METOD ZA SMANJENJE DEMOGRAFSKE PRISTRASNOSTI U MODELIMA PROCENE STAROSTI PUTEM OPTIMIZOVANE SKTRUKTURE SKUPOVA PODATAKA

By Nenad Panić

WORD COUNT 22231 TIME SUBMITTED 24-DEC-2025 08:45AM

PAPER ID 119637978



UNIVERZITET SINGIDUNUM

BEOGRAD

DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE

DOKTORSKA DISERTACIJA

METOD ZA SMANJENJE DEMOGRAFSKE PRISTRASNOSTI U MODELIMA PROCENE STAROSTI PUTEM OPTIMIZOVANE SKTRUKTURE SKUPOVA PODATAKA

MENTOR:

Prof. dr Marina Marjanović

STUDENT:

Nenad Panić

Broj indeksa:

480127/2022

Beograd, 2025.



Singidunum University

Department for postgraduate studies

PhD Thesis

METHOD FOR REDUCING DEMOGRAPHIC BIAS IN AGE ESTIMATION MODELS THROUGH OPTIMIZED DATASET COMPOSITION

MENTOR:

Prof. dr Marina Marjanović

STUDENT:

Nenad Panić

Index number:

480127/2022

Belgrade, 2025.

UNIVERZITET SINGIDUNUM

DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE

Danijelova 32, Beograd

Kandidat: Nenad Panić

Broj indeksa: 480127/2022

Studijski program: Inteligentno softversko inženjerstvo

Tema:

METOD ZA SMANJENJE DEMOGRAFSKE PRISTRASNOSTI U MODELIMA PROCENE STAROSTI PUTEM OPTIMIZOVANE SKTRUKTURE SKUPOVA PODATAKA

Mentor:

Prof. dr Marina Marjanović, redovni profesor

Univerzitet Singidunum, Beograd

Članovi komisije:

Prof. dr Marko Tanasković, vanredni profesor

Univerzitet Singidunum, Beograd

Doc. dr Milan Čabarkapa, redovni profesor

Fakultet inženjerskih nauka, Univerzitet u Kragujevcu

Beograd, 2025. godine

METOD ZA SMANJENJE DEMOGRAFSKE PRISTRASNOSTI U MODELIMA PROCENE STAROSTI PUTEM OPTIMIZOVANE SKTRUKTURE SKUPOVA PODATAKA

Apstrakt: Pristrasnost u sistemima za prepoznavanje lica često dovodi do nejednakih performansi među demografskim grupama. Ova studija se bavi ovim problemom istraživanjem kako sastav skupa podataka utiče na performanse i pristrasnost modela za procenu starosti u različitim etničkim grupama. Prilagodili smo unapred trenirane konvolucione neuronske mreže (CNN), kao što je VGG19, na raznovrsnim skupovima podataka UTKFace (23.705 uzoraka: 10.078 belih, 4.526 crnih, 3.434 azijskih) i APPA-REAL (7.591 uzorak: 6.686 belih, 231 crnih, 674 azijskih). Naš pristup je uključivao prilagođavanje sastava skupova podataka prekomernim uzorkovanjem manjinskih grupa ili smanjenjem uzoraka iz prekomerno zastupljenih grupa radi ublažavanja pristrasnosti. Sproveli smo eksperimente kako bismo identifikovali optimalan sastav skupa podataka koji minimizira razlike u performansama među etničkim grupama. Primarna metrika performansi bila je srednja apsolutna greška (MAE), koja meri prosečan iznos grešaka u predikciji. Takođe smo analizirali standardnu devijaciju MAE među etničkim grupama kako bismo procenili konzistentnost i pravičnost performansi. Naši rezultati pokazuju da jednostavno prekomerno uzorkovanje manjinskih grupa ne osigurava pravične performanse. Umesto toga, sistematska prilagođavanja, uključujući smanjenje uzoraka iz prekomerno zastupljenih grupa, dovela su do uravnoteženijih performansi i niže standardne devijacije MAE među etničkim grupama. Ovi nalazi ističu značaj prilagođenih podešavanja skupa podataka i sugerišu istraživanje naprednih metoda obrade podataka i algoritamskih prilagođavanja kako bi se poboljšala pravičnost i tačnost u tehnologijama za prepoznavanje lica.

Ključne reči: prepoznavanje lica; procena starosti; konvoluciona neuronska mreža (CNN); etnička pristrasnost

2

Naučna oblast: Tehničko-tehnološke nauke

Uža naučna oblast: Elektrotehnika i računarstvo

METHOD FOR REDUCING DEMOGRAPHIC BIAS IN AGE ESTIMATION MODELS THROUGH OPTIMIZED DATASET COMPOSITION

Abstract: Bias in facial recognition systems often results in unequal performance across demographic groups. This study addresses this by investigating how dataset composition affects the performance and bias of age estimation models across ethnicities. We fine-tuned pre-trained Convolutional Neural Networks (CNNs) like VGG19 on the diverse UTKFace dataset (23,705 samples: 10,078 White, 4526 Black, 3434 Asian) and APPA-REAL (7591 samples: 6686 White, 231 Black, 674 Asian). Our approach involved adjusting dataset compositions by oversampling minority groups or reducing samples from overrepresented groups to mitigate bias. We conducted experiments to identify the optimal dataset composition that minimizes performance disparities among ethnic groups. The primary performance metric was Mean Absolute Error (MAE), measuring the average magnitude of prediction errors. We also analyzed the standard deviation of MAE across ethnic groups to assess performance consistency and equity. Our findings reveal that simple oversampling of minority groups does not ensure equitable performance. Instead, systematic adjustments, including reducing samples from overrepresented groups, led to more balanced performance and lower MAE standard deviations across ethnicities. These insights highlight the importance of tailored dataset adjustments and suggest exploring advanced data processing methods and algorithmic tweaks to enhance fairness and accuracy in facial recognition technologies.

Keywords: facial recognition; age estimation; convolutional neural network (CNN); ethnicity bias

Scientific field: Technical and technological sciences

Narrow scientific field: Electrical engineering and computing Science

Sadržaj

1. UVOD 9

1.1. PROBLEM ISTRAŽIVANJA – PRISTRASNOST MODELA PROCENE STAROSTI 9
1.2. CILJEVI I ISTRAŽIVAČKA PITANJA 9
1.3. HIPOTEZE ISTRAŽIVANJA 10
1.4. PREGLED RELEVANTNE LITERATURE 11
1.4.1. EMPIRIJSKA ZAPAŽANJA O PRISTRASNOSTI U MODELIMA PREPOZNAVANJA LICA 11
1.4.2. PREGLED METODA AUTOMATSKE PROCENE STAROSTI 12
1.4.3. PREGLED PRETHODNIH ISTRAŽIVANJA I NJIHOVIH NEDOSTATAKA 13
1.5. NAUČNA I DRUŠTVENA OPRAVDANOST 21
1.6. SADRŽAJ DISERTACIJE 22

2. BAZE PODATAKA ZA PROCENU STAROSTI 23

2.1. PREGLED DOSTUPNIH BAZA PODATAKA ZA PROCENU STAROSTI 23
2.1.1. ULOGA DATASETOVA U REŠAVANJU ZADATKA PROCENE STAROSTI 32
2.1.2. KRITERIJUMI ZA IZBOR DATASETOVA 33
2.2. DETALJAN OPIS KORIŠĆENIH SKUPOVA PODATAKA 35
2.2.1. UTKFACE DATASET 36
2.2.2. APPA – REAL DATASET 39
2.3. KOMPARATIVNA ANALIZA KORIŠĆENIH DATASETOVA 43
2.4. PREDPROCESOVANJE DATASETOVA 44
2.4.1. SKALIRANJE, FORMATIRANJE, ČIŠĆENJE; STANDARDIZACIJA ULAZA 45
2.4.2. KOMBINOVANJE I FILTRIRANJE GRUPA ZA EKSPERIMENT 45
2.5. MOGUĆE PRIMENE DATASETOVA I SMERNICE ZA BUDUĆA ISTRAŽIVANJA 46

3. ARHITEKTURA I IZBOR MODELA 47

3.1. KORIŠĆENE CNN ARHITEKTURE 47
3.1.1. VGG16 47
3.1.2. VGG19 47
3.1.3. RESNET50 48
3.1.4. MOBILENETV2 48

3.2. OBRAZLOŽENJE IZBORA CNN MODELA 48
3.2.1. EFIKASNOST EKSTRAKCIJE KARAKTERISTIKA I OTPORNOST NA INTRA-KLASNE VARIJACIJE 49
3.2.2. PRILAGOĐENOST ZADATKU PROCENE STAROSTI 49
3.3. UPOREDNA ANALIZA PERFORMANSI ARHITEKTURA 50
3.4. ARHITEKTURA IZABRANOG MODELA (VGG19) 50

4. METODOLOGIJA 52

4.1. TRANSFER UČENJE I FINO PODEŠAVANJE MODELA 52
4.2. PODEŠAVANJE HIPERPARAMETARA PUTEM PRETRAGE MREŽOM 53
4.3. STRATEGIJE AUGMENTACIJE I REGULARIZACIJE MODELA 55
4.4. DIZAJN EKSPERIMENTA I EVALUACIJA PRAVIČNOSTI 55

5. REZULTATI 57

5.1. METODOLOGIJA EVALUACIJE PERFORMANSI 57
5.2. REZULTATI NA APPA-REAL SKUPU PODATAKA 58
5.3. REZULTATI NA UTKFACE SKUPU PODATAKA 68
5.4. UPOREDNA ANALIZA: NAŠ MODEL NASPRAM SLIČNIH RADOVA 78

6. ZAKLJUČAK 79

6.1. KLJUČNI CILJEVI I METODOLOŠKI PRISTUP 79
6.2. EFEKTI MANIPULACIJE PODACIMA NA PRAVIČNOST 79
6.3. REZULTATI MODELA NA UTKFACE I APPA-REAL SKUPOVIMA 79
6.4. ANALIZA MAPA KARAKTERISTIKA I RAZLIKE U PONAŠANJU MODELA 80
6.5. TUMAČENJE I SLOŽENOST PRAVIČNIH PERFORMANSI 82
6.6. STRATEGIJE ZA POBOLJŠANJE PRAVIČNOSTI 83
6.7. OGRANIČENJA I MOGUĆA POBOLJŠANJA 83
6.8. POREĐENJE SA PRETHODNIM ISTRAŽIVANJIMA 84
6.9. PREPORUKE ZA PRAKTIČNE SISTEME I BUDUĆA ISTRAŽIVANJA 85
6.10. ZAVRŠNA RAZMATRANJA 86

LITERATURA 87

SPISAK SLIKA **90**

SPISAK TABELA **91**

1. Uvod

1.1. Problem istraživanja – pristrasnost modela procene starosti

Pristrasnost u sistemima za prepoznavanje lica, uključujući i modele za procenu starosti predstavlja kritičan problem koji direktno utiče na pravičnost, transparentnost i tačnost. Ova pristrasnost može se ispoljiti u različitim oblicima, uključujući razlike u starosti, polu i etničkoj pripadnosti, često kao rezultat inherentnih pretpostavki i procesa donošenja odluka ugrađenih u arhitekturu modela.

Takva neravnoteža može dovesti do sistematskih netačnosti i diskriminatornih rezultata, čime se narušavaju osnovna prava i poverenje u automatizovane sisteme. Rešavanje pristrasnosti u sistemima za prepoznavanje lica ključno je kako bi se osiguralo da ove tehnologije budu pravične i pravedne, poštujući prava i dostojanstvo svih pojedinaca, uz istovremeno usklađivanje sa različitim regulativama koje imaju za cilj sprečavanje diskriminacije. Ovo uključuje Univerzalnu deklaraciju o ljudskim pravima, Evropsku konvenciju o ljudskim pravima i Opštu uredbu o zaštiti podataka (GDPR) [1,2].

Javni skupovi podataka za treniranje često pogoršavaju problem pristrasnosti jer su u velikoj meri usmereni ka određenim etničkim grupama, posebno licima belih osoba. Ovaj nedostatak raznolikosti može rezultirati manje preciznim prepoznavanjem pojedinaca iz nedovoljno zastupljenih etničkih grupa. Modeli trenirani na ovakvim pristrasnim skupovima podataka ne uspevaju dobro da generalizuju performanse na različite demografske grupe, što dovodi do sistematskih netačnosti i nepravednih ishoda.

1.2. Ciljevi i istraživačka pitanja

Kako automatska procena starosti postaje sve češće korišćena u različite svrhe poput forenzike [3] i nadzora, ovaj podzadatak prepoznavanja lica privukao je značajnu pažnju istraživača. Ova studija se fokusira na unapređenje pravičnosti modela za procenu starosti, s posebnim fokusom na rasnu pristrasnost.

Koristeći skupove podataka UTKFace i APPA-REAL, odabrane zbog njihove demografske raznovrsnosti i dostupnosti oznaka poput stvarne starosti i etničke pripadnosti, koje su ključne za ovo istraživanje, analiziramo uticaj neuravnoteženih podataka za treniranje na performanse i pristrasnost modela.

Primarna istraživačka pitanja su:

1. Kako prilagođavanje sastava skupa podataka utiče na performanse modela za procenu starosti?
2. Može li rebalansiranje skupova podataka ublažiti pristrasnost među različitim etničkim grupama?

Ova pitanja su ključna za razumevanje načina na koji se mogu razviti pravedniji i tačniji sistemi za prepoznavanje lica. Umesto fokusa na postizanje boljih celokupnih rezultata u odnosu na prethodna istraživanja na ova dva skupa podataka, cilj nam je:

- Analizirati odnos između sastava skupa podataka i ukupnih, kao i etnički specifičnih performansi modela;
- Kvantifikovati u kojoj meri rebalansiranje skupova podataka može ublažiti pristrasnost u modelima za procenu starosti;
- Utvrditi da li je samo rebalansiranje skupova podataka dovoljno ili treba biti kombinovano sa drugim tehnikama za ublažavanje pristrasnosti.

Baveći se ovim ciljevima, ova studija teži da doprinese razvoju pravednijih i tačnijih modela za procenu starosti, čime se unapređuje pouzdanost i pravičnost tehnologije za prepoznavanje lica.

1.3. Hipoteze istraživanja

Opšta hipoteza od koje se polazi u istraživanju je: „Pristrasnost u modelima za procenu starosti može se ublažiti optimizacijom sastava skupa podataka“.

Posebna hipoteza koja proizilazi iz opšte je: „Smanjenje pristrasnosti modela za procenu starosti može se postići prilagođavanjem distribucije uzoraka u trening skupu, bez značajnog narušavanja tačnosti modela“.

Pojedinačne hipoteze koje se testiraju u okviru istraživanja su:

- Pristrasnost modela u proceni starosti može se kvantifikovati analizom razlika u performansama među etničkim grupama.
- Sastav skupa podataka utiče na srednju apsolutnu grešku (MAE) i konzistentnost modela među različitim demografskim grupama.

- Prekomerno uzorkovanje manjinskih grupa ne osigurava nužno pravičnije performanse modela.
- Sistematsko smanjenje uzoraka prekomerno zastupljenih grupa može poboljšati ravnotežu u performansama modela među različitim etničkim grupama.
- Postoje napredne metode obrade podataka i algoritamskih podešavanja koje mogu dodatno smanjiti pristrasnost u modelima za procenu starosti.

1.4. Pregled relevantne literature

1.4.1. Empirijska zapažanja o pristrasnosti u modelima prepoznavanja lica

Literatura potvrđuje da sistemi za prepoznavanje lica i pridružene komponente kao što je procena starosti pokazuju sistematske razlike u tačnosti preko pola, starosti i etničke pripadnosti. Na primer, studije ukazuju na slabije performanse na licima dece [4,5] i na to da se žene češće pogrešno klasifikuju, delimično zbog većeg pokrivanja lica dužom kosom [6,7].

Dodatno, pristrasnost po rasnoj osnovi pokazuje visok stepen nekonzistentnosti između različitih skupova podataka, pri čemu na performanse modela mogu uticati i vanrasni faktori kao što su kvalitet slike, osvetljenje, ugao posmatranja ili izrazi lica. Šminka, na primer, identifikovana je kao značajan činilac koji može doprineti rodnoj pristrasnosti u evaluaciji [8].

Posebnu ulogu u oblikovanju performansi modela ima izbor i kompozicija skupa podataka. Većina javno dostupnih setova je demografski neuravnotežena, sa dominantnom zastupljenošću belih lica, što otežava validnu generalizaciju na nedovoljno zastupljene grupe. Iako su datasetovi poput FairFace dizajnirani da adresiraju ovaj problem, njihova primena još uvek ne rešava sve aspekte pristrasnosti, naročito u kontekstu regresije starosti i uticaja neuravnoteženog treninga [9].

Pored strukturalnih aspekata skupa podataka, radovi koji se bave razlikom između stvarne i očigledne starosti ukazuju na dodatne izazove u domenu pravičnosti. Atributi kao što su pol, rasa, facijalni izrazi i prisustvo šminke dokazano utiču na predikcije modela [10,11], čime se naglašava da i „okolišni“ i anotacioni faktori imaju ključan uticaj na evaluaciju pravičnosti u modelima računarskog vida.

1.4.2. Pregled metoda automatske procene starosti

Procena starosti može se rešavati različitim metodama, uključujući tehnike ručne ekstrakcije karakteristika i modele dubokog učenja, poput konvolucionih neuronskih mreža (CNN-ova). Istraživači su detaljno proučavali faktore koji utiču na starenje lica, koji mogu biti intrinzični (genetski) ili ekstrinzični (okolišni). Različite metode za reprezentaciju slika i modelovanje starosti su istražene, uključujući antropometrijske modele, modele aktivnih oblika (ASMs), modele aktivnog izgleda (AAMs), podprostorne obrasce starenja (AGES), starosne mnogostrukosti, modele izgleda i hibridne modele.

Pored ovih, tehnike ekstrakcije karakteristika, poput Gaborovih filtera, linearne diskriminantne analize (LDA), lokalnih binarnih obrazaca (LBPs), lokalnih direkcionih obrazaca (LDPs), lokalnih ternarnih obrazaca (LTPs), matrice ko-pojavnih nivoa sive boje (GLCM), prostorno fleksibilnih zakrpa (SFPs), Grassmannovih mnogostrukosti i biološki inspirisanih karakteristika (BIFs) takođe su istraživane [12].

Angulu i saradnici (2018) su sproveli pregled različitih tehnika procene starosti, sumirajući srednju apsolutnu grešku (MAE) i kumulativni skor (CS) različitih modela za procenu starosti ili starosnih grupa. Otkrili su da hibridni pristupi, koji kombinuju klasifikaciju i regresiju, generalno daju bolje rezultate u poređenju sa korišćenjem samo jedne od ovih metoda. Štaviše, metode dubokog učenja, posebno CNN-ovi, pokazale su obećavajuće rezultate, često nadmašujući tradicionalne metode [12].

ELKarazle i saradnici (2022) podržali su ove nalaze pružajući sveobuhvatan pregled tehnika mašinskog učenja za procenu starosti na osnovu slika lica. Izdvojili su izazove u ovom zadatku, poput varijacija u obrascima starenja među pojedincima, koje su rezultat genetike, načina života, zdravstvenog stanja i faktora okoline. Takođe su istakli ograničenu dostupnost raznovrsnih i visokokvalitetnih skupova podataka sa slikama lica, koji obuhvataju širok raspon starosnih doba, polova i etničkih grupa, kao i varijacije u osvetljenju, pozama i izrazu lica, koje utiču na ukupnu tačnost i performanse procene starosti među različitim etničkim grupama.

ELKarazle i saradnici zaključili su da modeli dubokog učenja, naročito oni zasnovani na transfernom učenju, generalno nadmašuju ručno izrađene modele zbog svoje sposobnosti da automatski uče složene karakteristike [13].

Ovaj zaključak dodatno podržava još jedna studija u kojoj su tehnike ručne ekstrakcije karakteristika primenjene na skupove podataka za procenu starosti lica i UTK Face skup podataka. Dobijene filtrirane slike konvertovane su u skalare i unete u Random Forest klasifikator i Support Vector klasifikator. U poređenju sa CNN-om, CNN je nadmašio tradicionalne tehnike mašinskog učenja za klasifikaciju starosti za čak 40% [14]. Ovaj dokazani uspeh CNN-ova je razlog zašto smo ih odabrali za potrebe našeg istraživanja.

1.4.3. Pregled prethodnih istraživanja i njihovih nedostataka

Kao što je ranije pomenuto, problem procene starosti privukao je značajnu pažnju istraživača. Detaljno poređenje studija o proceni starosti, kao i o ublažavanju neuravnoteženosti skupova podataka, prikazano je u Tabeli 1.

