jer je prikaz na ekranu nekog uređaja odvojen od njenog fizičkog zapisa u memoriji. Dakle, fizička reprezentacija digitalne slike u ovom se kontekstu odnosi na zapis u memoriji računala koji se u najčešće korištenom, rasterskom obliku sastoji od zaglavlja i matrica slike. Zaglavlje sadrži informacije o osnovnim parametrima kao što su rezolucija, format i sažimanje, a matrice se sastoje od podataka dobivenih procesom uzorkovanja i kvantizacije koji opisuju pojedinačne grafičke elemente, odnosno piksele. Odgovarajuća fizička reprezentacija važan je preduvjet za digitalnu obradu slike, ali budući da nam ne govori ništa o onome što slika prikazuje, presudan korak za učinkovitu klasifikaciju slika prema sadržaju izrada je njene logičke reprezentacije.

2.2.2 Logička reprezentacija slike

Logička reprezentacija slike odnosi se na značenje podataka sadržanih u slici kao što su njen opis, osobine objekata koje sadrži te njihovi međusobni odnosi. Objekt sadržan u slici definira

se kao dio slike sa određenim značenjem (primjerice, „mačka”, „stablo” ili „trokut”), a može se sastojati od jednog ili više različitih područja slike.

Prema modelu kojeg predlažu Grosky i Stanchev (2002) logička reprezentacija slike sastoji se od meta značajki, semantičkih značajki, te značajki boje, oblika, teksture i prostornih značajki.

Meta značajke odnose se na proces stvaranja slike kao što su datum, naziv, uređaj kojim je slika napravljena i slično. Semantičke značajke subjektivne su informacije koje na određeni način opisuju sliku, a obično ih određuje odgovarajući domenski stručnjak. Primjer semantičke značajke relevantne za klasifikaciju je oznaka klase.

Boja čini vrlo važne značajke korištene za pretraživanje i klasifikaciju slika, a primjer najčešće korištene metode izdvajanja ovih značajki je histogram boja.

Značajke oblika dijelimo na globalne i lokalne. Globalne značajke oblika kao što su opseg, površina ili središnja os odnose se na čitavu sliku. Lokalne značajke vezane su za pojedinačne objekte na slici, a njihov tipičan primjer su rubovi ili vrhovi kutova.

Tekstura je osobina površine koju ljudi obično povezuju s osjećajem dodira, a proizlazi iz međusobnog uređenja sastavnica materijala kojeg promatramo. Kada govorimo o teksturi u kontekstu digitalne slike ona se prikazuje kao karakterističan prostorni raspored piksela koji čini prepoznatljiv obrazac vrijednosti sivih tonova.

Prostorne značajke govore nam o prostornim međudodnosima objekata unutar slike, a mogu se izraziti na brojne načine od kojih autori navode ukupno sedam:

1. Topološki skup odnosa među objektima kao što su odvojenost, dodirivanje ili preklapanje
2. Vektorski skup relativnih prostornih odnosa (npr. lijevo od, istočno od)
3. Skup relativnih metričkih prostornih odnosa, odnosno udaljenosti (npr. blizu, daleko, jako blizu)
4. Dvodimenzionalni nizovi simboličkih reprezentacija objekata unutar slike. Slika se smatra matricom simbola pri čemu simboli odgovaraju slikovnim objektima, a dvodimenzionalni niz dobivamo projekcijom ovih simbola na vodoravnoj i okomitoj osi.
5. Metoda θR niza, odnosno strukture za predstavljanje prostornih odnosa temeljena na geometriji primjenom algoritma prostorne sličnosti (Gudivada 1998)
6. Graf prostorne orijentacije - prostorni odnosi simbola slikovnih objekata predstavljeni su kao rubovi u težinskog grafu (Gudivada i Raghavan 1995)

7. Prostorni raspored značajnih točaka temeljen na pristupu četverostrukog stabla (eng. *quad-tree*). Slika se rekurzivno dijeli na kvadrante sve dok najmanji kvadranti ne sadrže najviše jednu značajku, a njihovi prostorni odnosi sadržani su u strukturi četverostrukog stabla (Ahmad i Grosky 1997)

Značajke navedene u sklopu logičke reprezentacije slike uključuju najvažnije elemente njenog sadržaja, ali ne mogu u potpunosti obuhvatiti informacijsko bogatstvo slike. Kontekst, odnosno skup pozadinskih informacija potrebnih za tumačenje određene situacije ili događaja, od iznimno je visoke važnosti za razumijevanje svijeta koji nas okružuje. Prepoznavanje objekata sadržanih u slici često ovisi o informacijama izvedenima iz različitih vrsta konteksta koje su djelomično pokrivene prostornim značajkama, a njihova upotreba u klasifikaciji slika zasebno je područje istraživanja (Elezi 2020; Ferrando, Gállego, i Costa-jussà 2022).

2.3 Kontekstualne informacije

Poznavanje tipičnih konfiguracija objekata u nekoj određenoj sceni veoma je važan element ljudske vizualne percepcije, te je iz tog razloga čest predmet istraživanja na području ljudskog i računalnog vida. Na temelju klasifikacije odnosa objekta s njegovom okolinom proizašle iz proučavanja ljudske vizualne percepcije Galleguillos i Belongie (2010) navode tri vrste konteksta na temelju kojih je moguće izdvajanje kontekstualnih značajki slike relevantnih za prepoznavanje njenog sadržaja, a to su semantički, prostorni i kontekst odnosa veličina.

Semantički kontekst temelji se na tipičnoj povezanosti pojave nekog objekta s drugim objektima ili scenama. Primjerice, štednjak se najčešće pojavljuje u kontekstu kuhinje, a ista scena obično sadrži i druge kućanske aparate. Rani sustavi ovakve su informacije dobivali putem unaprijed definiranih pravila proizašlih iz znanja domenskih stručnjaka, te su bili ograničeni na usko specijalizirano područje primjene. Suvremeni pristupi temelje se na kodiranju implicitnih semantičkih međudnosa putem značajki piksela ili matrica zajedničkog pojavljivanja.

Prostorni kontekst proizlazi iz vjerojatnosti pronalaženja nekog objekta u određenom položaju u odnosu na druge objekte unutar scene. Dakle, osim samog zajedničkog pojavljivanja objekata, značajke prostornog konteksta sadrže informacije o njihovom prostornom međudnosu. Budući da je prostorni kontekst neodvojiv od semantičkog, sustavi koji se služe ovim značajkama u određenoj mjeri implicitno koriste informacije semantičkog konteksta.

Informacije prostornog konteksta također mogu biti eksplicitno kodirane ili dobivene implicitno različitim statističkim metodama.

Kontekst odnosa veličina ističe važnost prethodnog znanja o očekivanoj veličini objekata koji mogu biti pronađeni u određenoj sceni. Budući da različite veličine objekata mogu predstavljati problem modelima za njihovo prepoznavanje korisno je znati na koje veličine sustav može suziti potragu. Također se iz ovog konteksta mogu izdvojiti informacije o prostornim međuodnosima, posebice vezanih za dubinu.

Najčešće eksplicitno korištene u sustavima za prepoznavanje objekata su upravo značajke konteksta odnosa veličina i prostornog konteksta, dok su značajke semantičkog konteksta korištene implicitno budući da su već sadržane u drugim dvjema vrstama konteksta.

2.4 Izdvajanje značajki slike

Odabir odgovarajućeg algoritma strojnog učenja svakako je važan korak kod izrade učinkovitog modela za klasifikaciju slika, ali od presudne važnosti odabir je informativnih značajki pomoću kojih će algoritam učinkovito raspoznavati pripadnike različitih klasa. Određivanje semantičkih i meta značajki može biti mukotrpan, ali relativno jednostavan proces. Značajke boje, oblika i tekture najčešće su eksplicitno korištene značajke u klasifikaciji, a njihovo izdvajanje predstavlja poseban izazov za koji su razvijeni brojni algoritmi sa svojim tipičnim prednostima i nedostacima.

Navedene značajke mogu biti globalne i lokalne. Globalne značajke odnose se na čitavu sliku te nam omogućavaju prikaz njene suštine, ali bez dovoljno detalja da bi nam omogućile visoku preciznost prepoznavanja sadržaja. Lokalne značajke pružaju detaljnije informacije jer opisuju pojedine dijelove slike značajne za klasifikaciju (Lisin i ostali 2005).

Na primjerima značajki oblika navedenima u prethodnom poglavlju vidjeli smo kako su neke značajke inherentno lokalne ili globalne, ali raščlanimo li prethodno sliku na sastavne dijelove globalne značajke možemo vrlo lako učiniti lokalnima. Primjerice, histogram boja može se izraditi za čitavu sliku ili samo jedan dio, odnosno može biti globalan ili lokalan (Narwade Manoorkar i Kumar 2016).

2.4.1 Značajke boje

Boje su često prve vizualne značajke koje primjećujemo kao promatrači, a u kontekstu analize i obrade digitalne slike izvor su važnih informacija za obavljanje raznih zadataka kao što

sudohvaćanje, klasifikacija i segmentacija. Kvalitativne i kvantitativne značajke boje moguće je izdvojiti na različite načine koje karakterizira različita razina informativnosti.

Najjednostavniji primjer je histogram boja koji prikazuje distribuciju različitih boja, a njegov osnovni oblik dobiva se jednostavnim prebrojavanjem piksela iste boje unutar matrice slike.

Rezultat je vektor oblika

$$H = \{H[0], H[1], H[2] \dots H[n]\},$$

gdje n označava boju, a $H[n]$ broj piksela određene boje.

Uspoređujemo li slike različitih rezolucija histogram je potrebno normalizirati, odnosno podijeliti sa ukupnim brojem piksela na slici. Ovisno o korištenom prostoru boja, za svaki kanal koristi se poseban histogram (Murala, Gonde, i Maheshwari 2009).

Prednosti histograma boja su jednostavnost, intuitivnost, otpornost na rotaciju, translaciju i promjene odnosa veličina. Glavni nedostatak je što ne uzima u obzir nikakve prostorne informacije nego samo govori u kolikoj mjeri je određena boja prisutna u slici pa vizualno jako različite slike mogu imati gotovo identične histograme. Primjerice, jednak broj piksela zelene boje može biti raspršen po čitavoj slici ili koncentriran na jednom manjem području, stoga je razvijeno nekoliko srodnih metoda koje uključuju prostorne informacije.

(Murala, Gonde, i Maheshwari 2009; Tian 2013).

Za razliku od histograma koji nam prikazuje samo distribuciju pojedinih boja korelgoram je znatno informativniji jer pruža informacije o njihovom prostornom međuodnosu, odnosno za piksele određene boje izračunava se vjerojatnost da se na unaprijed definiranim udaljenostima nalaze pikseli iste boje.

Prednost mu je što sadrži određene prostorne informacije, ali s druge strane zahtijeva značajne računalne resurse te je osjetljiv na rotaciju, promjene odnosa veličina i šum (Huang i ostali 1997).

Vektor koherencije boja nešto je drugačija metoda ekstrakcije prostornih informacija, a temelji se na podjeli piksela jednake boje na koherentne i nekoherentne. Ukoliko je piksel određene boje dio kontinuiranog područja slike iste boje on je koherentan, a ako je izoliran onda je nekoherentan. Dakle, osim broja piksela pojedine boje vektor koherencije boja uključuje informaciju o tome da li je piksel dio veće površine iste boje čime se prevladava jedan od glavnih nedostataka histograma. Izdvajanje ove značajke odvija se u nekoliko koraka. Prvi je zamućenje slike tako da svaki piksel primi prosječnu vrijednost 8 susjednih piksela, a zatim se boje koje se pojavljuju na slici kvantiziraju tako da različite nijanse iste

boje poprima istu vrijednost. Nakon toga možemo pristupiti klasificiranju svakog pojedinog piksela kao koherentnog ili nekoherentnog, ovisno o tome da li je povezan s prethodno definiranim brojem piksela iste boje (primjerice 1% ukupnog broja piksela sadržanih u slici). Broj koherentnih i nekoherentnih piksela za svaku boju sprema se u vektor koherencije boja (Pass, Zabih, i Miller 1997).

Momenti boja statističke su mjere koje proizlaze iz tretiranja distribucije piksela kao vjerojatnosne distribucije pa shodno tome mogu biti karakterizirane jedinstvenim statističkim momentima kao što su srednja vrijednost, standardna devijacija i mjera asimetričnosti (eng. *skewness*) koje se zasebno izračunavaju za svaki kanal prostora boja. Momenti boja znatno su informativniji nego histogram, ali također ne uključuju prostorne informacije (Tian 2013).

2.4.2 Značajke oblika

Metode izdvajanja značajki oblika grubo se mogu podijeliti na one temeljene na konturi i temeljene na regiji. Metode temeljene na konturi fokusiraju se na izračunavanje značajki vezanih za granice među različitim oblicima dok se one temeljene na regiji odnose na značajke čitave regije. U svrhu pretraživanja i pronalaženja slika češće se koriste jednostavni deskriptori značajki oblika kao što su površina, momenti, cirkularnost i ekscentričnost, najčešće u kombinaciji kako bi se postigla veća robusnost. Učinkovita klasifikacija znatno je zahtjevnija zadaća, stoga se u tu svrhu koriste složenije značajke kao što su, primjerice, rubovi ili kutovi.

Detekcija rubova uključuje razne matematičke metode koje imaju za cilj identificirati rubove, odnosno krivulje u digitalnoj slici na kojima se uočavaju diskontinuiteti svjetline, odnosno intenziteta slike. Oštre promjene intenziteta najčešće se događaju na granicama između sjena ili različitih površina, a važne su značajke za prepoznavanje sadržaja zato što najčešće ukazuju na diskontinuitete dubine, orijentacija površine, svojstava materijala ili varijacije osvjetljenja fotografirane scene. Njihovom identifikacijom znatno se olakšava rekonstrukcija trodimenzionalnog svijeta prikazanog u dvije dimenzije (Barrow i Tenenbaum 1981).

Detektirani rubovi mogu biti ovisni ili neovisni o točki gledišta, odnosno mogu se mijenjati ili ostati isti s promjenom perspektive. Neovisni rubovi obično su vezani za inherentna svojstva trodimenzionalnih objekata kao što su bridovi ili plohe. Rubovi koji se mijenjaju promjenom točke gledišta najčešće su vezani za geometriju scene tj. relativne položaje objekata.

Detekcija rubova obično se odvija u četiri koraka. Prvi je zaglađivanje slike kojoj je cilj potisnuti što više šuma bez oštećivanja stvarnih rubova: Drugi je primjena filtera za

izoštavanje slike kojim se poboljšava kvaliteta rubova. Sama detekcija odvija se u četvrtom koraku. Temeljna ideja ovog je koraka određivanje je praga koji služi kao kriterij koje od rubnih piksela treba zadržati, a koje odbaciti kao šum. Posljednji korak je lokalizacija, odnosno određivanje točne lokacije ruba (Cui i ostali 2013).

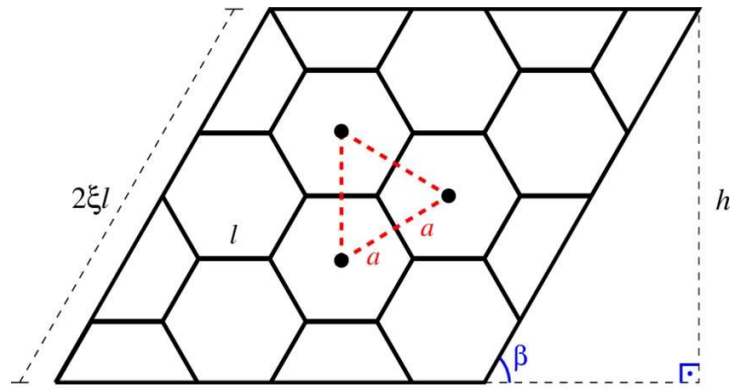
Rubove na digitalnim slikama najčešće detektiramo aproksimacijom parcijalnih derivacija funkcije koja izražava promjene intenziteta slike, primjerice primjenom konvolucijskog algoritma i odgovarajućeg filtra ili kernela. Kernel je matrica manjih dimenzija (primjerice 3x3) čiji su elementi unaprijed izračunati tako da rezultat konvolucije s matricom slike budu istaknuti rubovi slike. Neki od uobičajenih kernela korištenih u ovu svrhu su Prewitt, Sobel i Laplace operator (Reuter i ostali 2009; Cui i ostali 2013; Sobel 2014).

2.4.3 Značajke teksture

Tekstura kao taktilna karakteristika površine na digitalnoj se slici manifestira repetitivnim obrascem vrijednosti boja i svjetline. Analizi teksture možemo pristupiti različitim metodama, kao što su strukturne, statističke, transformacijske i metode temeljene na modelu.

Strukturni pristup podrazumijeva promatranje teksture kao skupa elemenata, tzv. teksela (eng. texel, izvedenica od „texture element“). Slično kao što je reprezentacija digitalne slike mapa piksela od kojih je sastavljena, tekstura se tada predstavlja svojom teksturnom mapom, odnosno mapom teksela. Struktura teksture tako može biti ovisna o veličini elemenata (npr. "gruba ili "fina"), njihovoj orijentaciji te stupnju pravilnosti prostornog rasporeda.

Uobičajena metoda ekstrakcije teksela je Voronojevateselacija, odnosno podjela je površine na diskretne poligonalne elemente koji se međusobno dodiruju. Elementi se definiraju kao skupovi točaka koje su najbliže pridruženom prototipnom vektoru, a granice između elemenata kao okomice na dužine koje spajaju susjedne vektore (Shapiro i Stockman 2001; Armi i Fekri-Ershad 2019). Primjer teselacije prikazan je na Slici 3.



Slika 3 Primjer teselacije (Albuquerque i ostali 2021, 21)

Strukturni pristup primjeren je za umjetne teksture zbog njihove relativne pravilnosti, ali prirodne teksture obično pokazuju višu razinu nasumičnosti pa je u njihovom slučaju ekstrakcija teksela znatno otežana.

Statističke metode pristupaju problemu pomoću različitih kvantitativnih mjera rasporeda vrijednosti intenziteta na određenom području, zahtijevaju znatno manje računalnih resursa te su podjednako primjenjive neovisno o vrsti teksture. Jedna od popularnijih statističkih metoda je metoda matrice istodobnog pojavljivanja razine sive (GLCM), koja uključuje stvaranje matrice učestalosti pojavljivanja različitih parova razina sive na slici. GLCM se zatim može koristiti za izračun različitih statističkih mjera povezanih s karakteristikama teksture, kao što su kontrast, energija i homogenost. Također često korištena statistička metoda je metoda matrice duljine niza razine sive (GLRLM), koja uključuje stvaranje matrice učestalosti pojavljivanja različitih duljina niza piksela s danom vrijednošću razine sive boje. GLRLM se zatim može koristiti za izračun različitih statističkih mjera, kao što su neujednačenost ili postotak duljine niza (Albregtsen, Nielsen, i Danielsen 2000).

Metode temeljene na transformaciji proizlaze iz ideje da se tekstura slike može opisati frekvencijskim sadržajem njezinih piksela. Tipičan primjer je metoda Fourierove transformacije, koja uključuje transformaciju slike iz prostorne u frekvencijsku domenu pomoću Fourierove transformacije. Frekvencijske komponente slike tada se mogu koristiti kao značajke teksture. Druga popularna metoda koja se temelji na transformaciji je metoda valične transformacije (eng. wavelettransform), koja na sličan način uključuje rastavljanje slike na različite frekvencijske podpojase pomoću valičke transformacije. Koeficijenti podpojasa tada se mogu koristiti kao karakteristike teksture (Wirsing 2020).

Izdvajanje značajki ključan je korak za učinkovitu klasifikaciju slika jer one moraju biti dovoljno informativne da omoguće razlikovanje pripadnika različitih klasa, a istovremeno dovoljno općenite da se mogu primijeniti na prethodno neviđene slike. Kvalitetne značajke također bi trebale biti otporne na varijacije proizašle iz razlika u osvjetljenju, kutu snimanja i ostalim faktorima koji predstavljaju posebne izazove za različite zadatke vezane za područje računalnog vida, pa tako i klasifikaciju.

2.5 Ostali izazovi vezani za klasifikaciju slika

Odabir odgovarajućih značajki u velikoj mjeri ovisi o sadržaju slika jer, primjerice, neki se objekti jasnije razlikuju svojim oblikom, dok je za razlikovanje drugih puno važnija boja. Iste značajke također mogu proizvesti nešto drugačije rezultate klasifikacije ovisno o odabranom klasifikacijskom algoritmu. Međutim, optimalan odabir značajki i metode strojnog učenja nisu jedini izazovi kada govorimo o klasifikaciji slika. Značajni problemi mogu proizaći iz samih karakteristika skupova podataka, a mogu se podijeliti u dvije osnovne skupine:

- visoka unutarklasna varijanca (visoka različitost predstavnika iste klase)
- niska međuklasna varijanca (velika sličnost predstavnika različitih klasa)

Izražene razlike unutar klasa mogu biti uzrokovane inherentnim osobinama objekata, odnosno njihovim različitim pojavnim oblicima. Primjerice, ribe mogu biti različitih boja i oblika pa tako klasa "ribe" može sadržavati slike veoma različitih značajki. Međutim, većina problema proizlazi iz samog procesa snimanja, a u literaturi se navodi njih ukupno šest: varijacija točke gledišta, razlike u osvjetljenju, zaklonjenost, varijacije veličine, deformacije i pozadinska pretrpanost (eng. *background clutter*).

Varijacija točke gledišta proizlazi iz različitih kutova snimanja koji, zbog jednostavne činjenice da su fotografije dvodimenzionalni prikazi trodimenzionalnih predmeta iz stvarnog svijeta, mogu proizvesti vrlo različite snimke istog predmeta.

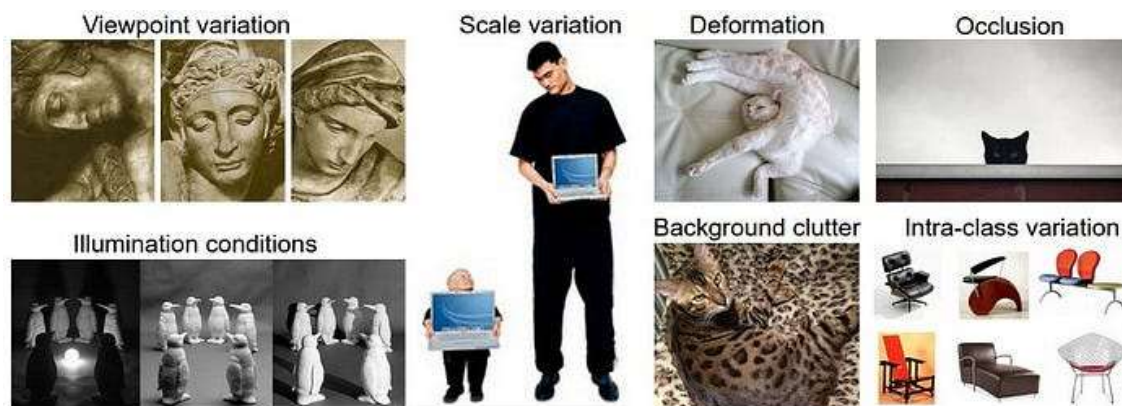
Razlike u osvjetljenju također mogu izazvati značajne probleme sustavima za prepoznavanje slika. Primjerice, značajke oblika temeljene na detekciji rubova prepoznaju iste kao nagle promjene vrijednosti intenziteta, što znači da osvjetljenost istog objekata iz različitih kutova može rezultirati različitim značajkama.

Zaklonjenost objekata na slici čest je problem jer predmeti koje tražimo obično su dio kompleksnih scena iz stvarnog života pa često nisu prikazani u cijelosti, odnosno djelomično su zaklonjeni ostalim entitetima unutar slike.

Varijacija odnosa veličina proizlazi iz činjenice da veličina objekata može znatno varirati, ne samo zbog njihovih inherentnih karakteristika poput stvarne veličine nego različitih položaja u prostoru, odnosno dubini koja se u dvodimenzionalnom prikazu manifestira kao razlika u veličini prikazanih predmeta.

Deformacija objekata na slici posebno je izražena kada se radi o objektima promjenjivog oblika. Primjerice, slika čovjeka ili životinje u trku može se značajno razlikovati ovisno o tome koji položaj tijela je prikazan na pojedinoj slici.

Pozadinska pretrpanost (eng. *background clutter*) odnosi se na kompleksne scene koje sadrže mnoštvo zbijenih entiteta, čime se znatno otežava izdvajanje objekta od interesa. (G. Wang 2011; L. Liu i ostali 2020). Tipični primjeri opisanih izazova prikazani su na Slici 4.



Slika 4 Tipični primjeri izazova vezanih za klasifikaciju slika (Stanford University bez dat.)

Navedenim problemima možemo pristupiti na razne načine, primjerice jednostavnim povećanjem količine relevantnih podataka u trening skupu ili razbijanjem klase na podklase pomoću klasteringa (Bai i ostali 2017; Pilarczyk i Skarbek 2019; Carbonnelle i Vleeschouwer 2022).

3 Strojno učenje

Računalo možemo eksplicitno programirati isključivo za rješavanje algoritamskih problema, odnosno izvršavanje onih zadataka koje se mogu svesti na konačan niz strogo definiranih instrukcija. Problemi koje susrećemo u stvarnom životu nisu uvijek algoritamski, ali to ne znači da se ne mogu rješavati uz pomoć računala. Strojno učenje stoga se često neformalno definira poznatom parafrazom Arthura Samuela kao učenje računala da rješava određene probleme bez da ga se eksplicitno programira.

Alpaydin (2020) nudi nešto strožu definiciju strojnog učenja kao programiranja računala za optimizaciju odabranog kriterija učinka korištenjem oglednih podataka ili prijašnjeg iskustva. Prema Mitchellu (1997) strojno učenje je programiranje računala na način da iz iskustva E uči rješavati problem T pod uvjetom da učinkovitost P raste sa iskustvom E .

Algoritmi strojnog učenja služe se različitim metodama, poput onih iz domene računarske statistike ili matematičke optimizacije, a rezultat njihove primjene na zadanom skupu podataka je model koji može biti prediktivan ili deskriptivan, odnosno može se upotrijebiti u svrhu predviđanja vezanih za instance podataka s kojima se još nije susreo ili ekstrakciju korisnog znanja iz skupa podataka pomoću kojeg je izgrađen. Prediktivne modele dobivamo metodama nadziranog učenja koje se provodi na način da odgovarajući domenski stručnjaci djeluju kao nadglednici koji korigiraju preslikavanje ulaza na odgovarajuće izlazne vrijednosti. Deskriptivni modeli opisuju podatke pronalaženjem pravilnosti ili obrazaca te ih pomažu pretvoriti u oblik čitljiv ljudskim subjektima ili prikladan za daljnju obradu, a u pravilu se dobivaju metodama nenadziranog strojnog učenja. Tipični primjeri nadziranog strojnog učenja su klasifikacija i regresija, a nenadziranog grupiranje (eng. *clustering*).

Strojno učenje tradicionalno se temelji na relativno jednostavnim modelima kao što su stabla odlučivanja, KNN ili metoda potpornih vektora, ali u današnje vrijeme koje karakterizira izobilje različitih vrsta nestrukturiranih podataka ključnu ulogu preuzimaju metode dubokog učenja, odnosno kompleksni modeli temeljeni na umjetnim neuronskim mrežama s velikim brojem slojeva. Prednosti dubokog učenja posebno su izražene na područjima vezanima za klasifikaciju slika, jer za razliku od tradicionalnih modela ne zahtijevaju zasebno izdvajanje značajki, omogućavaju rad s velikim količinama podataka, te su primjereniji za kompleksne zadatke kao što je obrada podataka s velikim udjelom šuma (Sharma i Kumar 2017; Mahony i ostali 2020).

3.1 Najvažniji tipovi zadataka u strojnom učenju

Strojno učenje možemo koristiti za rješavanje različitih problema s mnoštvom algoritama od kojih svaki može biti primjenjiv na jednu ili više zadataka. Neke od temeljnih i često korištenih zadataka strojnog učenja su klasifikacija, regresija i grupiranje (*clustering*).

3.1.1 Klasifikacija

Najčešće korištena metoda nadziranog učenja svakako je klasifikacija, čija je zadaća temeljem analize dovoljne količine podataka koje su domenski stručnjaci podijelili na dvije ili više klasa izvesti aproksimaciju funkcije koja klasificira instance podataka s kojima se nije prethodno susrela. Primjerice, primjenom algoritma strojnog učenja na skup fotografija prethodno klasificiranih prema sadržaju možemo izvesti funkciju koja s određenim stupnjem vjerojatnosti klasificira slike, odnosno prepoznaje njihov sadržaj. Ovisno o odabranom algoritmu, izlaz klasifikacijske funkcije može biti numerička vrijednost koja samo označava pripadnost određenoj klasi ili procijenjenu vjerojatnost pripadnosti nekoj klasi (Goodfellow, Bengio, i Courville 2016).

Klasifikacija slika fundamentalan je problem za sve zadatke koje možemo svesti na prepoznavanje ili uspoređivanje njihovog sadržaja, kao što su računalni vid, detekcija objekata, automatska anotacija slika ili pretraživanje i dohvaćanje, stoga predstavlja posebno područje interesa kojim ćemo se detaljnije baviti u jednoj od sljedećih cjelina.

3.1.2 Regresija

Regresijom nazivamo skup statističkih metoda kojima određujemo međuodnos zavisne (izlazne) varijable s jednom ili više nezavisnih (ulaznih) varijabli. Dijeli određene sličnosti s klasifikacijom, ali izlaz regresijske funkcije ne označava pripadnost klasi nego predviđa vrijednost neke varijable koja ovisi o ulaznim podacima, odnosno njena krivulja ne dijeli različite klase nego optimalno odgovara ulaznim podatkovnim točkama.

Najjednostavniji primjer je linearna regresija koja aproksimira linearnu funkciju ili, puno češće, složeniju linearnu kombinaciju koja najbliže odgovara podatkovnim točkama ovisno o određenom matematičkom kriteriju (Alpaydin 2020).