Tabela 1. Poređenje prethodnih radova.

Rad	Metodologije	Prednosti	Mane	Glavni nalazi
Analiza rasne i rodne pristrasnosti u modelima duboke procene starosti [8]	WideResNet (UTKFace, IMDB-WIKI) i FaceNet (IMDB-WIKI)	Koristi unapred trenirane modele; sveobuhvatna evaluacija pomoću MAE i RMSE	Inkozistentnost performansi između skupova podataka; mogući uticaj nerasnih faktora poput kvaliteta slike i poziranja	Muški subjekti imaju tačniju procenu starosti; inkozistentna rasna pristrasnost između skupova podataka; sugeriše da šminka može biti faktor za rodnu pristrasnost
FairFace: Skup podataka o atributima lica za uravnoteženu rasu, pol i starost [9]	ResNet-34, evaluiran na FairFace, UTKFace, LFWA+, CelebA	Uravnotežen skup podataka među sedam rasnih grupa; poboljšane performanse na različitim skupovima podataka	Potencijalne pristrasnosti u spoljnim skupovima podataka; ograničene metrike evaluacije	FairFace model pokazuje dosledne performanse sa razlikom u tačnosti manjom od 1% između klasifikacija muških/ženskih i belih/ne-belih osoba

Procena starosti i pola sa slika lica korišćenjem pažljivih konvolucionih mreža [15]	Residual Attention Network (RAN) and ResNet	Mehanizmi pažnje za važne regije lica; ansambl model poboljšava performanse	Zahteva složeno treniranje modela; potencijalno prekomerno prilagođavanje (overfitting)	Ansambl model nadmašuje pojedinačne modele u tačnosti predikcije starosti i pola
Inteligentan okvir za predikciju ljudske starosti sa slike lica zasnovan na algoritmima dubokog učenja [16]	Duboka konvoluciona neuronska mreža (DCNN) sa Cuckoo Search (CS) algoritmom	Visoka tačnost i računarska efikasnost; nova upotreba CS za optimizaciju	Kompleksnost u prethodnoj obradi; zahteva obimne računarske resurse	DCNN-CS model nadmašuje druge metode poput CNN, DNN, LSTM i SVR u tačnosti i efikasnosti
Procena starosti na ljudskim slikama lica korišćenjem Support Vector Regression (SVR) i karakteristika zasnovanih na teksturi [17]	SVR sa LBP, LPQ i BSIF za ekstrakciju karakteristika	Efikasne metode ekstrakcije karakteristika; kombinuje više karakteristika zasnovanih na teksturi	Ograničeno na karakteristike zasnovane na teksturi; visok računarski trošak	Kombinacija BSIF i LPQ karakteristika sa PCA dimenzijom od 70 postiže najbolju srednju apsolutnu grešku (MAE)
Procena starosti na osnovu slika lica: pregled [12]	Različite strategije validacije; učenje metrika sa više mnogostrukosti	Sveobuhvatan pregled; naglašava strategije validacije	Ne pruža nove eksperimente; ograničene praktične primene	Ističe važnost strategija validacije kako bi se izbeglo prekomerno prilagođavanje (overfitting) i poboljšala generalizacija
Od očigledne do stvarne starosti: analiza pristrasnosti prema polu, starosti, etničkoj pripadnosti, šminki i izrazu lica u proceni stvarne starosti [10]	CNN sa korekcijom pristrasnosti u predikcijama	Bavi se korekcijom pristrasnosti; poboljšava procenu stvarne starosti	Ograničeno na specifične pristrasnosti; zahteva obiman skup podataka	Korišćenje očiglednih oznaka za trening poboljšava procenu stvarne starosti; korekcija pristrasnosti poboljšava performanse

Dijagnostikovanje modela dubokog učenja za preciznu procenu starosti sa slike [18]	Sistemska dijagnostika sa modelima dubokog učenja; arhitekture za učenje više zadataka	Sveobuhvatna evaluacija procedura treniranja; pristup sa više zadataka	Visoka kompleksnost; zahteva velike skupove podataka	Pristup zasnovan na regresiji sa MAE gubitkom je preferiran; arhitektura za učenje više zadataka nadmašuje druge modele

O uticaju pristrasnosti u percepciji starosti na regresiju stvarne starosti [11]	VGG16 sa integracijom atributa lica	Inkorporira pristrasnosti ljudske percepcije; poboljšava tačnost procene starosti	Zahteva složenu adaptaciju modela; obimno treniranje	Inkorporacija atributa lica poboljšava procenu starosti; analiza zasnovana na atributima otkriva uticaj pola, rase, sreće i šminke
Pregled metoda za upravljanje klasifikacijom i rešavanje problema neuravnoteženosti podataka [19]	Povećavanje uzoraka (SMOTE, ADASYN), smanjivanje uzoraka (RUS, T-Link), SVM, KNN, Naivni Bayes, Odlučujuće stablo, Bagging, Boosting (AdaBoost, RUSBoost, SMOTEBoost)	Ističe metode koje poboljšavaju predikciju manjinske klase i smanjuju uticaj neuravnoteženosti klasa; hibridne metode često daju bolju tačnost	Povećani računarski trošak; potencijalni gubitak informacija tokom undersamplinga; izazovi sa podacima visoke dimenzionalnosti	Neuravnoteženost podataka i dalje predstavlja ključni izazov koji utiče na performanse klasifikatora; hibridne i ansambl metode su efikasne, ali računarski zahtevne
Rešavanje problema neuravnoteženosti klasa u medicinskim skupovima podataka [20]	SMOTE oversampling, undersampling zasnovan na klasterima, odlučujuće stablo, Fuzzy Unordered Rule	Uravnotežava podatke; poboljšava predikciju manjinske klase; efikasno za medicinske skupove podataka	Povećani računarski trošak; kompleksnost u implementaciji	Modifikovana metoda undersamplinga zasnovana na klasterima nadmašuje tradicionalne metode; SMOTE

	Induction Algoritam			pokazuje dobre rezultate u klasifikaciji
Klasifikacija neuravnoteženih skupova podataka i rešenja: Pregled [21]	Tehnike na nivou podataka (SMOTE, ADASYN), Tehnike na nivou algoritma (učenje osetljivo na troškove), Ansambl metode (Bagging, Boosting)	Sveobuhvatan pregled tehnika; primenjivo u različitim domenima	Kompleksnost implementacije; nedostatak empirijske validacije za neke tehnike	Tehnike poput SMOTE-a i učenja osetljivog na troškove su efikasne; izbor metode treba da bude specifičan za skup podataka kako bi se postigli optimalni rezultati
Upravljanje neuravnoteženim skupovima podataka: Pregled [22]	Različite tehnike ponovnog uzorkovanja (slučajno oversampling, undersampling), učenje osetljivo na troškove, ansambl metode (boosting, bagging), selekcija karakteristika	Sveobuhvatan pregled; ističe više pristupa; praktična preporuka	Nema jedinstvene najbolje metode; neke tehnike mogu dovesti do prekomernog prilagođavanja (overfitting) ili gubitka korisnih podataka	Naglašava potrebu za prilagođenim pristupom zasnovanim na karakteristikama skupa podataka; preporučuje ansambl metode i učenje osetljivo na troškove

Prethodna istraživanja, kao što je analiza Puc-a i saradnika (2020) [8], istražuju performanse modela za procenu starosti prethodno obučenih na skupovima podataka kao što su UTKFace i APPA-REAL u različitim rasnim i polnim grupama. Oni su utvrdili da su modeli obično precizniji za muškarce nego za žene, što sugeriše na pristrasnost prema polu. Takođe, razlike u performansama između rasa pokazuju nekonzistentne varijacije između skupova podataka, što ukazuje na to da faktori poput kvaliteta slike i pozicije mogu takođe uticati na tačnost. Iako Puc i saradnici (2020) [8] priznaju potencijalnu

neuravnoteženost skupa podataka, oni aktivno ne manipulišu skupom podataka kako bi smanjili pristrasnost. Naše istraživanje se specifično bavi efektima ponovnog balansiranja skupova podataka kako bi se postigla ravnotežnija reprezentacija različitih rasnih grupa, pružajući konkretne dokaze da je ponovno balansiranje kao strategija smanjenja pristrasnosti efikasno.

Karkkainen i Joo (2019) [9] uvode FairFace skup podataka kako bi smanjili rasnu pristrasnost u skupovima podataka sa atributima lica. Međutim, oni ne istražuju uticaj neuravnoteženih podataka za obuku na performanse modela i pristrasnost, kao što to radimo mi. Njihovo korišćenje starosnih opsega umesto tačnih godina komplikuje direktnu uporedivost sa našim radom, koji se fokusira na preciznu regresiju starosti.

Iako su postigli relativno uporedive performanse među rasnim grupama, njihova sveukupna tačnost u klasifikaciji je oko 60%.

Abdolrashidi i saradnici (2020) [15] fokusiraju se na poboljšanje tačnosti predikcije starosti i pola korišćenjem ansambla pažljivih i rezidualnih konvolucionih neuronskih mreža. Iako koriste demografski raznovrsni UTKFace skup podataka, oni ne analiziraju uticaj neuravnoteženosti skupa podataka na performanse modela među etničkim grupama ili polovima. Njihovo istraživanje ističe izazove u predikciji starosti i pola zbog intra-klasnih varijacija, ali ne istražuje strategije smanjenja pristrasnosti. Naš rad proširuje njihova saznanja time što se specifično bavi neuravnoteženostima skupa podataka i istražuje tehnike za smanjenje pristrasnosti.

Sathyavathi i Baskaran (2023) [16] fokusiraju se na poboljšanje tačnosti predikcije starosti korišćenjem okvira dubokog učenja koji kombinuje Duboku Konvolucionu Neuronsku Mrežu (DCNN) sa Cuckoo Search (CS) algoritmom. Iako su njihovi skupovi podataka (UTKFace, FGNET, CACD) raznovrsni, oni ne analiziraju efekat sastava skupa podataka na performanse za specifične etničke grupe niti istražuju strategije smanjenja pristrasnosti. Njihovo istraživanje pominje potencijalne probleme sa prikupljanjem relevantnih karakteristika za tačnu procenu starosti, ali ne ulazi dublje u ove probleme niti istražuje njihove uzroke.

Amelia i Wahyono (2022) [17] teže poboljšanju tačnosti procene starosti korišćenjem karakteristika zasnovanih na teksturi i regresije podržanih vektora (SVR). Oni

priznaju ograničenja svog skupa podataka, koji se uglavnom sastoji od slika iz zapadnih zemalja i možda neće dobro funkcionisati na azijskim slikama zbog nedovoljnog predstavljanja. Ovo sugeriše potencijalnu pristrasnost, ali ovo nije dublje istraženo u njihovom radu.

Anketno istraživanje o proceni starosti koje su sproveli Angulu i saradnici (2018) [12] diskutuje izazove skupa podataka, ali ne istražuje duboko neuravnoteženost povezanu sa etničkim grupama. Naše istraživanje cilja na kritični problem neuravnoteženosti skupa podataka, obezbeđujući pravičnost i tačnost među različitim etničkim grupama.

Xing sa saradnicima (2019) [18] istražuje arhitekture modela, ali ne vrši opsežnu evaluaciju performansi među različitim etničkim grupama. Naše istraživanje pažljivo procenjuje degradaciju tačnosti u odnosu na etničke grupe, pružajući vredne uvide u algoritamske pristrasnosti u sistemima za prepoznavanje lica.

Studija Clapes i saradnika (2018) [10] analizira pristrasnosti u proceni očigledne starosti i koristi ih za poboljšanje procene stvarne starosti. Oni identifikuju ciljne pristrasnosti (pol, etnička pripadnost, šminka, facijalni izraz) i pristrasnosti u pretpostavkama (pristrasnosti koje uvode osobe koje prave procene). Iako oni eksplicitno ne istražuju uticaj neuravnoteženih podataka za obuku, njihov rad se bavi pitanjem pristrasnosti u modelima za procenu starosti.

Jacques i saradnici (2019) [11] fokusiraju se na poboljšanje procene stvarne starosti uključujući očiglednu starost i facijalne karakteristike u model dubokog učenja. Oni naglašavaju značaj razmatranja pristrasnosti povezanih sa facijalnim karakteristikama u proceni starosti, ali ne bave se direktno neuravnoteženostima u podacima ili tehnikama rebalansiranja. Njihov rad identifikuje nekoliko problema u postojećim metodama, kao što su pristrasnosti u percepciji starosti na koje utiču različiti faktori, uključujući pol, rasu, facijalne izraze i šminku, što utiče na tačnost modela za procenu starosti.

Anketni rad o metodama za upravljanje klasifikacijom i rešavanjem problema neuravnoteženosti podataka (2020) autora Hasib I saradnika [19] pruža opsežan pregled metoda koje se bave neuravnoteženošću klasa u skupovima podataka. Ovaj rad

kategorizuje metodologije u metode na nivou podataka, metode na nivou algoritma, metode ansambla i hibridne metode.

Metode na nivou podataka uključuju tehnike prekomernog uzorkovanja kao što su SMOTE i ADASYN, koje generišu sintetičke podatke za balansiranje distribucije klasa, i tehnike poduzorkovanja kao što su RUS i T-Link, koje smanjuju veličinu većinske klase kako bi balansirao skup podataka. Metode na nivou algoritma fokusiraju se na poboljšanje klasifikatora kao što su SVM, KNN, Naivni Bayes i stabla odlučivanja kako bi se nosili sa neuravnoteženim podacima. Metode ansambla, kao što su Bagging i Boosting, kombinuju više algoritama za poboljšanje performansi klasifikacije. Rad takođe ističe hibridne metode koje integrišu uzorkovanje podataka i poboljšanje algoritama kako bi efikasno rešavao problem neuravnoteženosti klasa. Studija naglašava računske izazove povezane sa ovim metodama, posebno u upravljanju podacima visokih dimenzija i obezbeđivanju minimalnog gubitka informacija tokom uzorkovanja. Rad zaključuje da, iako hibridne i ansambl metode pružaju značajna poboljšanja, one dolaze sa povećanim računskim troškovima i složenošću.

Rad rešavanje problema neuravnoteženosti klasa u medicinskim skupovima podataka autora Rahmana i Davisa (2013) [20] istražuje performanse tehnika oversamplinga i undersamplinga za balansiranje kardiovaskularnih podataka. Korišćene metodologije uključuju SMOTE tehniku oversamplinga i tehniku undersamplinga zasnovanu na klasterima. Prednosti ovog pristupa su balansiranje podataka, poboljšanje predikcije manjinskih klasa i efikasnost za medicinske skupove podataka. Međutim, postoje i nedostaci, kao što su povećani računski troškovi i složenost implementacije. Glavni zaključak rada je da modifikovana metoda undersamplinga zasnovana na klasterima nadmašuje tradicionalne metode, a SMOTE daje dobar rezultat u klasifikaciji.

U radu „Klasifikacija neuravnoteženih skupova podataka i rešenja: Pregled" (2014) Dr. D. Ramyachitre i P. Manikandana [21], autori pružaju sveobuhvatan pregled izazova i rešenja u vezi sa neuravnoteženim skupovima podataka. Istražuju različite metodologije, uključujući tehnike na nivou podataka poput SMOTE-a, algoritamske metode kao što je učenje sa troškovima i ansambl metode poput Bagging-a i Boosting-a. Prednosti njihovog pristupa uključuju detaljnu kategorizaciju metoda i njihovu efikasnost u različitim scenarijima. Međutim, rad takođe ističe složenost implementacije ovih metoda i

nedostatak empirijske validacije za neke tehnike. Glavni zaključak rada je da su tehnike poput SMOTE-a i učenja sa troškovima efikasne, ali izbor metode treba da bude prilagođen specifičnim karakteristikama skupa podataka kako bi se postigao optimalan rezultat.

Rad "Upravljanje neuravnoteženim skupovima podataka: Pregled" autora Kotsiantisa, Kanellopoulosa i Pintelasa (2006) [22] istražuje različite tehnike za rešavanje problema neuravnoteženih skupova podataka, koji su česti u stvarnim klasifikacionim problemima gde je jedna klasa značajno brojnija od druge. Autori kategorizuju rešenja u tehnike na nivou podataka i algoritamske tehnike. Na nivou podataka, rešenja uključuju različite oblike ponovnog uzorkovanja, kao što su nasumično ponavljanje, nasumično smanjenje, usmereno ponavljanje i usmereno smanjenje, zajedno sa selekcijom karakteristika prilagođenom neuravnoteženim skupovima podataka. Algoritamska rešenja uključuju prilagođavanje troškova pogrešnih klasifikacija i pragova odlučivanja, kao i primenu metoda za učenje sa jednom klasom. Pregled ističe upotrebu ansambl metoda, poput boosting-a i bagging-a, za poboljšanje klasifikacionih rezultata kombinovanjem više modela. Rad naglašava da, iako metode ponovnog uzorkovanja mogu rešiti problem neuravnoteženosti, one mogu uvesti i druge probleme poput prekomernog prilagođavanja (overfitting) ili gubitka korisnih podataka. Pregled takođe razmatra metrike za evaluaciju specifične za neuravnotežene skupove podataka, uključujući ROC krive i metrike preciznosti-povratka. Glavni zaključak je da nijedna pojedinačna metoda univerzalno ne prevazilazi druge, i da izbor tehnike često zavisi od specifičnih karakteristika skupa podataka i problema koji se rešava.

Kako bismo rešili problem neuravnoteženosti skupa podataka, istražili smo različite strategije za ublažavanje pristrasnosti u modelima procene starosti prilagođavanjem reprezentacije različitih rasnih grupa. Za razliku od prethodnih studija koje su koristile napredne tehnike i kombinacije poput SMOTE, ADASYN i ansambl metoda, naš pristup je sistematski smanjivao veličinu uzorka određenih rasnih grupa, dok su druge grupe ostale konstantne, kako bi se posmatrao uticaj na performanse modela.

Iz naših eksperimenata smo primetili da varijacije u skupu podataka značajno utiču na srednju apsolutnu grešku (MAE) i standardnu devijaciju među različitim etničkim

grupama. Na primer, smanjenje uzorka azijske grupe za 90% rezultiralo je ukupnim MAE od 5,4808 i standardnom devijacijom od 0,3015, što ukazuje na poboljšanje u doslednosti performansi u poređenju sa originalnim skupom podataka (Ukupni MAE: 4,8925, standardna devijacija: 0,7401). Međutim, smanjenje uzorka drugih grupa nije uvek dovelo do boljih rezultata, što ukazuje na složenost postizanja uravnoteženih performansi među svim demografijama.

U nekim slučajevima, balansiranje skupova podataka smanjenjem uzoraka iz prekomerno zastupljenih grupa dovelo je do nižih standardnih devijacija i pravednijih performansi među različitim rasnim grupama. Na primer, balansiranje bele grupe na 20% i 40% dalo je standardne devijacije od 0,0440 i 0,2451, što ukazuje na stabilniji učinak.

Naši nalazi sugerišu da, iako napredne tehnike poput SMOTE i ansambl metoda imaju svoje prednosti, sistematski i jednostavniji pristup prilagođavanju reprezentacije grupa takođe može pružiti vredne uvide i potencijalno poboljšati pravednost performansi modela. Ovaj novi pristup selektivnog smanjenja uzoraka iz određenih grupa može dopuniti postojeće tehnike i ponuditi nova gledišta za buduća istraživanja.

Zaključno, naše istraživanje pokazuje da balansiranje skupa podataka može efikasno ublažiti pristrasnost i poboljšati pravednost modela za procenu starosti. Sistematskom analizom uticaja različitih kompozicija skupa podataka, pružamo sveobuhvatan uvid u to kako različiti pristupi utiču na performanse modela i pravednost. Naši nalazi doprinose tekućim naporima u razvoju pravednijih i tačnijih tehnologija prepoznavanja lica, na kraju unapređujući ovu oblast prema pravednijim i nepristrasnijim rešenjima.

1.5. Naučna i društvena opravdanost

U eri sve šire primene veštačke inteligencije u osetljivim oblastima poput nadzora, bezbednosti i digitalne identifikacije, postaje ključno da razvijeni sistemi budu pravični, pouzdani i odgovorni. Posebno u zadacima kao što je automatska procena starosti na osnovu slike lica, pristrasnosti u performansama modela prema određenim etničkim ili demografskim grupama mogu imati ozbiljne posledice po pojedince i društvo u celini.

Ova disertacija se bavi jednom od ključnih prepreka za etičku i pravednu upotrebu AI sistema – načinom na koji sastav skupa podataka utiče na performanse i demografsku

pravičnost modela. Pitanje uravnoteženosti i reprezentativnosti podataka nije samo tehnički izazov, već i fundamentalni etički zahtev, čije zanemarivanje može dovesti do diskriminatornih praksi, pa čak i kršenja regulatornih okvira.