3.1.3 Grupiranje (eng. *clustering*)

Grupiranje ili *clustering* primjer je nenadziranog strojnog učenja koji, za razliku od klasifikacije gdje su klase unaprijed definirane, samostalno grupira instance podataka u grupe (klastere) sukladno sličnosti njihovih atributa. Sličnost koja određuje pojedine grupe ne može se jednoznačno definirati, stoga postoje brojni algoritmi grupiranja koji problemu pristupaju na različite načine (Estivill-Castro 2002).

3.2 Osnovni problemi kod strojnog učenja

Usprkos brojnim prednostima, primjena metoda strojnog učenja nije bez svojih izazova od kojih jedan dio svakako proizlazi iz same prirode ovih metoda, ali značajan udio problema vezan je za osobine skupova podataka prikupljenih iz stvarnog svijeta. Smetnje u podacima, vrijednosti koje nedostaju, prevelik broj dimenzija, irelevantni i redundantni atributi, mješovite vrste atributa, neuravnoteženi skupovi podataka, pretjerana ili nedovoljna ugođenost modela (*overfitting* i *underfitting*), podešavanje hiperparametara, interpretabilnost modela, odabir modela, te njihova vremenska i prostorna složenost samo su neki od najvažnijih izazova s kojima se susrećemo u praksi.

Smetnje u podacima proizlaze iz različitih čimbenika koji utječu na proces prikupljanja na način da povećavaju udio šuma u podacima. Smetnje mogu dolaziti iz različitih izvora, kao što su nesavršenosti instrumenata za prikupljanje, ljudska pogreška, različite interferencije signala i drugi nepovoljni utjecaji. Na sličan način javlja se problem vrijednosti koje nedostaju, odnosno onih vrijednosti u skupu podataka koje nisu unesene zbog različitih razloga, poput ljudske pogreške ili prekida u radu instrumenata. Osim činjenice da je većina algoritama strojnog učenja dizajnirana za rad s potpunim skupovima podataka, vrijednosti koje nedostaju mogu uzrokovati slične probleme kao pogrešne vrijednosti. Postoje brojne tehnike za rješavanje ovog problema, poput odbacivanja instanci s nepotpunim vrijednostima ili popunjavanja nekim oblikom srednje vrijednosti, ali one su uvijek u određenoj mjeri problematične jer mogu značajno utjecati na performanse i pristranost konačnog modela (Elhassan i ostali 2022).

Prevelik broj dimenzija, također poznat pod nazivom kletva dimenzionalnosti (eng. *curse of dimensionality*) događa se kada broj dimenzija (atributa) znatno nadmašuje broj uzoraka. Budući da treniranje modela koji u zadovoljavajućoj mjeri generalizira zahtijeva da trenažni skup sadrži više uzoraka sa svakom kombinacijom vrijednosti, s porastom broja dimenzija

dramatično raste potreba za količinom podataka potrebnih za treniranje kvalitetnog modela. Navedeni problem moguće je ublažiti različitim tehnikama redukcije dimenzionalnosti ili odabira atributa, a posebno je izražen kod tradicionalnih metoda strojnog učenja dok se duboke neuronske mreže u pravilu znatno bolje nose s visokodimenzionalnim podacima (Poggio i ostali 2017; Dhal i Azad 2022).

Pitanje nevažnih ili redundantnih atributa povezano je s problemom kletve dimenzionalnosti, a odnosi se na attribute koji ne pridonose kvaliteti modela u dovoljnoj mjeri da bi njihova prisutnost u skupu podataka bila opravdana, stoga razmatranje takvih značajki nije korisno i obično dovodi do slabijih performansi modela. Uklanjanjem irelevantnih i suvišnih značajki smanjuje se dimenzionalnost podataka i kompleksnost modela, te ubrzava proces njegova treniranja (Al-Tashi i ostali 2020).

Rad s miješanim vrstama atributa još je jedan izazov u strojnom učenju koji proizlazi iz osobine ulaznih podataka. Naime, skupovi podataka mogu sadržavati mješavinu kategoričkih, numeričkih ili tekstualnih značajki, koje zahtijevaju različite tehnike preprocesiranja. Primjerice, kategoričke attribute moguće je kodirati u njihove numeričke reprezentacije te na taj način zaobići problem neujednačenosti tipova varijabli (Hsu i ostali 2021).

Nadalje, čest problem za brojna područja praktične primjene predstavljaju neuravnoteženi skupovi podataka u kojima se broj uzoraka u pojedinoj klasi može značajno razlikovati. Tipičan primjer je medicinska dijagnostika kod koje je često potrebno detektirati relativno rijetke patološke promjene među velikim brojem zdravih primjeraka. Modeli trenirani na takvim skupovima podataka mogu biti pristrani prema većinskoj klasi, stoga je izrazito važno posvetiti pažnju ovom problemu. Uobičajene tehnike njegovog rješavanja uključuju odabir odgovarajućih mjera uspješnosti koje uzimaju u obzir neuravnoteženost klasa, ponderiranje uzoraka prema zastupljenosti klase kojoj pripadaju, te različite metode manipulacije brojem uzoraka u pojedinoj klasi kao što je uklanjanje viška uzoraka iz većinskih klasa ili povećanje broja uzoraka manjinskih klasa njihovom replikacijom, generiranjem umjetnih primjeraka ili nekom drugom metodom (Pereira i Saraiva 2020).

Prekomjerna ugođenost modela (eng. *overfitting*) događa se u slučajevima kada se model u tolikoj mjeri prilagodi trenažnom skupu podataka da nauči i njegove nevažne elemente poput šuma, pa rezultira visokim performansama na istom skupu, ali slabom generalizacijom na podacima s kojima se prethodno nije susreo. Suprotan problem predstavlja nedovoljna ugođenost modela (eng. *underfitting*) koja se događa kada model u nedovoljnoj mjeri obuhvati temeljne obrasce trening skupa podataka pa rezultira znatno nižim performansama

nego što bi ih proizveo optimalno trenirani model (Montesinos López, Montesinos López, i Crossa 2022, 109).

Podešavanje hiperparametara modela, odnosno onih parametara koje model ne uči iz podataka nego ih podešava korisnik, također može predstavljati značajan izazov, posebice kada se radi o modelima s velikim brojem hiperparametara kao što su duboke neuronske mreže. Razvijene su brojne metode kojima je moguće olakšati i unaprijediti njihovo podešavanje, u rasponu od različitih načina pretraživanja raspona mogućih vrijednosti do automatizacije ovog procesa primjenom algoritama strojnog učenja (Gencoglu i ostali 2019).

Važno pitanje kada govorimo o primjeni metoda strojnog učenja predstavlja i interpretabilnost modela. Naime, iako postoje algoritmi koji proizvode modele za koje je proces kojim dolaze do odluke relativno lako objašnjiv, poput različitih stabala odlučivanja, značajan dio modela funkcionira na principu „crne kutije”, odnosno nije moguće jednostavno ili u zadovoljavajućoj mjeri objasniti način na koji dolaze do konačne odluke. Interpretabilnost modela važan je čimbenik povjerenja korisnika jer ukoliko nije moguće na jasan način prikazati korake koji dovode do konačnog rezultata njihove primjene, povjerenje ljudi može biti narušeno. U slučajevima kada su od presudne važnosti samo visoke performanse modela nedostatak interpretabilnosti nije kritičan, ali može biti izrazito problematičan za područja primjene u kojima odluke mogu imati nesagledive posljedice, poput medicine, financija ili pravosuđa. Interpretabilnost modela također može biti važna u svrhu njihovog unapređenja, stoga su razvijene tehnike pomoću kojih u određenoj mjeri možemo interpretirati i one modele koji nisu inherentno interpretabilni, poput različitih metoda vizualizacije ili atribucije (Dunn, Mingardi, i Zhuo 2021; Schlegel i Keim 2021).

Budući da je na određenom skupu podataka moguće razviti beskonačan broj modela koji u različitoj mjeri zadovoljavaju naše zahtjeve za performansama, kritično pitanje kod primjene metoda strojnog učenja upravo je odabir odgovarajućeg modela. Mnogi su pristupi ovom problemu, a kada govorimo o odabiru prediktivnih modela najčešće korištena u praksi je neka varijanta unakrsne validacije. Osnovni koncept je vrlo jednostavan, a temelji se na nasumičnoj podjeli skupa podataka na skup za trening modela i skup za validaciju na kojem testiramo njegove performanse. Proces se ponavlja nekoliko puta na različito razdijeljenim skupovima, a kao konačan model se uzima onaj sa najboljom prosječnom mjerom uspješnosti (Ding, Tarokh, i Yang 2018).

Posljednji, ali ne najmanje važan problem na koji ćemo se osvrnuti pitanje je vremenske i prostorne kompleksnosti, odnosno vremena i memorijskih resursa koje određeni algoritam zahtijeva za svoje izvršavanje. Navedeni problem posebno je izražen kada se radi o velikim

količinama podataka i ograničenim računalnim resursima, stoga je optimizacija u svrhu smanjenja vremenske i prostorne kompleksnosti jedno od ključnih pitanja za brojna područja praktične primjene metoda strojnog i dubokog učenja (Montaha i ostali 2022).

3.3 Duboko učenje

Pojam dubokog učenja kao podvrste strojnog učenja odnosi se na primjenu različitih oblika dubokih neuronskih mreža, odnosno umjetnih neuronskih mreža velikog broja slojeva. Duboko učenje donijelo je značajan napredak ovog područja u proteklih dvadesetak godina, ponajviše zbog porasta dostupne količine podataka i računalnih resursa koji predstavljaju ključan faktor njihove uspješne primjene. Međutim, povijest neuronskih mreža započela je već sredinom dvadesetog stoljeća, tijekom kojeg je doživjela nekoliko uspona i padova. Detaljnije razumijevanje dubokih neuronskih mreža teško je razdvojiti od povijesti njihovog razvoja, stoga ćemo na sljedećih nekoliko stranica obuhvatiti neke od ključnih koncepata koji su prethodili njihovoj uspješnoj praktičnoj primjeni.

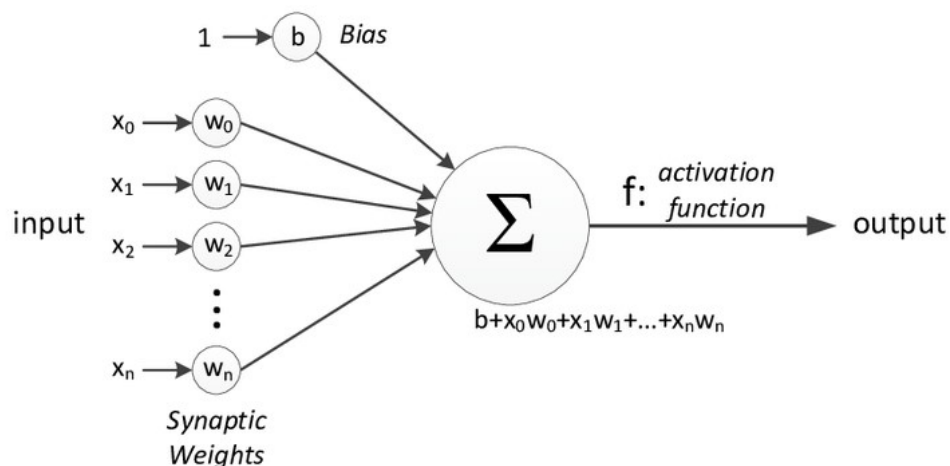
3.3.1 Umjetne neuronske mreže

Umjetne neuronske mreže (eng. *Artificial Neural Networks* - ANN) su strukture u određenoj mjeri nadahnute biološkim neuronskim mrežama, a mogu se opisati kao usmjereni težinski grafovi čiji su čvorovi organizirani u više međusobno povezanih slojeva. Čvorove (neurone) ovakve mreže čine matematičke funkcije koje, zbog određene sličnosti s aktivacijom bioloških neurona, nazivamo aktivacijskim funkcijama. Sinapse kojima su povezani ovakvi umjetni neuroni čine usmjerene težinske veze čija težinska vrijednost označava njihovu snagu.

Treniranje neuronske mreže vrši se manipulacijom težinskih vrijednosti čiji je konačni cilj mapiranje ulaznih vrijednosti na odgovarajuće vrijednosti izlaznog sloja. Primjerice, kada govorimo o problemu klasifikacije izlazna vrijednost je oznaka klase, odnosno vjerojatnosna vrijednost pripadnosti određenoj klasi (Mitchell 1997; Goodfellow, Bengio, i Courville 2016). Temeljne koncepte na kojima počiva treniranje neuronskih mreža najlakše je prikazati na primjeru, počevši od njihovog najjednostavnijeg oblika.

3.3.2 Perceptron

Rani oblik jednostavne umjetne neuronske mreže bio je perceptron kojeg su predstavili McCulloch i Pitts, a njegovu prvu funkcionalnu implementaciju izradio je Frank Rosenblatt 1958. u aeronautičkom laboratoriju Cornell. Perceptron je jednostavan model neuronske mreže koja se sastoji od jednog sloja umjetnih neurona povezanih s jednim izlaznim neuronom putem sinapsi s pridruženim težinskim vrijednostima koje određuju snagu veze (Slika 5). Ulazni neuroni primaju ulazne podatke i vrše izračun, te ga proslijeđuju izlaznom neuronu koji izvodi konačno predviđanje. Izračun koji izvodi perceptron linearna je kombinacija vektora ulaza i težinskih vrijednosti, koja se proslijeđuje funkciji praga, pomoću koje se odlučuje hoće li se izlazni neuron aktivirati ili ne. Najčešće korištena funkcija praga je funkcija koraka, koja je definirana kao $f(x) = 1$ ako je $x > 0$ i $f(x) = 0$ ako je $x \leq 0$.



Slika 5 Perceptron (Fountas 2023, 12)

Prilagođavanje težinskih vrijednosti tipično se obavlja nekom varijantom gradijentnog spusta, optimizacijskog algoritma za pronalaženje minimuma funkcije s više varijabli. Minimum ciljne funkcije pronalazi se postepenim ažuriranjem njezinih parametara u negativnom smjeru derivacije te funkcije. Funkcija koju u ovom slučaju optimiziramo je funkcija gubitka, poznata još kao funkcija troška ili pogreške, koja mjeri koliko loše mreža vrši predviđanja na ulaznom skupu podataka. Parametri ove funkcije su težine povezane s njenim ulazima.

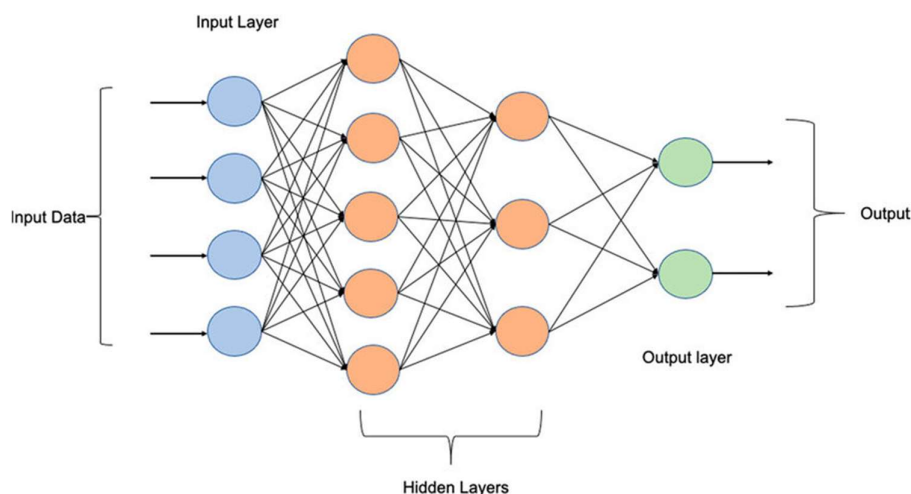
Najčešće korištene funkcije gubitka za perceptrone su srednja kvadratna pogreška (eng. *mean squared error*) i unakrsna entropija (eng. *cross entropy*).

Neke od prednosti perceptrona su relativno jednostavna implementacija i razumijevanje, te brzo treniranje zbog visoke računalne učinkovitosti. Glavno ograničenje je što nisu u

možnosti klasificirati podatke koji nisu linearno odvojivi, odnosno ne mogu naučiti složene, nelinearne odnose. Također se mogu koristiti samo za zadatke binarne klasifikacije, što znači da nisu prikladni za slučajeve kada izlaz može poprimiti višestruke vrijednosti kao što su klasifikacija s više klasa ili regresija. Navedene nedostatke možemo nadvladati dodavanjem tzv. skrivenih slojeva između ulaza i izlaza, čime dobivamo arhitekturu neuronske mreže poznatu pod nazivom višeslojni perceptron (Alpaydin 2020, 275).

3.3.3 Višeslojni perceptron (MLP)

Arhitektura višeslojnog perceptrona temelji se na klasičnom perceptronu, ali se za razliku od njega sastoji od ulaznog, izlaznog te jednog ili više skrivenih slojeva neurona. Pojam dubokog učenja vezuje se upravo za neuronske mreže, a proizlazi iz koncepta njihove dubine, odnosno broja skrivenih slojeva. Jednostavan primjer višeslojnog perceptrona prikazan je na Slici 6.

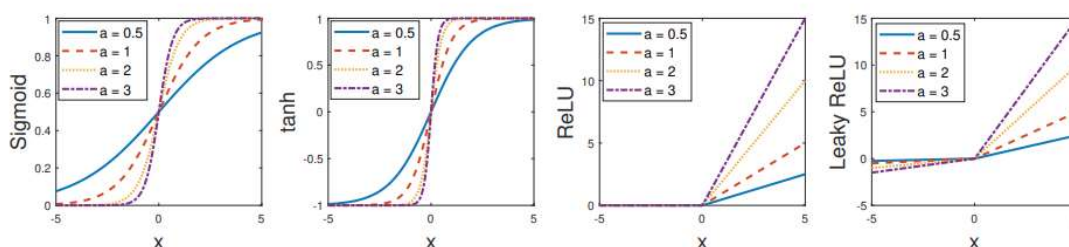


Slika 6 Jednostavni višeslojni perceptron sa dva skrivena sloja neurona (Afan i ostali 2021, 1424)

Ulazni sloj višeslojnog perceptrona prima ulazne podatke koje zatim obrađuju neuroni skrivenih slojeva pri čemu izlaz prethodnog sloja postaje ulaz sljedećeg. Konačno predviđanje generira se u izlaznom sloju koji može imati različit broj neurona ovisno o specifičnostima zadatka, kao što je broj mogućih klasa u korištenom skupu podataka.

Jedna od ključnih prednosti višeslojnog perceptrona mogućnost je rješavanja složenih nelinearnih problema. Osim povećane kompleksnosti proizašle iz većeg broja slojeva umjetnih neurona, ovoj sposobnosti doprinosi i korištenje nelinearnih aktivacijskih funkcija

koje omogućuju mreži da nauči složene, nelinearne odnose između ulaznih i izlaznih podataka. Ponajviše zbog svoje jednostavnosti i učinkovitosti, danas najčešće korištena aktivacijska funkcija je tzv. zglobnica ili funkcija ispravljene linearne jedinice (ReLU), definirana kao $f(x) = \max(0, x)$. Također se u istu svrhu mogu koristiti druge aktivacijske funkcije kao što su sigmoidna, tanh ili leakyReLU funkcija (Slika 7).



Slika 7 Sigmoidna, tanh, ReLU i LeakyReLU funkcija (Jagtap i Karniadakis 2019, 4)

Treniranje višeslojnog perceptrona odvija se kroz dvije glavne faze – propagacija podataka prema naprijed i propagacija pogreške u natrag. Propagacija prema naprijed odnosi se na prolazak podataka od ulaznog prema izlaznom sloju. Propagacija pogreške u natrag je algoritam pomoću kojeg prilagođavamo njene težinske vrijednosti sukladno veličini odstupanja dobivenog od stvarnog rezultata. Odstupanje mjerimo pomoću funkcije gubitka kojoj se prosljeđuje konačno predviđanje. Dakle, neuronska mreža za vrijeme treniranja počinje ulaznim slojem, a završava funkcijom gubitka. Matematička osnova ovog algoritma temelji se na računanju parcijalne derivacije funkcije gubitka s obzirom na svaku težinu u mreži, čime se dobivaju informacije o tome kako prilagoditi težinu da bi se pogreška svela na minimum. Budući da je matematička funkcija čitave mreže zapravo kompozicija funkcija svakog pojedinog sloja u kojoj posljednju funkciju čini funkcija gubitka, algoritam na ovaj način radi unatrag kroz mrežu koristeći se pravilom ulančavanja - kreće od izračunavanja gradijenta s obzirom na težine izlaznog sloja, zatim koristi te informacije za izračunavanje gradijenta pogreške s obzirom na težine prethodnog sloja, i tako dalje, sve dok cijela mreža ne bude obrađena (Goodfellow, Bengio, i Courville 2016, 163).

Višeslojni perceptron primjenjiv je na vrlo širok raspon problema, ali može biti izrazito računalno zahtjevan te sklon *overfittingu*, posebice ako se radi o arhitekturi velikog broja slojeva (dubine) ili neurona u pojedinim slojevima (širine). *Overfitting* se događa kada kapacitet modela znatno nadmašuje kompleksnost zadaće. Promatramo li model kao matematičku funkciju koja se mijenja procesom njegova treniranja, kapacitet modela

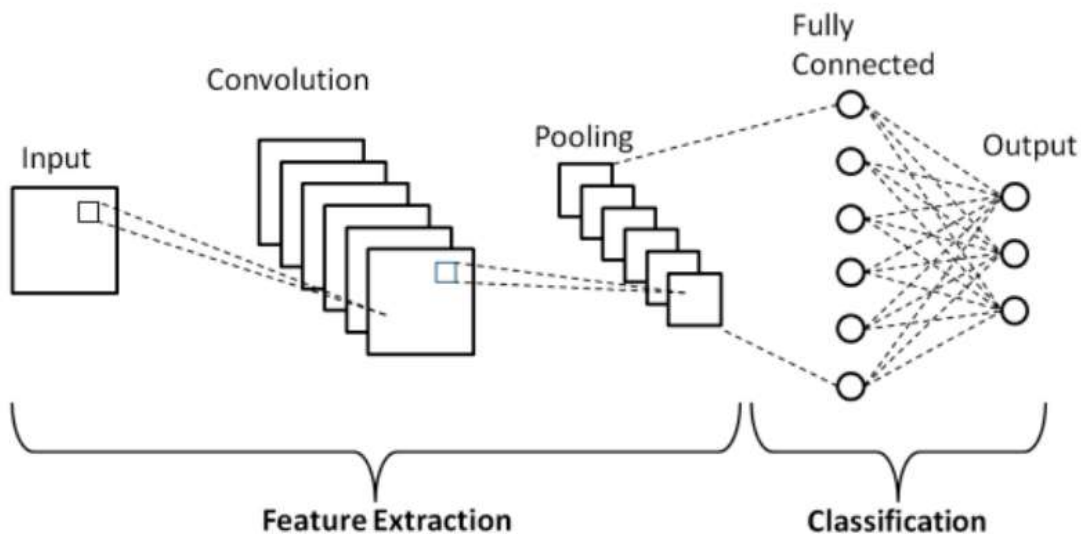
definiramo kao skup mogućih funkcija čiji oblik model može poprimiti. Kada je ovaj skup prevelik naš model može, osim korisnih atributa ulaznih podataka, uzeti u obzir i šum te na taj način ostvarivati jako dobra predviđanja na trening skupu podataka, ali znatno lošije generalizirati na podatke s kojima se prethodno nije susreo, stoga se u svrhu detekcije ovog problema treniranje klasifikacijskih modela tipično provodi korištenjem većeg skupa za treniranje i manjeg za testiranje. U svrhu izbjegavanja overfittinga primjenjuju se različite metode regularizacije kao što su neki oblici ograničavanja težinskih vrijednosti ili tzv. rano zaustavljanje, odnosno prekidanje procesa treniranja prije nego nastupi *overfitting* (Goodfellow, Bengio, i Courville 2016; Alpaydin 2020).

Višeslojni perceptroni sa više skrivenih slojeva načelno pripadaju metodama dubokog učenja, ali kada govorimo o klasifikaciji slika dijele nedostatak karakterističan za klasične metode strojnog učenja, a to je potreba za često izrazito računalno zahtjevnim izdvajanjem značajki ulaznog skupa podataka čija učinkovitost znatno ovisi o korištenju odgovarajućih metoda, odnosno vještini osobe koja vrši odabir značajki. Nadilaženje ovog nedostatka omogućila je nešto složenija arhitektura konvolucijskih neuronskih mreža razvijena upravo u svrhu obrade slikovnih podataka.

3.3.4 Konvolucijske neuronske mreže

Uobičajena metoda izdvajanja značajki slike prije primjene klasičnih metoda strojnog učenja upotreba je raznih vrsta konvolucijskih filtera, a ideja konvolucijskih neuronskih mreža objedinjuje ove korake u jedinstvenu arhitekturu.

Primjer osnovne struktura CNN-a sastoji se od ulaznog sloja, jednog ili više konvolucijskih slojeva, jednog ili više slojeva sažimanja i jednog ili više potpuno povezanih slojeva (Slika 8).



Slika 8 Jednostavna konvolucijska neuronska mreža (Phung i Rhee 2019, 3)

Ulazni sloj prima ulaznu sliku koju potom obrađuju konvolucijski slojevi. Konvolucijski slojevi koriste filtere za izdvajanje značajki kao što su oblici i teksture. Dobivene značajke se zatim kombiniraju u mapu značajki koja se propušta kroz slojeve sažimanja. Prednost mape značajki naspram vektora dobivenih tradicionalnim metodama leži u tome da ona zadržava prostornu hijerarhiju čime se omogućuje mreži da nauči lokalne i globalne značajke ulazne slike te u određenoj mjeri implicitno iskoristi kontekstualne informacije sadržane u slici. Dodatna prednost konvolucijskih slojeva leži u tome što za razliku od potpuno povezanih slojeva sadrže znatno manje veza među neuronima susjednih slojeva, čime se značajno smanjuje kompleksnost, a ujedno i računalna zahtjevnost čitave mreže (Goodfellow, Bengio, i Courville 2016; Islam, Jia, i Bruce 2020).

Slojevi sažimanja ubrzavaju proces treniranja na način da dodatno smanjuju dimenzije skrivenog sloja kombiniranjem izlaza skupine neurona na prethodnom sloju (tzv. prozora sažimanja) u jedan neuron u sljedećem sloju. Primjerice, najčešće korištena metoda sažimanja maksimalne vrijednosti (eng. *max pooling*) odabire najveću vrijednost iz prozora sažimanja koju prenosi u neuron sljedećeg sloja zanemarujući pritom ostale vrijednosti, čime se uklanjaju redundantne informacije ulaznih značajki. Osim što smanjuje složenost neuronske mreže, opisana metoda povećava robustnost modela u odnosu na translacije objekata na slikama (Zeiler i Fergus 2013).

Nakon prolaska kroz slojeve sažimanja višedimenzionalna mapa značajki se vektorizira te propušta kroz jedan ili više potpuno povezanih slojeva koji izvode konačno predviđanje.

Potpuno povezani slojevi slični su onima u tradicionalnom višeslojnom perceptronu (MLP), te obično koriste nelinearne aktivacijske funkcije kao što je funkcija ispravljene linearne jedinice (ReLU). Aktivacijska funkcija odabire se sukladno prirodi zadaće, a posebnost ReLU leži u tome da negativne vrijednosti pretvara u nulu pa na određeni način oponaša aktivaciju bioloških neurona koji poznaju samo dva stanja, odnosno mogu se aktivirati ili ne. Samo treniranje mreže također se temelji na propagaciji pogreške u natrag optimizacijom funkcije gubitka pomoću neke varijante stohastičkog gradijentnog spusta.

Važno je skrenuti pažnju na činjenicu da proces treniranja konvolucijskih neuronskih mreža ne uključuje samo ugađanje težinskih vrijednosti veza u potpuno povezanim slojevima nego se njime prilagođavaju i vrijednosti konvolucijskih filtera. Na ovaj se način mogu postići značajno bolji rezultati jer, umjesto ljudskog subjekta koji ručno odabire predefinirane konvolucijske filtere, algoritam generira odgovarajuće vrijednosti filtera procesom učenja iz ulaznih podataka. Također je moguće „prenijeti znanje” CNN-a, odnosno mape značajki prethodno izrađene treniranjem na velikim skupovima podataka primijeniti kao sredstvo izdvajanja značajki manjih skupova, čime se omogućava postizanje znatno boljih rezultata nego kada bismo istu mrežu od temelja trenirali isključivo na malom skupu podataka. Važna prednost konvolucijskih neuronskih mreža leži u mogućnostima paralelizacije te iskorištavanja moćnih grafičkih procesorskih jedinica, čime se omogućava relativno brzo treniranje na vrlo velikim skupovima podataka, a veličina trenažnog skupa često može biti važnija od odabira odgovarajućeg algoritma strojnog učenja (Banko i Brill 2001; Krizhevsky 2014).

Konvolucijske neuronske mreže svoj glavni nedostatak dijele s ostalim oblicima neuronskih mreža, a to je sklonost *overfittingu* koja raste s njenom dubinom, odnosno širinom pa je najčešće nužno primijeniti neku od metoda regularizacije. Također zbog svoje kompleksnosti nisu primjenjive na svaku vrstu zadaća pa, usprkos dramatičnom povećanju učinkovitosti klasifikacije, nisu u potpunosti istisnule tradicionalne metode koje još uvijek pronalaze svoju primjenu na jednostavnije probleme za koje metode dubokog učenja mogu biti pretjerane ili prezahtjevne (Mahony i ostali 2020).

4 Klasifikacija slika metodama strojnog i dubokog učenja

Klasifikaciji slika tradicionalno se pristupa različitim algoritmima klasičnog strojnog učenja, ali posljednjih desetak godina značajno je porasla primjena metoda dubokog učenja na ovom zahtjevnom području. U ovom kratkom pregledu osvrnuti ćemo se na prednosti i nedostatke navedenih metoda, čimbenike koji su utjecali na njihov razvoj, te istaknuti neka od područja uspješne primjene.

4.1 Klasični modeli strojnog učenja u klasifikaciji slika

Tradicionalni pristupi klasifikaciji slika oslanjaju se na ručni odabir značajki kojeg odlikuje niz nedostataka. Osim što u značajnoj mjeri ovisi o znanju i vještini ljudskih subjekata, sam proces izdvajanja često je računalno zahtjevan, a jednom izdvojene značajke zbog svoje se specifičnosti mogu primijeniti samo na slične skupove podataka.

Usprkos tome, klasičnimodeli strojnog učenja, kao što su metoda potpornih vektora (eng. *Support Vector Machines - SVM*), metoda najbližih susjeda (eng. *K-Neares tNeighbors - KNN*) ili nasumična šuma (eng. *Random Forest - RF*), relativno se uspješno primjenjuju u različitim kontekstima, posebice kada se radi o malim skupovima podataka i ograničenim računalnim resursima u vidu nedostatka grafičkih procesorskih jedinica.

Primjerice, Chandra i Bedi (2021) istražili su primjenu SVM-a na klasifikaciju slika iz različitih domena primjene, kao što su medicinska dijagnostika, hiperspektralna analiza, otkrivanje defekata u proizvodnji, prepoznavanje listova biljaka, kemijskih uzoraka i prepoznavanje lica. Zaključak istraživanja je da SVM predstavlja robusan algoritam za klasifikaciju slika na skupovima podataka male ili srednje veličine, ali naglašavaju kako uspješnost njegove primjene značajno ovisi o mnogim drugim čimbenicima, poput kvalitete ulaznih podataka i pravilnog odabira informativnih značajki. Lin i ostali (2011) također ističu problem izrazite računalne zahtjevnosti u radu s velikim skupovima podataka, te sukladno tome predlažu metodu za brže izdvajanje značajki i treniranje SVM modela koja omogućava relativno uspješnu klasifikaciju slika velikih razmjera.

Sheykhmousa i ostali (2020) su kroz sustavni pregled literature usporedili primjenu SVM i RF algoritama na području klasifikacije slika dobivenih sustavima za daljinsko očitavanje (eng. *remote sensing*). Zaključak istraživanja koje uključuje 251 studiju je da su oba algoritma podjednako uspješna u izvršavanju ove zadaće, ali rezultati mogu značajno ovisiti o

karakteristikama ulaznih podataka. Primjerice, RF se pokazao boljim izborom za rad sa slikama niske rezolucije, dok SVM pokazuje bolje rezultate u radu sa slikama više rezolucije i skupovima s velikim brojem značajki.

KNN metoda također je pronašla uspješnu upotrebu na raznim područjima klasifikacije slika. Primjerice, Amato, Falchi, i Gennaro (2015) predstavljaju za to vrijeme brzu i učinkovitu metodu prepoznavanja spomenika na digitalnim fotografijama. Navedeno istraživanje također ističe važnost odabira odgovarajućih značajki usporedbom njihove uspješnosti na ovom specifičnom zadatku kroz niz eksperimenata. Tradicionalne metode strojno učenja uspješno su korištene na području medicinske dijagnostike pa tako Rajini i Bhavani (2011) predlažu KNN metodu za klasifikaciju snimaka magnetske rezonance u svrhu otkrivanja moždanih lezija koja rezultatima nadmašuje jednostavnu umjetnu neuronsku mrežu.

Kao što vidimo, klasični pristupi klasifikaciji slika metodama strojnog učenja pokazali su impresivne rezultate na različitim područjima primjene, ali zbog svojih ograničenja značajno zaostaju za suvremenim metodama dubokog strojnog učenja koje kroz proteklih desetak godina preuzimaju primat, posebice kada govorimo o klasifikaciji slika velikih razmjera.

4.2 Duboko učenje u klasifikaciji slika

Primjena dubokog učenja na klasifikaciju slika najčešće se odnosi na primjenu dubokih konvolucijskih neuronskih mreža koje karakterizira automatizirano učenje značajki iz ulaznog skupa podataka. Prva funkcionalna konvolucijska neuronska mreža bila je LeNet-5, ali revolucija ovog područja započinje tek nekoliko godina kasnije pobjedom AlexNet modela na ILSVRC¹ natjecanju u klasifikaciji slika. Osim same arhitekture ovog modela, ogroman utjecaj na popularizaciju primjene metoda dubokog učenja na različite probleme iz domene računalnog vida imala je činjenica da je u okvirima ovog pothvata uspješno demonstrirana primjena za to vrijeme moćnih grafičkih procesorskih jedinica, čime je dramatično skraćeno vrijeme treniranja modela (Lecun i ostali 1998; Krizhevsky, Sutskever, i Hinton 2012). Daljnji razvoj arhitektura dubokih neuronskih mreža usko je vezan za napredak tehnologije grafičkih procesorskih jedinica koje su omogućile praktičnu primjenu dubljih i kompleksnijih arhitektura, kao što su VGGNet ili ResNet obitelj modela (He i ostali 2015; Simonyan i Zisserman 2015).

¹<https://image-net.org/challenges/LSVRC/2012/results.html>

Različiti CNN modeli još uvijek dominiraju područjem klasifikacije slika, ali njihova su ograničenja potakla razvoj dubokog učenja u novim smjerovima. Primjerice, neki koncepti koji godinama dominiraju područjem obrade prirodnog jezika uspješno su prilagođeni klasifikaciji slika, pa tako obitelj VisionTransformer modela predstavlja obećavajuće područje istraživanja. Također se sve češće eksperimentira s modelima koji objedinjuju elemente konvolucijskih neuronskih mreža i enkoder – transformer arhitektura, a postoje naznake da bi ovakvi hibridni pristupi u budućnosti mogli proizvesti značajno bolje rezultate (Dosovitskiy i ostali 2021; Shao, Bi, i Chen 2022).

Neovisno o tome da li u svrhu klasifikacije slika koristimo neki od tradicionalnih ili suvremenih modela dubokog učenja, nezaobilazan korak pri njihovom odabiru predstavlja evaluacija uspješnosti koju je moguće provesti temeljem nekoliko različitih kriterija.

4.3 Evaluacija uspješnosti klasifikacijskih modela

Formalne definicije koje nude Mitchell (1997) i Alpaydin (2020) spomenute na početku poglavlja navode kriterij uspješnosti kao jedan od ključnih elemenata strojnog učenja. Kada govorimo o klasifikaciji, postoji nekoliko različitih kriterija s naglascima na različite komponente uspješnosti, a svi proizlaze iz koncepta pozitivnih i negativnih prediktivnih vrijednosti kojima se opisuje izvedba dijagnostičkih testova ili nekog drugog statističkog prediktivnog modela pa tako rezultat klasifikacije može imati jedan od četiri moguća osnovna ishoda:

Stvarno pozitivno (eng. true positive – TP) predviđanje ishod je klasifikacije kod kojeg pozitivna pripadnost klasi predviđena modelom odgovara stvarnoj pripadnosti.

Lažno pozitivno (eng. false positive – FP) predviđanje označava ishod klasifikacije kod kojeg negativna pripadnost klasi predviđena modelom ne odgovara stvarnoj pripadnosti.

Lažno negativno (eng. false negative – FN) predviđanje označava ishod klasifikacije kod kojeg negativna pripadnost klasi predviđena modelom ne odgovara stvarnoj pripadnosti.

Stvarno negativno (eng. true negative – TN) predviđanje označava ishod klasifikacije kod kojeg negativna pripadnost klasi predviđena modelom odgovara stvarnoj pripadnosti.

(Vujovic 2021)

Najjednostavnija mjera uspješnosti klasifikacije njena je točnost (eng. *accuracy*), a računa se jednostavni dijeljenjem broj točnih sa ukupnim brojem predviđanja:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Njena najveća prednost su jednostavnost i intuitivnost, ali može biti varljiva jer nam ne govori ništa o zastupljenosti pojedinih klasa u predviđanju pa tako modeli s relativno visokom točnosti predviđanja mogu u potpunosti zanemariti određene klase. Ovaj problem posebno je izražen kod neuravnoteženih skupova podataka, odnosno onih skupova čije klase sadrže znatno različit broj instanci.

Dvije osnovne mjere uspješnosti klasifikacije koje pristupaju ovom problemu su preciznost (eng. *precision*) i odziv (eng. *recall*). Preciznost nam odgovara na pitanje koliki je udio pozitivnih predviđanja bio ispravan, a računa se dijeljenjem broja stvarno pozitivnih predviđanja sa ukupnim brojem pozitivnih predviđanja:

$$Precision = \frac{TP}{(TP + FP)}$$

Odziv odgovara na pitanje koliki udio stvarno pozitivnih instanci je zaista prepoznato klasifikacijom. Računa se kao omjer broja stvarno pozitivnih i zbroja stvarno pozitivnih i lažno negativnih predviđanja:

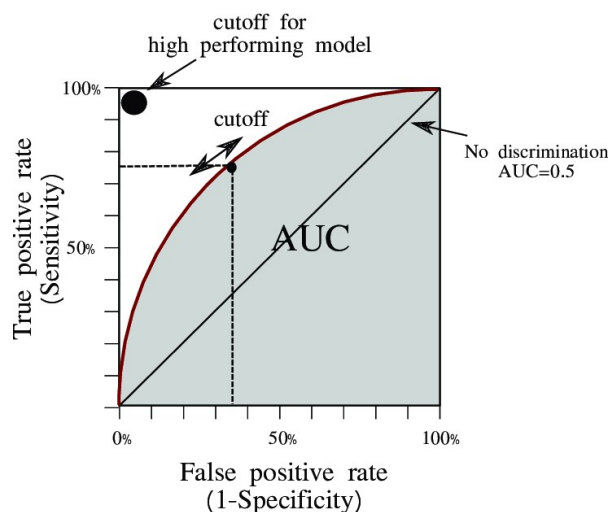
$$Recall = \frac{TP}{(TP + FN)}$$

Da bismo napravili potpunu evaluaciju uspješnosti određenog modela potrebno je uzeti u obzir obje mjere, a stvar se može znatno pojednostaviti korištenjem neke od mjera koje ujedinjaju preciznost i odziv, kao što je F1 mjera, koja se računa kao harmonijska sredina preciznosti i odziva:

$$F1 = \frac{2 \cdot (precision \cdot recall)}{(precision + recall)}$$

F1 mjera može primiti vrijednost u rasponu od 0 do 1 pri čemu 1 označava savršenu klasifikaciju, a 0 njenu suprotnost. Jedan od glavnih nedostataka F1 mjere leži u činjenici da jednaku važnost pridaje preciznosti i odzivu, što u praksi često nije slučaj (Hand, 2017).

Druga mjera uspješnosti kojom možemo nadopuniti točnost je površina ispod ROC (eng. *Receiver Operating Characteristic*) krivulje, također poznata kao ROC AUC (eng. *Area Under the Curve*). Krivulja se sastoji od dva parametra, a to su stopa stvarno pozitivnih i stopa lažno pozitivnih predviđanja (Slika 9).



Slika 9 ROC AUC krivulja (Rodríguez-Hernández, Pruneda, i Rodríguez-Díaz 2021, 8)

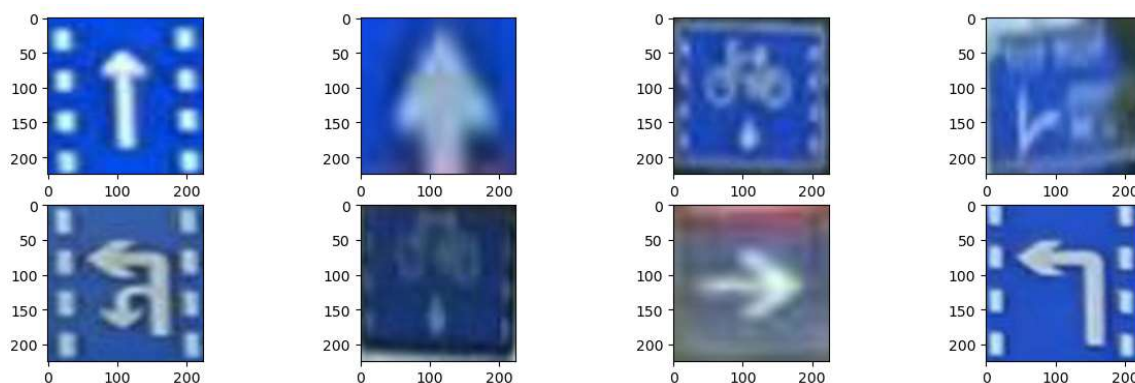
Uspješnost klasifikacije mjeri se površinom ispod krivulje, a može poprimiti vrijednosti u rasponu od 0 do 1, pri čemu od 0.5 označava uspješnost jednaku ili slabiju od nasumičnog pogađanja, kada govorimo o binarnoj klasifikaciji vrijednosti (Powers 2020; Vujovic 2021).

Preciznost, odziv, F1 mjera i ROC AUC u svom su osnovnom obliku zamišljene kao mjere uspješnosti binarne klasifikacije, ali se mogu jednostavno proširiti na probleme klasifikacije s više od dvije klase na način da se izračunavaju posebno za svaku klasu pri čemu sve ostale tretiramo kao jednu zajedničku klasu. Kao globalne mjere uspješnosti klasifikacije tada se uzima njihova aritmetička ili ponderirana aritmetička sredina.

5 Istraživački dio (eksperimenti s klasifikacijom slika)

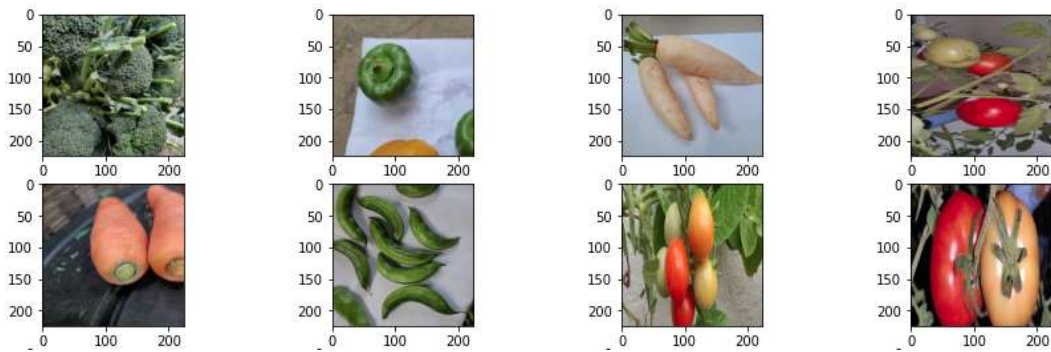
Svrha istraživanja je primijeniti, usporediti i evaluirati učinkovitost tradicionalnih naspram suvremenih metoda dubokog učenja na problem klasifikacije slika, kao tipičnog primjera nestrukturiranih podataka. U svim eksperimentima skupovi podataka su sljedeći:

1. "Traffic Sign Classification and Recognition" sadrži primjere deformacija, varijacija u osvjetljenju i djelomične zaklonjenosti objekata, a sastoji se od 6349 slika prometnih znakova fotografiranih u različitim uvjetima formata 224x224 piksela. Datoteke su podijeljene na trening skup od 6025 i testni skup od 324 primjerka. Također su ručno označene sa 10 različitih oznaka klase („Traffic Sign Classification and Recognition“, Kaggle, 2. Svibanj 2022.). Primjeri nekoliko slika iz ovog skupa prikazani su na Slici 10.



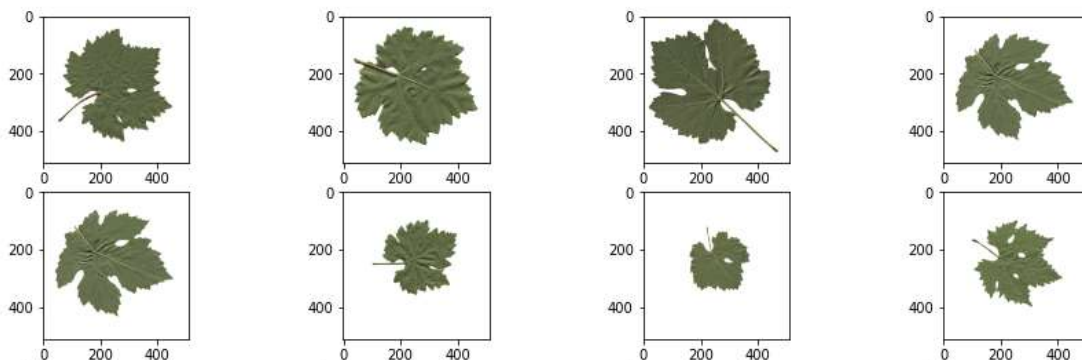
Slika 10 Nasumični primjerci iz skupa podataka "Traffic Sign Classification and Recognition"

2. "Vegetable image dataset" sadrži primjere zaklonjenosti i pozadinske pretrpanosti, a sastoji se od 21 000 slika povrća dimenzija 224x224 piksela podijeljenih na trening (15000 slika) i test (3000 slika). Slike su označene sa 15 različitih oznaka klase (Ahmed et al. 2021; („Vegetable image dataset“, Kaggle, 2. Svibanj 2022.). Primjeri nekoliko slika iz ovog skupa prikazani su na Slici 11.



Slika 11 Nasumični primjerci iz skupa podataka "Vegetable image dataset"

3. "Grapevine leaves image dataset" sastoji se od 500 slika dimenzija 511x511 piksela podijeljenih u 5 klasa, a tipičan je primjer skupa niske međuklasne varijacije, odnosno visoke sličnosti pripadnika različitih klasa (Kokluetal 2022; („Grapevine leaves image dataset“, Kaggle, 2. Svibanj 2022.). Primjeri nekoliko slika iz ovog skupa prikazani su na Slici 12.



Slika 12 Nasumični primjerci iz skupa podataka "Grapevine leaves imagedataset"

Pregled osnovnih značajki navedenih skupova podataka zbirno su prikazani u Tablici 1 na sljedećoj stranici.

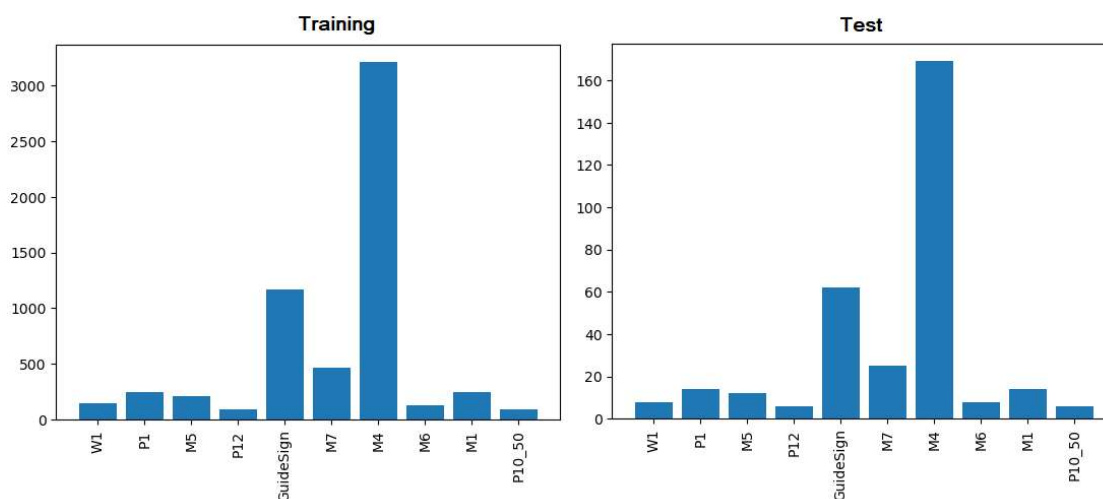
Tablica 1 Skupovi podataka korišteni u eksperimentima

| Naziv skupa podataka | Dimenzi je slika | Broj slika trening skupa | Broj slika validacij skog skupa | Broj slika testnog skupa | Broj klasa | Veličina u MB | Preuzeto [URL podataka] |
|---|------------------|--------------------------|---------------------------------|--------------------------|------------|---------------|---|
| Traffic Sign Classification and Recognition | 224x224 | 6025 | - | 324 | 10 | 26.96 | https://www.kaggle.com/datasets/wjybuqi/traffic-sign-classification-and-recognition |
| Vegetable image dataset | 224x224 | 15000 | 3000 * | 3000 | 15 | 571.74 | https://www.kaggle.com/datasets/misrahmed/vegetable-image-dataset |
| Grapevine leaves image dataset | 511x511 | 400 | - | 100** | 5 | 114.54 | https://www.kaggle.com/datasets/c051c23362365cde2917b6a726363087db1229b94b9391ebd0b8bd237ef4062f |

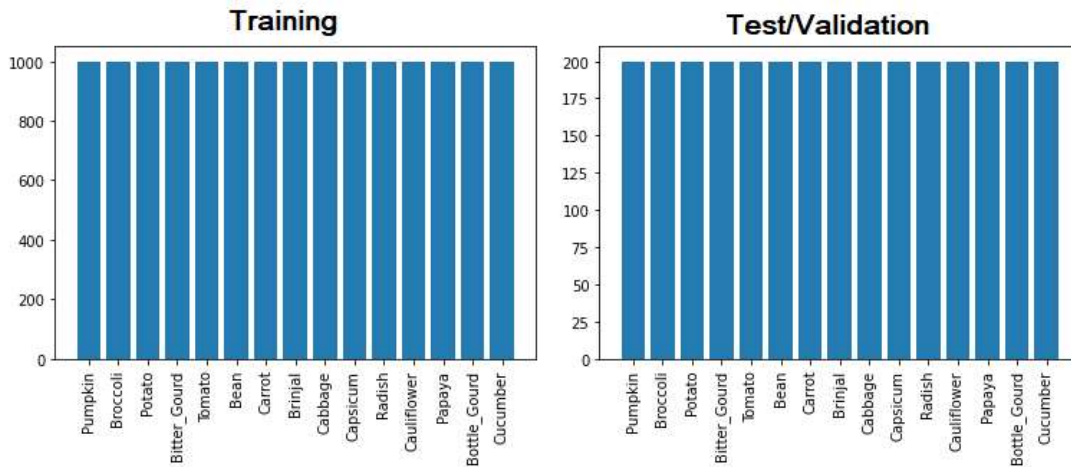
* validacijski skup je korišten samo u eksperimentima s metodama dubokog učenja

** razdioba na testni i trening skup je napravljena naknadno budući da je skup u izvornom obliku nepodijeljen

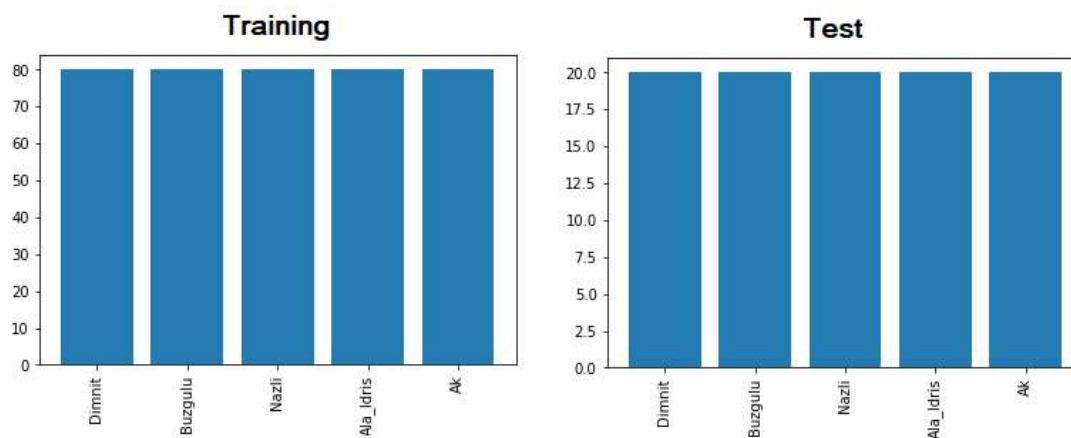
"Vegetable image dataset" i "Grapevine image dataset" su savršeno uravnoteženi skupovi, dok "Traffic Sign Classification and Recognition" skup podataka predstavlja primjer neuravnoteženog skupa. Distribucija po klasama prikazana je na Slikama 13 – 15.



Slika 13 Distribucija "Traffic Signs Classification and Recognition" skupa podataka



Slika 14 Distribucija "Vegetable image dataset" skupa podataka



Slika 15 Distribucija "Grape vine leaves image dataset" skupa podataka

5.1 Eksperimenti s tradicionalnim modelima strojnog učenja

U okviru eksperimenta primijenjena su dva tradicionalna modela strojnog učenja koja su se pokazala učinkovitim na širokom spektru problema iz domene klasifikacije slika, a to su nasumična šuma (eng. *Random Forest - RF*) i metoda potpornih vektora (eng. *Support Vector Machine - SVM*) (Horning 2010; Sheykhmousa i ostali 2020; Chandra i Bedi 2021).

Budući da SVM zbog visoke računalne kompleksnosti u svom klasičnom obliku nije pogodan za velike skupove podataka, u svrhu eksperimenta korišten je ThunderSVM, novija implementacija istog algoritma koja učinkovito iskorištava snagu grafičkih procesorskih jedinica čime se višestruko smanjuje vrijeme trajanja treniranja uz rezultate usporedive izvornoj implementaciji (Wen i ostali 2018).

5.1.1 Podešavanje hiperparametara

Izrazita računalna zahtjevnost SVM algoritma predstavlja značajnu prepreku opsežnijem podešavanju hiperparametara u radu s velikim skupovima podataka (Mantovani i ostali 2015). ThunderSVM implementacija višestruko ubrzava proces treniranja modela korištenjem GPU-a, ali računalna zahtjevnost i dalje predstavlja značajan problem, stoga su u eksperimentima korištene preporučene vrijednosti koje u većini slučajeva omogućavaju dobivanje zadovoljavajućih rezultata (ThunderSVM Developers bez dat.)

RandomForest se u literaturi često navodi kao jedan od primjera algoritama koji daju izrazito dobre rezultate s preporučenim vrijednostima hiperparametara pa su iz navedenog razloga one također ostale nepromijenjene, izuzev *n_estimators* (u literaturi se spominje i kao *number_of_trees*) parametra koji se ne podešava u klasičnom smislu nego bi trebao biti postavljen na dovoljno visoku vrijednost pod uvjetom da je dozvoljavaju dostupni računalni resursi (Probst i Boulesteix 2017; Probst, Wright, i Boulesteix 2019; scikit-learn developers bez dat.)

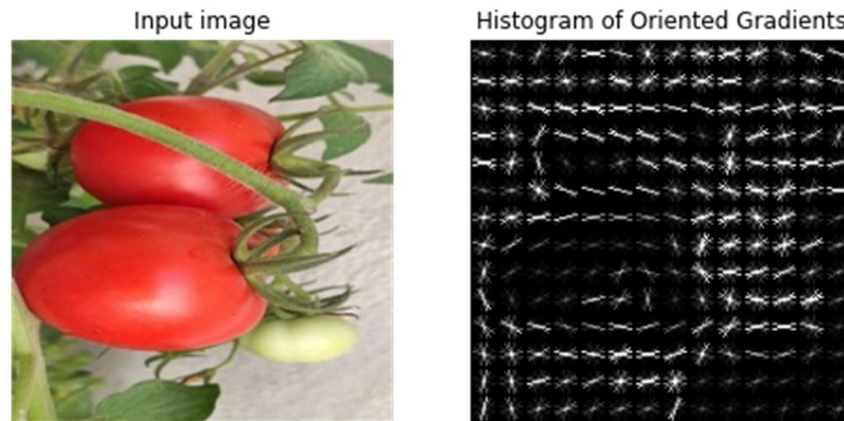
5.1.2 Izdvajanje značajki

Budući da ne sadrže informacije o samom sadržaju slike, sirovi slikovni podaci nisu primjereni za treniranje klasifikacijskog modela, stoga je potrebno prethodno izdvojiti dovoljno informativne značajke, odnosno one koje u odgovarajućoj mjeri omogućavaju razlikovanje pripadnika različitih klasa.

Kao metode izdvajanja značajki primijenjeni su HOG (histogram orijentiranih gradijenata, eng. *histogram of oriented gradients*) i LBP (lokalni binarni uzorci, eng. *local binary patterns*) algoritmi. Navedena kombinacija značajki pokazala se učinkovitom na širokom spektru slikovnih podataka (Azam, Riya, i Ahmed 2021; V i Gripsy 2020).