Naučni doprinos ove disertacije ogleda se u sistematskoj analizi uticaja različitih distribucija podataka na performanse modela za procenu starosti. Rad pokazuje da jednostavne, ali promišljene strategije balansiranja mogu dovesti do stabilnijih i pravičnijih rezultata, bez potrebe za prekomerno složenim modelima ili algoritamskim intervencijama. Time se doprinosi razvoju prakse koja je ujedno tehnički efikasna, regulatorno usklađena i društveno odgovorna.

1.6. Sadržaj disertacije

Disertacija je strukturirana tako da prati tok istraživanja, od postavljanja problema do eksperimentalne evaluacije i zaključaka. Korišćene su savremene istraživačke metode u oblasti računarskog vida i dubokog učenja, uključujući eksperimentalnu analizu, evaluaciju modela, kvantitativnu obradu rezultata i uporednu analizu performansi u kontekstu demografske pravičnosti. U uvodnom delu rada definišu se predmet i ciljevi istraživanja, uz postavljanje teorijskog okvira koji obuhvata značaj pravičnosti u modelima veštačke inteligencije, posebno u zadacima koji uključuju predikciju starosti na osnovu slika lica. Drugo poglavlje posvećeno je detaljnoj analizi korišćenih skupova podataka sa fokusom na njihovu demografsku strukturu, distribuciju starosnih grupa, etničku raznovrsnost i pripremu podataka za obuku modela. Treće poglavlje daje prikaz izabranih konvolucionih neuronskih mreža, zajedno sa obrazloženjem izbora arhitektura i analizom njihovih performansi na osnovu zadatka regresije stvarne starosti. U četvrtom poglavlju detaljno se opisuje metodologija istraživanja: primena transfer učenja, strategije finog podešavanja modela, optimizacija hiperparametara putem pretrage mrežom, strategije augmentacije i balansiranja podataka, kao i dizajn eksperimenta sa naglaskom na evaluaciju pravičnosti i standardnih metrika performansi. Peto poglavlje prikazuje rezultate eksperimenata, uključujući efekte manipulacije distribucijom podataka po etničkim grupama, uticaj na srednju apsolutnu grešku (MAE), kao i na druge metrike. Završno, šesto poglavlje sadrži zaključke do kojih se došlo na osnovu sprovedenih analiza, kao i diskusiju o izazovima, ograničenjima i potencijalnim pravcima budućih istraživanja. Poseban akcenat stavljen

je na implikacije dobijenih rezultata za izradu pravičnih i pouzdanih sistema za automatsku procenu starosti u praktičnim primenama.

2. Baze podataka za procenu starosti

2.1. Pregled dostupnih baza podataka za procenu starosti

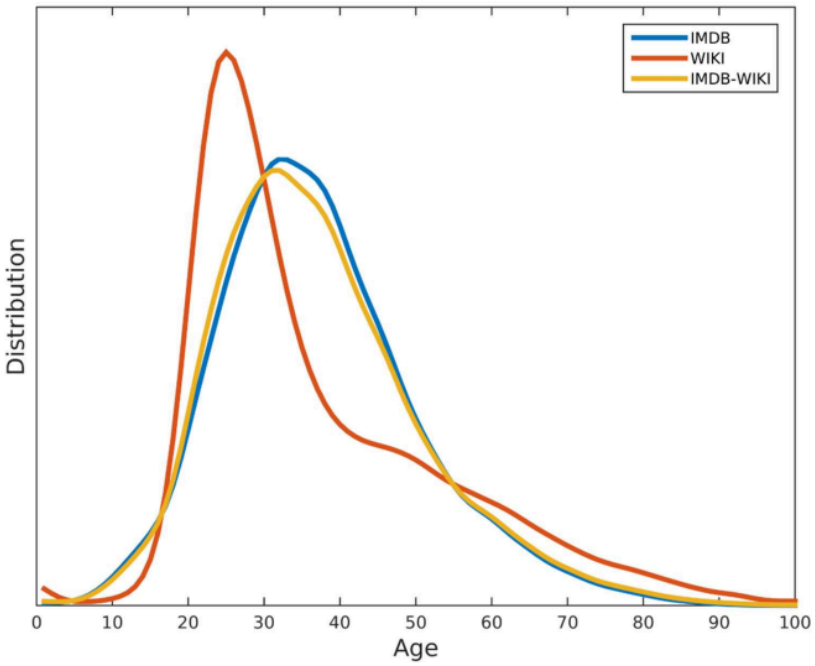
Razvoj i evaluacija modela za automatsku procenu starosti oslanjaju se na raspoložive baze podataka, čije karakteristike direktno utiču na performanse i pravičnost modela. U ovoj oblasti pravi se jasna razlika između stvarne (hronološke) starosti, koja se temelji na pouzdanim metapodacima kao što su datumi rođenja, i očigledne starosti, koja odražava subjektivni utisak posmatrača o starosti osobe na slici. Obe vrste oznaka imaju svoju primenu: stvarna starost pogodna je za tačne regresione zadatke, dok se očigledna koristi u kontekstima gde je relevantna ljudska percepcija starosti. Skupovi podataka se dalje razlikuju prema uslovima prikupljanja slika od strogo kontrolisanih okruženja sa standardizovanim osvetljenjem i pozom, do “in-the-wild” datasetova koji sadrže slike sa velikim varijacijama u osvetljenju, poziciji, kvalitetu i izrazu lica. Takvi faktori mogu značajno uticati na tačnost modela i na distribuciju grešaka između različitih demografskih grupa. Pored toga, sve veći broj studija ukazuje na činjenicu da neuravnoteženost podataka po demografskim kategorijama, poput pola ili rase, dovodi do sistematskih pristrasnosti u rezultatima modela. Modeli obučeni na takvim podacima često pokazuju značajno bolju tačnost za dominantne grupe, dok za nedovoljno zastupljene grupe postižu lošije performanse. Zato je u kontekstu pravičnosti od ključne važnosti razmotriti osobine i pristrasnosti svakog skupa podataka. U nastavku je dat pregled najznačajnijih dostupnih datasetova za procenu starosti lica, njihovih osobina, prednosti i ograničenja.

IMDB-WIKI predstavlja jedan od najvećih javno dostupnih skupova slika lica sa informacijama o starosti i polu, prikupljenih sa IMDb i Wikipedia profila. Sadrži ukupno 523,051 sliku (oko 460 hiljada sa IMDb i 62 hiljade sa Wikipedie). Starost je određena kao hronološka (biološka) na osnovu datuma rođenja glumaca i vremenske oznake fotografije. Ipak zbog automatizovanog prikupljanja, oznake starosti u ovom datasetu sadrže šum, samim tim autori navode da ne mogu garantovati tačnost dobijenih godina. Ovaj šum može negativno uticati na performanse modela i njihovu stabilnost po

grupama. Uprkos tome, obim i raznovrsnost slika “iz divljine” čine IMDB-WIKi vrednim resursom, pre svega za pretreniranje dubokih modela ili kao dodatni izvor varijabilnosti prilikom obuke modela za procenu starosti. Primer par uzoraka iz ovog dataseta može se videti na Slici 1, kao i distribucija godina na Slici 2 [23].



Slika 1. Primeri slika iz IMDB-WIKI skupa podataka



Slika 2. Prikaz distribucije godina uzoraka iz IMDB-WIKI skupa podataka

FG-NET (Face and Gesture Recognition Network Aging Database) je skup manjih razmera, ali metodološki značajan jer obuhvata više fotografija istih osoba kroz različite faze života. Sadrži ukupno 1,002 slike od 82 subjekta, sa starosnim rasponom od novorođenčadi do 69 godina [24]. Oznake predstavljaju stvarne godine starosti u trenutku snimanja, što čini FG-NET pogodnim za proučavanje uzdužnih obrazaca starenja.

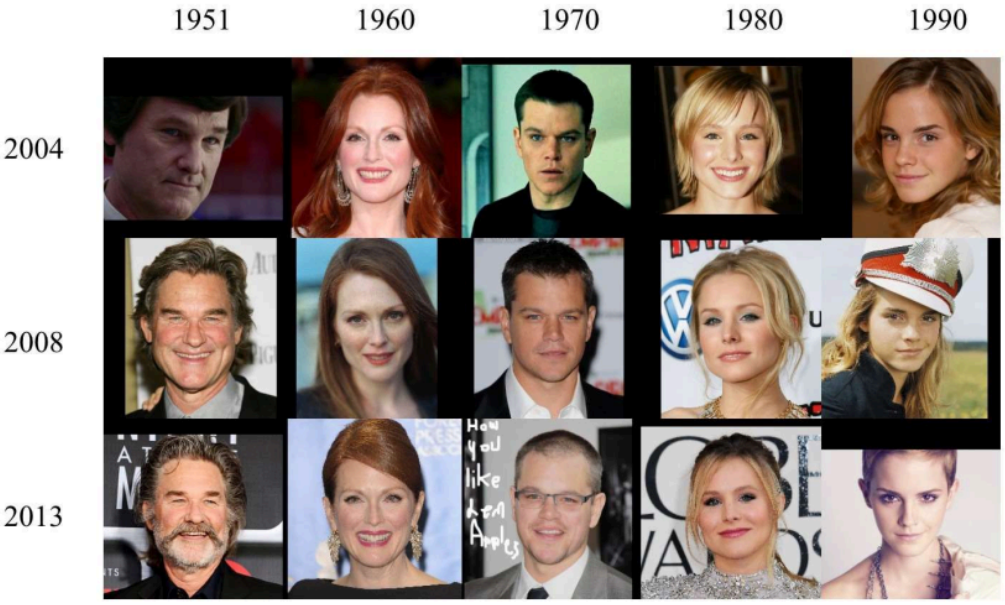
Međutim, njegova skromna veličina i ograničena demografska raznovrsnost umanjuju njegovu vrednost za treniranje savremenih dubokih modela i za fer evaluaciju. U praksi, FG-NET se češće koristi za proveru koncepta ili evaluaciju specifičnih metoda, nego za obuku opštih modela. Primer nekoliko uzoraka iz ovog skupa podataka može se videti na Slici 3.



Slika 3. Primer slika iz FG-NET skupa podataka

CACD (Cross-Age Celebrity Dataset) obuhvata veliki broj slika poznatih ličnosti raspoređenih kroz različite starosne dobi. Skup sadrži oko 163,446 slika 2,000 različitih slavnih osoba prikupljenih sa interneta. Fotografije su pretežno “iz divljine”, sa varijacijama u ekspresiji i uslovima, što daje bogatstvo varijanti. Međutim, pristrasnost skupa je izražena, subjekti su mahom poznate ličnosti uglavnom iz zapadnih zemalja, te populacija nije reprezentativna za širu javnost. Starost nije direktno zabeležena pri prikupljanju, umesto toga, procena godina izvedena je posredno korišćenjem godine fotografije i godine rođenja slavne osobe. Ovakav pristup može uvesti sistematsku grešku u oznakama i pristrasnost ka specifičnim starosnim obrascima poznatih. Autori CACD skupa naglašavaju da su starosne oznake grube procene i da skup pre svega služi za problem prepoznavanja lica kroz starosne promene, a ne za preciznu regresiju starosti. Stoga, iako značajan i raznovrstan, CACD se koristi sa oprezom u zadatku procene starosti, često kombinujući sa drugim podacima ili u scenarijima gde blaže greške u

oznakama nisu kritične. Nekoliko primeraka uzoraka iz ovog skupa podataka može se videti na Slici 4 [25].



Slika 4. Primer slika iz CACD skupa podataka

MORPH II je jedna od najčešćih referentnih baza u istraživanjima procene starosti i sadrži višestruke fotografije istih subjekata u kontrolisanim uslovima, u formatu takozvanih "mugshot" policijskih fotografija. Ovaj set podataka sadrži preko 55 hiljada slika od 13,000+ osoba, sa starosnim rasponom od 16 do 77 godina. Svaka slika ima pouzdano zabeleženu stvarnu starost, kao i oznake pola i etničke pripadnosti subjekta. Kvalitet i konzistentnost oznaka učinili su MORPH II čestim skupom za treniranje i benchmark poređenja modela. Ipak, demografska neuravnoteženost MORPH skupa je značajna, većina slika potiče od muškaraca afričkog porekla, oko 85% uzoraka su muški, a preko 77% pripada afroameričkoj populaciji. Ovakva struktura reflektuje izvorište podataka (policijske evidencije iz određenih područja) i dovodi do pristrasnosti koja ograničava prenosivost modela obučenih na MORPH-u. Takođe, uniformni "zatvorski" stil fotografija sa frontalnim prikazom, jednobojnom pozadinom i kontrolisanim osvetljenjem odudara od varijabilnosti realnih fotografija, što otežava fer poređenje sa rezultatima na skupovima iz realnog sveta. U kontekstu pravičnosti, modeli istrenirani isključivo na MORPH-u mogu pokazati pogoršane performanse na grupama koje su u

ovom datasetu malo zastupljene, pa se često pribegava kombinovanju ovo skupa sa drugim, raznovrsnijim izvorima podataka. Primer uzoraka može se videti na Slici 5 [26].

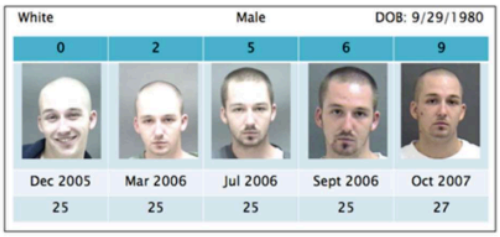


Fig. 1. Image Progression for White Male

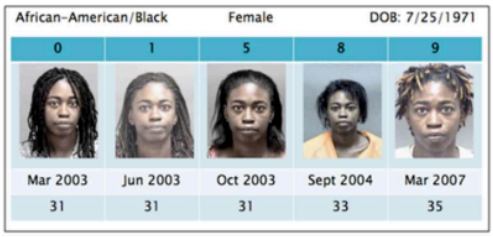


Fig. 2. Image Progression for African-American Female

Slika 5. Primer slika iz MORPH II skupa podataka

Adience skup je projektovan za klasifikaciju lica po starosnim opsezima umesto po tačnoj dobi, uz naglasak na raznovrsne i realistične uslove slikanja. Sadrži oko 26,850 fotografija prikupljenih od 2,284 osoba, podeljenih u 8 starosnih kategorija. Slike su prikupljene iz javno dostupnih Flickr albuma snimljenih mobilnim telefonima, bez filtriranja ili poziranja, čime Adience predstavlja izazovan skup sa velikim varijacijama u uglu slikanja, izrazu lica, osvetljenju i kvalitetu. Ovakav dizajn ima prednosti za praktične primene, često je dovoljno odrediti starosnu kategoriju, na primer za personalizovane oglase po starosnim grupama, umesto precizne godine. Kako Adience skup koristi starosne kategorije umesto stvarnih godina, evaluacija modela se vrši preko klasifikacione tačnosti, što onemogućava primenu standardnih regresionih metrika poput MAE i time otežava direktno poređenje sa modelima treniranim za preciznu procenu stvarne starosti. Ipak, Adience je dragocen za evaluaciju modela u uslovima realističnog šuma, jer svaki model koji postiže dobre rezultate na ovom skupu demonstrira otpornost na varijacije koje se javljaju u nekontrolisanim okruženjima. Primer nekoliko uzoraka iz ovog seta podataka može se videti na Slici 6 [27].



Slika 6. Primer slika iz Adience skupa podataka

AgeDB je relativno manji skup, ali sa ručno verifikovanim godinama života za svaki uzorak. Sastoji se od 16,488 slika različitih javnih ličnosti, ukupno 568 identiteta, pri čemu svaka slika ima anotaciju stvarne starosti, pola i identiteta osobe. Ključna prednost AgeDB skupa jeste da je prikupljen "ručno" i pažljivo proveren, čime su oznake tačne do nivoa godine i praktično bez šuma. Za razliku od polu-automatizovanih kolekcija poput IMDB-WIKI ili CACD gde se potkradaju greške u oznakama, AgeDB nudi pouzdanost za evaluaciju modela. Autori su pokazali da ova karakteristika čini AgeDB pogodnim za merenje performansi algoritama za procenu starosti, ali i za zadatke verifikacije identiteta kroz starosne razlike, jer pruža čist skup podataka nezagađen greškama u oznakama. Zbog ograničene veličine, AgeDB se ređe koristi za treniranje modela, ali je čist izbor za validaciju i testiranje. U savremenim radovima, modeli istrenirani na velikim skupovima često se ocenjuju na AgeDB da bi se izmerila njihova apsolutna tačnost na jednom od

najpouzdanijih skupova sa stvarnim godinama. Primer nekoliko uzoraka iz ovog seta podataka može se videti na Slici 7 [28].



Slika 7. Primer slika iz AgeDB skupa podataka

MegaAge-Asian je skup dizajniran da ciljano pokrije istočnoazijsku populaciju sa oznakama stvarnih godina. Obuhvata 40,000 slika lica isključivo azijskih osoba, sa starosnim rasponom od 0 do 70 godina. Ovakav regionalni fokus omogućava preciznije istraživanje regionalnih obrazaca starenja. S druge strane, usko demografsko pokriće znači da se zaključci dobijeni na ovom datasetu ne mogu lako generalizovati na globalnu populaciju. Modeli trenirani ili testirani isključivo na ovom skupu mogu pokazati pristrasnost kada se primene na lica drugih etničkih grupa. Iz tog razloga MegaAge-Asian naglašava potrebu pažljivog tretmana zbog demografske specifičnosti, idealno je koristiti ga u kombinaciji sa drugim skupovima kako bi se obezbedila fer evaluacija koja uzima u obzir različite populacije. Primer nekoliko uzoraka može se videti na Slici 8 [29].



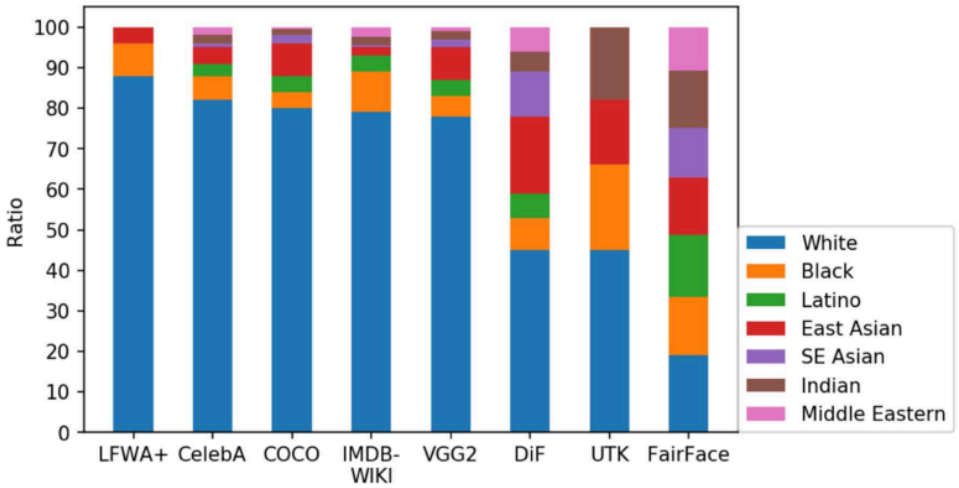
Slika 8. Primer slika iz MegaAge-Asian skupa podataka

FairFace je noviji skup prvenstveno kreiran radi promovisanja balansa po rasnim i etničkim grupama u treniranju i evaluaciji modela. Sadrži 108,501 sliku lica prikupljenih iz velike Flickr kolekcije (YFCC-100M), pažljivo izbalansiranih po sedam definisanih rasnih kategorija: bela, crna, indijska, istočnoazijska, jugoistočnoazijska, bliskoistočna i latino populacija. Uz to su za svaku sliku dostupne i oznake pola i starosne kategorije (starost nije data kao precizan broj godina, već kroz intervale, npr. 0-2, 3-9, 10-19, itd.). FairFace je koncipiran kao standardni resurs za merenje i ublažavanje pristrasnosti u algoritmima za analizu lica. Za razliku od ranijih skupova koji su bili pretežno sastavljeni od slika belaca, FairFace eksplicitno adresira tu neravnotežu uključujući podjednak broj uzoraka različitih rasa, čime omogućava obučavanje modela koji imaju konzistentniju tačnost preko grupa. U istraživanjima pravičnosti, FairFace se često koristi kao dopunski evaluacioni skup. Model istreniran na nekom standardnom skupu može se testirati na FairFace kako bi se izmerilo da li postoji značajna razlika u greškama predikcije između lica različitih etničkih kategorija. Takođe, prilikom treniranja, uključivanje FairFace slika može pomoći u smanjenju demografske pristrasnosti modela. Na Slici 9 se mogu videti primeri uzoraka iz ovog skupa podataka. Takođe se na Slici 10 može videti tabela poređenja demografske zastupljenosti FairFace skupa podataka sa različitim drugim

setovima, pokazujući svoju unapređenu uravnoteženost u odnosu na sve druge skupove podataka [9].