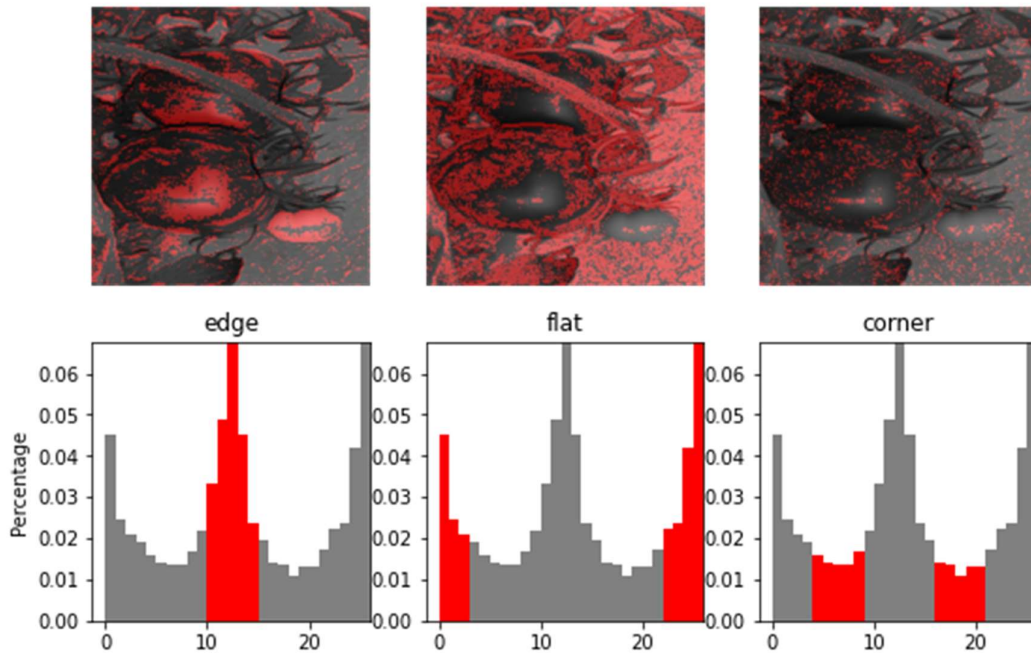
Histogram orijentiranih gradijenata (HOG) deskriptor je značajki oblika koji se od uobičajene detekcije rubova razlikuju po tome što ne detektira samo rubove na slici nego i njihov smjer, odnosno ne odgovara samo na pitanje da li je određen piksel dio ruba nego se pomoću vrijednosti intenziteta okolnih piksela izračunava vektor smjera i intenziteta, odnosno njegov gradijent i orijentacija. Na ovaj način dobiva se informativnija značajka jer osim samog ruba sadrži podatak o tome u kojem se smjeru i u kojoj mjeri povećavanju vrijednosti intenziteta piksela područja koja rub razdvaja. Rezultat ovog algoritma je histogram sa 9 stupaca od kojih svaki predstavlja intenzitet u rasponu od 20 stupnjeva orijentacije. Navedene značajke

izračunavaju se u lokaliziranim regijama pa se tako generira zaseban histogram za svaku pojedinu regiju slike (Dalal i Triggs 2005). Na Slici 16 prikazan je primjer HOG značajki za sliku iz skupa "Vegetable image dataset".



Slika 16 Primjer vizualizacije HOG značajki za sliku iz Vegetable image dataset skupa

Lokalni binarni uzorci (LBP) deskriptor je značajki teksture temeljen na ideji da se dvodimenzionalne površinske teksture mogu opisati dvjema komplementarnim mjerama – uzorkom i njegovom snagom, odnosno lokalnim prostornim uzorcima i kontrastom sivih tonova. Jednostavno rečeno, LBP promatra točke koje okružuju središnju točku i vraća binarni rezultat koji govori jesu li okolne točke većeg ili manjeg intenziteta od središnje točke. Kada su okolni pikseli svi crni ili potpuno bijeli, tada je to područje slike ravno, odnosno bez značajki. Grupacije kontinuiranih crnih ili bijelih piksela smatraju se uniformiranim uzorcima koji se mogu protumačiti kao kutovi ili rubovi, a ukoliko se pikseli izmjenjuju između crnih i bijelih piksela, uzorak se smatra neuniformnim (Ojala, Pietikainen, i Maenpaa 2002). Slika 17 prikazuje rezultat LBP ekstrakcije na primjeru slike iz skupa „Vegetable image dataset”. Niz od tri slike redom prikazuje dijelove slike označene kao tub (eng. *edge*), ravno područje bez značajki (eng. *flat*) i područja na kojima su detektirani kutovi (eng. *corner*), sa pripadajućim histogramima.



Slika 17 Primjeri LBP značajki sa pripadajućim histogramima

Za potrebe eksperimenta upotrijebljene su implementacije HOG i LBP algoritama iz Pythonovog paketa scikit-image (thescikit-image development team bez dat.; bez dat.)

5.1.3 Evaluacija modela

U svrhu evaluacije modela korištene su sljedeće mjere uspješnosti: točnost (*accuracy*), F1-score i ROC AUC, dobivene korištenjem biblioteka paketa *sklearn*.

Točnost (eng. *accuracy*) je jednostavna i intuitivna mjera uspješnosti, ali može biti varljiva jer nam ne govori ništa o zastupljenosti pojedinih klasa u konačnom predviđanju, stoga je preporučljivo da se koristi u kombinaciji s nekom od mjera koje nadilaze ovaj nedostatak. F1-score i ROC AUC navedenom problemu pristupaju na nešto različite načine pa zajedno omogućavaju potpuniju evaluaciju uspješnosti modela.

5.2 Eksperimenti s modelima dubokog učenja

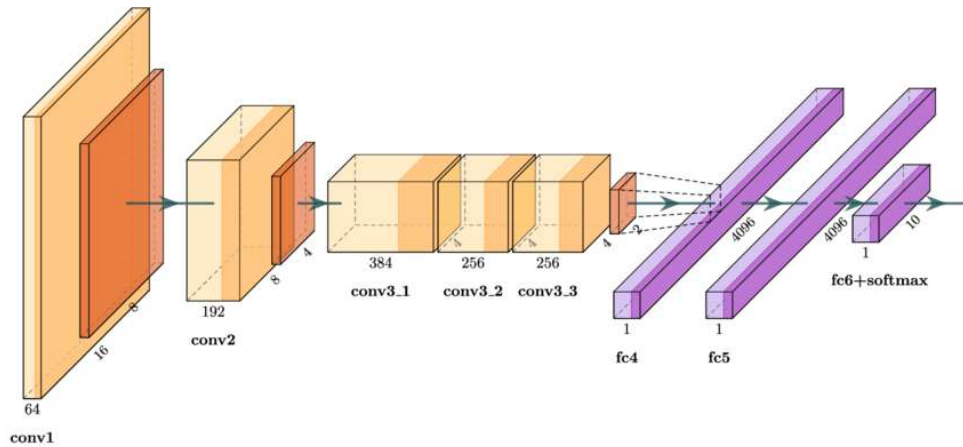
Korištena su ukupno četiri modela dubokog učenja, od kojih svaki predstavlja neke od važnijih napredaka područja istraživanja njihove primjene na problem klasifikacije slika. AlexNet (Krizhevsky, Sutskever, i Hinton 2012) je bila prva arhitektura koja je demonstrirala mogućnosti konvolucijskih neuronskih mreža na klasifikaciji slika velikih razmjera, VGGNet (Simonyan i Zisserman 2015) je dodatno razradila neke od koncepata kao što su povećanje dubine, ResNet arhitektura (He i ostali 2015) je ponudila učinkovita rješenja problema proizašlih iz povećanja dubine, a VisionTransformer (Dosovitskiy i ostali 2021) predstavlja napredak na ovom području uspješnom primjenom koncepata iz domene obrade prirodnog jezika.

Primijenjeni su modeli prethodno trenirani na ImageNet skupu podataka, odnosno demonstrirana je primjena prijenosa znanja naučenog na većem skupu (eng. *transfer learning*). U svim eksperimentima korištena je metoda peterostruke unakrsne validacije (eng. 5-fold cross-validation) u svrhu odabira modela, izuzev „Vegetable image dataset“ kod kojeg je korišten validacijski skup podataka.

1.5.1 Klasifikacija temeljena na AlexNet modelu

AlexNet je bila prva duboka konvolucijska mreža koja je pobijedila na ImageNet LSVRC natjecanju algoritama za klasifikaciju slika. Ovaj događaj predstavlja prekretnicu na području dubokog učenja i računalnog vida jer je pokazao potencijal primjene konvolucijskih neuronskih mreža na zadaće klasifikacije slika velikih razmjera.

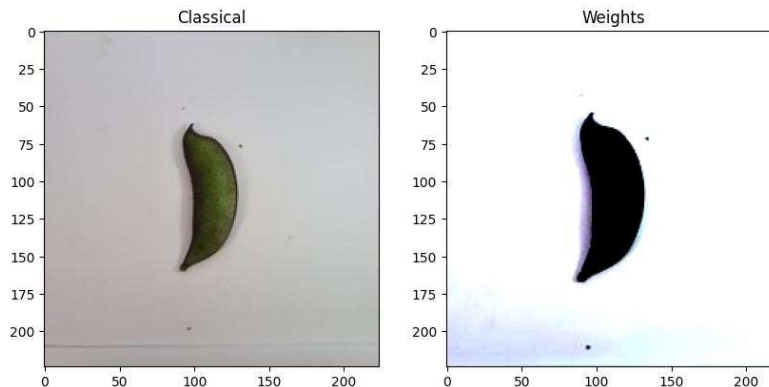
Izvorna arhitektura se sastoji od pet konvolucijskih slojeva i tri potpuno povezana sloja. Prvi sloj je konvolucijski sloj s 96 filtera veličine 11x11, a nakon njega slijedi sloj maksimalnog sažimanja s prozorom veličine 3x3. Drugi sloj je konvolucijski s 256 filtera veličine 5x5, nakon čega ponovo dolazi sloj maksimalnog sažimanja prozora veličine 3x3. Treći i četvrti sloj također su konvolucijski slojevi s 384 filtera veličine 3x3, a peti sloj sastoji se od 256 filtera veličine 3x3. Izlaz petog sloja prolazi kroz tri potpuno povezana sloja s 4096, 4096, odnosno 1000 neurona. Arhitektura ovog modela je detaljno prikazana na Slici 18.



Slika 18 Prikaz arhitekture AlexNet modela (Strisciuglio, Lopez Antequera, i Petkov 2020, 17969)

Kao aktivacijsku funkciju AlexNet koristi funkciju ispravljene linearne jedinice (ReLU) u svim slojevima osim posljednjeg koji koristi softmax, a kao mjera regularizacije primjenjena je u to vrijeme inovativna *dropout* metoda kojom mreža nasumično postavlja izlaze neurona na nulu sukladno predefiniciranoj vjerojatnosti (u slučaju izvorne AlexNet arhitekture vrijednost je $p=0.5$). Na ovaj način suzbija se složena međusobna adaptacija neurona na način da se mreža raščlanjuje na više podskupova neurona čime se konvolucijske slojeve prisiljava da nauče robusnije značajke (Krizhevskiyetal. 2012; Srivastavaetal. 2014).

Zaslужnost AlexNet arhitekture također leži u tome što je na velikoj pozornici demonstrirala neke od ključnih koncepata koji su obilježili daljnji razvoj područja, kao što je značaj dubine mreže za njene performanse te korištenje moćnih grafičkih procesorskih jedinica.



Slika 19 Slika predprocesirana klasičnom metodom i metodom transformacije u sklopu predtreniranog AlexNet modela iz TorchVision paketa

Model smo inicijalizirali težinskim vrijednostima dobivenima prethodnim treniranjem na ImageNet skupu, a zatim dodatno ugodili težinske vrijednosti klasifikatora na odabranim skupovima podataka. Da bismo ovo postigli bilo je potrebno „zamrznuti” slojeve mreže zadužene za izdvajanje značajki, odnosno naznačiti da se procesom treniranja ne mijenjanju, nakon čega se ponovo kreiraju slojevi zaduženi za klasifikaciju s prilagodbama sukladnim potrebama pojedinog skupa podataka.

Prije primjene ovakvog modela ciljani skup podataka potrebno je na odgovarajući način predprocesirati. Metoda predprocesiranja uvelike ovisi o korištenom modelu i načinu na koji je treniran, a paket TorchVision znatno olakšava ovaj korak nudeći odgovarajuće transformacije zajedno sa težinskim vrijednostima predtreniranih modela. Na Slici 19 prikazan je primjer slike predprocesirane klasičnom metodom transformacije i pomoću transformacije u sklopu predtreniranog AlexNet modela iz TorchVision paketa.

Struktura klasifikatora i hiperparametri modela ovise o specifičnosti pojedinog skupa podataka, kako slijedi:

"Vegetable image dataset" – izuzev prilagođenih dimenzija izlaznog sloja, korištena je izvorna struktura klasifikatora. U svrhu izbjegavanja overfittinga (uobičajenog problema ranije spomenutog u cjelini 3.2) razvijene su razne metode regularizacije, a u ovom slučaju korištena je dropout metoda uključena u izvornu arhitekturu, s nepromijenjenim vrijednostima ($p=0.5$) prema (Krizhevsky, Sutskever, i Hinton 2012). Zadovoljavajući rezultati klasifikacije postignuti su primjenom standardne CrossEntropyLoss funkcije gubitka i SGD (eng. *Stochastic Gradient Descent*) optimizatorom nepromjenjive stope učenja ($lr=0.001$). Također je korištena metoda ranog zaustavljanja (eng. *early stopping*) sa kriterijem stagnacije mjere *accuracy*.

"Traffic Sign Classification and Recognition" – struktura klasifikatora je izmijenjena dodavanjem dvaju *torch.nn.BatchNorm1d* slojeva u svrhu primjene normalizacije mini-grupe. Normalizacija mini-grupe ili normalizacija serije jedna je od tehnika treniranja dubokih neuronskih mreža kojom se normalizira izlaz nekog sloja za svaku grupu (eng. *batch*) ulaznih podataka. Primjenom ove metode izbjegava se problem unutarnjeg kovarijantnog pomaka, čime možemo postići značajnu stabilizaciju procesa učenja, te smanjenje broja epoha potrebnih za postizanje iste razine treniranosti (Ioffe i Szegedy 2015). Kao metoda regularizacije također je primijenjena ugrađena *dropout* metoda sa nešto višim vrijednostima od izvorne ($p=0.8$).

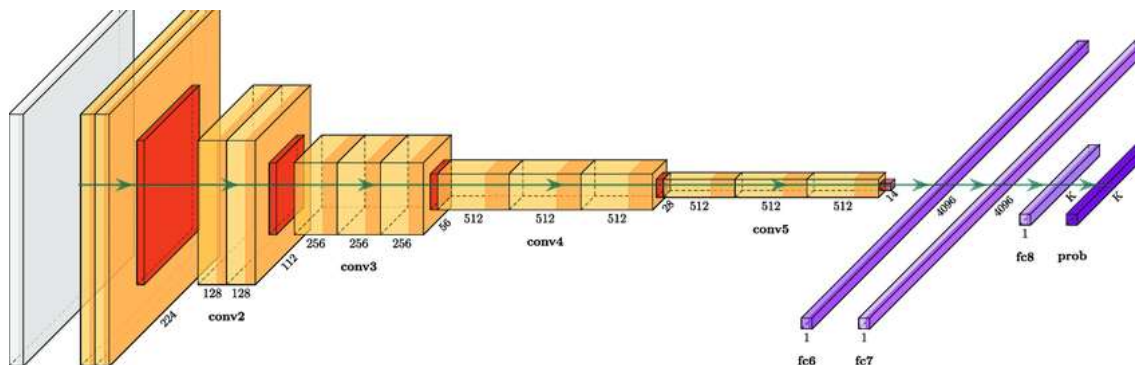
Korištena je CrossEntropyLoss funkcija gubitka i SGD optimizator sa nepromjenjivom stopom učenja ($lr=0.001$) određenom metodom postepenog smanjivanja vrijednosti unutar zadanog raspona, a budući da se radi o neuravnoteženom skupu podataka kao kriterij ranog prekidanja uzeta je vrijednost ponderiranog gubitka (eng. *weighted loss*).

"Grapevine leaves image dataset" – korištena je izvorna struktura klasifikatora sa ugrađenom *dropout* metodom regularizacije s povećanim vrijednostima ($p=0.7$) dobivenom postepenim povećavanjem do postizanja modela bez znakova overfittinga. Korištena je CrossEntropyLoss funkcija gubitka i SGD optimizator, a budući da fiksne vrijednosti stope učenja nisu davale zadovoljavajuće rezultate primijenjena je metoda cikličke stope učenja u rasponu od $lr=0.0025$ do $lr=0.01$. Radi se o naprednoj metodi promjenjive stope učenja kojom ciklički povećavamo i smanjujemo stopu kroz unaprijed zadani raspon. Metoda se javila kao odgovor na slučajeve kada optimizacijski algoritam zapne na nepovoljnom lokalnom minimumu funkcije gubitka. Naglim povećanjem stope učenja nakon što smo je kroz nekoliko koraka smanjivali može se postići da algoritam „iskoči” iz područja lokalnog minimuma, čime se olakšava pronalaženje nižeg lokalnog ili, u idealnom slučaju, globalnog minimuma. Navedeni raspon određen je metodom opisanom u (Smith 2017), odnosno probnim treniranjem modela uz postepeno povećavanje stope učenja dok vrijednosti gubitka ne počnu naglo rasti. Stopa učenja u trenutku prije naglog rasta, kada je vrijednost gubitka najmanja, predstavlja maksimalnu vrijednost stope učenja iz našeg raspona, a dijeljenjem ove vrijednost sa 3 ili 4 dobit ćemo minimalnu preporučenu vrijednost.

1.5.2 Klasifikacija temeljena na VGG16 modelu

VGG16 arhitektura nastala je kao jedan od odgovora na uspješno demonstriranje ranije spomenutih koncepata. Nadahnuti uspjehom AlexNet arhitekture, njeni su tvorci dizajnirali konvolucijsku neuronsku mrežu s puno većim brojem slojeva (varijanta korištena u eksperimentu sastoji se od 16 slojeva), ali su znatno pojednostavnili arhitekturu na način da svaki konvolucijski sloj koristi filtere dimenzija 3×3 s korakom 1. Nakon njih slijede slojevi maksimalnog sažimanja dimenzija prozora 2×2 . Prvi konvolucijski sloj ima 64 filtera, drugi 128 filtera, treći 256 filtera, a četvrti i peti po 512 filtera. Arhitektura završava s tri potpuno povezana sloja – prva dva sastoje se od po 4096 neurona, a treći u izvornoj varijanti 1000 jer se ImageNet skup podataka sastoji od 1000 klasa. Kao aktivacijska funkcija koristi se funkcija

ispravljene linearne jedinice (ReLU), osim u završnom sloju koji koristi softmax (Simonyan i Zisserman 2015). Struktura VGG16 modela prikazana je na Slici 14.



Slika 20 Struktura VGG16 modela (Blauch, Behrmann, i Plaut 2020, 4)

VGG model osvojio je tek drugo mjesto na ILSVRC natjecanju 2014. godine, ali se istaknuo svojom jednostavnošću, homogenom topologijom i dubinom znatno većom nego u ostalih CNN modela toga vremena. Njegov glavni nedostatak je iznimno visok broj parametara (čak 138 milijuna), što ga čini računalno zahtjevnim i skupim rješenjem, neprimjerenim za upotrebu na uređajima ograničene računalne moći.

Daljnji pokušaji povećanja dubine neuronskih mreža čiju je učinkovitost demonstrirala VGG obitelj modela razotkrili su nove izazove. Prvi je bio vezan za tzv. problem iščezavajućeg ili eksplodirajućeg gradijenta. Problem iščezavajućeg gradijenta događa se kada se gradijenti funkcije gubitka toliko približe nuli da njihova propagacija rezultira zanemarivim korekcijama težinskih vrijednosti, što može toliko usporiti proces učenja da mreža u praktičnom smislu prestane učiti. Eksplozirajući gradijenti predstavljaju suprotan problem, odnosno gradijenti postanu toliko veliki da korekcija više ne dovodi do smanjenja pogreške (Y. Bengio, Simard, i Frasconi 1994; Sun 2019). Drugi izazov proizlazi iz problema degradacije, odnosno opadanja performansi s povećanjem dubine mreže.

Inicijalizaciju modela, predprocesiranje podataka i "zamrzavanje" slojeva zaduženih za izdvajanje značajki napravili smo na jednak način kao kod primjene AlexNet modela.

Struktura modela (klasifikatora) i hiperparametri modela ovisni o specifičnosti pojedinog skupa podataka određeni su kako slijedi:

"Vegetable image dataset" - izuzev prilagođenih dimenzija izlaznog sloja, korištena je izvorna struktura klasifikatora. U svrhu regularizacije korišten je dropout metoda uključena u izvornu arhitekturu, s nepromijenjenim vrijednostima ($p=0.2$) prema (Simonyan i Zisserman 2015), te L2 regularizacija poznata kao kažnjavanje visokih težinskih vrijednosti (eng. *weight decay*). *Weight decay* za određeni faktor umanjuje vrijednosti težina u svakoj iteraciji te ih na taj način postepeno približava nuli, ali one nikada ne poprime vrijednost nule. Ovo se postiže tako da funkciji gubitka dodajemo zbroj kvadrata težina pomnožen sa vrijednošću *lambda* koju nazivamo stopom regularizacije (Goodfellow, Bengio, i Courville 2016).

Zadovoljavajući rezultati klasifikacije postignuti su primjenom standardne CrossEntropyLoss funkcije gubitka i Adam optimizatorom. Nepromjenjiva stopa učenja nije davala zadovoljavajuće rezultate, stoga je primijenjena ExponentialLR metoda promjenjive stope učenja koja u svakoj epohi smanjuje stopu za predefinirani faktor ($\gamma=0.9$). Temeljna je ideja ove metode da u početku učenja, dok su parametri mreže još uvijek daleko od optimalnih vrijednosti, možemo ubrzati proces primjenom relativno visoke stope učenja, a kako se pogreška mreže približava minimalnoj vrijednosti postepeno smanjivati stopu učenja. Također je korištena metoda ranog zaustavljanja (eng. *early stopping*) s kriterijem stagnacije mjere *accuracy*.

"Traffic Sign Classification and Recognition" – struktura klasifikatora je izmjenjena dodavanjem dvaju *torch.nn.BatchNorm1d* slojeva u svrhu primjene normalizacije mini-grupe. Kao metoda regularizacije također je primijenjena ugrađena *dropout* metoda sa znato višim vrijednostima od izvorne ($p=0.6$), te L2 regularizacija sa faktorom (*weight_decay*=0.01) određenim metodom postepenog povećavanja vrijednosti unutar zadanog raspona,.

Korištena je CrossEntropyLoss funkcija gubitka i Adam optimizator sa nepromjenjivom stopom učenja ($lr=0.001$) također određenom metodom postepenog smanjivanja vrijednosti unutar zadanog raspona, a budući da se radi o neuravnoteženom skupu podataka kao kriterij ranog prekidanja uzeta je vrijednost ponderiranog gubitka (eng. *weighted loss*).

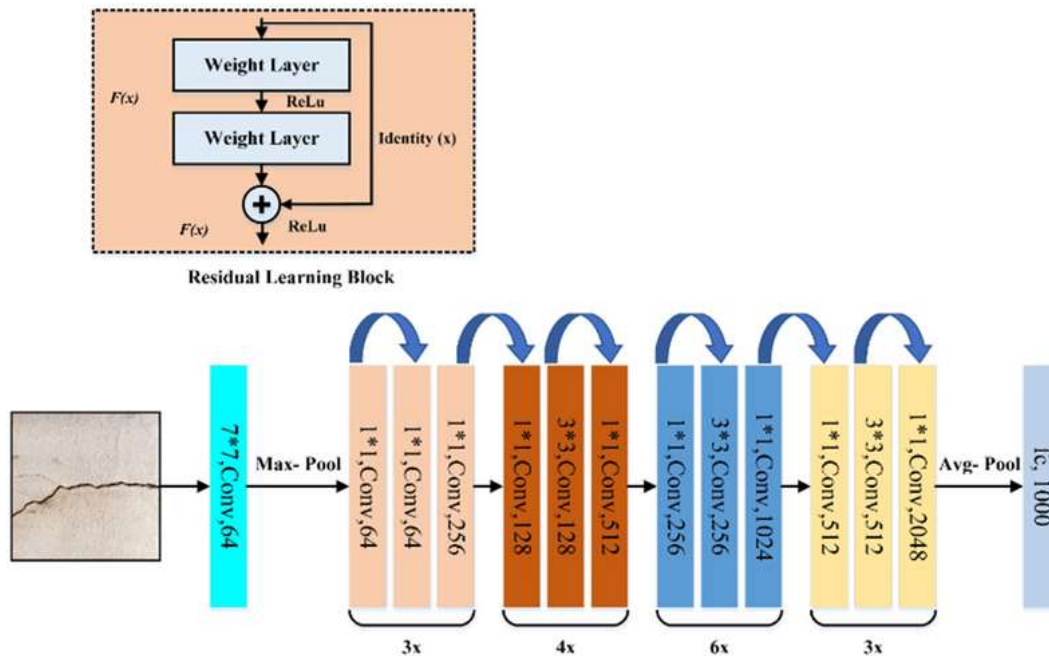
"Grapevine leaves image dataset" – korištena je izvorna struktura klasifikatora sa ugrađenom *dropout* metodom regularizacije s povećanim vrijednostima ($p=0.8$) i ($p=0.85$) dobivenima postepenim povećavanjem do postizanja modela bez znakova overfittinga. Korištena je CrossEntropyLoss funkcija gubitka i SGD optimizator, a budući da fiksne vrijednosti stope učenja nisu davale zadovoljavajuće rezultate primijenjena je metoda cikličke

stope učenja u rasponu od $lr=0.0025$ do $lr=0.01$. Također je korištena metoda ranog zaustavljanja (eng. *early stopping*) s kriterijem stagnacije mjere *accuracy*.

5.2.1 Klasifikacija temeljena na ResNet50 modelu

Problemi iščezavajućeg i eksplodirajućeg gradijenta mogu se zadovoljavajuće riješiti odgovarajućom inicijalizacijom težina, normalizacijom i primjenom ReLU aktivacijske funkcije, a ResNet arhitektura rješava oba problema uvođenjem rezidualnih blokova, odnosno gradivnih jedinica sastavljenih od dvaju ili više slojeva mreže s dodatnom vezom putem koje ulaz prvog sloja zaobilazi sve slojeve bloka i dodaje se izlazu zadnjeg sloja (eng. *skip connection*). Primjerice, označimo li ulaz kao x , a funkciju slojeva bloka kao $F(x)$, konačni izlaz rezidualnog bloka biti će $y=F(x) + x$ (He i ostali 2015). Točan način na koji ovakva arhitektura funkcionira još uvijek je predmet rasprave, a neka od predlaganih objašnjenja uključuju omogućavanje protoka informacija kroz mrežu na način da se blokovi koji ne doprinose učenju mogu jednostavno zaobići, slobodniji protok gradijenata prema plićim slojevima, pa čak i teza kako se uvođenjem zaobilaznih veza zapravo ne tvori jako duboka mreža nego svojevrsan ansambl nekoliko relativno plitkih mreža (Balduzzi i ostali 2018; Veit, Wilber, i Belongie 2016).

U eksperimentu je korištena ResNet50 varijanta ResNet modela s pedeset slojeva koja, za razliku od varijanti manje dubine koje su sastavljene od rezidualnih blokova sa dva konvolucijska sloja, koristi iste blokove sastavljene od tri konvolucijska sloja te tzv. blokove identiteta koji uz slojeve normalizacije sadrže konvolucijske slojeve dimenzije filtra 1×1 . Blokovi identiteta znatno smanjuju broj parametara mreže, čime ona postaje računalno učinkovitija usprkos povećanoj dubini (He i ostali 2015). Arhitektura ResNet50 modela i tipičnog rezidualnog bloka prikazana je na Slici 21.



Slika 21 Arhitektura ResNet50 modela i tipičnog rezidualnog bloka (L. Ali i ostali 2021, 9)

Inicijalizaciju modela, predprocesiranje podataka i „zamrzavanje“ slojeva zaduženih za izdvajanje značajki napravljeni su na jednak način kao kod primjene AlexNet modela.

Struktura klasifikatora i hiperparametri modela ovisni o specifičnosti pojedinog skupa podataka određeni su kako slijedi:

"Vegetable image dataset" - izuzev prilagođenih dimenzija izlaznog sloja, korištena je izvorna struktura klasifikatora. U svrhu regularizacije korištena je L2 regularizacija s vrijednosti $\text{weight_decay}=0.01$ dobivenom postepenim povećavanjem do izostanka *overfittinga*.

Zadovoljavajući rezultati klasifikacije postignuti su primjenom standardne CrossEntropyLoss funkcije gubitka i SGD optimizatorom. Nepromjenjiva stopa učenja nije davala zadovoljavajuće rezultate, stoga je primijenjena metoda cikličke stope učenja u rasponu od $\text{lr}=0.01$ do $\text{lr}=0.001$. Također je korištena metoda ranog zaustavljanja (eng. *early stopping*) sa kriterijem stagnacije mjere *accuracy*.

"Traffic Sign Classification and Recognition" – struktura klasifikatora je izmijenjena dodavanjem *torch.nn.BatchNorm1d* sloja u svrhu primjene normalizacije mini-grupe. Kao metoda regularizacije također je primijenjena ugrađena *dropout* metoda s vrijednošću ($p=0.4$) u svrhu koje je bilo potrebno dodati *dropout* sloj u klasifikator, te L2 regularizacija sa

faktorom ($\text{weight_decay}=0.01$) određenim metodom postepenog povećavanja vrijednosti unutar zadanog raspona. Korištena je `CrossEntropyLoss` funkcija gubitka i SGD optimizator. Traženje nepromjenjive vrijednosti stope učenja adaptivnom metodom nije dalo zadovoljavajuće rezultate, pa je primijenjena metoda cikličke stope učenja u rasponu od $\text{lr}=0.0025$ do $\text{lr}=0.01$ dobivenom metodom prema (Smith 2017). Budući da se radi o neuravnoteženom skupu podataka kao kriterij ranog prekidanja uzeta je vrijednost ponderiranog gubitka (eng. *weighted loss*).

"Grapevine leaves image dataset" – struktura klasifikatora je izmijenjena dodavanjem `torch.nn.BatchNorm1d` slojem u svrhu primjene normalizacije mini-grupe. Kao metoda regularizacije također je primijenjena *dropout* metoda s vrijednošću ($p=0.85$) u svrhu koje je bilo potrebno dodati *dropout* sloj u klasifikator. Vrijednost p dobivena je postepenim povećavanjem do postizanja modela bez znakova *overfittinga*. Korištena je `CrossEntropyLoss` funkcija gubitka i SGD optimizator s fiksnom stopom učenja ($\text{lr}=0.0005$) dobivenom metodom postepenog smanjivanja vrijednosti unutar zadanog raspona. Također je korištena metoda ranog zaustavljanja (eng. *early stopping*) sa kriterijem stagnacije mjere *accuracy*.