Slika 9. Primer slika iz FairFace skupa podataka



Slika 10. Poređenje demografske uravnoteženosti FairFace skupa sa ostalim skupovima podataka

Ukupno posmatrano, dostupni skupovi za procenu starosti predstavljaju različite kompromise između veličine, kvaliteta oznaka, realističnosti uslova snimanja i demografske pokrivenosti. Veći skupovi poput IMDB-WIKI i CACD donose obim i

varijabilnost neophodne za duboko učenje, ali često po cenu značajnijeg šuma u podacima i pristrasnosti ka dominantnim grupama. Manji skupovi poput FG-NET i AgeDB nude pouzdanije i preciznije oznake, ali ograničenu statističku snagu i slabiju zastupljenost različitih populacija. Za zadatke pravične procene starosti, jasno je da nijedan pojedinačni skup nije dovoljan sam po sebi. Kombinovanjem komplementarnih skupova na primer korišćenjem velikih baza za pretreniranje, a manjih provereno tačnih baza za validaciju i fino podešavanje može se iskoristiti najbolje od oba sveta. Takav pristup omogućava da se postigne visoka ukupna tačnost modela, a istovremeno prati i kontroliše distribucija grešaka po različitim starosnim, polnim i rasnim podskupovima, što je od suštinskog značaja za fer i pouzdanu primenu sistema za procenu starosti.

2.1.1. Uloga datasetova u rešavanju zadatka procene starosti

Efikasnost i tačnost modela za procenu starosti lica duboko su povezane sa kvalitetom, kvantitetom i reprezentativnošću podataka koji se koriste u fazama obučavanja i testiranja. Duboke konvolucione neuronske mreže (CNN) poseduju milione parametara i zahtevaju veliki broj raznovrsnih i pouzdano obeleženih uzoraka kako bi izbegle prekomerno prilagođavanje (overfitting) i obezbedile dobru generalizaciju na nove primere.

U domenu procene starosti, treniranje modela zahteva ne samo veliki broj slika već i odgovarajuću pokrivenost svih demografskih i starosnih grupa. Ako određene starosne kategorije, etničke grupe ili polovi nisu adekvatno zastupljeni, model neće uspeti da nauči stabilne obrasce za te slučajeve, što direktno utiče na pravičnost i robusnost sistema.

Istorijski gledano, rana istraživanja su se oslanjala na ograničene baze podataka kao što je FG-NET, koja sadrži svega 1.002 slike. Takva ograničenja su nužno uslovljavala izbor jednostavnijih modela i skromnije rezultate. Kasniji razvoj većih baza kao što je MORPH omogućio je povećanje obima uzoraka za više od 50 puta, čime su otvorene mogućnosti za primenu složenijih modela dubokog učenja. Međutim, i MORPH baza je imala izraženu demografsku pristrasnost, jer je oko 85% uzoraka predstavljalo muške osobe iz jedne dominantne rasne grupe, što je ograničavalo prenosivost modela na šire populacije.

Napredak u dubokom učenju u oblasti procene starosti nije bio moguć bez paralelnog napretka u razvoju datasetova. Veliki, precizno obeleženi i raznovrsni skupovi podataka omogućili su obučavanje modela koji prevazilaze tradicionalne pristupe zasnovane na ručno izdvojenim karakteristikama. Time se potvrđuje da dataset nije pomoćno sredstvo, već ključna komponenta u građenju uspešnih modela, jer upravo on definiše domen znanja i granice opšte primenljivosti svakog predikcionog sistema. Stoga, selekcija i priprema kvalitetnih i uravnoteženih skupova podataka predstavlja ne samo tehnički, već i etički preduslov za dalji napredak oblasti.

2.1.2. Kriterijumi za izbor datasetova

Prilikom izbora skupova podataka za probleme procene starosti lica, potrebno je razmotriti više kriterijuma kako bi se obezbedilo da skup odgovara zadatku i omogućava treniranje pouzdanog i nepristrasnog modela. Ključni kriterijumi uključuju:

Obim i reprezentativnost podataka: Skup treba da sadrži dovoljno veliki broj uzoraka koji pokrivaju širok spektar različitih osoba. Veći broj slika smanjuje rizik od prenaučenosti modela i poboljšava statističku pouzdanost prilikom treniranja. Pored pukog broja, bitno je da uzorci budu reprezentativni za varijabilnost koja postoji u realnom svetu, uključujući različite etničke grupe, starosne opsege, polove, stilove fotografisanja, rezolucije slike, itd. Raznovrsniji skup smanjuje pristrasnost modela ka dominantnim karakteristikama u podacima. Na primer, skup podataka u kome su većina fotografija visokog kvaliteta neće omogućiti modelu da nauči kako da se nosi sa niskokvalitetnim slikama. Slično tome, skup poput već pomenutog MORPH, iako vrlo obiman, pokazuje slabosti jer ne zadrži slike dece niti starijih iznad 70 godina. Model treniran na takvim podacima verovatno neće dobro procenjivati starost upravo za te kategorije. Idealno, izabrani dataset pokriva ceo raspon starosti od detinjstva do starosti, i obuhvata osobe različitog porekla kako bi se izbegla demografska pristrasnost.

Broj klasa i način anotacije starosti: U zavisnosti od pristupa proceni starosti (da li se problem formuliše kao klasifikacija u starosne kategorije ili kao regresija kontinuirane vrednosti), važno je kako su starosne oznake predstavljene. Neki skupovi obezbeđuju starost kao diskretnu klasu, dok drugi daju kontinuiranu oznaku. Fini granularitet omogućava precizniji model, ali zahteva i više podataka. Ukoliko se koriste kategorije na

primer “dete”, “odrastao”, “stariji”, skup treba da ima dovoljan broj uzoraka u svakoj kategoriji kako bi model naučio relevantne osobine. Broj klasa mora biti usklađen sa ciljevima istraživanja, previše klasa može otežati učenje ako nema dovoljno uzoraka po klasi, dok premalo klasa može biti suviše grubo i sakriti suptilne razlike u izgledu između bliskih starosnih dobi.

Tačnost i tip oznaka: Od presudne je važnosti da oznake starosti budu pouzdane. Hronološka ili stvarna starost je idealna oznaka za probleme procene starosti, to je stvarna starost osobe u trenutku snimanja fotografije. Međutim, nije uvek moguće doći do te informacije, naročito za slike prikupljene sa interneta. Neki skupovi zato koriste procenu starosti od strane ljudskih posmatrača, gde se više ljudi izjašnjava koliko osoba na slici izgleda staro, pa se računa prosečna ili medijalna procena. Ovakve oznake reflektuju subjektivnu percepciju starosti i mogu odstupati od stvarnih godina, ali su korisne za treniranje modela koji imitiraju ljudske procene. Pri odabiru skupa, treba razmotriti da li su oznake starosti dobijene direktno (npr. iz službenih dokumenata, baze podataka ili lično od subjekta) ili posredno (algoritamskom ili ljudskom procenom). Neprecizne oznake unose šum u podatke i mogu ograničiti maksimalnu tačnost modela. Stoga se prednost daje skupovima sa verifikovanim starosnim oznakama.

Dodatne anotacije i meta-podaci: Za sveobuhvatnu analizu mogu biti korisne i druge oznake u skupu podataka, pre svega pol i etnička pripadnost ili rasa. Ove oznake omogućavaju ispitivanje da li model jednako dobro radi za različite demografske grupe, odnosno detekciju eventualnih pristrasnosti. Takođe, one omogućuju izgradnju multitask modela (npr. istovremena procena starosti i pola) ili primenu tehnika prilagođavanja domenu radi poboljšanja tačnosti za određene podskupove. Pri izboru skupa podataka, poželjno je da poseduje takve informacije ili da istraživač može sam da ih prikupi. Naravno, dodatne anotacije imaju smisla samo ako su dovoljno tačne i ako postoji razumna zastupljenost različitih kategorija.

Ukratko, idealan dataset za procenu starosti trebalo bi da bude velik, raznovrstan, balansiran i pouzdano anotiran. Kompromisi su često neizbežni, neki skupovi su veoma veliki ali sadrže šum u oznakama ili su pristrani ka određenoj populaciji, dok drugi imaju precizne oznake ali su malog obima. Zadatak istraživača je da od dostupnih baza podataka odabere one koje najbolje ispunjavaju navedene kriterijume u skladu sa

ciljevima istraživanja, ili da kombinovanjem i dopunom skupova te nedostatke umanji. U nastavku poglavlja opisani su konkretni skupovi podataka korišćeni u ovom radu, uz osvrt na ove kriterijume.

2.2. Detaljan opis korišćenih skupova podataka

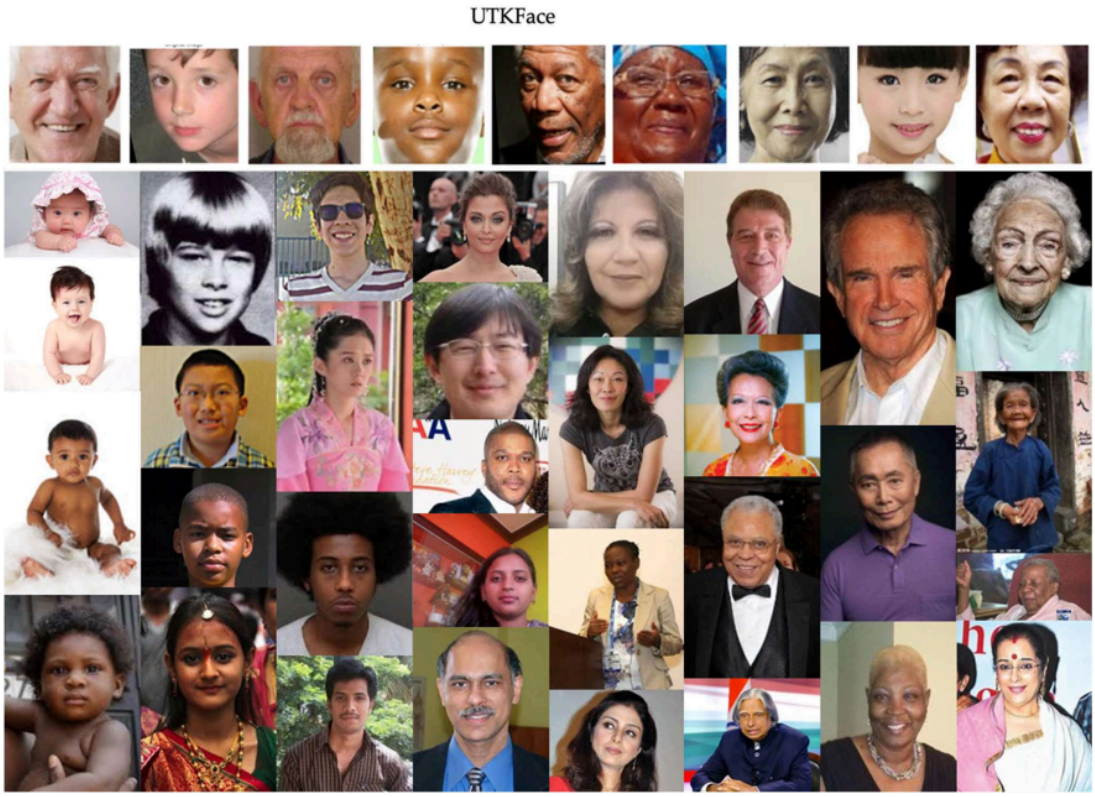
U ovom radu korišćena su dva javno dostupna skupa podataka za procenu starosti lica: UTKFace [30] i APPA-REAL [31]. Obe baze predstavljaju skupove fotografija lica prikupljenih u nekontrolisanim uslovima, ali se razlikuju po načinu anotacije starosti i demografskim karakteristikama. U narednim pod poglavljima dat je detaljan opis strukture ovih skupova, načina na koji su prikupljeni, kao i njihovih prednosti i ograničenja.

Izbor UTKFace i APPA-REAL skupa proizilazi iz kriterijuma postavljenih u prethodnoj sekciji (2.1.2), uključujući raznovrsnost podataka, pouzdanost oznaka, dostupnost demografskih atributa, prisustvo varijabilnih uslova snimanja i pogodnost za regresione evalucione metrike. Ova dva skupa zajedno čine metodološki balansiran izbor, koji omogućava istovremeno treniranje i evaluaciju modela sa ciljem ispitivanja preciznosti i pravičnosti.

U poređenju sa drugim dostupnim skupovima, kombinacija UTKFace i APPA-REAL predstavlja konzistentan kompromis između obima, kvaliteta i raznovrsnosti. Na primer, skupovi poput IMDB-WIKI i CACD, iako brojčano dominantni, trpe od značajnog šuma u starosnim oznakama i selekcione pristrasnosti usled pretežne zastupljenosti slavnih osoba sa Zapada. MORPH II, iako tačno anotiran, karakteriše izrazita demografska neuravnoteženost i stilizovana estetika koja umanjuje njegovu prenosivost na realne aplikacije. Adience i FairFace ne nude kontinualne oznake starosti, već samo kategorije, što ograničava njihovu primenu u preciznim regresionim zadacima. AgeDB, iako veoma pouzdan, nije dovoljne veličine za ozbiljno treniranje, dok je MegaAge-Asian regionalno ograničen i fokusiran isključivo na istočnoazijsku populaciju, čime je njegova generalizacija ograničena. Iz tog razloga kombinacija UTKFace i APPA-REAL najbolje odgovara istraživačkim pitanjima ove disertacije.

2.2.1. UTKFace dataset

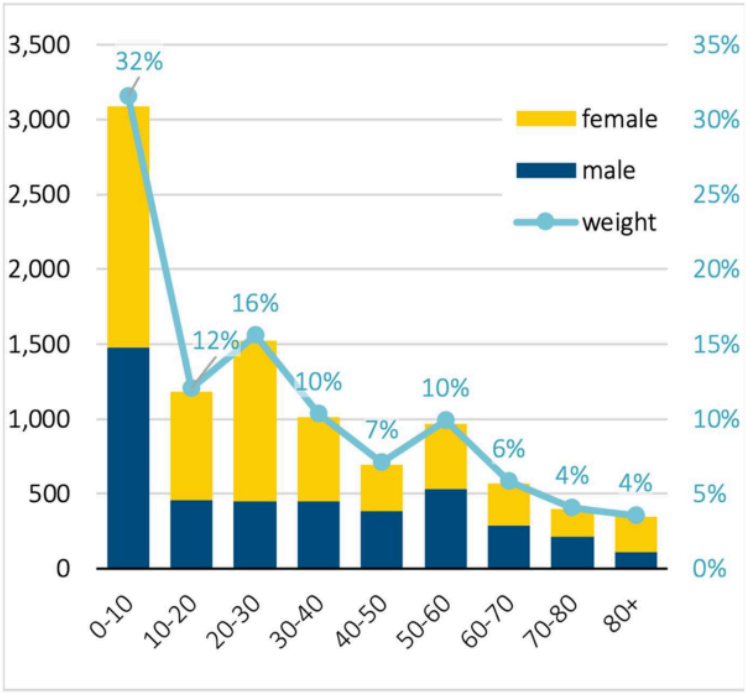
UTKFace je jedan od najpoznatijih skupova podataka za probleme procene starosti, prepoznavanje pola i analize etničke pripadnosti na osnovu slika lica. To je velik skup slika lica sa širokim rasponom starosti. Pokriva uzraste od 0 do 116 godina i ukupno sadrži 23,705 fotografija jedinstvenih osoba. Svaki uzorak u sebi uključuje metapodatke. Tako, za svaku sliku dostupna je oznaka starosti (integer od 0 do 116), oznaka pola (0 – muški, 1 - ženski) i oznaka rase/etničke grupe (0 – bela, 1 - crna, 2 – istočnoazijska, 3 – indijska, 4 – ostali, što obuhvata latino, bliskoistočne itd.). Fotografije u UTKFace setu snimljene su u različitim uslovima i prikazuju lica u raznovrsnim pozama, izražajima, osvetljenjima i rezolucijama, uključujući i delimične okretaje glave, naočare i ostalo. Na Slici 11 ispod dati su primeri iz ovog skupa koji ilustruju raznovrsnost uzoraka [32].



Slika 11. Primeri slika iz UTKFace skupa podataka

Primeri obuhvataju različite starosne dobi, polove i etničke grupe. Ovaj skup slika „iz prirodnog okruženja“ sadrži kako fotografije dece i adolescenata, tako i odraslih i starijih osoba, uz značajne varijacije u uslovima snimanja.

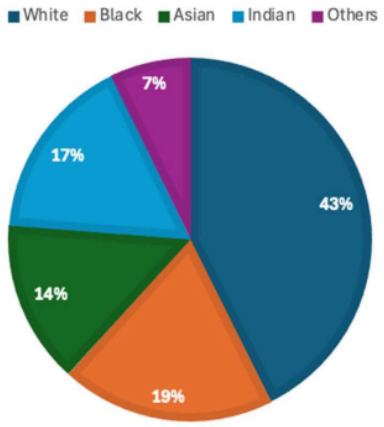
Kao što se iz strukture može videti, UTKFace eksplicitno pokriva tri važne dimenzije informacija o svakom licu, starost, pol i rasu, što ga čini veoma korisnim za treniranje modela koji moraju da generalizuju na više demografskih grupa i istovremeno omogućava istraživanje tih faktora na performanse. Broj slika po starosnim godinama relativno je ujednačen za veći deo opsega, uz blag trend opadanja broja uzoraka za najstarije dobi. Ipak, primećeno je da postoji blaga neuravnoteženost u korist najmlađih uzrasta – oko 32% svih slika prikazuje decu uzrasta do 10 godina. Detaljan pregled distribucije godina može se videti na Slici 12.



Slika 12. Distribucija uzoraka UTKFace skupa podataka po grupi godina

Ovaj disbalans znači da su lica dece i beba nešto prekomerno zastupljena u odnosu na ostale grupe, što može uticati na treniranje modela da bude za nijansu precizniji za mlađa lica. Što se tiče rasne strukture, UTKFace obezbeđuje znatno širu zastupljenost ne-belih populacija nego raniji skupovi, oko 57% lica su osobe koje nisu

bele puti (crna populacija oko 19%, Azijati 15%, Indijci 17%, ostali 7%). Vizuelna reprezentacija ove podele može se videti na Slici 13.



Slika 13. Vizuelna reprezentacija podele uzoraka prema etničkoj pripadnosti unutar UTKFace skupa podataka

Iako ni ovaj skup nije savršeno izbalansiran, UTKFace se ističe po raznovrsnosti u poređenju sa mnogim alternativama. U njemu su značajno zastupljene azijske i indijske etničke grupe, koje su u nekim drugim bazama retke. Polna distribucija je blizu ravnomerne, približno 55% uzoraka su lica žena, a 45% muškaraca, što znači da model treniran na UTKFace nema izrazitu pristrasnosnt ka jednom polu po pitanju brojnosti podataka. Detaljna kompozicija ovog skupa može se videti u Tabeli 2.

Tabela 2. Pregled UTKFace skupa podataka

Ime	Ukupan broj uzoraka	Pol		Grupa					Opseg godina
		Muški	Ženski	Bela	Crna	Azijska	Indijska	Drugi	
UTKFace	23,705	12,391	11,314	10,078	4526	3434	3975	1692	1–116

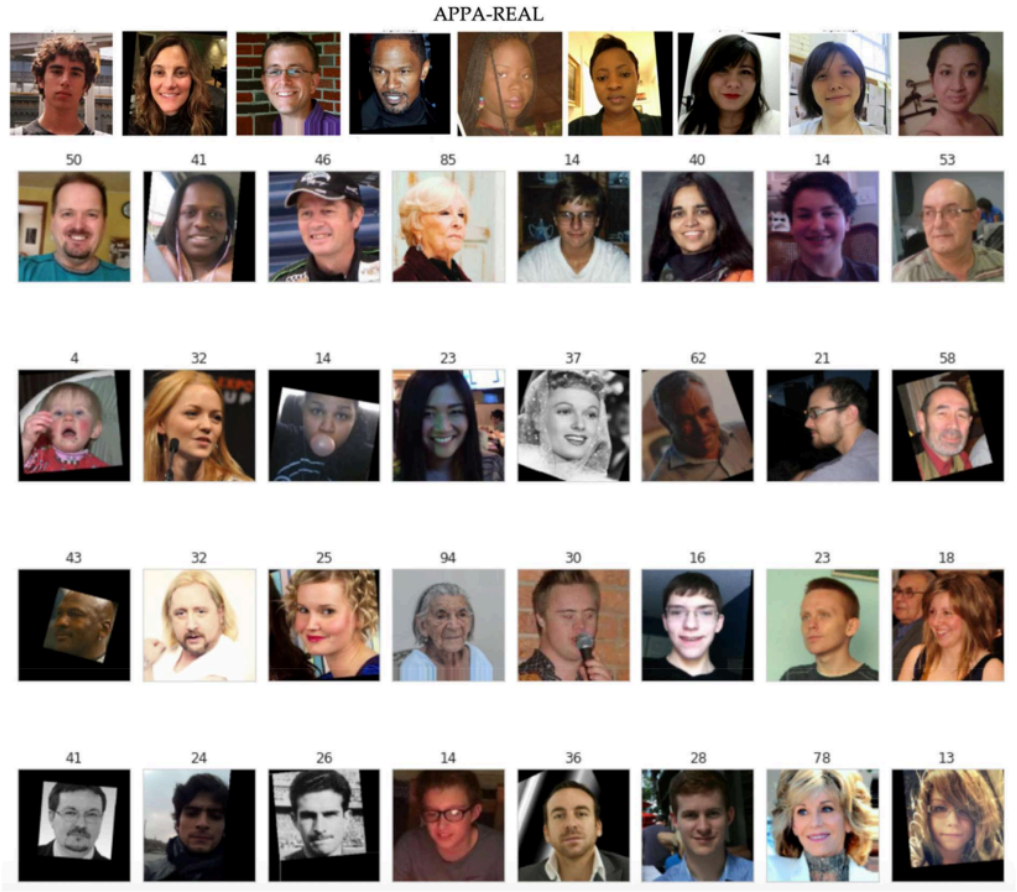
UTKFace je sastavljen metodologijom “web-scraping”, slike lica su prikupljene sa interneta uz pomoć postojeće aplikacije za prepoznavanje starosti. Autori skupa su iskoristili algoritam za procenu starosti DEX, treniran na IMDB-WIKI bazi kako bi inicijalno odredili starost prikazanih osoba. Potom je svaka takva automatska oznaka ručno proverena i po potrebi ispravljena od strane ljudskog anotatora kako bi se povećala tačnost. Ovakav pristup omogućio je relativno brzo anotiranje desetina hiljada slika, ali

uvodi i određena ograničenja. Oznake starosti u UTKFace nisu garantovano potpuno precizne kao što bi bile da su dobijene direktno od subjekata ili njihove dokumentacije. U suštini, one predstavljaju procene (kombinaciju algoritamske i ljudske), moguće je da za pojedine slike starost nije tačno pogođena, naročito u slučajevima gde je fotografija neuobičajena ili osoba izgleda znatno mlađe/starije od svojih godina. Takođe, kategorije etničke pripadnosti su pojednostavljene na pet grupa, što ne odražava pun spektar demografske raznovrsnosti, posebno kategorija "ostali" koja objedinjuje više različitih manjinskih grupa koje se međusobno mogu razlikovati po izgledu. Još jedno praktično ograničenje UTKFace skupa jeste da su određene starosne grupe malobrojne uprkos ukupnom velikom broju slika. Ekstremno stari (npr. preko 90 godina) prisutni tu tek u tragovima, što može ograničiti pouzdanost modela u tom opsegu godina. S druge strane, velika prednost UTKFace je upravo njegov realističan karakter, slike nisu snimane u kontrolisanom okruženju već potiču iz svakodnevnih situacija, što znači da model treniran na njima uči da se nosi sa različitim osvetljenjima, pozadinama, uglovima snimanja i ostalim varijacijama. Tako stečena robusnost je korisna za primenu modela u praksi. UTKFace je iz tog razloga postao standardni izbor za treniranje i testiranje algoritama za procenu starosti i srodne zadatke.

2.2.2. APPA-REAL dataset

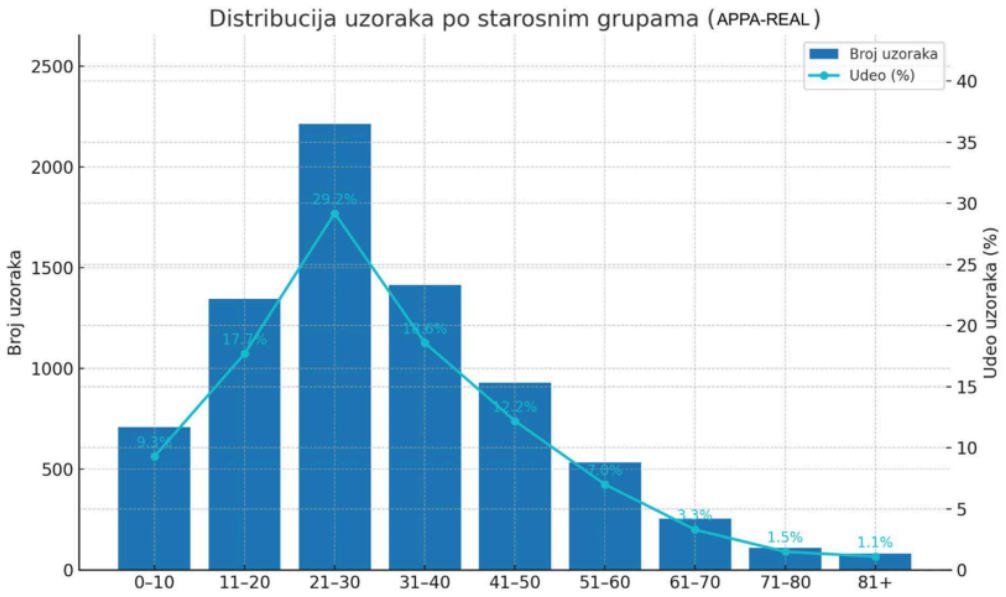
APPA-REAL (skraćeno od Apparent and Real Age) je posebna baza podataka kreirana sa ciljem proučavanja razlike između hronološke i percipirane starosti posmatrane na fotografiji. Ovaj skup sadrži ukupno 7,591 fotografiju lica različitih osoba, pri čemu je za svaku fotografiju prikupljena dvojaka anotacija starosti: (1) stvarna starost subjekta u trenutku fotografisanja, i (2) procena starosti koju su dali ljudski posmatrači. Stvarna starost izražena je u godinama i predstavlja referentnu vrednost. Dobijena je direktno od samih učesnika ili iz njihovih ličnih podataka tokom prikupljanja. Nasuprot tome, procenjena ili percipirana starost određena je putem glasanja velikog broja ljudi. Za svaku sliku sakupljeno je u proseku oko 38 nezavisnih procena posmatrača o tome koliko osoba na slici izgleda staro. Ti glasovi (kojih je ukupno prikupljeno skoro 300,000 za ceo skup) agregirani su računanjem proseka, čime se dobija fina procena koja simulira

kolektivnu presudu ljudi o starosti prikazanog lica. Nekoliko uzoraka može se videti na slici 14 [31].



Slika 14. Primeri slika iz APPA-REAL skupa podataka

Starosni raspon obuhvaćen bazom je od 1 do 100 godina. Time APPA-REAL, slično UTKFace-u, pokriva čitav životni vek, s tim što je srednji opseg (odrasle osobe) najbrojniji, naročito grupa od 21 do 30 godina koja iznosi 29.2% svih uzoraka. Takođe slično kao kod UTKFace skupa stariji opsezi godina imaju dosta manju zastupljenost. Detaljan pregled distribucije godina može se videti na Slici 15.



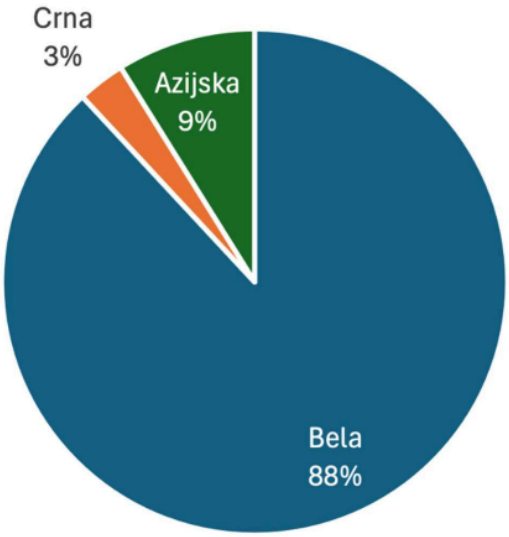
Slika 15. Distribucija uzoraka APPA-REAL skupa podataka po grupi godina

Ukupan broj slika je nešto manji u poređenju s UTKFace, ali kvalitet anotacija je izuzetno visoka. APPA-REAL je jedinstven po tome što pruža pouzdanu informaciju o stvarnoj starosti uz paralelni uvid u ljudsku percepciju starosti. Ovakva dvostruka anotacija omogućava istraživanje koliko se i zbog čega razlikuje procena automatskog modela od ljudske, kao i ispitivanje koji faktori utiču na to da osoba izgleda mlađe ili starije nego što jeste. Veličina skupa dovoljna je za treniranje modernih modela uz adekvatne tehnike, a istovremeno nije preterano opterećujuća za manuelno prikupljanje podataka od ljudi, što je u ovom slučaju bilo ključno zbog prikupljanja glasova.

APPA-REAL je nastao kao rezultat ciljane kampanje prikupljanja fotografija i ocena, koju su sproveli istraživači u sklopu izazova ChaLearn 2016. Autori su razvili posebnu web aplikaciju za prikupljanje, koristeći integraciju sa Facebook API-jem radi prikupljanja fotografija različitih korisnika, i angažovali su platformu Amazon Mechanical Turk za prikupljanje ocena starosti od velikog broja posmatrača. Ovakav pristup omogućio je da se dobije raznovrstan skup lica u realnim uslovima [31].

Glavno ograničenje APPA-REAL skupa je njegova relativno skromna veličina u poređenju sa mega-skupovima kao što su IMDB-WIKI ili čak UTKFace. Sa oko 6,000 slika u trening podskupu, modelima treniranim isključivo na APPA-REAL može nedostajati

dovoljno uzoraka da postignu vrhunske performanse, naročito dubokim modelima sa mnogo parametara. Stoga se APPA-REAL često koristi u kombinaciji sa drugim skupovima. Drugo važno ograničenje jeste demografska neuravnoteženost, iako je prikupljanje vršeno putem otvorene aplikacije, u bazi preovlađuju osobe bele rase dok su ostale rase slabije zastupljene. Vizuelna reprezentacija ove podele može se videti na Slici 16.



Slika 16. Vizuelna reprezentacija podele uzoraka prema etničkoj pripadnosti unutar APPA-REAL skupa podataka

Ova neuravnoteženost verovatno odražava krug dostupnih učesnika u prikupljanju i predstavlja potencijalni izvor pristrasnosti u modelima treniranim na APPA-REAL; oni mogu biti vrlo precizni za većinsku grupu, ali sa većom greškom za lica drugačijeg etničkog porekla. Što se tiče pola, skup je otprilike izbalansiran sa 50% muškaraca i 50% žena, pa po polu nema izrazite pristrasnosti. Detaljna kompozicija ovog dataseta može se videti u tabeli 3.

Ime	Ukupan broj uzoraka	Pol		Grupa					Opseg godina
		Muški	Ženski	Bela	Crna	Azijska	Indijska	Drugi	
APPA-REAL	7591	3818	3773	6686	231	674			1–100

Tabela 3. Pregled APPA-REAL skupa podataka

2.3. Komparativna analiza korišćenih datasetova

Nakon pojedinačnog opisa karakteristika skupova UTKFace i APPA-REAL, u ovom delu ćemo uporediti ove baze podataka iz perspektive ključnih aspekata za zadatak procene starosti. Cilj je da se istaknu komplementarnosti i izazovi oba skupa, što će poslužiti kao osnova za odluke o načinu pripreme podataka i dizajnu modela u narednim poglavljima.

Oba posmatrana dataseta sadrže informacije o polu i etničkoj pripadnosti subjekata, što omogućava analizu demografske strukture. Već je u prethodnim sekcijama napomenuto da UTKFace i APPA-REAL imaju značajno različitu etničku raspodelu. UTKFace je konstruisan sa ciljem da uključi približno uporedive veličine podskupova različitih etničkih grupa. Dataset koristi pet etničkih kategorija, koje obuhvataju lica bele rase, azijskog porekla, afričkog porekla, porekla sa indijskog potkontinenta, kao i kategoriju „ostalih”. Ova poslednja uključuje Hispanoamerikance, Latinoamerikance i populaciju sa Bliskog istoka. Iako distribucija nije u potpunosti uravnotežena, UTKFace sadrži nekoliko hiljada uzoraka u svim kategorijama, pri čemu najveći pojedinačni udeo imaju slike belaca. Ovakva raznolikost čini UTKFace jednim od demografski najuravnoteženijih javno dostupnih skupova u oblasti procene starosti lica.

Nasuprot tome, APPA-REAL pokazuje izrazitu dominaciju jedne grupe – preko 88% svih lica pripada beloj rasi. Ostale kategorije su znatno manje zastupljene: svega nekoliko procenata uzoraka čine lica tamne puti, dok je oko 9% azijskog porekla. Dataset eksplicitno diferencira samo tri etničke kategorije: belu, crnu i azijsku, bez dodatnih podkategorija koje bi mogle obuhvatiti druge manjinske grupe.

UTKFace, zahvaljujući višoj raznolikosti i eksplicitnijoj klasifikaciji etničkih kategorija, pruža bogatiju osnovu za analizu među-grupnih razlika. Njegova struktura omogućava pouzdaniju evaluaciju modela u kontekstu pravičnosti u poređenju sa APPA-

REAL skupom, koji je značajno neuravnoteženi u pogledu reprezentacije različitih demografskih grupa.

Sa aspekta polne strukture, oba skupa su približno uravnotežena i tu ne postoji dramatična razlika. UTKFace ima blagu premoć ženskih fotografija (~55%), dok APPA-REAL ima gotovo tačno 50% ženskih i 50% muških subjekata. Iz tog razloga, ni jedan ni drugi skup ne pate od izražene pristrasnosti po pitanju pola. Ipak, zanimljivo je razmotriti kako pol utiče na procenu starosti: poznato je da modeli (kao i ljudi) mogu različito grešiti u proceni starosti muškaraca u odnosu na žena, obično usled razlika u obrascima starenja na licu ili šminke kod žena. U tom smislu, UTKFace i APPA-REAL daju dobru osnovu za treniranje i testiranje modela bez straha da će jedan pol biti zapostavljen u podacima.

Naredna važna tačka poređenja između UTKFace i APPA-REAL skupa jeste priroda uslova u kojima su fotografije nastale, odnosno stepen kontrole nad slikama. Oba skupa spadaju u tzv. **"**in-the-wild**"** baze. Slike su prikupljene iz realnog okruženja, nevezano za laboratorijske uslove. To znači da postoji velika varijacija u pozama lica, izražajima, osvetljenju, pozadini, kvalitetu kamere, rezoluciji itd. Na primer, u UTKFace možemo naći i profesionalne portrete i amaterske fotografije niske rezolucije; lica su nekad direktno okrenuta ka kameri, a nekad blago pod uglom ili u profilu; svetlo može biti prirodno ili veštačko, različitog intenziteta. Slično, APPA-REAL je prikupljen preko društvenih mreža i u njemu su zastupljene fotografije iz svakodnevnog života, od selfi snimaka do porodičnih fotografija sa različitim uslovima. Nijedan od ovih skupova nije sniman u studiju, za razliku od nekih ranijih baza.

2.4. Predprocesovanje datasetova

Efikasna priprema podataka predstavlja ključni korak u izgradnji robusnog modela za procenu starosti. U ovom radu sprovedene su dve faze predprocesovanja: (1) tehničko formatiranje i standardizacija slika u skladu sa zahtevima dubokih modela, i (2) filtriranje i strukturisanje demografskih grupa radi analize pravičnosti i obezbeđivanja fer poređenja između različitih datasetova.

2.4.1. Skaliranje, formatiranje, čišćenje; standardizacija ulaza

Za sve slike iz oba skupa podataka primenjeni su identični tehnički koraci predprocesovanja kako bi se obezbedila konzistentnost ulaznih podataka. Slike su skalirane na dimenzije 224 × 224 × 3, što odgovara zahtevima ulaza za modele korišćene u ovom radu. Primenjena je njihova funkcija preprocess_input, koja centrira kanale boja na nulu u skladu sa statistikom ImageNet dataseta na kojem su pretrenirani i konvertuje slike iz RGB u BGR format. Ovim pristupom omogućeno je efikasno korišćenje prenosa učenja, čime se značajno ubrzava obuka i poboljšava generalizacija modela čak i pri ograničenom broju dostupnih uzoraka.

Zahvaljujući sposobnosti konvolucionih neuronskih mreža (CNN) da automatski uče relevantne reprezentacije iz sirovih vizuelnih podataka, eliminisana je potreba za ručnim inženjeringom karakteristika. CNN arhitekture poput VGG19 prepoznaju obrasce kao što su teksture, oblici i ivice na različitim nivoima apstrakcije, što ih čini posebno pogodnim za zadatke kao što je procena starosti sa slika lica. Korišćenjem prenosa učenja iskorišćeni su niži slojevi prethodno obučenih mreža za detekciju opštih karakteristika, dok su viši slojevi dodatno trenirani u skladu sa specifičnostima problema procene starosti.

2.4.2. Kombinovanje i filtriranje grupa za eksperiment

Uprkos zajedničkoj orijentaciji ka realističnim, “in-the-wild” uslovima, UTKFace i APPA-REAL skupovi podataka pokazuju značajne razlike u rasnoj zastupljenosti. Oba skupa su demografski dominantno orijentisana ka beloj populaciji, ali se međusobno razlikuju po širini i distribuciji drugih etničkih grupa.

U cilju usklađivanja analitičkog okvira i obezbeđivanja direktne uporedivosti između skupova, izvršena je rekategorizacija oznaka u UTKFace skupu podataka. Konkretno, crna i indijska grupa objedinjene su u jedinstvenu kategoriju “crna”, dok su primerci označeni kao “ostali” odbačeni iz dalje analize. Ova odluka je doneta kako bi se izbegla konfuzija izazvana mešovitim sastavom kategorije “ostali”, koja sadrži teško interpretabilan miks različitih etničkih identiteta koji ne mogu biti dosledno mapirani na strukturu APPA-REAL skupa. Nakon ovog koraka ukupan broj uzoraka u UTKFace skupu je 22,013, sa sledećim proporcijama: Bela grupa 45,8%, Crna grupa 38.6% i Azijska grupa

15.6%. Ovakav pristup omogućava standardizovanu rasnu segmentaciju između skupova i doprinosi fer evaluaciji modela u pogledu demografske pravičnosti.

2.5. Moguće primene datasetova i smernice za buduća istraživanja

Opisani skupovi podataka, UTKFace i APPA-REAL, predstavljaju dragocene resurse ne samo za potrebe ovog rada već i za širu istraživačku zajednicu, te otvaraju više pravaca primene i nadogradnje. Moguće primene ovih datasetova obuhvataju brojne zadatke u računarskoj viziji i u praksi za unapređenje modela za procenu starosti. Kao standardni trening skupovi, UTKFace i APPA-REAL mogu poslužiti za obučavanje sve preciznijih modela za procenu starosti lica. Takođe moguća primena može biti za multitask učenje, zbog toga što UTKFace sadrži i informacije o polu i etnicitetu, može se koristiti za istovremeno treniranje modela koji predviđa više atributa sa jedne slike lica, npr. starost, pol i rasu zajedno. Naravno, u skladu sa tim i kako su oba skupa javno dostupna i pokrivaju različite demografske grupe, oni su pogodni za evaluaciju pravičnosti algoritama za procenu starosti. Modeli istrenirani na ovim skupovima mogu naći praktične primene, a time i povratno uticati na razvoj datasetova. Na primer, u industriji zabave i oglašavanja već se koriste algoritmi za procenu starosti radi personalizacije sadržaja ili verifikacije godina. Korišćenjem UTKFace i APPA-REAL obezbeđuje se da takvi algoritmi budu obučeni na raznolikim primerima. Buduća istraživanja mogla bi proširiti datasetove uključivanjem novih demografskih grupa koje trenutno nisu dobro pokrivene kako bi globalne aplikacije bile pravedne za sve korisnike. Jedna specifična primena datasetova za procenu starosti jeste u medicini i biometrici, praćenje kako lice jedne osobe stari kroz vreme. Iako UTKFace i APPA-REAL nisu longitudinalni (svaka osoba se pojavljuje tipično samo jednom), metodologije razvijene na njima mogu se primeniti na skupove gde imamo više slika iste osobe tokom različitih godina. Buduća istraživanja bi mogla kombinovati modele obučene na UTKFace/APPA-REAL sa takvim skupovima (npr. FG-NET ili noviji longitudinalni skupovi) da bi se izgradili sistemi sposobni da predviđaju ne samo trenutnu starost već i tok starenja, ili da detektuju ubrzano starenje kao potencijalni zdravstveni indikator.