5.2.2 Klasifikacija temeljena na Visiontransformer modelu

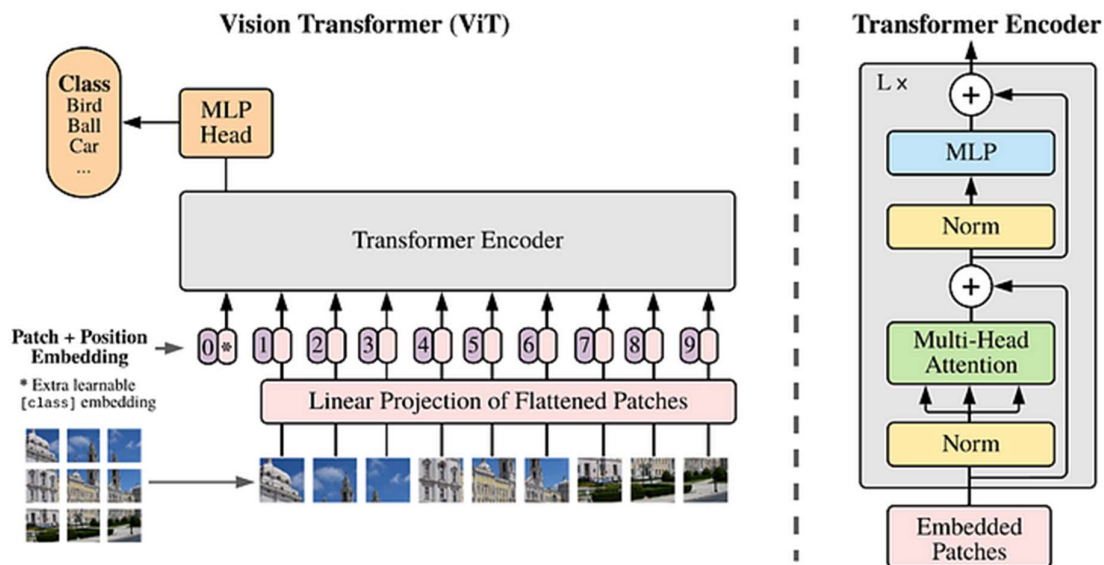
Dok su se drugi autori bavili problemom konvolucijskih neuronskih mreža, odnosno rješavanjem problema proizašlih iz povećanja njihove dubine, istraživači istraživačkog odjela tvrtke Google odlučili su se za drugačiji pristup.

VisionTransformer (ViT) uspješan je pokušaj primjene koncepta koji godinama dominira područjem obrade prirodnog jezika. Transformeri su modeli dubokog učenja koji pomoću mehanizma samo-pažnje različito ponderiraju dijelove ulaznih podataka, što ih čini posebno pogodnima za obradu prirodnog jezika. Naime, umjesto fokusiranja na značenje pojedinih riječi transformeri mogu kao ulaz uzeti čitave rečenice i dodijeliti različite težinske vrijednosti pojedinim riječima ovisno o njihovom položaju u rečenici, čime se zadržavaju kontekstualne informacije neophodne za ispravno razumijevanje jezika (Vaswani i ostali 2017). ViT model na sličan način tretira slikovne podatke tako da rastavlja slike na mozaik pojedinačnih dijelova (u izvornoj arhitekturi dimenzija 16×16 piksela) koje tada tretira kao riječi u rečenici, odnosno dodjeljuje im težinske vrijednosti sukladno njihovim prostornim međuodnosima. Posebnost ovakvog pristupa leži u tome da eksplicitno iskorištava kontekstualne informacije

slike za razliku od, primjerice, CNN modela koji prostorne informacije koriste implicitno kao svojevrsnu nuspojavu mapa značajki.

Sama arhitektura ViT modela nešto je složenija nego što je to slučaj sa CNN modelima, a njegov ključni dio je tzv. transformer-ekoder sastavljen od dva segmenta kojima prethodi po jedan sloj normalizacije. Prvi je više-glavni sloj sam-opažnje (eng. *multi-headself-attention layer*) koji omogućuje modelu da se posveti različitim dijelovima slike i odvagne njihovu relativnu važnost. Nakon njega slijedi višeslojni perceptron s linearnom jedinicom Gaussove pogreške (GeLU) kao aktivacijskom funkcijom. Za svaki segment postoji rezidualna zaobilazna veza slično kao kod ResNet obitelji modela.. Arhitektura čitavog modela sastavljena je od nekoliko transformer-ekoder blokova nakon kojih dolazi klasifikacijska glava u obliku višeslojnog perceptrona (eng. *MLP head*), jednako kao kod CNN klasifikacijskih modela (Dosovitskiy i ostali 2021).

Ovisno o broju transformer-ekoder blokova, ViT modeli mogu biti različite dubine, a za potrebe eksperimenata odabrali pliću varijantu od 12 slojeva (ViT-B/16). Struktura ViT-B/16 modela prikazana je na Slici 22.



Slika 22 Arhitektura ViT-B/16 modela (Dosovitskiy i ostali 2021, 3)

Inicijalizaciju modela, predprocesiranje podataka i "zamrzavanje" slojeva zaduženih za izdvajanje značajki napravili smo na jednak način kao kod primjene AlexNet modela.

Struktura klasifikatora i hiperparametri modela ovisni o specifičnosti pojedinog skupa podataka određeni su kako slijedi:

"**Vegetable image dataset**" - izuzev prilagođenih dimenzija izlaznog sloja, korištena je izvorna struktura klasifikatora. Zadovoljavajući rezultati klasifikacije postignuti su primjenom standardne CrossEntropyLoss funkcije gubitka i SGD optimizatorom. Nepromjenjiva stopa učenja nije davala zadovoljavajuće rezultate, stoga je primijenjena metoda cikličke stope učenja u rasponu od $lr=0.1$ do $lr=0.001$. Također je korištena metoda ranog zaustavljanja (eng. *early stopping*) a kriterijem stagnacije mjere *accuracy*.

"**Traffic Sign Classification and Recognition**" – struktura klasifikatora je izmijenjena dodavanjem *dropout* sloja s vrijednošću $p=0.6$ dobivenom postepenim povećavanjem do postizanja modela bez overfittinga.

Korištena je CrossEntropyLoss funkcija gubitka i SGD optimizator. Traženje nepromjenjive vrijednosti stope učenja gridsearch metodom nije dalo zadovoljavajuće rezultate, pa je primijenjena metoda cikličke stope učenja u rasponu od $lr=0.001$ do $lr=0.004$ dobivenom metodom prema (Smith 2017). Budući da se radi o neuravnoteženom skupu podataka kao kriterij ranog prekidanja uzeta je vrijednost ponderiranog gubitka (eng. *weighted loss*).

"**Grapevine leaves image dataset**" – struktura klasifikatora je izmijenjena dodavanjem *torch.nn.BatchNorm1d* slojem u svrhu primjene normalizacije mini-grupe. Kao metoda regularizacije također je primijenjena *dropout* metoda s vrijednošću ($p=0.8$) u svrhu koje je bilo potrebno dodati *dropout* sloj u klasifikator. Vrijednost p dobivena je postepenim povećavanjem do postizanja modela bez znakova *overfittinga*. Korištena je CrossEntropyLoss funkcija gubitka i SGD optimizator sa cikličkom stopom učenja u rasponu od $lr=0.0017$ do $lr=0.068$ dobivenom metodom prema (Smith 2017). Također je korištena metoda ranog zaustavljanja (eng. *early stopping*) sa kriterijem stagnacije mjere *accuracy*.

5.3 Dodatna razmatranja uz podešavanje hiperparametara, evaluaciju i izbor modela

Premda su naznačeni korišteni pristupi uz izbor hiperparametara, evaluaciju i izbor modela u provedenim eksperimentima korisno je dodatno razjasniti i teorijski potkrijepiti te elemente.

5.3.1 Regularizacija

Uobičajen problem kod primjene metoda strojnog, a posebice dubokog učenja, je *overfitting*, odnosno pretjerano ugađanje modela. Vjerojatnost *overfittinga* može se znatno umanjiti primjenom metoda regularizacije, koje Goodfellow, Bengio, i Courville (2016) definiraju kao svaku izmjenu koju napravimo na algoritmu učenja koja je namijenjena smanjenju pogreške generalizacije, ali ne i pogreške na skupu za trening. Najčešće se radi o različitim metodama koje umanjuju kompleksnost ili kapacitet modela. U okvirima eksperimenata koristili smo tri metode regularizacije: rano prekidanje (eng. *early stopping*), L2 regularizacija poznata i kao kažnjavanje visokih težinskih vrijednost (eng. *weight decay*) i ispadanje neurona (eng. *dropout*).

Rano prekidanje najjednostavnija je metoda, a temelji se na prekidanju procesa treniranja prije nego on postane dovoljno kompleksan da nastupi *overfitting*. Kriteriji za prekidanje mogu biti razni, ali najčešće se radi o stagnaciji određene mjere uspješnosti modela kroz određen broj epoha. Budući da ovaj jednostavan pristup posjeduje vrlo skromne mogućnosti obično ga je potrebno koristiti u kombinaciji s nešto složenijim metodama (Goodfellow, Bengio, i Courville 2016). Kriteriji korišteni u eksperimentima su *accuracy* za uravnotežene skupove (Vegetable i Grapevineleavesdataset), dok smo za neuravnoteženi Traffic signs dataset koristili ponderirani gubitak (eng. *weighted loss*).

L2 regularizacija za određeni postotak umanjuje vrijednosti težina u svakoj iteraciji te ih na taj način postepeno približava nuli, ali one nikada ne poprime vrijednost nule. Ovo se postiže tako da funkciji gubitka dodajemo zbroj kvadrata težina pomnožen sa vrijednošću *lambda* koju nazivamo stopom regularizacije (Goodfellow, Bengio, i Courville 2016).

Ispadanje neurona (eng. *dropout*) sastoji se od nasumičnog postavljanja ulaza neurona na nulu s predodređenom vjerojatnosti, čime se oni isključuju iz procesa propuštanja podataka prema naprijed i propagacije pogreške unatrag. Na ovaj način mreža "zaboravlja" određeni dio težinskih vrijednosti, čime se umanjuje vjerojatnost *overfittinga*, te složene međusobne adaptacije neurona pa se konvolucijske slojeve prisiljava da nauče robusnije značajke (Srivastava i ostali 2014).

Izvorna arhitektura AlexNet i VGG16 modela sadrži ugrađene *dropout* slojeve pa smo u njihovom slučaju postepenim povećanjem pronašli optimalnu p vrijednost, dok je kod ostalih modela bilo potrebno dodavati *dropout* slojeve.

5.3.2 Stopa učenja (eng. *learning rate*)

Stopa učenja je parametar koji u algoritmu optimizacije određuje veličinu koraka u svakoj iteraciji dok se kreće prema minimumu funkcije gubitka. Odaberemo li previsoku vrijednost može se dogoditi da algoritam jednostavno "preskoči" minimum funkcije čime, osim otežanog pronalaženja minimuma, dobijemo visoke oscilacije vrijednosti gubitka mreže. Preniska vrijednost može znatno usporiti proces učenja zbog presporog spuštanja prema minimumu funkcije ili može uzrokovati da algoritam pronađe samo najnepovoljniji lokalni minimum. Budući da se radi o izrazito važnom parametru, Goodfellow, Bengio, i Courville (2016) navode stopu učenja kao jedini parametar kojeg vrijedi podešavati ukoliko raspolažemo ograničenim vremenom i resursima.

Postoje brojne metode podešavanja ovog hiperparametra od kojih smo, ovisno o razini izazova kojeg je određen model predstavljao, koristili četiri. Prva metoda poznata je pod šaljivim nazivom "Grad Student Descent", a sastoji se od intuitivnog podešavanja vrijednosti dok ne dobijemo zadovoljavajuće rezultate ili utrošimo raspoloživo vrijeme i resurse (Gencoglu i ostali 2019). Ovakav je pristup u iznimnim slučajevima dao dovoljno dobre rezultate uz prihvatljivu potrošnju vremena i resursa, ali za većinu modela bilo je potrebno primijeniti neku od naprednijih metoda.

Najčešće korištena metoda određivanja fiksne vrijednosti stope učenja bila je postepeno povećavanje vrijednosti unutar zadanog raspona koja se sastoji od testiranja nekoliko unaprijed određenih vrijednosti iz određenog raspona, nakon čega odabiremo vrijednost koja daje najbolji rezultat. Navedena metoda također sadrži određene elemente intuitivnog nagađanja, ali je za razliku od prethodne strukturirana.

Neovisno o metodi pomoću koje je odabrana, fiksna vrijednost stope učenja ne daje uvijek zadovoljavajuće rezultate pa je potrebno primijeniti neki od pristupa promjenjive stope učenja. Jedna od jednostavnijih metoda sastoji se od postepenog smanjivanja vrijednosti kroz proces treniranja. Temeljna je ideja da u početku učenja, dok su parametri mreže još uvijek daleko od optimalnih vrijednosti, može ubrzati proces primjenom relativno visoke stope učenja, a kako se pogreška mreže približava minimalnoj vrijednosti postepeno se smanjuje stopa učenja. U našem slučaju koristili smo ExponentialLR funkciju iz PyTorch okvira, koja smanjuje stopu učenja za određeni faktor u svakoj epohi.

Nešto složeniji pristup ovom problemu je ciklička stopa učenja (eng. *cyclic learning rate*) kojom ciklički povećavamo i smanjujemo stopu učenja kroz određeni raspon. Metoda se javila kao odgovor na slučajeve kada optimizacijski algoritam zapne na nepovoljnom

lokalnom minimumu funkcije gubitka. Naglim povećanjem stope učenja nakon što smo je kroz nekoliko koraka smanjivali može se postići da algoritam "iskoči" iz područja lokalnog minimuma, čime se olakšava pronalaženje nižeg lokalnog ili, u idealnom slučaju, globalnog minimuma.

Kod primjene ove metode važno je pronaći odgovarajući raspon vrijednosti, a u ovu svrhu primijenili smo jednostavan pristup kojeg predlažu njeni autori. Pristup se sastoji od probnog treniranja modela uz postepeno povećavanje stope učenja dok vrijednosti gubitka ne počnu naglo rasti. Stopa učenja u trenutku prije naglog rasta, kada je vrijednost gubitka najmanja, predstavlja maksimalnu vrijednost stope učenja iz našeg raspona. Dijeljenjem ove vrijednost sa 3 ili 4 dobit ćemo minimalnu preporučenu vrijednost (Smith 2017; 2018).

5.3.3 Normalizacija mini-grupe (eng. *batch normalization*)

Normalizacijom u općem smislu nazivamo postupak promjene vrijednosti numeričkih varijabli kojom one poprimaju odgovarajuću vrijednost iz unaprijed određenog raspona (primjerice, između 0 i 1). Normalizacija mini-grupe ili normalizacija serije jedna je od tehnika treniranja dubokih neuronskih mreža kojom se normalizira izlaz nekog sloja za svaku grupu (eng. *batch*) ulaznih podataka. Primjenom ove metode izbjegava se problem unutarnjeg kovarijantnog pomaka, čime možemo postići značajnu stabilizaciju procesa učenja, te smanjenje broja epoha potrebnih za postizanje iste razine treniranosti. Metoda također u određenoj mjeri djeluje regularizacijski pa može, primjerice, omogućiti primjenu nižih vrijednosti kod primjene metode ispadanja neurona, a u nekim slučajevima potpuno ukloniti potrebu primjene *dropout* (Ioffe i Szegedy 2015).

5.3.4 Validacija, evaluacija i izbor modela

Kako je već navedeno u cjelini 5.1, izrazita računalna zahtjevnost korištenih tradicionalnih metoda onemogućila je bilo kakvo značajnije podešavanje hiperparametara, stoga je ovaj dio eksperimenata proveden s preporučenim vrijednostima bez potrebe za posebnim izborom modela.

Metode validacije, evaluacije i izbora modela u okvirima eksperimenta s modelima dubokog učenja odabrane su sukladno osobinama pojedinog skupa podataka. "**Vegetable image dataset**" u izvornom je obliku podijeljen na trening, validacijski i testni skup podataka, stoga

je u njegovom slučaju validacija provedena na validacijskom skupu, a test odabranog modela na testnom skupu. Odabran je model sa najboljim mjerama uspješnosti na validacijskom skupu.

Budući da je "**Traffic Sign Classification and Recognition**" neuravnotežen i sastoji se od znatno manjeg broja uzoraka, za njegovu je validaciju odabrana stratificirana peterostruka unakrsna validacija (eng. *stratified 5-fold cross-validation*), koja zadržava omjere veličina pojedinih klasa. Odabran je model sa najboljim srednjim vrijednostima mjera uspješnosti dobivenima peterostrukom unakrsnom validacijom.

"**Grapevine Leaves Image Dataset**" je savršeno uravnotežen skup, ali sadrži najmanji broj instanci, te je u njegovom slučaju također primijenjena stratificirana peterostruka unakrsna validacija, s obzirom da je blaga neuravnoteženost veličine klasa kojom obično rezultira konvencionalan pristup unakrsnoj validaciji bila znatno izraženija zbog malog broja uzoraka. Odabran je model sa najboljim srednjim vrijednostima mjera uspješnosti dobivenima peterostrukom unakrsnom validacijom.

Kao mjere evaluacije modela korištene su točnost (accuracy), F1-score i ROC AUC.

6 Rezultati i rasprava

Na sljedećim stranicama prezentirati ćemo, usporediti i raspraviti rezultate eksperimenata provedenih korištenjem tradicionalnih metoda strojnog učenja i suvremenih modela dubokog učenja na klasifikaciju slika.

6.1 Usporedba tradicionalnih modela

6.1.1 Rezultati klasifikacije "Vegetable image dataset" skupa podataka

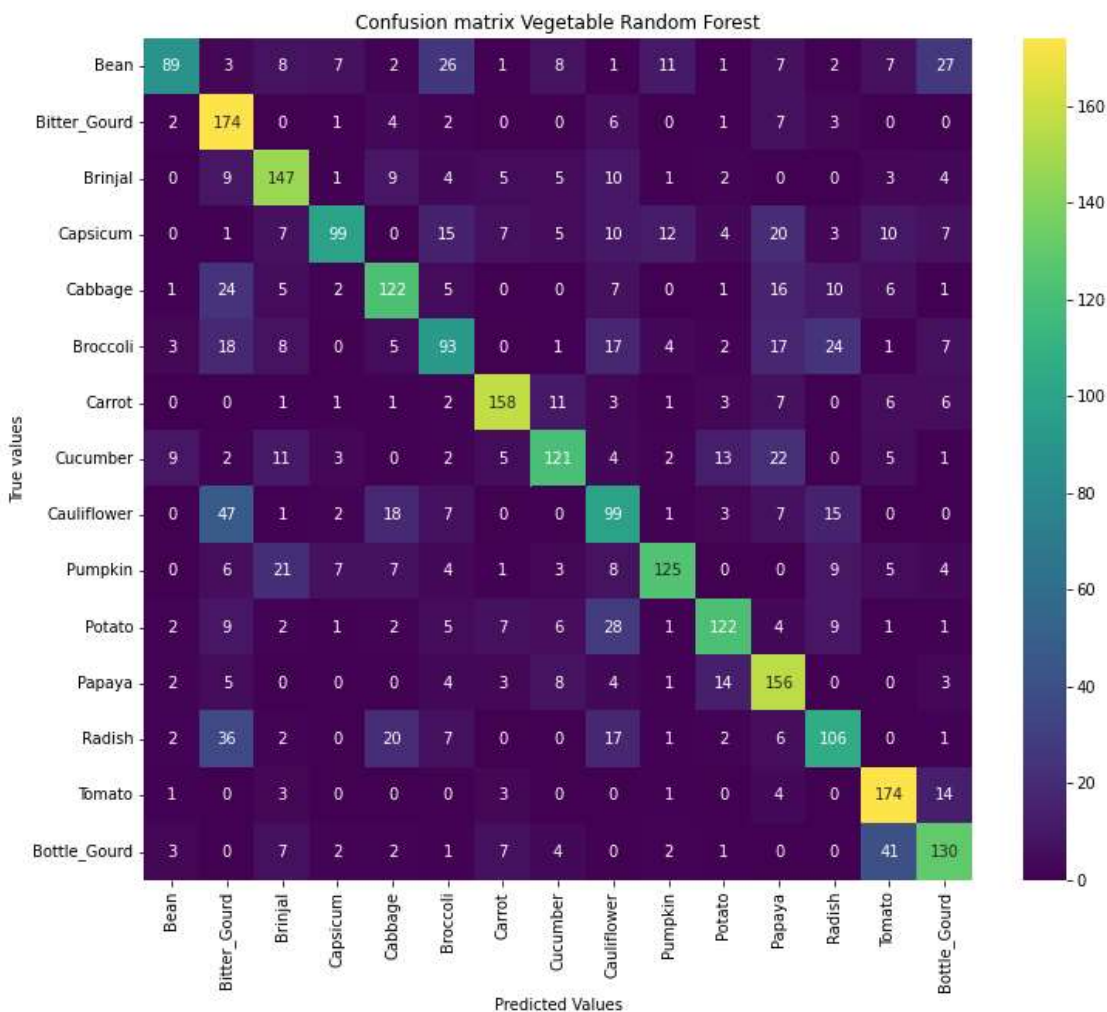
"Vegetable image dataset" najveći je od tri korištena skupa, te iz tog razloga predstavlja poseban izazov u radu s računalno zahtjevnim metodama kao što su ekstrakcija HOG i LBP značajki i treniranje SVM modela. Ekstrakcija značajki pokazala se izrazito memorijski zahtjevnom, što vjerojatno najbolje ilustrira činjenica da je njen konačni rezultat *numpy array* dimenzija (15000, 76420) ukupne veličine 8.54 GB. Treniranje SVM modela na takvom skupu podataka trajalo bi pretjerano dugo s obzirom na raspoložive resurse, stoga je za ovaj dio eksperimenta primijenjena ThunderSVM biblioteka koja znatno ubrzava proces korištenjem grafičkih procesorskih jedinica. Kako je ranije spomenuto, usprkos značajnom ubrzanju ipak je treniranje modela na ovom skupu podataka bilo dovoljno dugotrajno da onemogućiti bilo kakvo značajnije podešavanje hiperparametara. Treniranje RandomForest modela proteklo je znatno brže, odnosno za razliku od ThunderSVM modela za čije su treniranje bili potrebni sati, RF model je bio spreman za nekoliko minuta.

Rezultati klasifikacije daju značajnu prednost ThunderSVM modelu prema svim mjerama uspješnosti, izuzev ROC AUC (Tablica 2).

Tablica 2 Prikaz rezultata ThunderSVM modela i RF modela pri kvalifikaciji Vegetable Image skupa podataka

| ThunderSVM | | | | | RandomForest | | | | |
|----------------|--------------------|--------|----------|---------|----------------|--------------------|--------|----------|---------|
| | Precision | recall | f1-score | support | | Precision | recall | f1-score | support |
| Bean | 0.81 | 0.70 | 0.75 | 200 | Bean | 0.78 | 0.45 | 0.57 | 200 |
| Bitter_Gourd | 0.71 | 0.81 | 0.75 | 200 | Bitter_Gourd | 0.52 | 0.87 | 0.65 | 200 |
| Brinjal | 0.73 | 0.82 | 0.77 | 200 | Brinjal | 0.66 | 0.73 | 0.70 | 200 |
| Capsicum | 0.76 | 0.68 | 0.72 | 200 | Capsicum | 0.79 | 0.49 | 0.61 | 200 |
| Cabbage | 0.66 | 0.70 | 0.68 | 200 | Cabbage | 0.64 | 0.61 | 0.62 | 200 |
| Broccoli | 0.63 | 0.73 | 0.68 | 200 | Broccoli | 0.53 | 0.47 | 0.49 | 200 |
| Carrot | 0.84 | 0.86 | 0.85 | 200 | Carrot | 0.80 | 0.79 | 0.80 | 200 |
| Cucumber | 0.74 | 0.70 | 0.72 | 200 | Cucumber | 0.70 | 0.60 | 0.65 | 200 |
| Cauliflower | 0.66 | 0.61 | 0.64 | 200 | Cauliflower | 0.46 | 0.49 | 0.48 | 200 |
| Pumpkin | 0.79 | 0.75 | 0.77 | 200 | Pumpkin | 0.77 | 0.62 | 0.69 | 200 |
| Potato | 0.77 | 0.78 | 0.78 | 200 | Potato | 0.72 | 0.61 | 0.66 | 200 |
| Papaya | 0.71 | 0.81 | 0.76 | 200 | Papaya | 0.57 | 0.78 | 0.66 | 200 |
| Radish | 0.81 | 0.65 | 0.72 | 200 | Radish | 0.59 | 0.53 | 0.56 | 200 |
| Tomato | 0.85 | 0.88 | 0.87 | 200 | Tomato | 0.67 | 0.87 | 0.76 | 200 |
| Bottle_Gourd | 0.89 | 0.81 | 0.85 | 200 | Bottle_Gourd | 0.63 | 0.65 | 0.64 | 200 |
| accuracy | | | 0.75 | 3000 | accuracy | | | 0.64 | 3000 |
| macroavg | 0.76 | 0.75 | 0.75 | 3000 | macroavg | 0.65 | 0.64 | 0.64 | 3000 |
| weightedavg | 0.76 | 0.75 | 0.75 | 3000 | weightedavg | 0.65 | 0.64 | 0.64 | 3000 |
| ROC_AUC_SCORE: | 0.4661172619047619 | | | | ROC_AUC_SCORE: | 0.9360444047619046 | | | |

Usporedba modela temelji se na tri mjere uspješnosti koje na zadovoljavajući način oslikavaju njihove ukupne performanse (accuracy, F1 Score i ROC AUC), ali važno je napomenuti kako paket ScikitLearn omogućava znatno detaljniji pregled rezultata klasifikacije pa tako gornji primjer prikazuje relevantne mjere uspješnosti za svaku pojedinu klasu pomoću modula *sklearn.metrics.classification_report*. Navedeni prikaz moguće je po potrebi nadopuniti korištenjem *sklearn.metrics.confusion_matrix* modula koji daje prikaz pripadajuće matrice konfuzije koja rezultate klasifikacije daje u vidu matrice stvarnih pripadnosti klasi i onih koje je model predvidio. Radi se o korisnom alatu koji omogućava detaljan uvid u broj ispravno i neispravno klasificiranih uzoraka za svaku pojedinu klasu. Primjerice, iz matrice konfuzije prikazane na Slici 23 vidljivo je da je model temeljen na RandomForest algoritmu ispravno klasificirao 125 uzoraka klase *Pumpkin*, ali je isto tako 21 uzorak iste klase pogrešno klasificirao kao *Brinjal*.



Slika 23 Matrica konfuzije rezultata klasifikacije RandomForest modela na VegetableImageskupu podataka

6.1.2 Rezultati klasifikacije "Traffic Sign Classification and Recognition" skupa podataka

"Traffic Sign Classification and Recognition" značajno je manji u usporedbi sa Vegetableimagedataset skupom, ali je također predstavljao izazov u radu s računalno zahtjevnim metodama kao što su ekstrakcija HOG i LBP značajki i treniranje SVM modela. Rezultat ekstrakcije značajki u ovom je slučaju *numpy array* dimenzija (6024, 76420) ukupne veličine 3.43 GB. Rezultati klasifikacije daju značajnu prednost ThunderSVM modelu prema svim mjerama uspješnosti, izuzev ROC AUC (Tablica 3). Budući da se radi o neuravnoteženom skupu podataka mjera *accuracy* vrlo je ograničene korisnosti, stoga pri izboru modela treba obratiti pažnju na ostale mjere uspješnosti ovisno o zahtjevima koje pred njega postavlja praktična upotreba.

Tablica 3 Prikaz rezultata ThunderSVM modela i RF modela pri kvalifikaciji Traffic Sign Classification and Recognition skupa podataka

| ThunderSVM | | | | | Random Forest | | | | |
|----------------|--------------------|--------|----------|---------|----------------|--------------------|--------|----------|---------|
| | Precision | recall | f1-score | support | | Precision | recall | f1-score | support |
| GuideSign | 0.78 | 0.98 | 0.87 | 62 | GuideSign | 0.82 | 0.95 | 0.88 | 62 |
| M1 | 0.85 | 0.79 | 0.81 | 14 | M1 | 1.00 | 0.14 | 0.25 | 14 |
| M4 | 0.94 | 0.92 | 0.93 | 169 | M4 | 0.81 | 0.97 | 0.88 | 169 |
| M5 | 0.80 | 0.67 | 0.73 | 12 | M5 | 1.00 | 0.33 | 0.50 | 12 |
| M6 | 1.00 | 0.25 | 0.40 | 8 | M6 | 1.00 | 0.12 | 0.22 | 8 |
| P12 | 1.00 | 0.84 | 0.91 | 25 | P12 | 0.94 | 0.68 | 0.79 | 25 |
| P1 | 0.87 | 0.93 | 0.90 | 14 | P1 | 1.00 | 0.57 | 0.73 | 14 |
| M7 | 1.00 | 1.00 | 1.00 | 6 | M7 | 1.00 | 0.83 | 0.91 | 6 |
| P10_50 | 1.00 | 1.00 | 1.00 | 6 | P10_50 | 1.00 | 1.00 | 1.00 | 6 |
| W1 | 0.88 | 0.88 | 0.88 | 8 | W1 | 1.00 | 0.62 | 0.77 | 8 |
| Accuracy | | | 0.90 | 324 | accuracy | | | 0.84 | 324 |
| macro avg | 0.91 | 0.82 | 0.84 | 324 | macro avg | 0.96 | 0.62 | 0.69 | 324 |
| weightedavg | 0.90 | 0.90 | 0.89 | 324 | weightedavg | 0.86 | 0.84 | 0.81 | 324 |
| ROC_AUC_SCORE: | 0.6582919307875934 | | | | ROC_AUC_SCORE: | 0.9516291840907206 | | | |

6.1.3 Rezultati klasifikacije "Grapevine Image" skupa podataka

"Grapevine Leaves Image Dataset" pokazao se vrlo problematičnim ponajviše zbog malog broja uzoraka i visoke međuklasne sličnosti. Računalna zahtjevnost bila je znatno manje izražena, što je vidljivo iz rezultata ekstrakcije značajki koji je u ovom slučaju *numpyarray* dimenzija (400, 76420) ukupne veličine 232.2MB. Rezultati klasifikacije prema svim mjerama uspješnosti ne zadovoljavaju kriterije nijedne praktične primjene, neovisno o odabranom modelu (Tablica 4).