Smernice za buduća istraživanja u domenu skupova za procenu starosti mogu se sažeti na sledeći način: potrebno je težiti ka sve većoj raznolikosti i obuhvatnosti baza (po demografiji i uslovima snimanja), uz očuvanje kvaliteta oznaka. Kombinovanje različitih

izvora podataka, od Internet fotografija, preko kontrolisanih snimanja, do crouwdsourcing platformi pokazalo se uspešnim pristupom, te će verovatno budući projekti ići u tom smeru. Na kraju, otvorenost i dostupnost ovih skupova od ključnog je značaja za napredak, UTKFace i APPA-REAL su javno dostupni, što je omogućilo njihov širok uticaj. Nastavak takve prakse kroz objavljivanje novih skupova i verzija omogućiće istraživačima da zajednički unapređuju modele za procenu starosti i da ih čine sve pouzdanijim i pravednijim u različitim uslovima i za sve korisnike.

3. Arhitektura i izbor modela

3.1. Korišćene CNN arhitekture

U ovom radu korišćeni su široko priznati CNN modeli koji se često pominju u savremenim istraživanjima: VGG16 [33], VGG19 [33], ResNet50 [34] i MobileNetV2 [35], svi prethodno obučeni na skupu podataka ImageNet [36]. U nastavku su ukratko opisane njihove ključne karakteristike i razlozi uključivanja u uporednu analizu.

3.1.1. VGG16

VGG16 [33] je duboka CNN arhitektura sa ukupno 16 slojeva sa trenirajućim težinama, od čega 13 konvolucionih i 3 potpuno povezana (dense) sloja. Pooling slojevi nisu uključeni u ovaj broj jer ne sadrže težine. Karakteristika VGG modela jeste dosledna upotreba malih 3×3 filtera u svim konvolucionim slojevima, uz redukciju prostorne dimenzionalnosti pomoću 2×2 max pooling slojeva. Ovakav uniformni i sekvencijalni dizajn omogućava postepeno formiranje hijerarhijskih vizuelnih reprezentacija, uz zadržavanje relativne jednostavnosti u poređenju sa dubljim modelom VGG19. I pored toga, VGG16 ostvaruje visoke performanse u brojnim zadacima računarskog vida, što ga je učinilo pouzdanim referentnim modelom u našoj uporednoj analizi.

3.1.2. VGG19

VGG19 [33] predstavlja proširenu varijantu VGG16 arhitekture, sa ukupno 19 slojeva sa trenirajućim težinama (16 konvolucionih i 3 potpuno povezana sloja, ne računajući pooling slojeve). Arhitektura zadržava osnovni dizajn VGG modela: sekvencijalni niz konvolucija 3×3 sa ReLU aktivacijom i 2×2 max pooling slojevima za smanjenje prostornih dimenzija. Međutim, VGG19 uvodi dodatne konvolucione slojeve u

treći, četvrti i peti blok mreže, čime postiže veću dubinu. Ova dodatna složenost omogućava modelu da nauči detaljnije hijerarhijske reprezentacije karakteristične za starenje lica, što je doprinelo tome da u našim eksperimentima ostvari najbolje rezultate. Zbog toga je VGG19 izabran kao konačni model za dalje eksperimente u ovom radu.

3.1.3. ResNet50

ResNet50 [34] je 50-slojna CNN arhitektura koja uvodi koncept rezidualnog učenja kroz skip veze. Unutar rezidualnih blokova, izlaz jednog sloja dodaje se izlazu sloja koji se nalazi nekoliko nivoa dublje u mreži. Ova skip veza omogućava efikasniji protok gradijenta i uspešno treniranje veoma dubokih mreža bez problema nestajanja gradijenta. ResNet50 je dokazano uspešan na više referentnih zadataka prepoznavanja slika zbog čega smo ga odabrali kao snažnog kandidata i u domenu procene starosti lica. Očekivali smo da će njegova sposobnost učenja složenih vizuelnih karakteristika biti korisna za naš zadatak.

3.1.4. MobileNetV2

MobileNetV2 [35] je CNN arhitektura projektovana za efikasnost na uređajima ograničenih resursa (mobilni i ugrađeni sistemi). Koristi dubinske separabilne konvolucije kako bi drastično smanjila broj parametara i računarske operacije, a uvodi i invertovane rezidualne blokove sa linearnim bottleneck slojevima radi očuvanja reprezentativne moći modela. Uprkos svom kompaktnom dizajnu, MobileNetV2 postiže konkurentne performanse na zadacima klasifikacije slika. Uvrstili smo ga u uporednu analizu da bismo ispitali odnos između složenosti modela i performansi, posebno u kontekstu praktične primene gde su memorija i procesorska snaga ograničeni. MobileNetV2 je u eksperimentima pokazao solidnu tačnost u proceni starosti uz znatno manji model i brže izvođenje, demonstrirajući prednosti svoje optimizovane arhitekture.

3.2. Obrazloženje izbora CNN modela

Odabir navedenih CNN modela vođen je ključnim kriterijumima. Prvi kriterijum jeste efikasnot ekstrakcije karakteristika lica i otpornost na varijacije unutar klase (intra-klasne varijacije među uzorcima), i drugi jeste prilagođenost modela samom zadatku procene starosti lica. U nastavku su objašnjena ova dva aspekta izbora modela.

3.2.1. Efikasnost ekstrakcije karakteristika i otpornost na intra-klasne varijacije

Duboke CNN arhitekture koje smo odabrali dokazano su efikasne u izdvajanju relevantnih vizuelnih karakteristika sa slika lica. Zahvaljujući velikom broju konvolucionih slojeva, ovi modeli mogu da nauče više nivoe apstrakcije, od osnovnih ivica i tekstura kože do složenijih obrazaca, što obezbeđuje da se informacije bitne za određivanje starosti uspešno ekstraktuju iz slike. Na primer, VGG i ResNet arhitekture dostižu zapaženu dubinu i kapacitet reprezentacije, hvatajući finese u strukturi lica koje korelišu sa godinama, dok MobileNetv2 uprkos manjoj složenosti efikasno fokusira kapacitet na najbitnije karakteristike lica.

Pored efikasnosti ekstrakcije obeležja, odabrani modeli pokazuju i robustnost na intra-klasne varijacije, razlike među instancama iste klase. U kontekstu procene starosti, to podrazumeva varijacije među licima koja imaju istu ili sličnu starost (npr. različite osobe istih godina mogu izgledati prilično različito) kao i varijacije usled poza, osvetljenja ili izraza lica unutar iste starosne kategorije. Naši eksperimenti su pokazali da svi razmatrani modeli imaju približno sličnu stabilnost performansi uprkos ovakvim varijacijama unutar klase; nijedna arhitektura se nije izdvojila sa posebno većim padom tačnosti usled ovih promena, Drugim rečima, intra-klasna varijabilnost uticala je na rezultate modela VGG16, VGG19, ResNet50 i MobileNetV2 u sličnoj meri, što ukazuje da svi poseduju dobru generalizacionu sposobnost prema varijacijama među sličnim uzorcima.

3.2.2. Prilagođenost zadatku procene starosti

Svi izabrani modeli inicijalno su trenirani na opštem problemu klasifikacije, klasifikovanjem milion slika u 1000 klasa iz ImageNet skupa, ali se mogu efektivno prilagoditi specifičnom zadatku procene starosti lica. Njihova duboka arhitektura pruža bogatstvo predstavljenih karakteristika koje, nakon prenošenja učenja na naš domen, omogućavaju modelu na nauči obeležja direktno povezana sa starenjem. Modeli kao što su VGG19 i ResNet50 već su uspešno korišćeni u srodnim oblastima, uključujući prepoznavanje identiteta i klasifikaciju atributa lica, što potvrđuje njihovu sposobnost da modeluju fine vizuelne obrasce povezane sa procesom starenja. Fleksibilnost koju

omogućava pretrenirani pristup čini ih pogodnim za dalju adaptaciju putem finog podešavanja, čime se povećava tačnost predikcija. Detaljna strategija prilagođavanja modela ovom zadatku opisana je u sekciji 4, Metodologija.

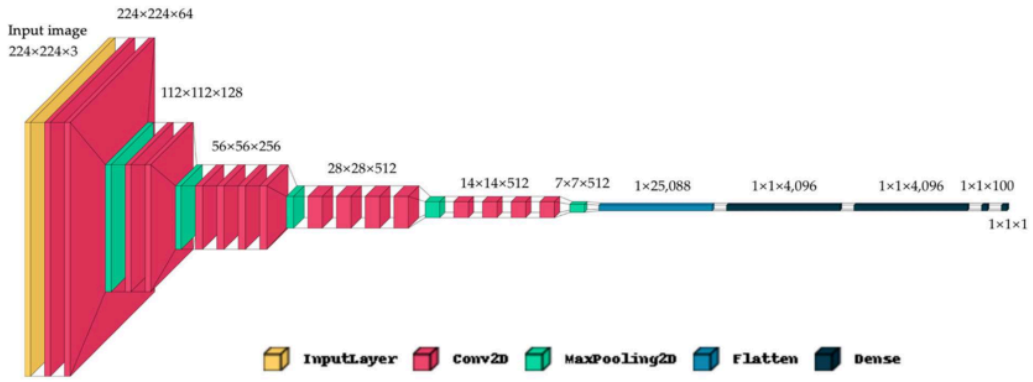
3.3. Uporedna analiza performansi arhitektura

U uporednoj analizi performansi korišćenih modela, VGG19 se izdvojio kao najuspešniji. Ostvario je najnižu srednju apsolutnu grešku (MAE) pri proceni starosti kako na originalnom skupu podataka, tako i na verziji skupa sa izbalansiranim brojem uzoraka. ResNet50 i MobileNetV2 takođe su pokazali konkurentne rezultate, dok je VGG16 očekivano za nijansu slabiji od svoje dublje varijante VGG19. Važno je napomenuti da su svi modeli zahtevali određeni stepen dotreniranja (fine-tuning) kako bi postigli optimalne performanse na našem zadatku. Konkretni slojevi koje je bilo neophodno otvoriti za treniranje kako bi se postigle optimalne performanse razlikovali su se između modela. Korišćenje modela isključivo kao ekstraktora karakteristika bez dodatnog treniranja dubokih slojeva pokazalo se nedovoljnim. S druge strane, potpuno treniranje svih slojeva od samog početka pogoršalo je rezultate zbog prekomernog prilagođavanja modela trening skupu. Ovi uvidi potvrđuju značaj pažljivo dizajnirane strategije finog podešavanja, koja je detaljnije objašnjena u Sekciji 4.

U zaključku ove analize, VGG19 je, zahvaljujući svojoj dubini i hijerarhijskom hvatanju karakteristika lica, pružio najtačniju procenu starosti u našem eksperimentalnom okruženju. Zbog ovih performansi, odabrali smo upravo VGG19 kao bazni model za dalje eksperimente. Buduća istraživanja mogla bi da ispitaju ansambl modele (kombinovanje više modela) ili nove napredne arhitekture poput EfficientNet, koje bi se potencijalno bolje nosile sa demografskom raznolikošću podataka i dodatno poboljšale pravičnosti i tačnost predviđanja.

3.4. Arhitektura izabranog modela (VGG19)

Nakon identifikacije najbolje performirajućeg modela u početnoj uporednoj evaluaciji, u našem slučaju VGG19, svi naredni eksperimenti fokusirani su na ovu arhitekturu kao bazu. Grafički prikaz kompletne arhitekture VGG19 modela dat je na Slici 17, radi lakšeg uvida u strukturu slojeva i veze među njima.



Slika 17. Reprezentacija arhitekture VGG-19 modela korišćenog u ovom radu.

Kao što je prikazano na Slici 17, naš model prima ulaznu veličinu od 224 × 224 × 3. Arhitektura počinje blokom koji sadrži dva konvoluciona sloja, nakon kojih sledi max pooling sloj. Max pooling je tehnika smanjenja uzorka koja se koristi za smanjenje prostornih dimenzija ulazne reprezentacije, što pomaže u smanjenju računarskog opterećenja, kontroli overfitting-a i povećanju robusnosti mreže. Ovaj korak je ključan u CNN arhitekturama kako bi se pojednostavila reprezentacija, a pri tome sačuvale važne karakteristike.

Zatim imamo još jedan blok sa dva konvoluciona sloja i max pooling slojem. Nakon toga sledi treći blok sa četiri konvoluciona sloja i max pooling slojem. Četvrti i peti blok se sastoje od po četiri konvoluciona sloja i max pooling sloja. Ukupno, mreža sadrži 16 konvolucionih slojeva.

Nakon konvolucionih i pooling slojeva, sloj za spljoštavanje (flattening) pretvara višedimenzionalni tenzor u jednosmerni vektor. Ova transformacija je ključna jer omogućava da izlaz iz konvolucionih i pooling slojeva bude prosleđen u guste (dense) slojeve za finalnu klasifikaciju ili regresiju, efikasno spajajući oba dela mreže.

Nakon sloja za spljoštavanje, slede dva gusta sloja sa po 4.096 čvorova, zatim još jedan gusti sloj sa 100 čvorova i na kraju izlazni sloj sa jednim čvorom. Skriveni konvolucioni i gusti slojevi koriste ReLU (Rectified Linear Unit) aktivacionu funkciju, koja je široko korišćena u CNN-ovima. ReLU uvodi nelinearnost u model, omogućavajući mu

da uči složene obrasce i funkcije. Ona je računarski efikasna, jer uključuje jednostavnu operaciju postavljanja negativnih vrednosti na nulu, dok pozitivne vrednosti ostaju nepromenjene, a takođe pomaže u ublažavanju problema nestajanja gradijenta. Pošto se bavimo regresionim zadatkom, aktivaciona funkcija koja se koristi u izlaznom sloju je linearna aktivacija. Linearna aktivaciona funkcija ne transformiše svoj ulaz, što znači da je izlaz neurona direktno jednak njegovom ulazu.

4. Metodologija

4.1. Transfer učenje i fino podešavanje modela

Da bismo maksimalno iskoristili prednosti navedenih složenih modela na relativno ograničenim skupovima podataka, primenili smo transfer učenje i pažljivo osmišljene strategije finog podešavanja modela (fine-tuning). U nastavku je objašnjen princip transfer učenja i naš pristup prilagođavanju modela za zadatak automatske procene starosti, nakon čega su predstavljeni postupci podešavanja hiperparametara i dizajn eksperimenta.

Transfer učenje predstavlja metodologiju kojom se znanje stečeno treniranjem modela na jednom velikom skupu podataka koristi za ubrzanje i poboljšanje učenja na drugom, srodnom zadatku. U našem radu iskoristili smo CNN modele prethodno obučene na velikom skupu slika ImageNet i prilagodili ih zadatku procene starosti. Ovaj pristup značajno smanjuje potrebnu količinu podataka i vreme obuke za novi zadatak, jer pretreniran model već poseduje bogat skup korisnih karakteristika naučenih iz ogromne baze slika. Umesto treniranja složene mreže od nule, transfer učenje nam omogućava da upotrebimo postojeću, dobro uhodanu arhitekturu i fokusiramo se samo na prilagođavanje njenih parametara specifičnostima našeg domena. Još jedna prednost ovog pristupa je mogućnost korišćenja veoma dubokih modela sa velikim brojem parametara bez potrebe za izuzetno velikim računarskim resursima, budući da su glavni obrasci već naučeni u inicijalnoj obuci. Konkretno, rani slojevi pretreniranog CNN modela tipično izdvajaju opšte karakteristike (ivice, oblike, teksture), te ih možemo zadržati uz minimalne izmene, dok se kasniji slojevi mogu specifično dotrenirati za prepoznavanje karakteristika vezanih za starenje lica.

Pri primeni transfer učenja, ključno je odrediti koliki deo modela će se dotrenirati na novom zadatku, to jest kakva će biti strategija finog podešavanja. Naš pristup sastojao se iz postepenog otključavanja slojeva za ponovnu obuku. U početku smo većinu parametara osnovnog modela držali zamrznutim i obučavali samo novi izlazni sloj prilagođen zadatku regresije starosti. Potom smo sukcesivno uključivali dodatne slojeve u proces obuke, počevši od najviših (onih najbližih izlazu) konvolucionih slojeva pa postepeno ka sve ranijim slojevima, prateći pri svakom koraku uticaj na performanse na validacionom skupu.

Eksperimentalno smo ustanovili da je selektivni fine-tuning nužan za postizanje najboljih rezultata kod svih posmatranih arhitektura. Svi modeli zahtevali su da se barem neki od prvobitno zamrznutih slojeva ponovo uključe u proces obuke kako bi se dostigao minimum greške, ali se optimalan broj otključanih slojeva razlikovao po modelu, a tako i po datasetovima na kojima su modeli trenirani. Treniranje isključivo novododatog izlaznog sloja pokazalo se nedovoljnim. Istovremeno, potpuno otključavanje svih slojeva mreže od samog početka treninga pokazalo se kontraproduktivno, ako bismo odmah pokušali da treniramo celu mrežu, performanse bi opale zbog prevelike prilagodljivosti modela (overfitting) na trening podatke. Zbog toga smo primenili promišljeno, postepeno fino podešavanje. Ovakav pristup omogućio je zadržavanje korisnih opštih obrazaca naučenih u ranim slojevima, uz istovremenu adaptaciju dubljih slojeva specifičnostima zadatka procene starosti. Kao rezultat, svaki model je, uz optimalan obim fine-tuninga, ostvario značajno bolje performanse nego u scenarijima bez ikakvog dotreniranja ili sa preagresivnim dotreniranjem.

4.2. Podešavanje hiperparametara putem pretrage mrežom

Nakon definisanja arhitektura i strategije treninga, kako bi optimizovali modele za predikciju stvarne starosti koristeći UTKFace i APPA-REAL skupove podataka, sproveli smo pretragu mrežom (grid search), varirajući hiperparametre unutar empirijski opravdanih opsega. Pretraga mrežom omogućila nam je sistematsko istraživanje kombinacija hiperparametara kako bismo identifikovali skup koji daje najbolje performanse modela. Konkretno, testirali smo različite stope učenja, veličine partija (batch size) i broj epoha unutar specificiranih opsega. Pretraga mrežom je uključivala

obuku više instanci svakog modela sa različitim kombinacijama ovih hiperparametara i izbor kombinacija koje su dale najnižu srednju apsolutnu grešku na validacionom skupu.

Stopa učenja je varirana u opsegu od 0.1 do 0.000001. Ovaj širok opseg je odabran kako bi se balansirala brzina konvergencije i finiji nivoi podešavanja. Veće stope učenja mogu dovesti do brže konvergencije, ali mogu preskočiti optimalne vrednosti, što rezultira lošim performansama. Niže stope učenja omogućavaju finije podešavanje, poboljšavajući sposobnost modela da se konvergira ka optimalnijem rešenju. Ovaj opseg nam omogućava da istražimo različite nivoe ažuriranja težina modela, osiguravajući da ne propustimo optimalne tačke koje mogu nastati na različitim skalama. Optimalna stopa učenja za oba skupa podataka bila je 0.0001, što je minimiziralo MAE na validacionom skupu.

Veličine partija su istraživane u opsegu od 16 do 128. Ovaj opseg je odabran da bi se balansirala računarska efikasnost i generalizacija modela. Manje veličine partija, iako računski skuplje, nisu dovele do bolje generalizacije u poređenju sa veličinom partije od 64. Veće veličine partija efikasnije su koristile GPU kapacitete, ali su ponekad dovodile do lošije generalizacije. Optimalna veličina partije od 64 dala je najbolju tačnost, balansirajući računarsku efikasnost i generalizaciju modela, i minimizirala MAE tokom pretrage mrežom.

Broj epoha variran je u opsegu od 30 do 100. Ovaj opseg je odabran kako bi se pronašla tačka gde model postiže optimalne performanse bez prekomernog fitovanja. Više epoha pomaže modelu da nauči složenije obrasce u podacima, ali takođe povećava rizik od prekomernog fitovanja, posebno sa ograničenim podacima. Ovaj opseg je obično dovoljan da model nauči složene obrasce, dok nam pruža fleksibilnost da zaustavimo obuku kada se greška na validacionom skupu stabilizuje. Obuka za 60 epoha bila je dovoljna za konvergenciju na APPA-REAL skupu podataka, dok je 50 epoha bilo optimalno za UTKFace skup podataka. Nakon tih tačaka, dodatne epohe nisu značajno smanjile grešku na validacionom skupu i ponekad su dovele do prekomernog fitovanja.