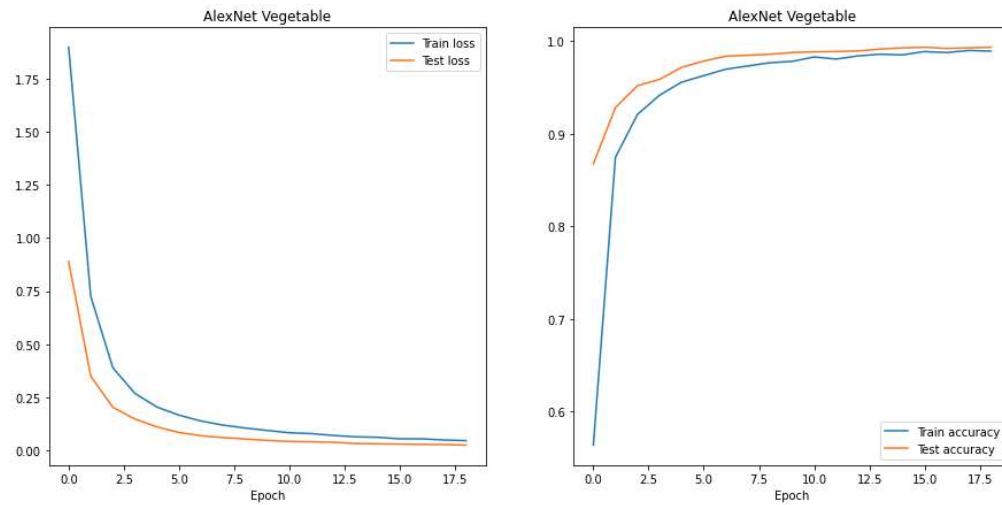
Tablica 4 Prikaz rezultata ThunderSVM modela i RF modela pri kvalifikaciji Grapevine Image skupa podataka

| ThunderSVM | | | | | Random Forest | | | | |
|----------------|-----------|--------|----------|---------|----------------|-----------|--------|----------|---------|
| | Precision | recall | f1-score | support | | Precision | recall | f1-score | support |
| Ala_Idris | 0.19 | 0.15 | 0.17 | 20 | Ala_Idris | 0.55 | 0.30 | 0.39 | 20 |
| Ak | 0.59 | 0.50 | 0.54 | 20 | Ak | 0.50 | 0.50 | 0.50 | 20 |
| Nazli | 0.41 | 0.35 | 0.38 | 20 | Nazli | 0.43 | 0.45 | 0.44 | 20 |
| Buzgulu | 0.42 | 0.65 | 0.51 | 20 | Buzgulu | 0.40 | 0.50 | 0.44 | 20 |
| Dimnit | 0.11 | 0.10 | 0.10 | 20 | Dimnit | 0.48 | 0.55 | 0.51 | 20 |
| accuracy | | | 0.35 | 100 | accuracy | | | 0.46 | 100 |
| macro avg | 0.34 | 0.35 | 0.34 | 100 | macro avg | 0.47 | 0.46 | 0.46 | 100 |
| weightedavg | 0.34 | 0.35 | 0.34 | 100 | weightedavg | 0.47 | 0.46 | 0.46 | 100 |
| ROC_AUC_SCORE: | 0.456125 | | | | ROC_AUC_SCORE: | 0.7866875 | | | |

6.2 Rezultati DL modela

6.2.1 Rezultati klasifikacije temeljene na AlexNet modelu

"Vegetable image dataset" - treniranje konačnog AlexNet modela trajalo je ukupno 18 epoha s vrlo stabilnim opadanjem vrijednosti gubitka i porastom mjere accuracy, a rezultiralo je dobro ugođenim modelom. Krivulje za loss i accuracy po epohama su prikazane na Slici 23.



Slika 24 Loss i accuracy krivulje za Vegetable image skupu podataka (rezultati za treningi testni skup podataka)

Rezultati klasifikacije na testnom skupu podataka gotovo su savršeni prema svim mjerama uspješnosti (accuracy, F1 Score i ROC AUC iznose iznad 0.99):

Tablica 5 Zbirni prikaz rezultata kvalifikacije na Vegetable Image skupu podataka za model temeljen na AlexNet modelu

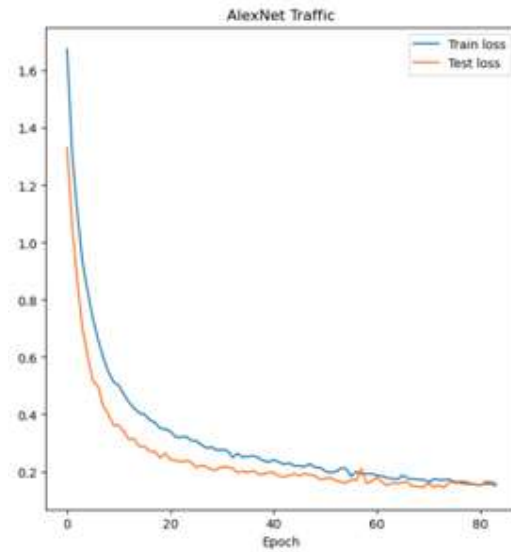
```

-----
AlexNetVegetabledataset
-----
ROC_AUC:0.9999134523809523
F1Score: 0.9933363409808055
Accuracy: 0.9933333333333333

```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Bean | 0.98 | 1.00 | 0.99 | 200 |
| Bitter_Gourd | 1.00 | 0.99 | 0.99 | 200 |
| Bottle_Gourd | 1.00 | 0.99 | 1.00 | 200 |
| Brinjal | 0.99 | 1.00 | 0.99 | 200 |
| Broccoli | 0.99 | 0.98 | 0.99 | 200 |
| Cabbage | 0.99 | 0.99 | 0.99 | 200 |
| Capsicum | 0.99 | 0.99 | 0.99 | 200 |
| Carrot | 1.00 | 1.00 | 1.00 | 200 |
| Cauliflower | 0.99 | 0.99 | 0.99 | 200 |
| Cucumber | 1.00 | 0.98 | 0.99 | 200 |
| Papaya | 0.98 | 0.99 | 0.99 | 200 |
| Potato | 1.00 | 1.00 | 1.00 | 200 |
| Pumpkin | 0.99 | 0.98 | 0.99 | 200 |
| Radish | 1.00 | 1.00 | 1.00 | 200 |
| Tomato | 0.99 | 0.98 | 0.99 | 200 |
| accuracy | | | 0.99 | 3000 |
| macroavg | 0.99 | 0.99 | 0.99 | 3000 |
| weightedavg | 0.99 | 0.99 | 0.99 | 3000 |

"Traffic Sign Classification and Recognition" - za treniranje konačnog AlexNet modela bile su potrebne 83 epohe. Vrijednost ponderiranog gubitka (*weighted loss*) je u stabilnom padu tijekom čitavog vremena treniranja, koje je rezultiralo dobro ugođenim modelom. Budući da se radi o neuravnoteženom skupu, razmatrali smo samo krivulju *weightedloss* vrijednosti po epohama prikazanu na Slici 24.



Slika 25 Loss krivulja za Traffic Sign Classification and Recognition dataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka proizvela je nešto slabije, ali još uvijek izvrsne rezultate prema svim mjeramauspješnosti (Tablica 6):

Tablica 6 Zbirni prikaz rezultata klasifikacije na Traffic Signs Classification and Recognition skupu podataka za model temeljen na AlexNet modelu

```

-----
AlexNetTrafficdataset
-----
ROC_AUC:0.9919630615344639
F1Score: 0.9380787023230053
Accuracy: 0.9691358024691358

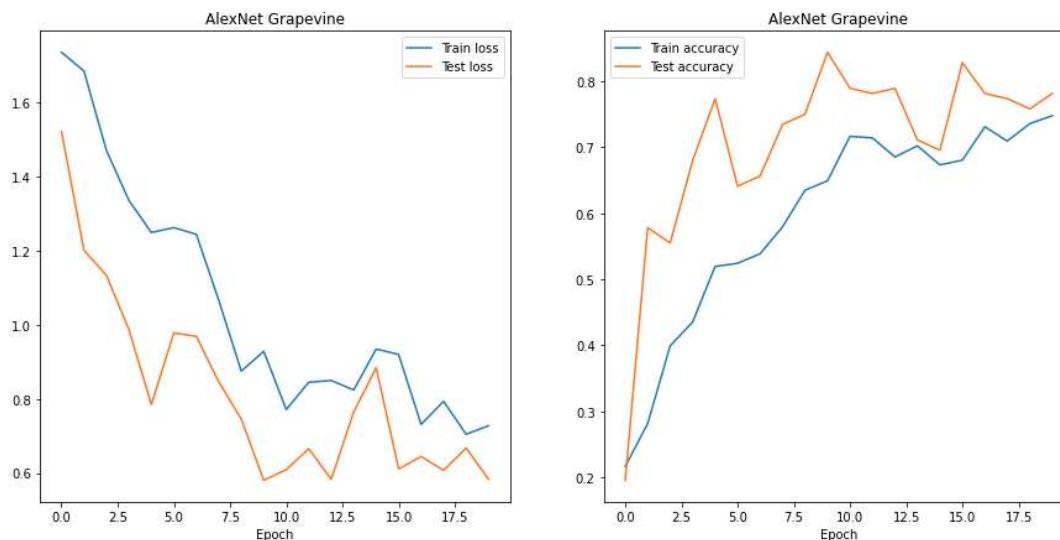
      precision    recall  f1-score   support

GuideSign      1.00      1.00      1.00         62
      M1         0.65      0.93      0.76         14
      M4         0.99      0.98      0.99        169
      M5         0.92      0.92      0.92         12
      M6         1.00      0.62      0.77          8
      M7         1.00      0.96      0.98         25
      P1         0.93      1.00      0.97         14
      P10_50     1.00      1.00      1.00          6
      P12         1.00      1.00      1.00          6
      W1         1.00      1.00      1.00          8

accuracy                0.97         324
macroavg                0.95         324
weightedavg            0.98         324

```


"Grapevine leaves image dataset" – ovaj skup podataka karakterizira izrazito mali broj uzoraka i visoka međuklasna sličnost, što znatno otežava postizanje dobre ugođenosti modela. Treniranje konačnog AlexNet modela trajalo je 19 epoha s relativno nestabilnim, ali sigurnim opadanjem vrijednosti gubitka i porastom mjere accuracy. Nestabilnost bi se mogla dijelom pripisati visokoj maksimalnoj stopi učenja. Ugođenost modela je vidljivo slabija nego za prethodna dva skupa podataka. Krivulje su prikazane na Slici 25.



Slika 26 Loss i accuracy krivulje za Grape vine leaves image dataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka bila je nešto manje uspješna u usporedbi sa ostalim skupovima, prema svim mjerama uspješnosti (Tablica 7):

Tablica 7 Zbirni prikaz rezultata klasifikacije na GrapevineImageDataskupu podataka za model temeljen na AlexNet modelu

```

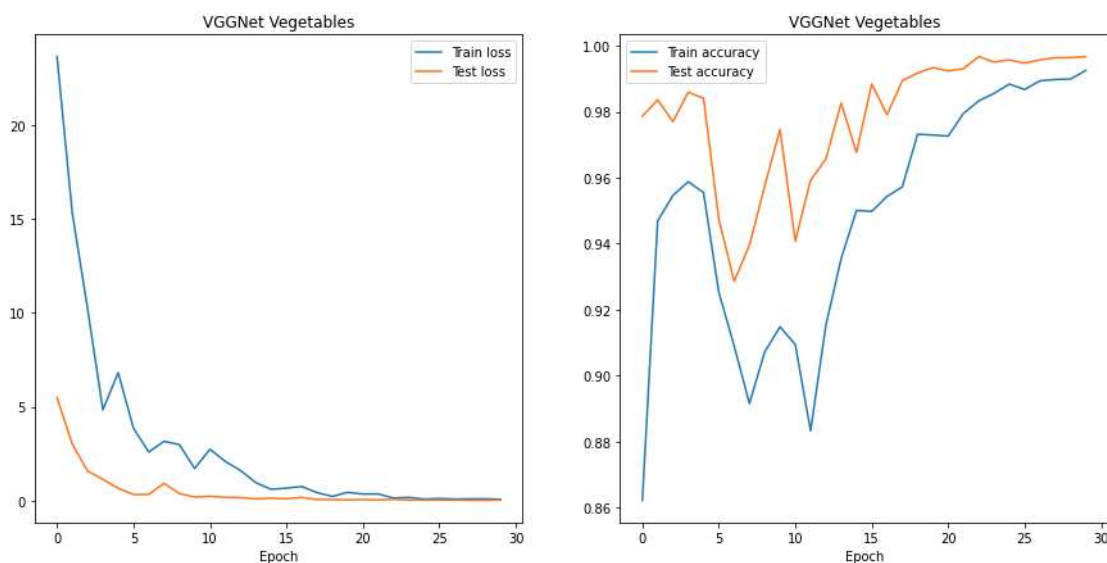
-----
AlexNetGrapevinedataset
-----
ROC_AUC:0.98475
F1Score: 0.9194544948543815
Accuracy: 0.92

```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Ak | 0.95 | 0.90 | 0.92 | 20 |
| Ala_Idris | 0.94 | 0.80 | 0.86 | 20 |
| Buzgulu | 0.83 | 0.95 | 0.88 | 20 |
| Dimnit | 0.95 | 0.95 | 0.95 | 20 |
| Nazli | 0.95 | 1.00 | 0.98 | 20 |
| accuracy | | | 0.92 | 100 |
| macro avg | 0.92 | 0.92 | 0.92 | 100 |
| weighted avg | 0.92 | 0.92 | 0.92 | 100 |

6.2.2 Rezultati klasifikacije temeljene na VGG16 modelu

"Vegetable image dataset" - treniranje konačnog VGG16 modela trajalo je ukupno 28 epoha sa nestabilnim opadanjem vrijednosti gubitka i porastom mjere accuracy kroz prvu trećinu procesa treniranja, ali se postepeno stabilizira i rezultira dobro ugođenim modelom. Uzmemo li u obzir da je u ovom slučaju korišten ExponentialLR, nestabilnost u početku treniranja možemo pripisati visokoj početnoj stopi učenja. Krivulje za loss i accuracy po epohama su prikazane na Slici 27.



Slika 27 Loss i accuracy krivulje za VegetableImageDataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka daje gotovosavršene rezultate prema svim mjerama uspješnosti (accuracy, F1 Score i ROC AUC iznose iznad 0.99) (Tablica 8).

Tablica 8 Zbirni prikaz rezultata kvalifikacije na VegetableImage skupu podataka za model temeljen na VGG16 modelu

```

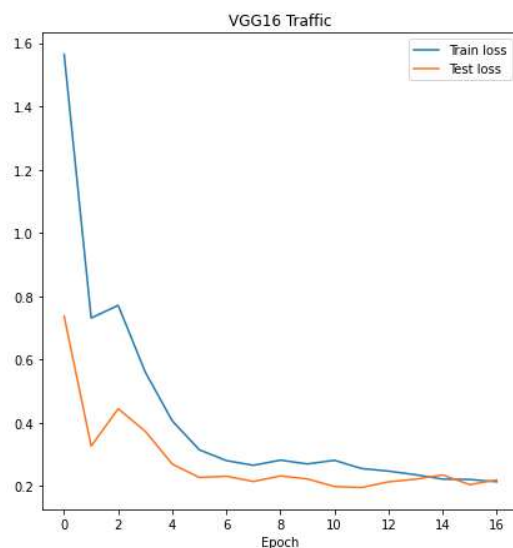
-----
VGG16 Vegetabledataset
-----
ROC_AUC: 0.9999669047619048
F1Score: 0.9956682249748947
Accuracy: 0.9956666666666667

precisionrecall  f1-score  support
      Bean      0.99      0.99      0.99      200
Bitter_Gourd    1.00      0.99      0.99      200
Bottle_Gourd    0.99      1.00      0.99      200
  Brinjal      0.99      0.99      0.99      200
   Broccoli    1.00      0.98      0.99      200
    Cabbage    1.00      1.00      1.00      200
   Capsicum    1.00      0.99      1.00      200
     Carrot    1.00      1.00      1.00      200
Cauliflower    1.00      1.00      1.00      200
   Cucumber    0.99      0.98      0.99      200
    Papaya    0.99      0.99      0.99      200
    Potato    1.00      1.00      1.00      200
   Pumpkin    1.00      1.00      1.00      200
    Radish    1.00      1.00      1.00      200
    Tomato    1.00      1.00      1.00      200

accuracy
macro avg      1.00      1.00      1.00      3000
weighted avg    1.00      1.00      1.00      3000

```

"Traffic Sign Classification and Recognition" - za treniranje konačnog VGG16 modela bilo je potrebno 16 epoha. Vrijednost ponderiranog gubitka (*weighted loss*) je u relativno stabilnom padu tijekom čitavog vremena treniranja, koje je rezultiralo dobro ugođenim modelom. Budući da se radi o neuravnoteženom skupu, razmatrali smo samo krivulju *weighted loss* vrijednosti po epohama prikazanu na Slici 28.



Slika 28 Loss krivulja za Traffic Sign Classification and Recognition dataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka proizvela je nešto slabije, ali još uvijek jako dobre rezultate prema svim mjerama uspješnosti (Tablica 9).

Tablica 9 Zbirni prikaz rezultata kvalifikacije na Traffic Signs Classification and Rescognition skupu podataka za model temeljen na VGG16 modelu

```

-----
VGG16 Traffic dataset
-----

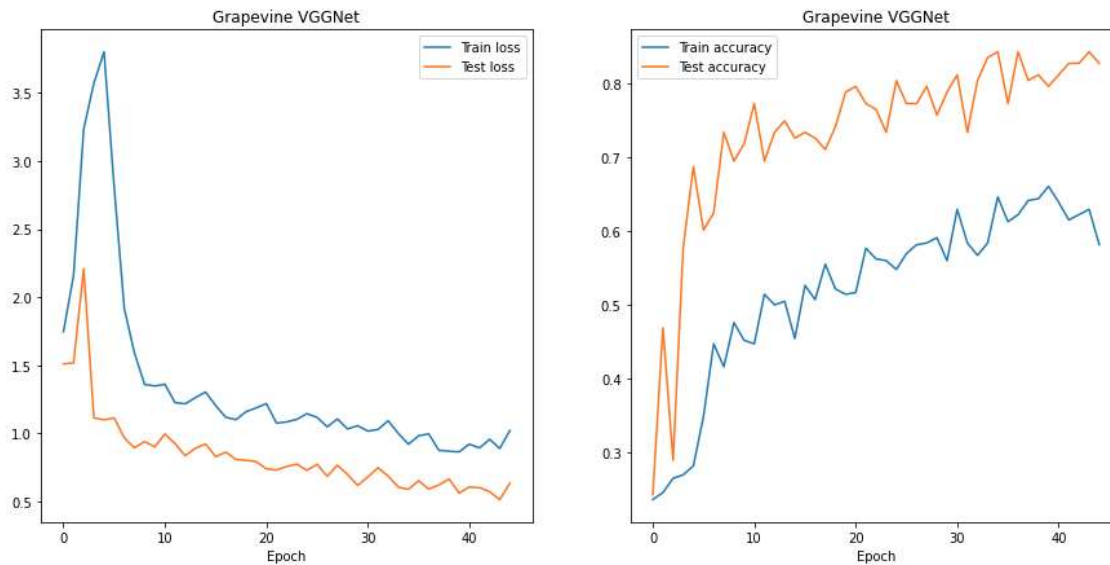
ROC_AUC:0.9871363275907182
F1Score: 0.925306852620921
Accuracy: 0.9567901234567902

Precision    recall    f1-score    support
GuideSign    0.98      0.95      0.97      62
M1            0.73      0.79      0.76      14
M4            0.95      0.99      0.97      169
M5            0.92      0.92      0.92      12
M6            1.00      0.62      0.77      8
M7            1.00      0.92      0.96      25
P1            1.00      1.00      1.00      14
P10_50       1.00      0.83      0.91      6
P12          1.00      1.00      1.00      6
W1           1.00      1.00      1.00      8

accuracy     0.96      0.96      324
macro avg    0.96      0.90      0.93      324
weighted avg 0.96      0.96      0.96      324

```

"Grapevine Leaves Image Dataset" – treniranje modela na ovom skupu otežano neovisno o odabranom modelu, ali VGG16 pokazao se izrazito problematičnim. Uzrok pripisujemo činjenici kako se radi o najkompleksnijem od korištenih modela (sastoji se od ukupno 138 milijuna parametara) što ga čini nepovoljnim za rad s malim skupovima podataka. Treniranje konačnog modela trajalo je 44 epohe s relativno nestabilnim opadanjem vrijednosti gubitka i porastom mjere accuracy. Ugođenost modela nije optimalna s vidnim znakovima *underfittinga* kojeg pripisujemo agresivnoj regularizaciji (*dropout* s visokim p vrijednostima). Krivulje su prikazane na Slici 29.



Slika 29 Loss i accuracy krivulje za Grapevine Leaves Image dataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka proizvela je nešto slabije rezultate prema svim mjerama uspješnosti (Tablica 10).

Tablica 10 Zbirni prikaz rezultata klasifikacije na GrapevineImageDataset skupu podataka za model temeljen na VGG16 modelu

```

-----
VGG16 Grapevinedataset
-----
ROC_AUC:0.9566250000000001
F1Score: 0.7970771281485568
Accuracy: 0.8

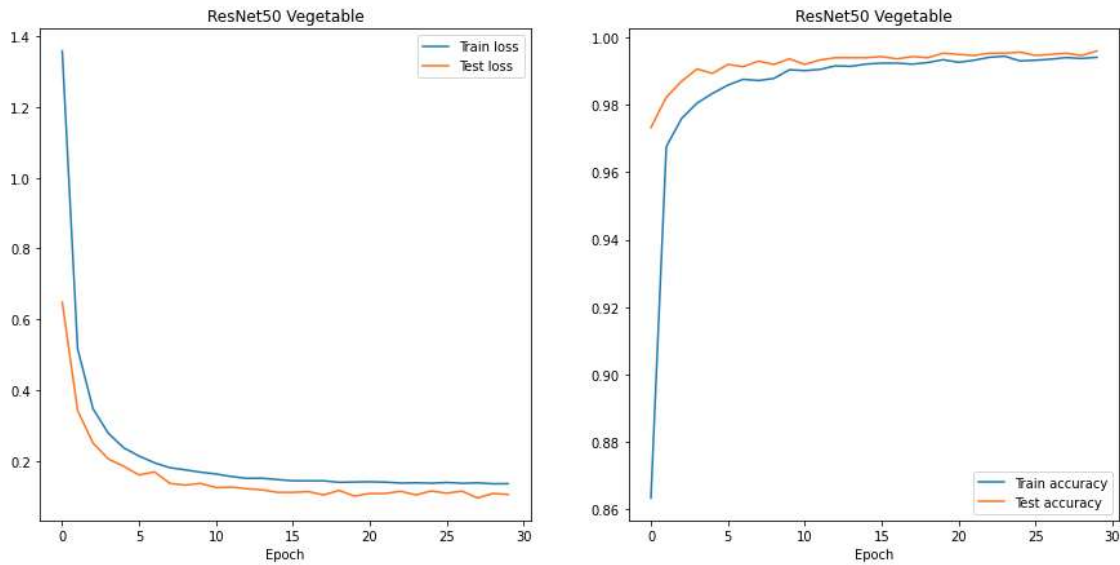
precisionrecall  f1-score  support
      Ak         0.62     0.90     0.73     20
    Ala_Idris    0.92     0.55     0.69     20
      Buzgulu    0.80     0.80     0.80     20
      Dimnit     0.88     0.75     0.81     20
      Nazli      0.91     1.00     0.95     20

  accuracy
macro avg    0.83     0.80     0.80     100
weighted avg 0.83     0.80     0.80     100

```

6.2.3 Rezultati klasifikacije temeljene na ResNet50 modelu

"Vegetable image dataset" - treniranje konačnog ResNet50 modela rezultiralo je dobro ugođenim modelom, a trajalo je 29 epoha sa stabilnim opadanjem vrijednosti gubitka i porastom mjere accuracy. Krivulje za loss i accuracy po epohama su prikazane na Slici 30.



Slika 30 Loss i accuracy krivulje za Vegetable Images dataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka daje gotovo savršene rezultate prema svim mjerama uspješnosti (accuracy, F1 Score i ROC AUC iznose iznad 0.99) (Tablica 11).

Tablica 11 Zbirni prikaz rezultata klasifikacije na Vegetable Images dataset skupu podataka za model temeljen na ResNet50 modelu

```

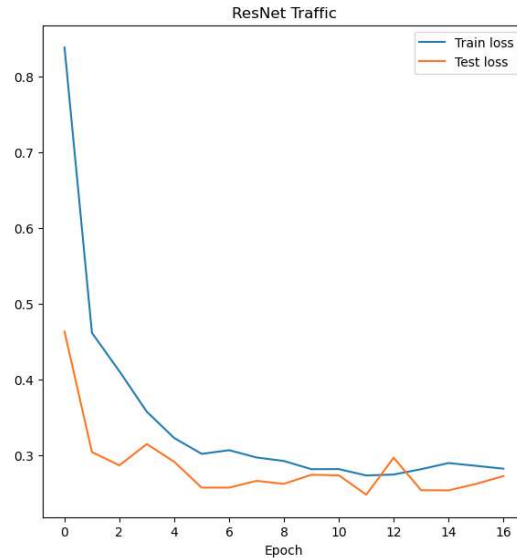
-----
ResNet50 Vegetabledataset
-----
ROC_AUC:0.9999530952380952
F1Score: 0.997667481223684
Accuracy: 0.9976666666666667
Precision    recall    f1-score    support
Bean         1.00     1.00     1.00     200
Bitter_Gourd 1.00     0.99     0.99     200
Bottle_Gourd 1.00     0.99     1.00     200
Brinjal      1.00     1.00     1.00     200
Broccoli     1.00     1.00     1.00     200
Cabbage      1.00     0.99     1.00     200
Capsicum     1.00     1.00     1.00     200
Carrot       1.00     1.00     1.00     200
Cauliflower  1.00     1.00     1.00     200
Cucumber     0.99     0.99     0.99     200
Papaya       0.99     0.99     0.99     200
Potato       1.00     1.00     1.00     200
Pumpkin      1.00     1.00     1.00     200
Radish       1.00     1.00     1.00     200
Tomato       1.00     1.00     1.00     200

accuracy                    1.00     3000
macro avg                   1.00     1.00     1.00     3000
weighted avg                 1.00     1.00     1.00     3000

```

"Traffic Sign Classification and Recognition" - za treniranje konačnog ResNet50 modela bilo je potrebno 16 epoha. Vrijednost ponderiranog gubitka (*weighted loss*) jeu relativno

stabilnom padu tijekom čitavog vremena treniranja, koje je rezultiralo dobro ugođenim modelom. Budući da se radi o neuravnoteženom skupu, razmatrali smo samo krivulju weightedloss vrijednosti po epohama prikazanu na Slici 31.



Slika 31 Loss krivulja za Traffic Sign Classification and Recognitiondataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka i u ovom je slučaju proizvela nešto slabije rezultate u usporedbi sa Vegetable Image Datasetom, ali još uvijek mogu biti opisani kao vrlo dobri prema svim mjerama uspješnosti (Tablica 12).

Tablica 12 Zbirni prikaz rezultata klasifikacije na Traffic Signs Classification and Recognition dataset skupu podataka za model temeljen na ResNet50 modelu

```

-----
ResNet50 Trafficdataset
-----

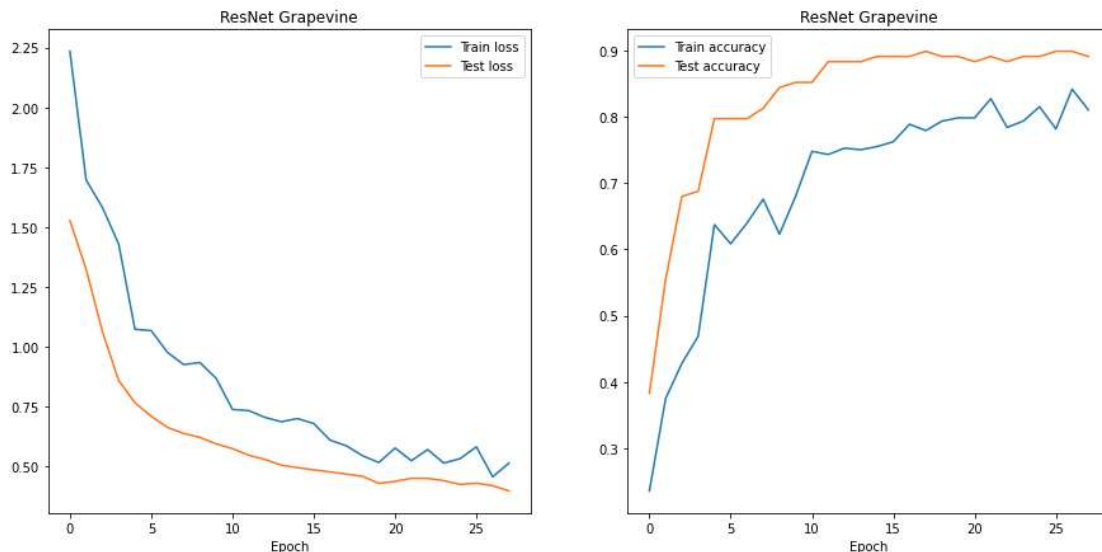
ROC_AUC:0.9890917731314136
F1Score: 0.9027282115951355
Accuracy: 0.9320987654320988

Precision    recall    f1-score    support
GuideSign    0.91      0.98      0.95        62
M1            0.71      0.71      0.71        14
M4            0.97      0.94      0.95        169
M5            0.92      1.00      0.96        12
M6            0.83      0.62      0.71        8
M7            0.95      0.84      0.89        25
P1            0.82      1.00      0.90        14
P10_50       1.00      1.00      1.00        6
P12          1.00      1.00      1.00        6
W1           0.89      1.00      0.94        8

accuracy     0.93      324
macro avg    0.90      0.91      0.90      324
weighted avg 0.93      0.93      0.93      324

```

"Grapevine Leaves Image Dataset" – treniranje konačnog modela trajalo je 27 epoha s relativno stabilnim opadanjem vrijednosti gubitka i porastom mjere accuracy, a rezultiralo je dobro ugođenim modelom. Krivulje su prikazane na Slici 32.



Slika 32 Loss i accuracy krivulje za Grapevine Leaves Image dataset (rezultati za trening i testni skup podataka)

Klasifikacija na testnom skupu podataka proizvela je nešto slabije rezultate u usporedbi sa ostalim skupovima podataka, prema svim mjerama uspješnosti (Tablica 13).

Tablica 13 Zbirni prikaz rezultata klasifikacije na Grapevine Image Dataset skupu podataka za model temeljen na ResNet50 modelu

```

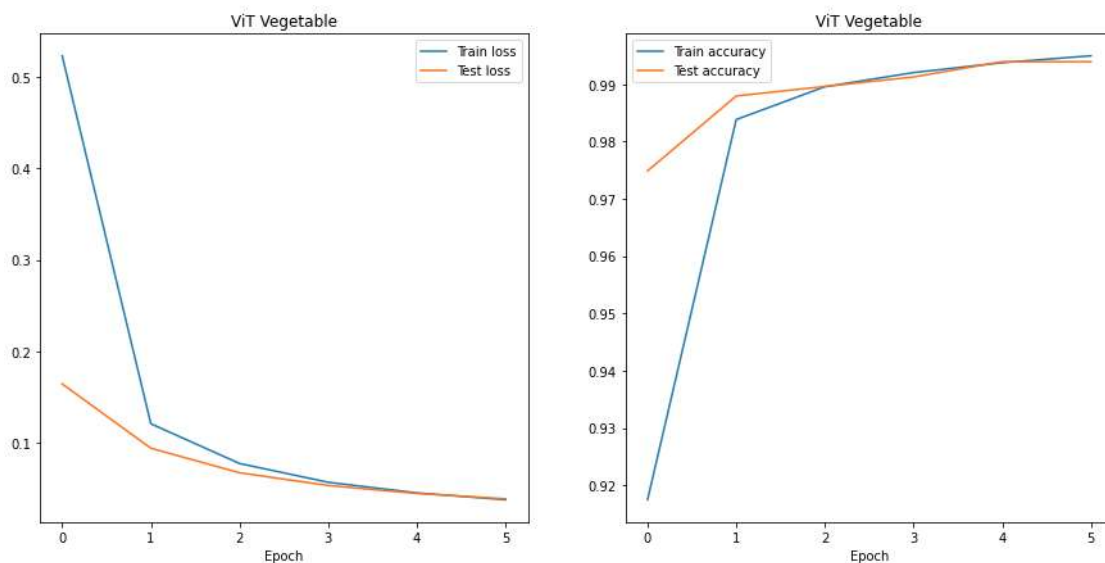
-----
ResNet50 Grapevine dataset
-----
ROC_AUC:0.9652499999999999
F1Score: 0.8684920111749381
Accuracy: 0.87

```

| | Precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Ak | 0.81 | 0.85 | 0.83 | 20 |
| Ala_Idris | 0.80 | 0.80 | 0.80 | 20 |
| Buzgulu | 0.88 | 0.75 | 0.81 | 20 |
| Dimnit | 0.95 | 0.95 | 0.95 | 20 |
| Nazli | 0.91 | 1.00 | 0.95 | 20 |
| accuracy | | | 0.87 | 100 |
| macro avg | 0.87 | 0.87 | 0.87 | 100 |
| weighted avg | 0.87 | 0.87 | 0.87 | 100 |

6.2.4 Rezultati klasifikacije temeljene na Vision Transformer modelu

"Vegetable image dataset" - treniranje ViT modela rezultiralo je solidno ugođenim modelom već nakon 5 stabilnih epoha. Krivulje za loss i accuracy po epohama su prikazane na Slici 33.



Slika 33 Loss i accuracy krivulje za Vegetable Image dataset (rezultati za trening i testni skup podataka)

Jednako kao kod ostalih modela primijenjenih na ovaj skup podataka, rezultati klasifikacije na testnom skupu gotovo su savršeni (accuracy, F1 Score i ROC AUC iznose iznad 0.99) (Tablica 14).

Tablica 14 Zbirni prikaz rezultata klasifikacije na Vegetable Image Dataset skupu podataka za model temeljen na Vision Transformer modelu

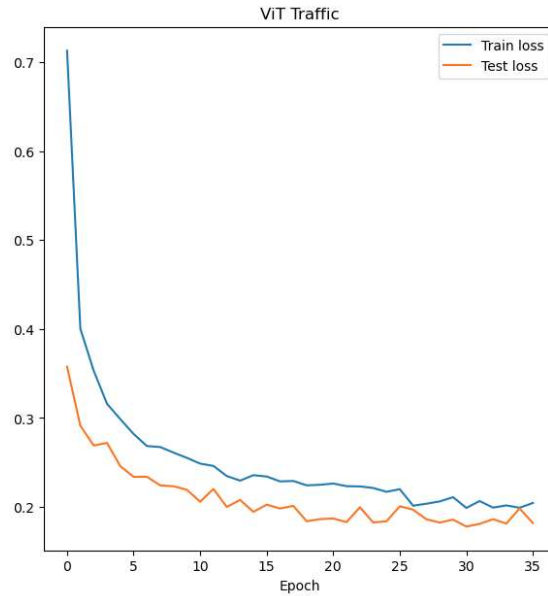
```

-----
ViTVegetabledataset
-----
ROC AUC:0.9999561904761904
F1Score: 0.9943297947654772
Accuracy: 0.9943333333333333

```

| | Precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Bean | 1.00 | 1.00 | 1.00 | 200 |
| Bitter_Gourd | 1.00 | 0.99 | 1.00 | 200 |
| Bottle_Gourd | 1.00 | 1.00 | 1.00 | 200 |
| Brinjal | 0.99 | 1.00 | 0.99 | 200 |
| Broccoli | 1.00 | 1.00 | 1.00 | 200 |
| Cabbage | 1.00 | 0.99 | 1.00 | 200 |
| Capsicum | 0.99 | 0.99 | 0.99 | 200 |
| Carrot | 1.00 | 1.00 | 1.00 | 200 |
| Cauliflower | 0.99 | 0.99 | 0.99 | 200 |
| Cucumber | 0.98 | 0.98 | 0.98 | 200 |
| Papaya | 0.99 | 0.97 | 0.98 | 200 |
| Potato | 1.00 | 0.99 | 1.00 | 200 |
| Pumpkin | 1.00 | 0.99 | 0.99 | 200 |
| Radish | 1.00 | 1.00 | 1.00 | 200 |
| Tomato | 1.00 | 1.00 | 1.00 | 200 |
| accuracy | | | 0.99 | 3000 |
| macro avg | 0.99 | 0.99 | 0.99 | 3000 |
| weighted avg | 0.99 | 0.99 | 0.99 | 3000 |

"Traffic Sign Classification and Recognition" – treniranje ViT modela rezultiralo je solidno ugođenim modelom, a trajalo je 35 epoha sa relativno stabilnim padom vrijednosti ponderiranog gubitka (*weighted loss*). Budući da se radi o neuravnoteženom skupu, razmatrali smo samo krivulju *weightedloss* vrijednosti po epohama prikazanu na Slici 34.



Slika 34 Loss krivulja za Traffic Sign Classification and Recognition dataset (rezultati za trening i testni skup podataka)

U usporedbi sa Vegetable Image Datasetom rezultati su nešto slabiji, ali vrlo dobri prema svim mjerama uspješnosti (Tablica 15).

Tablica 15 Zbirni prikaz rezultata klasifikacije na Traffic Signs Classification and Recognition skupu podataka za model temeljen na Vision Transformer modelu

```

-----
ViT Traffic dataset
-----

ROC_AUC:0.9884397198796251
F1Score: 0.9222774907056858
Accuracy: 0.9444444444444444

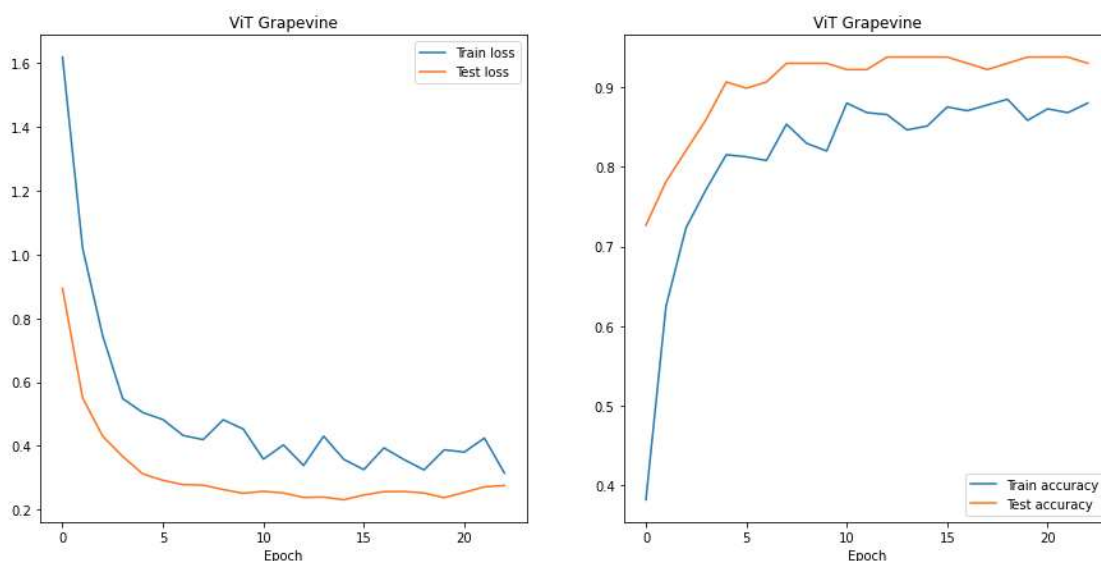
      Precision    recall    f1-score    support
-----
GuideSign    0.91     0.97     0.94         62
  M1         0.69     0.79     0.73         14
  M4         0.96     0.96     0.96        169
  M5         1.00     1.00     1.00         12
  M6         1.00     0.50     0.67          8
  M7         1.00     0.92     0.96         25
  P1         0.93     1.00     0.97         14
 P10_50     1.00     1.00     1.00          6
  P12       1.00     1.00     1.00          6
  W1         1.00     1.00     1.00          8

 accuracy                0.94         324
 macro avg             0.95     0.91     0.92         324
 weighted avg          0.95     0.94     0.94         324

```

"Grapevine Leaves Image Dataset" – treniranje ViT modela trajalo je 22 epohe s relativno stabilnimopadanjem vrijednosti gubitka i porastom mjere accuracy. Vidljive znakove

underfittinga pripisujemo agresivnoj regularizaciji (*dropout* s visokim p vrijednostima). Krivulje su prikazane na Slici 35.



Slika 35 Loss i accuracy krivulje za Grapevine Leaves Image dataset (rezultati za trening i testni skup podataka)

Rezultati klasifikacije na testnom skupu podataka nešto su slabiji u usporedbi s ostalim skupovima podataka, ali ovaj model dao je najbolje rezultate na Grapevine Image Dataset skupu (Tablica 16).

Tablica 16 Zbirni prikaz rezultata klasifikacije na Grapevine Image Dataset skupu podataka za model temeljen na ViT modelu

```

-----
ViT Grapevine dataset
-----

ROC_AUC:0.98475
F1Score: 0.9194544948543815
Accuracy: 0.92

  precision    recall  f1-score   support

     Ak         0.95     0.90     0.92         20
  Ala_Idris     0.94     0.80     0.86         20
   Buzgulu     0.83     0.95     0.88         20
   Dimnit      0.95     0.95     0.95         20
   Nazli       0.95     1.00     0.98         20

 accuracy         0.92         100
  macro avg       0.92         0.92         100
 weighted avg     0.92         0.92         100

```

6.3 Usporedba rezultata tradicionalnih i DL modela

Eksperimenti provedeni na Vegetabledataset skupu podataka proizveli su gotovo savršene rezultate primjenom modela dubokog učenja (sve mjere uspješnosti iznose iznad 0.99 neovisno o primijenjenom modelu). U usporedbi s izvornim istraživanjem Ahmed et al. (2021), provedenim primjenom VGG16, Inception V3, MobileNet i ResNet modela, također prethodno treniranih na ImageNet skupu, dobili smo gotovo identične rezultate. Primjena klasičnih modela ukupno je proizvela značajno slabije rezultate (srednja vrijednost za sve tri mjere iznosi 0.70). Usporedimo li međusobno rezultate klasičnih modela, RandomForestjedao dvostruko viši ROC AUC score (0.93 naspram 0.46) dok su preostale dvije mjere identične za oba modela (acc = 0.75 i F1-score 0.64). - **Error! Reference source not found.**

Tablica 17 Rezultati usporedbe DL modela i tradicionalnih modela na Vegetable image dataset skupu podataka

| | Modeli dubokog učenja | | | | Klasični modeli | | | |
|-----------------|-----------------------|-------|----------|------|--------------------|----------------|--------------|--------------------|
| | AlexNet | VGG16 | ResNet50 | ViT | Srednja vrijednost | TSVM + HOG/LBP | RF + HOG/LBP | Srednja vrijednost |
| ROC AUC | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.46 | 0.93 | 0.70 |
| F1-score | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.75 | 0.64 | 0.70 |
| Accuracy | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.75 | 0.64 | 0.70 |

Modeli dubokog učenja primijenjeni na Traffic Signs Classification and Recognition skupu podataka proizveli su nešto slabije rezultate nego na prethodnom skupu, ali još uvijek dovoljno dobre za manje zahtjevne praktične primjene. Srednje vrijednosti za ROC AUC, F1-Score i accuracy bile su redom: 0.98, 0.92 i 0.95. Ukupno najbolji model dubokog učenja bio je AlexNet sa AUC = 0.99, F1-Score=0.94 i Accuracy=0.97. Najmanje uspješan model (ViT) proizveo je nešto slabije rezultate, odnosno AUC=0.98, F1-score=0.92 i accuracy=0.94.

Rezultati klasičnih modela očekivano su značajno slabiji, ali ipak nešto bolji nego na prethodnom skupu podataka (Vegetable Images dataset), usprkos višestrukom manjem broju primjeraka u trening skupu podataka. Srednje vrijednosti za AUC, F1-score i accuracy su redom: 0.80, 0.77 i 0.87. RandomForest ističe se značajno višim AUC rezultatom (0.95 naspram 0.65), dok je TSVM postigao značajno viši F1-score (0.84 naspram 0.69).

Budući da su se HOG, LBP i njihove kombinacije pokazale izrazito pogodnima za primjenu na klasifikaciju prometnih znakova, spomenuta razlika rezultata u usporedbi sa dobivenima na znatno većem Vegetabledataset skupom može se objasniti odabirom informativnijih značajki za ovaj tip slikovnih podataka (Chirakkal i Han 2015; Ngoc Do et al. 2017).

Tablica 18 Rezultati usporedbe DL modela i tradicionalnih modela na Traffic Signs skupu podataka

| | Modeli dubokog učenja | | | | Klasični modeli | | | |
|-----------------|-----------------------|-------|----------|------|--------------------|----------------|--------------|--------------------|
| | AlexNet | VGG16 | ResNet50 | ViT | Srednja vrijednost | TSVM + HOG/LBP | RF + HOG/LBP | Srednja vrijednost |
| ROC AUC | 0.99 | 0.98 | 0.98 | 0.98 | 0.98 | 0.65 | 0.95 | 0.80 |
| F1-score | 0.94 | 0.92 | 0.90 | 0.92 | 0.92 | 0.84 | 0.69 | 0.77 |
| Accuracy | 0.97 | 0.96 | 0.93 | 0.94 | 0.95 | 0.90 | 0.84 | 0.87 |

Grapevine images dataset najzahtjevniji je skup podataka, ponajviše zbog niske međuklasne varijacije i malog broja primjeraka ($n=500$). Modeli dubokog učenja postigli su solidne, ali značajno slabije rezultate nego što je to slučaj sa prethodna dva skupa podataka (srednje vrijednosti za AUC, F1-score i accuracy iznose redom: 0.96, 0.85 i 0.85). Najbolji rezultati postigli su se primjenom ViT (AUC=0.98, F1-score=0.92 i accuracy=0.92). Rezultati dobiveni tradicionalnim modelima koje smo primijenili u eksperimentu ispod su granice upotrebljivosti. Srednje vrijednosti za AUC, F1-score i accuracy iznose redom: 0.62, 0.40 i 0.40.

Tablica 19 Rezultati usporedbe DL modela i tradicionalnih modela na Grapevine leaves dataset skupu podataka

| | Modeli dubokog učenja | | | | Klasični modeli | | | |
|-----------------|-----------------------|-------|----------|------|--------------------|----------------|--------------|--------------------|
| | AlexNet | VGG16 | ResNet50 | ViT | Srednja vrijednost | TSVM + HOG/LBP | RF + HOG/LBP | Srednja vrijednost |
| ROC AUC | 0.96 | 0.95 | 0.96 | 0.98 | 0.96 | 0.45 | 0.78 | 0.62 |
| F1-score | 0.80 | 0.80 | 0.87 | 0.92 | 0.85 | 0.34 | 0.46 | 0.40 |
| Accuracy | 0.80 | 0.80 | 0.87 | 0.92 | 0.85 | 0.35 | 0.46 | 0.40 |

Rezultati svih eksperimenata sukladni su rezultatima dosadašnjih istraživanja koja su, osim u iznimnim slučajevima, pokazala značajne prednosti primjene konvolucijskih neuronskih mreža prethodno treniranih na većim skupovima podataka (Karypidiset al. 2022; Lai 2019; Hazarikaetal. 2022; Wanget al. 2021).

Zaključak

Vrtoglav porast količine raznih tipova nestrukturiranih podataka, među kojima ogroman udio otpada na slikovne, postavio je nove izazove za razna područja informacijskih znanosti. Ovaj diplomski rad istražio je razlike između tradicionalnog strojnog učenja i suvremenih metoda dubokog učenja, prvenstveno vezane za njihovu primjenu na zadaće klasifikacije slika. Pregledom literature stekli smo uvid u temeljne izazove i koncepte vezane za navedene metode, te različite načine na koje one mogu olakšati i unaprijediti rješavanje različitih problema iz domene informacijskih znanosti. Poseban naglasak stavili smo na slike kao tipične primjere nestrukturiranih podataka, odnosno njihovu fizičku i logičku reprezentaciju, te različite metode izdvajanja značajki.

Kako bismo usporedili izvedbu tradicionalnih algoritama strojnog učenja i modela dubokog učenja, proveli smo niz eksperimenata na tri različita skupa podataka od kojih svaki svojim karakteristikama predstavlja poseban primjer tipičnih izazova za zadaću klasifikacije slika. Koristili smo dva tradicionalna algoritma strojnog učenja (Support Vector Machine i Random Forest), te četiri modela zasnovanim na modelima dubokog učenja od kojih svaki predstavlja neku od važnih prekretnica u povijesti razvoja ovog područja. Modeli AlexNet, VGG16 i ResNet50 donijeli su određene inovacije važne za razvoj konvolucijskih neuronskih mreža. VisionTransformer predstavlja obećavajuću arhitekturu koja na problem klasifikacije slika uspješno primjenjuje koncepte koji godinama dominiraju područjem obrade prirodnog jezika. Modeli dubokog učenja proizveli su značajno bolje rezultate klasifikacije prema svim mjerama uspješnosti na sva tri korištena skupa podataka.

Kako je vidljivo iz provedenih eksperimenata i rezultata prethodnih istraživanja navedenih u pregledu literature, modeli dubokog učenja općenito nadmašuju tradicionalne algoritme strojnog učenja u različitim domenama primjene. Osim što omogućavaju procesiranje ogromnih količina podataka korištenjem specijaliziranog hardvera kao što su grafičke procesorske jedinice, superiornost modela dubokog učenja nad tradicionalnim metodama može se pripisati i njihovoj sposobnosti da automatski uče značajke iz sirovih slikovnih podataka, za razliku od ručnog odabira značajki tipičnog za pristupe temeljene na tradicionalnim algoritmima strojnog učenja. Modeli dubokog učenja na ovaj način mogu naučiti višestruke razine reprezentacija značajki, što im omogućuje znatno bolje prepoznavanje složenih obrazaca i odnosa u velikim skupovima podataka. Naučene značajke također mogu biti dovoljno općenite da ih je moguće uspješno primijeniti na klasifikaciju znatno manjih skupova podataka koji zbog svoje veličine ne dozvoljavaju učenje jednako

informativnih značajki. Primjena navedenog koncepta poznatog pod nazivom *transfer learning*, također je uspješno demonstriran na provedenim eksperimentima.

Tradicionalni algoritmi strojnog učenja kao što su RandomForest ili SVM mogu biti vrlo učinkoviti za neke specifične primjene, ali njihovu učinkovitost na području klasifikacije slika značajno ograničavaju određeni nedostaci od kojih se najvažniji odnosi na potrebu za ručnim odabirom značajki čija informativnost ovisi isključivo o vještini i iskustvu osoba koje vrše odabir. Drugi nedostatak proizlazi iz činjenice da su, osim u iznimno rijetkim slučajevima, ograničeni na korištenje znatno sporijih centralnih procesorskih jedinica.

Međutim, važno je napomenuti da metode dubokog učenja nisu bez svojih specifičnih izazova. Naime, radi se o kompleksnim modelima sklonima overfittingu koji su uz to izrazito osjetljivi na odabir hiperparametara čime njihovo treniranje može biti znatno otežano, posebice kada govorimo o manjim skupovima podataka. Također značajno ovise o dostupnosti specijaliziranog hardvera kao što su grafičke procesorske jedinice, što još uvijek može ograničiti njihovu primjenjivost u uvjetima ograničenih računalnih resursa.

Usprkos svojim nedostacima, modeli dubokog učenja pokazali su superiorne performanse naspram tradicionalnih metoda te neupitno predstavljaju revolucionaran korak na području klasifikacije slika kao tipičnih primjera nestrukturiranih podataka što može doprinijeti poboljšanju organizacije, reprezentacije, pretraživanja, dohvaćanja i diseminacije informacija.

Literatura

- Abgaz, Yalemisew, Renato Rocha Souza, Japesh Methuku, Gerda Koch, i Amelie Dorn. 2021. „A Methodology for Semantic Enrichment of Cultural Heritage Images Using Artificial Intelligence Technologies“. *Journal of Imaging* 7 (8): 121. <https://doi.org/10.3390/jimaging7080121>.
- Acharya, Tinku, i Ajoy K. Ray. 2005. *Image Processing: Principles and Applications*. John Wiley & Sons.
- Afan, Haitham, Ahmedbahaaldin Ibrahim Ahmed Osman, Yusuf Essam, Al-Mahfoodh Najah, Yuk Huang, Ozgur Kisi, Mohsen Sherif, Ahmed Sefelnasr, Kwok Chau, i Ahmed El-Shafie. 2021. „Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques“. *Engineering Applications of Computational Fluid Mechanics* 15 (rujan): 1420–39. <https://doi.org/10.1080/19942060.2021.1974093>.
- Ahmad, Imran, i William I. Grosky. 1997. „Spatial similarity-based retrievals and image indexing by hierarchical decomposition“. U *Proceedings of the 1997 international conference on International database engineering and applications symposium*, 269–78. IDEAS'97. USA: IEEE Computer Society.
- Albregtsen, F., B. Nielsen, i H.E. Danielsen. 2000. „Adaptive gray level run length features from class distance matrices“. U *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, 3:738–41 sv.3. <https://doi.org/10.1109/ICPR.2000.903650>.
- Albuquerque, Clarice, Giuliano La Guardia, Reginaldo Palazzo, Cátia Queiroz, i Vandenberg Vieira. 2021. *Euclidean and Hyperbolic Asymmetric Topological Quantum Codes*.
- Ali, Dilawar, Kenzo Milleville, Alec Van den broeck, i Steven Verstockt. 2022. „NewspAIper : AI-Based Metadata Enrichment of Historical Newspaper Collections“. U *DH Benelux 2022 : RE-MIX : Creation and Alteration in DH, Abstracts*. <http://hdl.handle.net/1854/LU-8755465>.
- Ali, Luqman, Fady Alnajjar, Hamad Jassmi, Munkhjargal Gochoo, Wasif Khan, i Mohamed Serhani. 2021. „Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures“. *Sensors* 21 (ožujak): 1688. <https://doi.org/10.3390/s21051688>.
- Alpaydin, Ethem. 2020. *Introduction to Machine Learning*. 4. izd. Adaptive Computation and Machine Learning Series. Cambridge, MA, USA: MIT Press.
- Al-Tashi, Qasem, Said Jadid Abdulkadir, Helmi Md Rais, Seyedali Mirjalili, i Hitham Alhussian. 2020. „Approaches to Multi-Objective Feature Selection: A Systematic Literature Review“. *IEEE Access* 8: 125076–96. <https://doi.org/10.1109/ACCESS.2020.3007291>.
- Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, i Laith Farhan. 2021. „Review of deep learning: concepts, CNN architectures, challenges, applications, future directions“. *Journal of Big Data* 8 (1): 53. <https://doi.org/10.1186/s40537-021-00444-8>.
- Amato, Giuseppe, Fabrizio Falchi, i Claudio Gennaro. 2015. „Fast Image Classification for Monument Recognition“. *Journal on Computing and Cultural Heritage* 8 (4): 18:1-18:25. <https://doi.org/10.1145/2724727>.
- Andreopoulos, Alexander, i John K. Tsotsos. 2013. „50 Years of Object Recognition: Directions Forward“. *Computer Vision and Image Understanding* 117 (8): 827–91. <https://doi.org/10.1016/j.cviu.2013.04.005>.

- Armi, Laleh, i Shervan Fekri-Ershad. 2019. „Texture image analysis and texture classification methods - A review“. arXiv. <https://doi.org/10.48550/arXiv.1904.06554>.
- Azam, Kazi Sultana Farhana, Farhin Farhad Riya, i Shah Tuhin Ahmed. 2021. „Leaf Detection Using Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Classifying with SVM Utilizing Claim Dataset“. U *Intelligent Data Communication Technologies and Internet of Things*, uredio Jude Hemanth, Robert Bestak, i Joy Iong-Zong Chen, 313–23. Lecture Notes on Data Engineering and Communications Technologies. Singapore: Springer. https://doi.org/10.1007/978-981-15-9509-7_27.
- Bai, Yan, Feng Gao, Yihang Lou, Shiqi Wang, Tiejun Huang, i Ling-Yu Duan. 2017. „Incorporating Intra-Class Variance to Fine-Grained Visual Recognition“. arXiv. <https://doi.org/10.48550/arXiv.1703.00196>.
- Bajaj, Anu, Tamanna Sharma, i Om Sangwan. 2020. „Information Retrieval in Conjunction With Deep Learning“. U , <https://www.igi-global.com/chapter/information>. <https://doi.org/10.4018/978-1-5225-9643-1.ch014>.
- Balduzzi, David, Marcus Frean, Lennox Leary, J. P. Lewis, Kurt Wan-Duo Ma, i Brian McWilliams. 2018. „The Shattered Gradients Problem: If resnets are the answer, then what is the question?“ arXiv. <https://doi.org/10.48550/arXiv.1702.08591>.
- Banko, Michele, i Eric Brill. 2001. „Scaling to Very Very Large Corpora for Natural Language Disambiguation“. U *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 26–33. Toulouse, France: Association for Computational Linguistics. <https://doi.org/10.3115/1073012.1073017>.
- Barrow, H. G., i J. M. Tenenbaum. 1981. „Interpreting Line Drawings as Three-Dimensional Surfaces“. *Artificial Intelligence* 17 (1): 75–116. [https://doi.org/10.1016/0004-3702\(81\)90021-7](https://doi.org/10.1016/0004-3702(81)90021-7).
- Basori, Ahmad Hoirul, Ahmad Luqman bin Abdul Hamid, Andi Besse Firdausiah Mansur, i Norazah Yusof. 2019. „IMars: Intelligent Municipality Augmented Reality Service for Efficient Information Dissemination Based on Deep Learning Algorithm in Smart City of Jeddah“. *Procedia Computer Science*, 16th Learning and Technology Conference 2019 Artificial Intelligence and Machine Learning: Embedding the Intelligence, 163 (siječanj): 93–108. <https://doi.org/10.1016/j.procs.2019.12.091>.
- Bengio, Y., P. Simard, i P. Frasconi. 1994. „Learning long-term dependencies with gradient descent is difficult“. *IEEE Transactions on Neural Networks* 5 (2): 157–66. <https://doi.org/10.1109/72.279181>.
- Bengio, Yoshua, Aaron Courville, i Pascal Vincent. 2013. „Representation Learning: A Review and New Perspectives“. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8): 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>.
- Blauch, Nicholas, Marlene Behrmann, i David Plaut. 2020. „Computational insights into human perceptual expertise for familiar and unfamiliar face recognition“. *Cognition* 208 (lipanj): 104341. <https://doi.org/10.1016/j.cognition.2020.104341>.
- Boukhers, Zeyd, Nada Beili, Timo Hartmann, Prantik Goswami, i Muhammad Arslan Zafar. 2021. „MexPub: Deep Transfer Learning for Metadata Extraction from German Publications“. U *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 250–53. <https://doi.org/10.1109/JCDL52503.2021.00076>.
- Carbonnelle, Simon, i Christophe De Vleeschouwer. 2022. „Intraclass Clustering: An Implicit Learning Ability That Regularizes DNNs“. U . <https://openreview.net/forum?id=tqOvYpjPax2>.
- Celebi, M. Emre. 2011. „Improving the Performance of K-Means for Color Quantization“. <https://doi.org/10.1016/j.imavis.2010.10.002>.