Pored toga, evaluirali smo različite optimizatore, kao što su Adam i SGD, procenjujući njihov uticaj na dinamiku obuke i brzinu konvergencije. Takođe smo varirali konfiguraciju potpuno povezanih slojeva na vrhu modela kako bismo optimizovali performanse procene starosti.

4.3. Strategije augmentacije i regularizacije modela

Oba krajnja modela VGG19 koristila su Adam optimizator i imala su koristi od finog podešavanja. Međutim, specifični slojevi otključani za obuku razlikovali su se: fino podešavanje je započelo od sloja 17 za APPA-REAL skup podataka, dok su optimalni rezultati za UTKFace skup podataka postignuti finim podešavanjem od sloja 7 nadalje. Kako bismo umanjili prekomerno fitovanje usled relativno malih veličina skupova podataka, primenili smo tehnike augmentacije podataka, uključujući rotaciju slika (u opsegu od 40 stepeni), pomeranje širine i visine (0.2), zumiranje (0.2) i horizontalno okretanje. Takođe, postavljen je slučajni broj 42 za oba, oversampling i augmentaciju podataka.

Sistematskim podešavanjem ovih parametara kroz grid search, naš cilj je bio da pronađemo najbolju kombinaciju koja minimizuje MAE na validacionom skupu, osiguravajući robusne i generalizovane performanse modela. Uvidi koje smo stekli u odnos između postavki parametara i performansi modela ključni su za optimizaciju dubokih učenja za zadatke procene starosti.

Naš primarni fokus nije bio na postizanju najviše moguće tačnosti, već na razumevanju odnosa između sastava skupa podataka i performansi modela kroz različite grupe. Korišćenjem grid search-a, dobili smo dobro performirajuće hiperparametre koji adekvatno zadovoljavaju naše istraživačke ciljeve. Iako grid search pruža sistematski pristup podešavanju hiperparametara, on je računarski skup. Budući radovi mogu uključiti naprednije tehnike kao što su Random Search ili Bayesian Optimization, koje uzimaju uzorke hiperparametara iz definisanih raspodela, što može dovesti do boljih performansi uz manji računarski trošak. Iako naš cilj nije bio postizanje vrhunskih ukupnih performansi, uvidi stečeni u odnose performansi među grupama su robusni sa trenutnim podešavanjima. Istraživanje ovih naprednih metoda podešavanja u budućim istraživanjima moglo bi dodatno poboljšati ukupnu tačnost i pravičnost.

4.4. Dizajn eksperimenta i evaluacija pravičnosti

Nakon što smo fino podesili arhitekturu i hiperparametre našeg osnovnog modela (VGG19), prešli smo na definisanje eksperimentalnog dizajna s ciljem temeljne procene

uticaja sastava skupa podataka na performanse modela, kako u ukupnom slučaju, tako i po demografskim grupama.

U prvom koraku, primenili smo oversampling na originalne verzije skupova UTKFace i APPA-REAL tako da broj uzoraka u manjinskim grupama (konkretno, osoba crnačkog i azijskog porekla) dostigne broj uzoraka u dominantnoj grupi (bele rase). Ovako izbalansirani skup podataka koristili smo kao referentni skup (baseline), uporedo sa originalnim, neizbalansiranim skupom, tokom svih kasnijih evaluacija.

Cilj izjednačavanja broja uzoraka po grupama bio je da se pripreme uslovi za kontrolisane eksperimente smanjenja, u kojima smo sistematski uklanjali određeni procenat uzoraka svake demografske grupe, od blagih redukcija (10% uklonjenih uzoraka) pa sve do potpunog uklanjanja cele grupe (100% uklonjenih uzoraka). Drugim rečima, kreirali smo verzije skupova podataka u kojima su slike iz pojedinih grupa (bele, crne i azijske populacije) sukcesivno uklanjane u koracima od po 10%, uključujući i slučajeve potpune eliminacije te grupe iz trening skupa.

Ovakva konstrukcija omogućila nam je da kvantifikujemo razlike u performansama modela za pojedinačne grupe, kako se broj uzoraka smanjuje, kao i da procenimo posredni uticaj tih promena na ukupnu tačnost i na performanse ostalih grupa. Rezultati dobijeni kroz ove eksperimente upoređeni su sa rezultatima modela treniranih na originalnom skupu, izbalansiranom (oversamplovanom) skupu, kao i svim međukonfiguracijama između te dve krajnosti.

Na ovaj način konstruisan eksperimentalni okvir omogućio nam je da identifikujemo koje varijacije u sastavu skupa podataka imaju najmanje negativan uticaj na performanse modela, kao i da utvrdimo da li postoji optimalna konfiguracija podataka koja minimizuje varijaciju u performansama između demografskih grupa, čime direktno doprinosi pravednosti modela. Nakon toga, prešli smo na formalnu analizu dobijenih rezultata.

5. Rezultati

5.1. Metodologija evaluacije performansi

U našoj analizi koristili smo Srednju apsolutnu grešku (MAE), standardnu devijaciju (SD), nejednaki uticaj (DI) i jednakost mogućnosti (EoO) kako bismo procenili performanse našeg VGG19 modela kroz različite kompozicije skupa podataka i etničke grupe. Formula za Srednju apsolutnu grešku (MAE) je data sa:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left| y_i - \hat{y}_i \right|$$

Gde y_i predstavlja stvarnu starost, $\hat{y}_i$ predstavlja predviđenu starost, dok n predstavlja ukupan broj uzoraka. Kako bismo procenili varijabilnost u performansama modela između etničkih grupa (Bela, Crna i Azijska), izračunali smo standardnu devijaciju MAE za ove grupe. MAE za svaku grupu je izračunat posebno, a zatim korišćen za određivanje standardne devijacije. Standardna devijacija (SD) MAE se izračunava prema formuli:

$$\mathrm{SD} = \sqrt{\frac{1}{3}\sum_{\mathrm{j}=1}^{3} (\mathrm{MAE_j} - \overline{\mathrm{MAE}})^2}$$

1. MAE_j predstavlja MAE za j-tu grupu (Bela, Crna ili Azijska).

2. $\overline{MAE}$ je prosečan MAE za sve tri grupe.

Izračunavanje standardne devijacije (SD) MAE-ova pomaže u razumevanju obima varijabilnosti performansi među različitim grupama, ističući bilo kakve potencijalne pristrasnosti ili nedoslednosti u predikcijama modela.

Pored toga, uključeni metrički pokazatelji pravičnosti, Nejednaki uticaj (Disparate Impact, DI) i Jednakost mogućnosti (Equality of Opportunity, EoO), dodatno procenjuju pravičnost našeg modela među različitim etničkim grupama. Ovi metrički pokazatelji su definisani na sledeći način:

Nejednaki uticaj:

$$DI = \frac{MAE_{grupa}}{MAE_{Belci}}$$

DI meri odnos MAE za specifičnu grupu u odnosu na MAE za belu grupu. DI vrednost od 1 označava savršen paritet, što sugeriše da nema nejednakog uticaja. Vrednosti manje od 1 ukazuju na potencijalnu nepovoljnost za tu grupu.

Jednakost mogućnosti:

$$EoO = 1 - \left|\frac{MAE_{grupa} - MAE_{Belci}}{MAE_{Belci}}\right|$$

EoO procenjuje koliko je MAE specifične grupe blizak MAE bele grupe, prilagođeno razmeri MAE bele grupe. EoO vrednost blizu 1 sugeriše da obe grupe imaju slične stope grešaka, što ukazuje na pravičnost u pogledu performansi modela. Korišćenjem ovih metrika, bolje možemo razumeti i ublažiti potencijalne pristrasnosti koje naš model može pokazati, osiguravajući da on bude pravičan u odnosu na različite demografske grupe.

5.2. Rezultati na APPA-REAL skupu podataka

APPA-REAL skup podataka, koji je veoma nebalansiran jer grupe crna i azijska zajedno čine manje od 10% dostupnih podataka, iznenađujuće pokazuje relativno dobre rezultate u pogledu varijacija MAE između ovih grupa, sa standardnom devijacijom od 0.19 na modelu obučenom na originalnom obliku skupa podataka. Može se pretpostaviti da bi oversamplovanje kako bi se izjednačili veličine grupa dovelo do jednakih performansi između ovih grupa ili barem smanjilo devijaciju MAE. Međutim, naši nalazi sugerišu suprotno. Model obučen sa 20% manje uzoraka iz bele grupe imao je najmanju standardnu devijaciju od 0.04, pri čemu su grupe bela i crna imale gotovo jednake performanse, sa MAE vrednostima od 7.1676 i 7.1663, dok je azijska grupa bila malo bolja sa 7.0736. Originalna kompozicija skupa podataka dala je samo sedmu najmanju standardnu devijaciju među grupama, dok je jednako oversamplovani skup podataka zauzeo osmo mesto.

Ovo pokazuje da je šest drugih kompozicija skupa podataka postiglo bolje rezultate u pogledu smanjenja varijacije MAE između grupa, pri čemu je dataset sa 20% smanjenom belom grupom postigao smanjenje standardne devijacije od 78.94% u poređenju sa originalnim skupom podataka i 80.95% u poređenju sa jednako oversamplovanim skupom podataka. Najbolji ukupni MAE postignut je sa originalnom kompozicijom skupa podataka (6.45), dok je model sa najizjednačenijim performansama

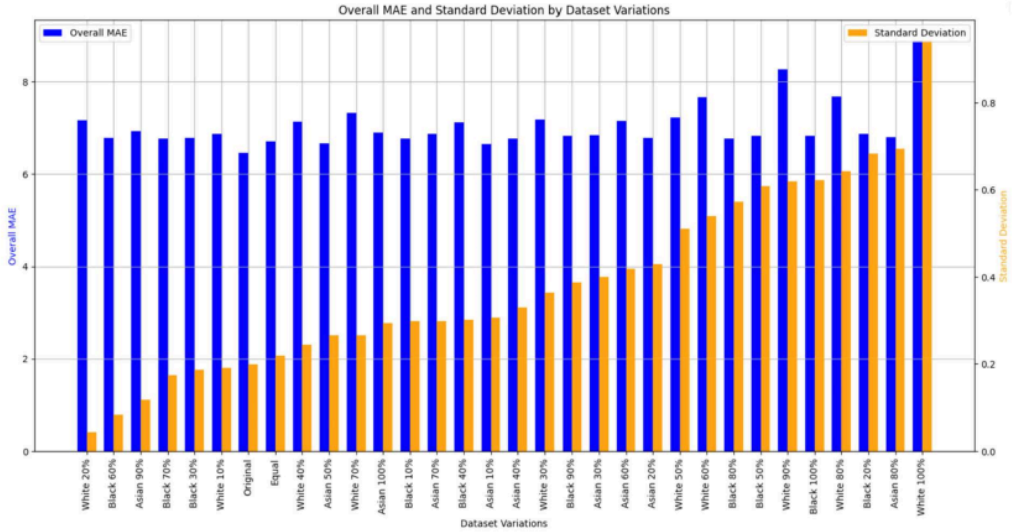
imao ukupni MAE od 7.16, što sugeriše da postizanje jednakih performansi među grupama dolazi sa malim troškom za ukupnu performansu. Najlošiji model je obučen sa kompozicijom skupa podataka u kojoj je bela grupa potpuno izostavljena, što je rezultiralo ukupnim MAE od 8.89 i standardnom devijacijom od 0.94. U poređenju s tim, model sa najizjednačenijim performansama imao je 95.74% manju varijaciju. Tabela 4 prikazuje ove rezultate, sortirane po najmanjoj standardnoj devijaciji.

Varijacije podataka	Ukupni MAE	MAE Bele Grupe	MAE Crne Grupe	MAE Azijske Grupe	Standardna Devijacija	DI Crna Grupa	EoO Crna Grupa	DI Azijska Grupa	EoO Azijska Grupa
Bela 20%	7.1603	7.1676	7.1663	7.0736	0.0440	0.9998	0.9998	0.9869	0.9869
Crna 60%	6.7788	6.764	6.7813	6.9478	0.0828	1.0026	0.9974	1.0272	0.9728
Azijska 90%	6.9338	6.956	6.9148	6.6874	0.1181	0.9941	0.9941	0.9613	0.9613
Crna 70%	6.7632	6.7431	7.1447	6.8198	0.1740	1.0595	0.9405	1.0114	0.9886
Crna 30%	6.7834	6.7537	7.2087	6.9303	0.1872	1.0673	0.9327	1.0261	0.9739
Bela 10%	6.8679	6.8687	7.1783	6.7177	0.1917	1.0451	0.9549	0.978	0.978
Original	6.4588	6.4241	6.4529	6.8593	0.1987	1.0045	0.9955	1.0677	0.9323
Jednaki	6.7129	6.4547	6.9672	6.5739	0.2189	1.0794	0.9206	1.0185	0.9815
Bela 40%	7.1286	7.153	7.3405	6.7528	0.2451	1.0262	0.9739	0.944	0.944
Azijska 50%	6.6626	6.6547	7.1479	6.5305	0.2666	1.0741	0.9259	0.9813	0.9813
Bela 70%	7.3275	7.3464	7.625	6.9741	0.2666	1.0379	0.9621	0.9493	0.9493
Azijska 100%	6.8969	6.8269	7.2377	7.5441	0.2938	1.0602	0.9398	1.1049	0.8951
Crna 10%	6.7724	6.77	7.2787	6.5685	0.2988	1.0751	0.9249	0.9703	0.9703
Azijska 70%	6.8689	6.9326	6.7004	6.2145	0.2991	0.9665	0.9665	0.8963	0.8963
Crna 40%	7.1176	7.0839	7.7754	7.2036	0.3017	1.0977	0.9023	1.0169	0.9831
Azijska 10%	6.6531	6.6385	7.2428	6.5518	0.3073	1.0910	0.909	0.9869	0.9869
Azijska 40%	6.7736	6.731	7.5072	6.927	0.3295	1.1152	0.8848	1.0291	0.9709
Bela 30%	7.1779	7.2144	7.49	6.6156	0.3650	1.0382	0.9621	0.9169	0.9169
Crna 90%	6.8216	6.8529	7.2218	6.2795	0.3877	1.0538	0.9462	0.9162	0.9162
Azijska 30%	6.8468	6.8523	7.4678	6.4999	0.3999	1.0898	0.9102	0.949	0.949

Azijska 60%	7.1463	7.1156	8.0009	7.1076	0.4192	1.1244	0.8756	0.9989	0.9989
Azijska 20%	6.7844	6.7327	7.733	6.9448	0.4303	1.1486	0.8514	1.0315	0.9685
Bela 50%	7.224	7.2357	7.9872	6.7413	0.5122	1.1039	0.8961	0.9317	0.9317
Bela 60%	7.6724	7.7254	8.1431	6.8498	0.5389	1.0541	0.9459	0.8866	0.8866
Crna 80%	6.7672	6.762	7.7404	6.3828	0.5719	1.1447	0.8553	0.9439	0.9439
Crna 50%	6.8294	6.845	7.7236	6.2425	0.6081	1.1285	0.8715	0.9119	0.9119
Bela 90%	8.2641	8.3888	8.039	6.9359	0.6191	0.9583	0.9583	0.8267	0.8267
Crna 100%	6.8285	6.7755	8.1289	6.8429	0.6227	1.2000	0.8000	1.0100	0.9900
Bela 80%	7.6813	7.7504	8.1856	6.6583	0.6424	1.0562	0.9438	0.8589	0.8589
Crna 20%	6.8741	6.8916	7.8781	6.2144	0.6831	1.1431	0.8569	0.9017	0.9017
Azijska 80%	6.7975	6.7792	8.0498	6.4349	0.6944	1.1873	0.8127	0.9491	0.9491
Bela 100%	8.8956	9.1258	7.3192	6.9751	0.9432	0.8022	0.8022	0.7646	0.7646

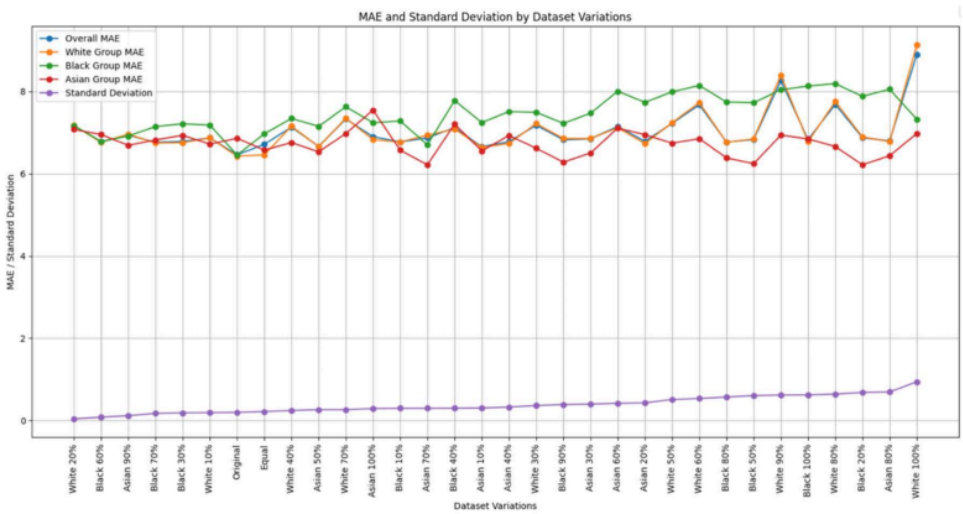
Tabela 4. Poređenje svih varijacija na APPA-REAL skupu podataka, sortirano prema najmanjoj standardnoj devijaciji između performansi grupa.

Radi bolje jasnoće promena u performansama modela, uključili smo Sliku 18, koja prikazuje ukupni MAE i standardnu devijaciju za različite varijacije skupa podataka.



Slika 18. Ukupni MAE i standardna devijacija prema varijacijama APPA-REAL skupa podataka

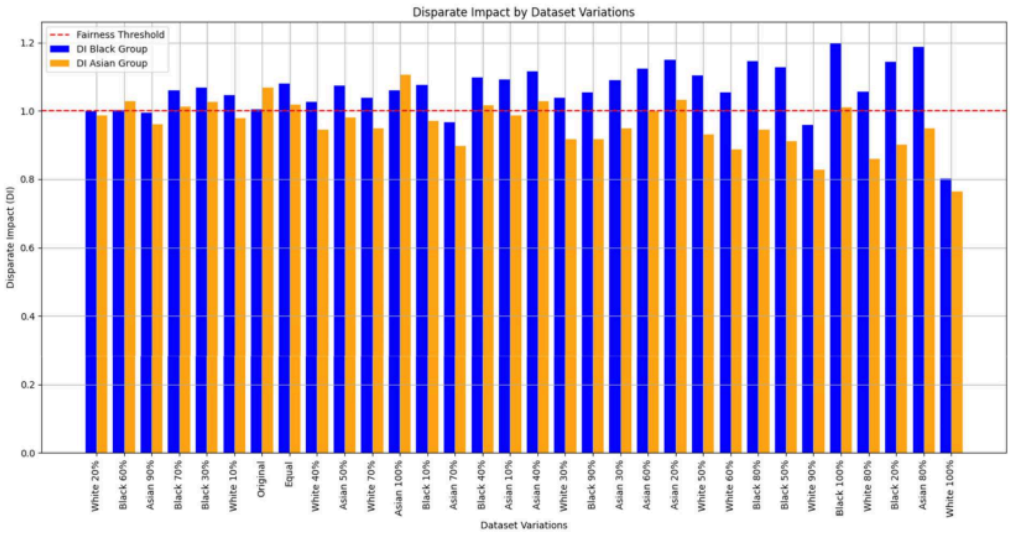
Slika 19 pruža detaljan prikaz trendova MAE i standardne devijacije za različite grupe i ukupno.



Slika 19. MAE i standardna devijacija prema varijacijama APPA-REAL skupa podataka.

Kako bismo bolje razumeli pravičnost ovih modela, analiziramo metrike nejednakog uticaja i jednakosti mogućnosti. Ove metrike pomažu u proceni kako različite demografske grupe bivaju pogođene performansama modela. Nejednaki uticaj je mera koja pokazuje da li je jedna grupa nepovoljno pogođena u poređenju sa drugom. DI vrednost blizu 1 ukazuje na to da model pravično tretira različite grupe, dok vrednost značajno različita od 1 ukazuje na potencijalnu pristrasnost.