- Chandha, Saikiran, Sucheth R, i Tirthankar Ghosal. 2023. „Setting the Scene: How Artificial Intelligence Is Reshaping How We Consume and Deliver Research“. *Upstream*, siječanj. <https://doi.org/10.54900/ebw4ce4-47p2tbr>.
- Chandra, Mayank Arya, i S. S. Bedi. 2021. „Survey on SVM and Their Application in Imageclassification“. *International Journal of Information Technology* 13 (5): 1–11. <https://doi.org/10.1007/s41870-017-0080-1>.
- Chen, Zhuyun, Konstantinos Gryllias, i Weihua Li. 2019. „Mechanical Fault Diagnosis Using Convolutional Neural Networks and Extreme Learning Machine“. *Mechanical Systems and Signal Processing* 133 (studeni): 106272. <https://doi.org/10.1016/j.ymsp.2019.106272>.
- Costa, Ruben, Paulo Figueiras, Ricardo Jardim-Gonçalves, José Ramos-Filho, i Celson Lima. 2017. „Semantic enrichment of product data supported by machine learning techniques“. U *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 1472–79. <https://doi.org/10.1109/ICE.2017.8280056>.
- Cui, Suxia, Yonghui Wang, Xiaoqing Qian, i Zhengtao Deng. 2013. „Image Processing Techniques in Shockwave Detection and Modeling“. *Journal of Signal and Information Processing* 4 (3): 109–13. <https://doi.org/10.4236/jsip.2013.43B019>.
- Dalal, N., i B. Triggs. 2005. „Histograms of Oriented Gradients for Human Detection“. U *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:886–93. San Diego, CA, USA: IEEE. <https://doi.org/10.1109/CVPR.2005.177>.
- Dhal, Pradip, i Chandrashekar Azad. 2022. „A Comprehensive Survey on Feature Selection in the Various Fields of Machine Learning“. *Applied Intelligence* 52 (4): 4543–81. <https://doi.org/10.1007/s10489-021-02550-9>.
- Ding, Jie, Vahid Tarokh, i Yuhong Yang. 2018. „Model Selection Techniques: An Overview“. *IEEE Signal Processing Magazine* 35 (6): 16–34. <https://doi.org/10.1109/MSP.2018.2867638>.
- Doan, AnHai, Jayant Madhavan, Pedro Domingos, i Alon Halevy. 2004. „Ontology Matching: A Machine Learning Approach“. U *Handbook on Ontologies*, uredio Steffen Staab i Rudi Studer, 385–403. International Handbooks on Information Systems. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-24750-0_19.
- Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, i ostali. 2021. „An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale“. *arXiv:2010.11929 [cs]*, lipanj. <http://arxiv.org/abs/2010.11929>.
- Dunn, Jack, Luca Mingardi, i Ying Daisy Zhuo. 2021. „Comparing interpretability and explainability for feature selection“. *arXiv*. <https://doi.org/10.48550/arXiv.2105.05328>.
- Elezi, Ismail. 2020. „Exploiting Contextual Information with Deep Neural Networks“. *arXiv*. <https://doi.org/10.48550/arXiv.2006.11706>.
- Elhassan, Ammar, Saleh M. Abu-Soud, Firas Alghanim, i Walid Salameh. 2022. „ILA4: Overcoming Missing Values in Machine Learning Datasets – An Inductive Learning Approach“. *Journal of King Saud University - Computer and Information Sciences* 34 (7): 4284–95. <https://doi.org/10.1016/j.jksuci.2021.02.011>.
- Ferrando, Javier, Gerard I. Gállego, i Marta R. Costa-jussà. 2022. „Measuring the Mixing of Contextual Information in the Transformer“. *arXiv*. <https://doi.org/10.48550/arXiv.2203.04212>.
- Fountas, Zafeirios. 2023. „Imperial College Spiking Neural Networks for Human-like Avatar Control in a Simulated Environment“, ožujak.

- Galleguillos, Carolina, i Serge Belongie. 2010. „Context based object categorization: A critical survey“. *Computer Vision and Image Understanding* 114 (lipanj): 712–22. <https://doi.org/10.1016/j.cviu.2010.02.004>.
- Gandomi, Amir, i Murtaza Haider. 2015. „Beyond the Hype: Big Data Concepts, Methods, and Analytics“. *International Journal of Information Management* 35 (2): 137–44. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.
- Gani, Abdullah, Aisha Siddiqa, Shahab Band, i Fariza Nasaruddin. 2015. „A survey on Indexing Techniques for Big Data: Taxonomy and Performance Evaluation“. *Knowledge and Information Systems* 46 (travanj). <https://doi.org/10.1007/s10115-015-0830-y>.
- Gencoglu, Oguzhan, Mark van Gils, Esin Guldogan, Chamin Morikawa, Mehmet Süzen, Mathias Gruber, Jussi Leinonen, i Heikki Huttunen. 2019. „HARK Side of Deep Learning -- From Grad Student Descent to Automated Machine Learning“. arXiv. <https://doi.org/10.48550/arXiv.1904.07633>.
- Gonzalez, Rafael C., i Richard Eugene Woods. 2018. *Digital Image Processing*. Pearson.
- Goodfellow, Ian, Yoshua Bengio, i Aaron Courville. 2016. *Deep Learning*. Adaptive Computation and Machine Learning Series. Cambridge, MA, USA: MIT Press.
- Grosky, William, i Peter Stanchev. 2002. „An Image Data Model“. U . https://doi.org/10.1007/3-540-40053-2_2.
- Gudivada, Venkat N. 1998. „ORZ_ \$Re\$ _Z-String: A Geometry-Based Representation for Efficient and Effective Retrieval of Images by Spatial Similarity“. *IEEE Transactions on Knowledge and Data Engineering* 10 (03): 504–12. <https://doi.org/10.1109/69.687982>.
- Gudivada, Venkat N., i Vijay V. Raghavan. 1995. „Design and evaluation of algorithms for image retrieval by spatial similarity“. *ACM Transactions on Information Systems* 13 (2): 115–44. <https://doi.org/10.1145/201040.201041>.
- Guo, Jiafeng, Yixing Fan, Liang Pang, Liu Yang, Qingyao Ai, Hamed Zamani, Chen Wu, W. Bruce Croft, i Xueqi Cheng. 2020. „A Deep Look into Neural Ranking Models for Information Retrieval“. *Information Processing & Management* 57 (6): 102067. <https://doi.org/10.1016/j.ipm.2019.102067>.
- Gupta, Jaya, Sunil Pathak, i Gireesh Kumar. 2022. „Deep Learning (CNN) and Transfer Learning: A Review“. *Journal of Physics: Conference Series* 2273 (1): 012029. <https://doi.org/10.1088/1742-6596/2273/1/012029>.
- Hasan Choudhury, Muntabir, Himarsha R. Jayanetti, Jian Wu, William A. Ingram, i Edward A. Fox. 2021. „Automatic Metadata Extraction Incorporating Visual Features from Scanned Electronic Theses and Dissertations“. U *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 230–33. <https://doi.org/10.1109/JCDL52503.2021.00066>.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren, i Jian Sun. 2015. „Deep Residual Learning for Image Recognition“. arXiv. <https://doi.org/10.48550/arXiv.1512.03385>.
- Horning, N. 2010. „Random Forests : An algorithm for image classification and generation of continuous fields data sets“. U . <https://www.semanticscholar.org/paper/Random-Forests-%3A-An-algorithm-for-image-and-of-data-Horning/9484240b4dfe10a40b80037d6869c4b035615be9>.
- Hsu, Chung-Chian, Wei-Cyun Tsao, Arthur Chang, i Chuan-Yu Chang. 2021. „Analyzing mixed-type data by using word embedding for handling categorical features“. *Intelligent Data Analysis* 25 (listopad): 1349–68. <https://doi.org/10.3233/IDA-205453>.
- Hu, Yingjie, Zhipeng Gui, Jimin Wang, i Muxian Li. 2022. „Enriching the metadata of map images: a deep learning approach with GIS-based data augmentation“. *International*

- Journal of Geographical Information Science* 36 (4): 799–821.
<https://doi.org/10.1080/13658816.2021.1968407>.
- Huang, Jing, S.R. Kumar, M. Mitra, Wei-Jing Zhu, i R. Zabih. 1997. „Image indexing using color correlograms“. U *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 762–68.
<https://doi.org/10.1109/CVPR.1997.609412>.
- Ioffe, Sergey, i Christian Szegedy. 2015. „Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift“. arXiv.
<https://doi.org/10.48550/arXiv.1502.03167>.
- Islam, Md Amirul, Sen Jia, i Neil D. B. Bruce. 2020. „How Much Position Information Do Convolutional Neural Networks Encode?“ arXiv.
<https://doi.org/10.48550/arXiv.2001.08248>.
- Ismael, Ahmed, i Dr Kashmar. 2019. „Construct a Strong and High Performance Algorithm to Generate Pseudorandom Number Generator (PRNG) for Stream Cipher“.
- Jagtap, Ameya, i George Karniadakis. 2019. „Adaptive activation functions accelerate convergence in deep and physics-informed neural networks“, lipanj.
- Javidi, Bahram. 2002. *Image Recognition and Classification: Algorithms, Systems, and Applications*. CRC Press.
- Karnani, Kevin, Joel Pepper, Yasin Bakiş, Xiaojun Wang, Henry Bart Jr., David E. Breen, i Jane Greenberg. 2022. „Computational Metadata Generation Methods for Biological Specimen Image Collections“. *International Journal on Digital Libraries*, studeni.
<https://doi.org/10.1007/s00799-022-00342-1>.
- Ke, Xiao, Jiawei Zou, i Yuzhen Niu. 2019. „End-to-End Automatic Image Annotation Based on Deep CNN and Multi-Label Data Augmentation“. *IEEE Transactions on Multimedia* 21 (8): 2093–2106. <https://doi.org/10.1109/TMM.2019.2895511>.
- Khoudja, Meriem Ali, Messaouda Fareh, i Hafida Bouarfa. 2018. „Ontology Matching using Neural Networks: Survey and Analysis“. U *2018 International Conference on Applied Smart Systems (ICASS)*, 1–6. <https://doi.org/10.1109/ICASS.2018.8652049>.
- Kim, Hee E., Alejandro Cosa-Linan, Nandhini Santhanam, Mahboubeh Jannesari, Mate E. Maros, i Thomas Ganslandt. 2022. „Transfer Learning for Medical Image Classification: A Literature Review“. *BMC Medical Imaging* 22 (1): 69.
<https://doi.org/10.1186/s12880-022-00793-7>.
- Krizhevsky, Alex. 2014. „One weird trick for parallelizing convolutional neural networks“. arXiv. <https://doi.org/10.48550/arXiv.1404.5997>.
- Krizhevsky, Alex, Ilya Sutskever, i Geoffrey Hinton. 2012. „ImageNet Classification with Deep Convolutional Neural Networks“. *Neural Information Processing Systems 25* (siječanj). <https://doi.org/10.1145/3065386>.
- Labsky, Martin, Miroslav Vacura, i Pavel Praks. 2005. „Web Image Classification for Information Extraction Web Image Classification for Information Extraction“, siječanj.
- Lakzaei, Batool, i Mehrnoush Shamsfard. 2021. „Ontology Learning from Relational Databases“. *Information Sciences* 577 (listopad): 280–97.
<https://doi.org/10.1016/j.ins.2021.06.074>.
- Lecun, Y., L. Bottou, Y. Bengio, i P. Haffner. 1998. „Gradient-based learning applied to document recognition“. *Proceedings of the IEEE* 86 (11): 2278–2324.
<https://doi.org/10.1109/5.726791>.
- Lin, Yuanqing, Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu, Liangliang Cao, i Thomas Huang. 2011. „Large-scale image classification: Fast feature extraction and SVM training“. U *CVPR 2011*, 1689–96.
<https://doi.org/10.1109/CVPR.2011.5995477>.

- Lisin, Dima, Marwan Mattar, Matthew Blaschko, Mark Benfield, i Erik Learned-Miller. 2005. „Combining Local and Global Image Features for Object Class Recognition“. U .
- Liu, Chunsheng, Shuang Li, Faliang Chang, i Yin Hai Wang. 2019. „Machine Vision Based Traffic Sign Detection Methods: Review, Analyses and Perspectives“. *Ieee Access* 7: 86578–96. <https://doi.org/10.1109/ACCESS.2019.2924947>.
- Liu, Li, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, i Matti Pietikäinen. 2020. „Deep Learning for Generic Object Detection: A Survey“. *International Journal of Computer Vision* 128 (2): 261–318. <https://doi.org/10.1007/s11263-019-01247-4>.
- Lucchese, Claudio, Franco Maria Nardini, Raffaele Perego, Roberto Trani, i Rossano Venturini. 2018. „Efficient and Effective Query Expansion for Web Search“. U *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1551–54. CIKM '18. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3269206.3269305>.
- Mahony, Niall O', Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco-Hernandez, Lenka Krpalkova, Daniel Riordan, i Joseph Walsh. 2020. *Deep Learning vs. Traditional Computer Vision*. Sv. 943. <https://doi.org/10.1007/978-3-030-17795-9>.
- Mantovani, Rafael G., André L. D. Rossi, Joaquin Vanschoren, Bernd Bischl, i André C. P. L. F. Carvalho. 2015. „To tune or not to tune: Recommending when to adjust SVM hyper-parameters via meta-learning“. U *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–8. <https://doi.org/10.1109/IJCNN.2015.7280644>.
- Meesad, Phayung. 2021. „Thai Fake News Detection Based on Information Retrieval, Natural Language Processing and Machine Learning“. *SN Computer Science* 2 (6): 425. <https://doi.org/10.1007/s42979-021-00775-6>.
- Mitchell, Tom M. 1997. *Machine Learning*. McGraw-Hill.
- Mohan, Amrita, Amit Kumar Singh, Basant Kumar, i Ramji Dwivedi. 2021. „Review on Remote Sensing Methods for Landslide Detection Using Machine and Deep Learning“. *Transactions on Emerging Telecommunications Technologies* 32 (7): e3998. <https://doi.org/10.1002/ett.3998>.
- Montaha, Sidratul, Sami Azam, A. K. M. Rakibul Haque Rafid, Sayma Islam, Pronab Ghosh, i Mirjam Jonkman. 2022. „A Shallow Deep Learning Approach to Classify Skin Cancer Using Down-Scaling Method to Minimize Time and Space Complexity“. *PLOS ONE* 17 (8): e0269826. <https://doi.org/10.1371/journal.pone.0269826>.
- Montesinos López, Osva Antonio, Abelardo Montesinos López, i Jose Crossa. 2022. „Overfitting, Model Tuning, and Evaluation of Prediction Performance“. U *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, uredio Osva Antonio Montesinos López, Abelardo Montesinos López, i José Crossa, 109–39. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_4.
- Murala, Subrahmanyam, Anil Balaji Gonde, i R. P. Maheshwari. 2009. „Color and Texture Features for Image Indexing and Retrieval“. U *2009 IEEE International Advance Computing Conference*, 1411–16. <https://doi.org/10.1109/IADCC.2009.4809223>.
- Narwade Manoorkar, Jyoti, i Dr Binod Kumar. 2016. „Local and Global Color Histogram Feature for Color Content-Based Image Retrieval System“. U , 293–300. https://doi.org/10.1007/978-981-10-0767-5_32.
- Ojala, T., M. Pietikainen, i T. Maenpaa. 2002. „Multiresolution gray-scale and rotation invariant texture classification with local binary patterns“. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7): 971–87. <https://doi.org/10.1109/TPAMI.2002.1017623>.

- Pal, Saurabh, Pijush Kanti Dutta Pramanik, Tripti Majumdar, i Prasenjit Choudhury. 2019. „A Semi-Automatic Metadata Extraction Model and Method for Video-Based e-Learning Contents“. *Education and Information Technologies* 24 (6): 3243–68. <https://doi.org/10.1007/s10639-019-09926-y>.
- Pass, Greg, Ramin Zabih, i Justin Miller. 1997. „Comparing images using color coherence vectors“. U *Proceedings of the fourth ACM international conference on Multimedia*, 65–73. MULTIMEDIA '96. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/244130.244148>.
- Pereira, Jeanne, i Filipe Saraiva. 2020. „A Comparative Analysis of Unbalanced Data Handling Techniques for Machine Learning Algorithms to Electricity Theft Detection“. U *2020 IEEE Congress on Evolutionary Computation (CEC)*, 1–8. <https://doi.org/10.1109/CEC48606.2020.9185822>.
- Phung i Rhee. 2019. „A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets“. *Applied Sciences* 9 (listopad): 4500. <https://doi.org/10.3390/app9214500>.
- Pilarczyk, Rafal, i Wladyslaw Skarbek. 2019. „On Intra-Class Variance for Deep Learning of Classifiers“. arXiv. <https://doi.org/10.48550/arXiv.1901.11186>.
- Poggio, Tomaso, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, i Qianli Liao. 2017. „Why and When Can Deep-but Not Shallow-Networks Avoid the Curse of Dimensionality: A Review“. *International Journal of Automation and Computing* 14 (5): 503–19. <https://doi.org/10.1007/s11633-017-1054-2>.
- Powers, David M. W. 2020. „Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation“. arXiv. <https://doi.org/10.48550/arXiv.2010.16061>.
- Probst, Philipp, i Anne-Laure Boulesteix. 2017. „To tune or not to tune the number of trees in random forest?“ arXiv. <https://doi.org/10.48550/arXiv.1705.05654>.
- Probst, Philipp, Marvin Wright, i Anne-Laure Boulesteix. 2019. „Hyperparameters and Tuning Strategies for Random Forest“. *WIREs Data Mining and Knowledge Discovery* 9 (3). <https://doi.org/10.1002/widm.1301>.
- Punj, Roopali, i Rakesh Kumar. 2021. „Data dissemination approach using machine learning techniques for WBANs“. *Concurrency and Computation: Practice and Experience* 34 (studeni). <https://doi.org/10.1002/cpe.6688>.
- Qader, Wisam, Musa M. Ameen, i Bilal Ahmed. 2019. „An Overview of Bag of Words; Importance, Implementation, Applications, and Challenges“. U , 200–204. <https://doi.org/10.1109/IEC47844.2019.8950616>.
- Rai, Rajeshwari. 2017. „Intricacies of Unstructured Data“. *ICST Transactions on Scalable Information Systems* 4 (rujan): 153151. <https://doi.org/10.4108/eai.25-9-2017.153151>.
- Rajini, N. Hema, i R. Bhavani. 2011. „Classification of MRI brain images using k-nearest neighbor and artificial neural network“. U *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 563–68. <https://doi.org/10.1109/ICRTIT.2011.5972341>.
- Reuter, Martin, Silvia Biasotti, Daniela Giorgi, Giuseppe Patanè, i Michela Spagnuolo. 2009. „Discrete Laplace–Beltrami Operators for Shape Analysis and Segmentation“. *Computers & Graphics*, IEEE International Conference on Shape Modelling and Applications 2009, 33 (3): 381–90. <https://doi.org/10.1016/j.cag.2009.03.005>.
- Rodríguez-Hernández, M., Rosa Pruneda, i Juan Rodríguez-Díaz. 2021. „Statistical Analysis of the Evolutive Effects of Language Development in the Resolution of Mathematical Problems in Primary School Education“. *Mathematics* 9 (svibanj): 1081. <https://doi.org/10.3390/math9101081>.

- Safder, Iqra, Saeed-Ul Hassan, Anna Visvizi, Thanapon Noraset, Raheel Nawaz, i Suppawong Tuarob. 2020. „Deep Learning-based Extraction of Algorithmic Metadata in Full-Text Scholarly Documents“. *Information Processing & Management* 57 (studeni): 102269. <https://doi.org/10.1016/j.ipm.2020.102269>.
- Schlegel, Udo, i Daniel A. Keim. 2021. „Time Series Model Attribution Visualizations as Explanations“. arXiv. <https://doi.org/10.48550/arXiv.2109.12935>.
- scikit-learn developers. bez dat. „Sklearn.Ensemble.RandomForestClassifier“. Scikit-Learn. Pristupljeno 23. ožujak 2023.a. <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
- . bez dat. „Sklearn.Metrics.Classification_report“. Scikit-Learn. Pristupljeno 23. ožujak 2023.b. https://scikit-learn/stable/modules/generated/sklearn.metrics.classification_report.html.
- . bez dat. „Sklearn.Metrics.Roc_auc_score“. Scikit-Learn. Pristupljeno 23. ožujak 2023.c. https://scikit-learn/stable/modules/generated/sklearn.metrics.roc_auc_score.html.
- Serrano-Guerrero, Jesus, Enrique Herrera-Viedma, Jose A. Olivas, Andres Cerezo, i Francisco P. Romero. 2011. „A Google Wave-Based Fuzzy Recommender System to Disseminate Information in University Digital Libraries 2.0“. *Information Sciences* 181 (9): 1503–16. <https://doi.org/10.1016/j.ins.2011.01.012>.
- Shahrivari, Vahid, Mohammad Mahdi Darabi, i Mohammad Izadi. 2020. „Phishing Detection Using Machine Learning Techniques“. arXiv. <https://doi.org/10.48550/arXiv.2009.11116>.
- Shao, Ran, Xiao-Jun Bi, i Zheng Chen. 2022. „A Novel Hybrid Transformer-CNN Architecture for Environmental Microorganism Classification“. *PLOS ONE* 17 (11): e0277557. <https://doi.org/10.1371/journal.pone.0277557>.
- Shapiro, Linda G., i George C. Stockman. 2001. *Computer Vision*. Prentice Hall.
- Sharma, Diksha, i Neeraj Kumar. 2017. „A Review on Machine Learning Algorithms, Tasks and Applications“ 6 (listopad): 2278–1323.
- Sheykhmousa, Mohammadreza, Masoud Mahdianpari, Hamid Ghanbari, Fariba Mohammadimanesh, Pedram Ghamisi, i Saeid Homayouni. 2020. „Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review“. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 6308–25. <https://doi.org/10.1109/JSTARS.2020.3026724>.
- Simonyan, Karen, i Andrew Zisserman. 2015. „Very Deep Convolutional Networks for Large-Scale Image Recognition“. *arXiv:1409.1556 [cs]*, travanj. <http://arxiv.org/abs/1409.1556>.
- Smith, Leslie N. 2017. „Cyclical Learning Rates for Training Neural Networks“. arXiv. <https://doi.org/10.48550/arXiv.1506.01186>.
- . 2018. „A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay“. arXiv. <https://doi.org/10.48550/arXiv.1803.09820>.
- Sobel, Irwin. 2014. „An Isotropic 3x3 Image Gradient Operator“. *Presentation at Stanford A.I. Project 1968*, veljača.
- Somodevilla García, María, Darnes Vilariño Ayala, Ivo Pineda, María Somodevilla García, Darnes Vilariño Ayala, i Ivo Pineda. 2018. „An Overview of Ontology Learning Tasks“. *Computación y Sistemas* 22 (1): 137–46. <https://doi.org/10.13053/cys-22-1-2790>.

- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, i Ruslan Salakhutdinov. 2014. „Dropout: A Simple Way to Prevent Neural Networks from Overfitting“. *Journal of Machine Learning Research* 15 (56): 1929–58. Stanford University. bez dat. „CS231n Convolutional Neural Networks for Visual Recognition“. Pristupljeno 26. ožujak 2023. <https://cs231n.github.io/classification/>.
- Strisciuglio, Nicola, Manuel Lopez Antequera, i Nicolai Petkov. 2020. „Enhanced robustness of convolutional networks with a push–pull inhibition layer“. *Neural Computing and Applications* 32 (prosinac): 1–15. <https://doi.org/10.1007/s00521-020-04751-8>.
- Sun, Ruoyu. 2019. „Optimization for deep learning: theory and algorithms“. arXiv. <https://doi.org/10.48550/arXiv.1912.08957>.
- the scikit-image development team. bez dat. „Histogram of Oriented Gradients — skimage v0.21.0rc0.dev0 docs“. Pristupljeno 23. ožujak 2023.a. https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_hog.html.
- . bez dat. „Local Binary Pattern for texture classification — skimage v0.20.0 docs“. Pristupljeno 23. ožujak 2023.b. https://scikit-image.org/docs/stable/auto_examples/features_detection/plot_local_binary_pattern.html.
- ThunderSVM Developers. bez dat. „ThunderSVM Parameters — ThunderSVM 0.1 documentation“. Pristupljeno 23. ožujak 2023. <https://thundersvm.readthedocs.io/en/latest/parameters.html>.
- Tian, Dongping. 2013. „A review on image feature extraction and representation techniques“. *International Journal of Multimedia and Ubiquitous Engineering* 8 (siječanj): 385–95.
- V, Greeshma K., i Dr J. Viji Gripsy. 2020. „Image Classification Using HOG and LBP Feature Descriptors with SVM and CNN“. *International Journal of Engineering Research & Technology* 8 (4). <https://doi.org/10.17577/IJERTCONV8IS04021>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, i Illia Polosukhin. 2017. „Attention Is All You Need“. *arXiv:1706.03762 [cs]*, prosinac. <http://arxiv.org/abs/1706.03762>.
- Veit, Andreas, Michael Wilber, i Serge Belongie. 2016. „Residual Networks Behave Like Ensembles of Relatively Shallow Networks“. arXiv. <https://doi.org/10.48550/arXiv.1605.06431>.
- Vieira, Pedro. 2021. „Hierarchical Classification of Lesions in Wireless Capsule Endoscopy Exams“. <https://doi.org/10.13140/RG.2.2.11263.28328>.
- Vladimir Estivill-Castro. 2002. „Why so many clustering algorithms: a position paper“. *ACM SIGKDD Explorations Newsletter* 4 (1): 65–75. <https://doi.org/10.1145/568574.568575>.
- Vujovic, Željko Đ. 2021. „Classification Model Evaluation Metrics“. *International Journal of Advanced Computer Science and Applications (IJACSA)* 12 (6). <https://doi.org/10.14569/IJACSA.2021.0120670>.
- Wang, Guanqi. 2011. „Study of Object Recognition and Identification Based on Shape and Texture Analysis“, listopad. <https://doi.org/10.25560/11732>.
- Wang, Yixin, i Michael I. Jordan. 2022. „Desiderata for Representation Learning: A Causal Perspective“. arXiv. <https://doi.org/10.48550/arXiv.2109.03795>.
- Wen, Zeyi, Jiashuai Shi, Qinbin Li, Bingsheng He, i Jian Chen. 2018. „ThunderSVM: A Fast SVM Library on GPUs and CPUs“. *Journal of Machine Learning Research* 19 (21): 1–5.
- Wirsing, Karlton. 2020. *Time Frequency Analysis of Wavelet and Fourier Transform. Wavelet Theory*. IntechOpen. <https://doi.org/10.5772/intechopen.94521>.
- Wu, Bichen, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, i Peter Vajda. 2020. „Visual

- Transformers: Token-based Image Representation and Processing for Computer Vision“. *arXiv*. <https://doi.org/10.48550/arXiv.2006.03677>.
- Xie, Lingxi, Richang Hong, Bo Zhang, i Qi Tian. 2015. „Image Classification and Retrieval are ONE“. U *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 3–10. ICMR '15. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2671188.2749289>.
- Yadav, Samir S., i Shivajirao M. Jadhav. 2019. „Deep Convolutional Neural Network Based Medical Image Classification for Disease Diagnosis“. *Journal of Big Data* 6 (1): 113. <https://doi.org/10.1186/s40537-019-0276-2>.
- Yao, Jing, Zhicheng Dou, Jun Xu, i Ji-Rong Wen. 2020. „RLPer: A Reinforcement Learning Model for Personalized Search“. U *Proceedings of The Web Conference 2020*, 2298–2308. WWW '20. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3366423.3380294>.
- . 2021. „RLPS: A Reinforcement Learning–Based Framework for Personalized Search“. *ACM Transactions on Information Systems* 39 (3): 27:1-27:29. <https://doi.org/10.1145/3446617>.
- Zantalis, Fotios, Grigorios Koulouras, Sotiris Karabetsos, i Dionisis Kandris. 2019. „A Review of Machine Learning and IoT in Smart Transportation“. *Future Internet* 11 (4): 94. <https://doi.org/10.3390/fi11040094>.
- Zeiler, Matthew D., i Rob Fergus. 2013. „Visualizing and Understanding Convolutional Networks“. *arXiv:1311.2901 [cs]*, studeni. <http://arxiv.org/abs/1311.2901>.
- Zhang, Chao, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, i Jiawei Han. 2018. „TaxoGen: Unsupervised Topic Taxonomy Construction by Adaptive Term Embedding and Clustering“. U *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2701–9. KDD '18. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3219819.3220064>.
- Zhang, Shuai, Lina Yao, Aixin Sun, i Yi Tay. 2019. „Deep Learning Based Recommender System: A Survey and New Perspectives“. *ACM Computing Surveys* 52 (1): 5:1-5:38. <https://doi.org/10.1145/3285029>.

TITLE: Application of machine learning on the classification of images as typical examples of unstructured data types

SUMMARY:

The increased amount of unstructured data has posed new challenges in various domains of information sciences, highlighting the need to develop innovative, efficient, and cost-effective processing methods, among which different machine learning methods occupy a prominent place. This master thesis investigated the differences between traditional machine learning and modern deep learning methods, primarily related to their application to image classification tasks as typical examples of unstructured data types. To compare the performance of traditional machine learning algorithms and deeplearning models, we conducted experiments on three different data sets ("Vegetable Image Dataset", "Traffic Signs Classification and Recognition" and "Grapevine Image Dataset"), each representing a particular example of typical challenges for the image classification task. Deep learning models produced better classification results on all three data sets.

KEYWORDS: machine learning; deep learning; image classification; unstructured data