Slika 20 prikazuje nejednaki uticaj za crnu i azijsku grupu kroz različite kompozicije skupa podataka.



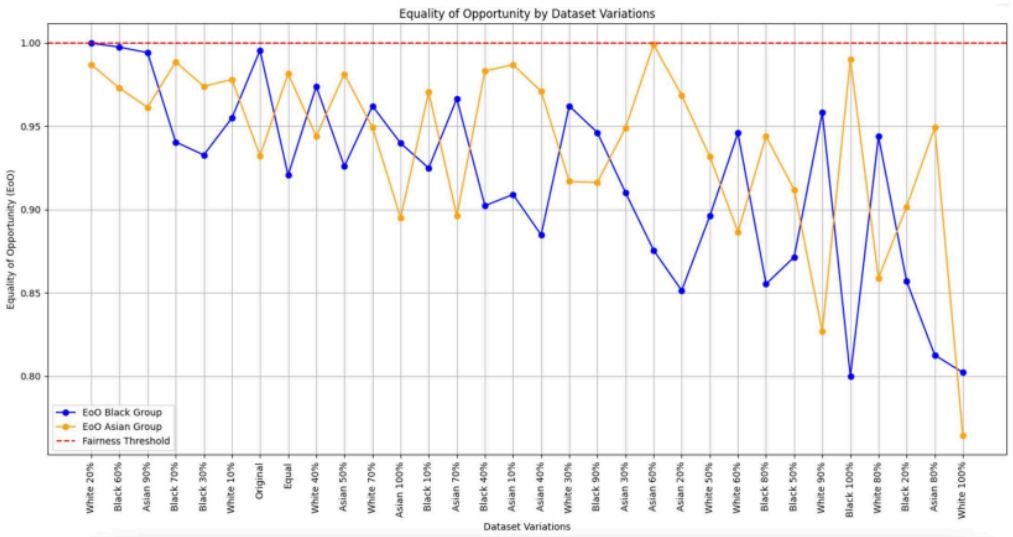
Slika 20. Nejednaki uticaj kroz varijacije APPA-REAL skupa podataka.

Za varijaciju sa smanjenjem 20% uzoraka Bele grupe, model pokazuje vrednosti DI od 0.9998 za obe grupe, Crnu i Azijsku, što sugeriše da ova varijacija postiže gotovo savršenu pravičnost među tim grupama. Kod varijacije sa 60% smanjenjem uzoraka Crne grupe, vrednosti DI iznose 1.0026 za Crnu grupu i 1.0272 za Azijsku grupu. Ove vrednosti su blizu 1, što ukazuje na minimalan različiti uticaj i sugeriše da i ova varijacija podržava pravičnost. Varijacija sa 90% uzoraka Azijske grupe ima vrednosti DI od 0.9941 za obe grupe, održavajući ravnotežu u tretmanu različitih demografskih grupa i osiguravajući minimalnu pristrasnost.

Zanimljivo je da, uprkos izjednačavanju veličina uzoraka, vrednosti DI za kod jednakog seta podataka iznose 1.0794 za Crnu grupu i 1.0185 za Azijsku grupu. Ove vrednosti ukazuju na blagu pristrasnost, što sugeriše da samo izjednačavanje proporcija skupa podataka ne eliminiše nužno nejednaki uticaj.

Jednakost mogućnosti meri da li različite grupe imaju iste šanse za postizanje povoljnog ishoda. Kao i DI, vrednost EoO bliska 1 ukazuje na pravedne performanse među grupama.

Slika 21 prikazuje vrednosti Jednakosti Mogućnosti za Crnu i Azijsku grupu kroz različite varijacije APPA-REAL skupa podataka.



Slika 21. Jednakost mogućnosti prema varijacijama APPA-REAL skupa podataka.

U varijaciji Bela 20%, vrednosti EoO za ovu varijaciju su 0.9998 za obe grupe, Crnu i Azijsku, što pokazuje da ovaj model pruža jednake mogućnosti za ove grupe. Za varijaciju Crna 60%, EoO vrednosti su 0.9974 za Crnu grupu i 0.9728 za Azijsku grupu, što ukazuje na dobru jednakost mogućnosti sa malim varijacijama. Varijacija Azijska 90% pokazuje EoO vrednosti od 0.9941 za obe grupe, što sugeriše da model ravnomerno funkcioniše u svim demografskim grupama.

Suprotno tome, EoO vrednosti za varijaciju jednaki dataset su 0.9206 za Crnu grupu i 0.9815 za Azijsku grupu. Ove vrednosti ukazuju na izvesnu nejednakost, naglašavajući da jednaka proporcija u skupu podataka ne znači automatski jednaku mogućnost.

Proučavanjem modela sa lošijim performansama, posebno onih na dnu tabele, dobijaju se dodatni uvidi. Na primer, varijacija Bela 100%, sa najvišom standardnom devijacijom (0.9432), ima DI vrednosti od 0.8022 za obe grupe Crnu i Azijsku, što ukazuje na značajan nejednaki uticaj. EoO vrednosti za ovu varijaciju su 0.7646, što pokazuje značajnu nejednakost mogućnosti. Ovo sugeriše da potpuno isključivanje određenih demografskih grupa može dovesti do značajnih pristrasnosti i nepravednih rezultata.

Drugi primer je varijacija Crna 100%, koja ima standardnu devijaciju od 0.6227. DI vrednosti su 1.2000 za Crnu grupu i 1.0100 za Azijsku grupu, što ukazuje na veći nejednaki uticaj, posebno za Crnu grupu. EoO vrednosti su 0.8000 za Crnu grupu i 0.9900 za Azijsku grupu, što dalje pokazuje značajne nejednakosti u mogućnostima. Ovi primeri ilustruju da varijacije u skupu podataka sa ekstremnim neravnotežama ili potpunim isključenjima određenih grupa obično pokazuju veće razlike u uticaju i mogućnostima. Stoga je važno uzeti u obzir i metrike performansi i mere pravičnosti prilikom procene i selektovanja varijacija skupa podataka za obuku modela. Cilj je postići ravnotežu koja minimizuje pristrasnost i promoviše pravedne rezultate za sve demografske grupe.

Duboko analizirajući zasebne grupe, možemo primetiti u Tabeli 5 kako MAE grupe reaguje na smanjenje uzoraka; prvo ćemo razmotriti belu grupu. Performanse bele grupe same po sebi slične su onima drugih grupa, najbolje sa originalnom kompozicijom skupa podataka sa MAE od 6.42, a druga po performansama je jednaka varijacija skupa podataka, gde su, ponovo, svi uzorci još uvek prisutni. Kao što se može videti u Tabeli 4, dolazi do očiglednog pogoršanja performansi što više uzoraka uklonimo. Povećanje MAE nije linearno. Reakcija na smanjenje uzoraka za 10% može varirati od povećanja MAE za 0.20% do 8.79%. Međutim, jasno je da smanjenje broja uzoraka za 10% ne znači automatski smanjenje performansi za 10%. Većinu vremena, stopa pogoršanja je mnogo manja od 10%, a nikada ne dostiže takvo trenutno smanjenje. Povećanje standardne devijacije takođe prati naš obrazac uklanjanja sve većeg broja uzoraka tokom obuke. Performanse bele grupe, kada se uopšte ne koristi za obuku, dovode do 42.06% višeg MAE tokom testiranja u poređenju sa originalnim skupom podataka i 41.38% u poređenju sa jednako oversamplovanim skupom podataka. Jasno je da prisustvo ovih uzoraka ima značaj.

Tabela 5. Promene performansi prilikom smanjenja uzorka bele grupe na APPA-REAL skupu podataka.

Varijacije seta podataka	Ukupan MAE	MAE Bele Grupe	MAE Crne Grupe	MAE Azijske Grupe	Standardna Devijacija	MAE Promena u odnosu na prošlu vrednost	MAE u odnosu na jednaki set podataka	MAE u odnosu na originalan set podataka	MAE Crne grupe u odnosu na originalan set podataka	MAE Crne grupe u odnosu na jednaki set podataka	MAE Azijske grupe u odnosu na originalan set podataka	MAE Azijske grupe u odnosu na jednaki set podataka
Original	6.4588	6.4241	6.4529	6.8593	0.1987							
Jednaki	6.7129	6.4547	6.9672	6.5739	0.2189	0.48%		0.48%	7.97%		−4.16%	
10%	6.8679	6.8687	7.1783	6.7177	0.1917	6.41%	6.41%	6.92%	11.24%	3.03%	−2.06%	2.19%
20%	7.1603	7.1676	7.1663	7.0736	0.0440	4.35%	11.04%	11.57%	11.06%	2.86%	3.12%	7.60%
30%	7.1779	7.2144	7.49	6.6156	0.3650	0.65%	11.77%	12.30%	16.07%	7.50%	−3.55%	0.63%
40%	7.1286	7.153	7.3405	6.7528	0.2451	−0.85%	10.82%	11.35%	13.76%	5.36%	−1.55%	2.72%
50%	7.224	7.2357	7.9872	6.7413	0.5122	1.16%	12.10%	12.63%	23.78%	14.64%	−1.72%	2.55%
60%	7.6724	7.7254	8.1431	6.8498	0.5389	6.77%	19.69%	20.26%	26.19%	16.88%	−0.14%	4.20%
70%	7.3275	7.3464	7.625	6.9741	0.2666	−4.91%	13.81%	14.36%	18.16%	9.44%	1.67%	6.09%
80%	7.6813	7.7504	8.1856	6.6583	0.6424	5.50%	20.07%	20.65%	26.85%	17.49%	−2.93%	1.28%
90%	8.2641	8.3888	8.039	6.9359	0.6191	8.24%	29.96%	30.58%	24.58%	15.38%	1.12%	5.51%
100%	8.8956	9.1258	7.3192	6.9751	0.9432	8.79%	41.38%	42.06%	13.42%	5.05%	1.69%	6.10%

Veoma je interesantno videti da se performanse crne grupe takođe pogoršavaju kako se sve više i više uzoraka bele grupe uklanja. Najveće povećanje MAE za tu grupu iznosi 26,85% u poređenju sa performansama na originalnom skupu podataka, kada je 80% uzoraka bele grupe uklonjeno. Zanimljivo je da performanse počinju da se poboljšavaju kada se ukloni 90% ili 100% uzoraka bele grupe, sa povećanjem MAE za crnu grupu u poređenju sa originalnim performansama od 13,42% kada je 100% uzoraka bele grupe uklonjeno. Crna grupa je manje reaktivna u poređenju sa njenim performansama na jednakom oversamplovanom skupu podataka, sa varijacijama u performansama koje nikada ne prelaze 20% povećanja, a obično su oko 5%. Kada je 100% uzoraka bele grupe uklonjeno, performanse crne grupe u poređenju sa jednakim sklopom podataka su lošije za samo 5,05%.

Performanse azijske grupe ostaju relativno iste, bez obzira na broj uklonjenih uzoraka bele grupe, sa povećanjem MAE u rasponu od −3,55% do 3,12% u poređenju sa originalnim performansama. Azijska grupa deluje nešto reaktivnije u poređenju sa performansama na jednakom skupu podataka, sa povećanjima MAE od oko 6%. Jasno je da su bele i crne grupe više povezane.

Ispitivanje performansi crne grupe, kako je prikazano u Tabeli 6, pokazuje primetno pogoršanje kako se postotak njenog uzorka uklanja. Iznenađujuće, ne postoji linearna korelacija između povećanog uklanjanja uzoraka i pogoršanja performansi. Na primer, uklanjanje 10% uzoraka rezultira povećanjem MAE od 12,80%, dok uklanjanje 20% dovodi do povećanja od 22,09% u poređenju sa performansama na originalnom skupu podataka. Međutim, nakon ovog trenutka, povećanje MAE varira između 5,09% i 20,49%, sugerišući da performanse postepeno stabilizuju ili čak poboljšavaju kada se ukloni više uzoraka nego samo 10% ili 20%. Ipak, samo kada je crna grupa potpuno isključena iz treninga, primećujemo najveće povećanje MAE, koje iznosi 25,97% u poređenju sa originalnim skupom podataka. Ovo povećanje je značajno niže od onog primećenog za belu grupu, što ukazuje na veću otpornost crne grupe na nedovoljnu zastupljenost uzoraka.

Tabela 6. Promene u performansama prilikom smanjenja uzorka crne grupe na APPA-REAL skupu podataka.

Varijacije dataseta	Ukupni MAE	MAE Bele Grupe	MAE Crne Grupe	MAE Azijske Grupe	Standardna Devijacija	MAE Promena u odnosu na prošlu vrednost	MAE u odnosu na jednak set podataka	MAE u odnosu na originalan set podataka	MAE Bele grupe u odnosu na original set podataka	MAE Bele grupe u odnosu na jednaki set podataka	MAE Azijske grupe u odnosu na originalan set podataka	MAE Azijske grupe u odnosu na jednaki set podataka
Original	6.4588	6.4241	6.4529	6.8593	0.1987							
Jednaki	6.7129	6.4547	6.9672	6.5739	0.2189	7.97%		7.97%	0.48%		−4.16%	
10%	6.7724	6.77	7.2787	6.5685	0.2988	4.47%	4.47%	12.80%	5.38%	4.88%	−4.24%	−0.08%
20%	6.8741	6.8916	7.8781	6.2144	0.6831	8.23%	13.07%	22.09%	7.28%	6.77%	−9.40%	−5.47%
30%	6.7834	6.7537	7.2087	6.9303	0.1872	−8.50%	3.47%	11.71%	5.13%	4.63%	1.04%	5.42%
40%	7.1176	7.0839	7.7754	7.2036	0.3017	7.86%	11.60%	20.49%	10.27%	9.75%	5.02%	9.58%

50%	6.8294	6.845	7.7236	6.2425	0.6081	−0.67%	10.86%	19.69%	6.55%	6.05%	−8.99%	−5.04%
60%	6.7788	6.764	6.7813	6.9478	0.0828	−12.20%	−2.67%	5.09%	5.29%	4.79%	1.29%	5.69%
70%	6.7632	6.7431	7.1447	6.8198	0.1740	5.36%	2.55%	10.72%	4.97%	4.47%	−0.58%	3.74%
80%	6.7672	6.762	7.7404	6.3828	0.5719	8.34%	11.10%	19.95%	5.26%	4.76%	−6.95%	−2.91%
90%	6.8216	6.8529	7.2218	6.2795	0.3877	−6.70%	3.65%	11.92%	6.67%	6.17%	−8.45%	−4.48%
100%	6.8285	6.7755	8.1289	6.8429	0.6227	12.56%	16.67%	25.97%	5.47%	4.97%	−0.24%	4.09%

Zanimljivo je da performanse crne grupe pokazuju veću otpornost u poređenju sa jednakim skupom podataka, sa prosečnim povećanjem MAE od oko 10% u različitim testovima, što dostiže maksimalnih 16,97% kada su svi uzorci uklonjeni tokom testiranja. Kada su u pitanju reakcije drugih grupa na smanjenje uzorka crne grupe, ne postoji jasna korelacija. Povećanje MAE za belu grupu kreće se od 4,97% do 10,27% i ne pokazuje dosledno veće povećanje kada se ukloni više uzoraka crne grupe. Azijska grupa pokazuje nešto reaktivniji odgovor na uklanjanje uzoraka crne grupe u poređenju sa uklanjanjem uzoraka bele grupe, iako ne u značajnoj meri. Uklanjanje uzoraka crne grupe zapravo poboljšava performanse azijske grupe, sa smanjenjima MAE koja se kreću od −0,58% do −9,40% u poređenju sa originalnim skupom podataka kada se ukloni 20% uzoraka crne grupe.

Kada je reč o azijskoj grupi, kao što je prikazano u Tabeli 7, ona se čini najmanje reaktivnom na smanjenje uzoraka. Zanimljivo je da uklanjanje 70% njenih uzoraka dovodi do smanjenja MAE od 9,40% u poređenju sa originalnim performansama, što sugeriše na poboljšanje tačnosti u nekim slučajevima. Samo kada su svi uzorci uklonjeni, MAE se povećava za 9,98% u poređenju sa originalnim skupom podataka i za 14,76% u poređenju sa jednakim modelom. Nasuprot tome, bela grupa pokazuje malu reakciju na smanjenje uzoraka azijske grupe, sa povećanjima MAE koja se dosledno kreću između 5% i 10%, bez obzira na broj uklonjenih uzoraka azijske grupe.

Tabela 7. Promene u performansama nakon smanjenja uzorka azijske grupe na APPA-REAL

Varijacije seta podataka	Ukupni MAE	MAE Bele Grupe	MAE Crne Grupe	MAE Azijske Grupe	Standardna devijacija	MAE Promena u odnosu na prethodnu vrednost	MAE u odnosu na jednaki set podataka	MAE u odnosu na originalni set podataka	MAE Bele grupe u odnosu na originalni set podataka	MAE Bele grupe u odnosu na jednaki set podataka	MAE Crne grupe u odnosu na originalni set podataka	MAE Crne grupe u odnosu na originalni set podataka
Original	6.4588	6.4241	6.4529	6.8593	0.1987							
Jednaki	6.7129	6.4547	6.9672	6.5739	0.2189	−4.16%		−4.16%	0.48%		7.97%	
10%	6.6531	6.6385	7.2428	6.5518	0.3073	−0.34%	−0.34%	−4.48%	3.34%	2.85%	12.24%	3.96%
20%	6.7844	6.7327	7.733	6.9448	0.4303	6.00%	5.64%	1.25%	4.80%	4.31%	19.84%	10.99%
30%	6.8468	6.8523	7.4678	6.4999	0.3999	−6.41%	−1.13%	−5.24%	6.67%	6.16%	15.73%	7.19%
40%	6.7736	6.731	7.5072	6.927	0.3295	6.57%	5.37%	0.99%	4.78%	4.28%	16.34%	7.75%
50%	6.6626	6.6547	7.1479	6.5305	0.2666	−5.72%	−0.66%	−4.79%	3.59%	3.10%	10.77%	2.59%
60%	7.1463	7.1156	8.0009	7.1076	0.4192	8.84%	8.12%	3.62%	10.76%	10.24%	23.99%	14.84%
70%	6.8689	6.9326	6.7004	6.2145	0.2991	−12.57%	−5.47%	−9.40%	7.92%	7.40%	3.84%	−3.83%
80%	6.7975	6.7792	8.0498	6.4349	0.6944	3.55%	−2.11%	−6.19%	5.53%	5.03%	24.75%	15.54%
90%	6.9338	6.956	6.9148	6.6874	0.1181	3.92%	1.73%	−2.51%	8.28%	7.77%	7.16%	−0.75%
100%	6.8969	6.8269	7.2377	7.5441	0.2938	12.81%	14.76%	9.98%	6.27%	5.77%	12.16%	3.88%

Suprotno tome, crna grupa pokazuje značajnije reakcije, sa povećanjima MAE od 3,84% do 24,75% kada je 80% uzoraka azijske grupe uklonjeno u poređenju sa performansama na originalnom skupu podataka. Zanimljivo je da performanse crne grupe ostaju stabilnije u poređenju sa jednako oversamplovanim skupom podataka, što sugeriše da oversamplovanje generalno pomaže u stabilizaciji performansi.

5.3. Rezultati na UTKFace skupu podataka

UTK Face skup podataka, iako je u početku uravnoteženiji u poređenju sa APPA REAL skupom podataka sa većim udelom crnih i azijskih grupa, iznenađujuće ne postiže pravičnije performanse među grupama kao modeli obučeni na APPA REAL skupu podataka. Generalno, MAE je niži na UTK skupu podataka zbog većeg broja uzoraka

dostupnih za obuku. Najbolji model ostaje onaj obučavan na originalnom skupu podataka, sa ukupnim MAE od 4,89, dok je model sa jednako oversamplovanim skupom podataka postigao MAE od 4,98.

Zanimljivo je da je standardna devijacija među grupama u originalnom skupu podataka 0,74. Suprotno očekivanjima, jednako oversamplovani skup podataka pokazuje povećanu varijaciju sa standardnom devijacijom od 0,85, zauzimajući 18. mesto u pogledu varijacije u poređenju sa originalnim skupom podataka koji je zauzeo 5. mesto. Ovo naglašava da samo izjednačavanje proporcija skupa podataka nije dovoljno za postizanje uravnoteženih performansi među demografskim grupama.

Model koji pokazuje najmanje odstupanje među grupama uključuje uklanjanje 90% uzoraka azijske grupe tokom obuke, što rezultira standardnom devijacijom od 0,30. Iako ovaj model pokazuje viši ukupni MAE od 5,48 u poređenju sa originalnim modelom (povećanje od 12,06%) i jednako oversamplovanim modelom (povećanje od 10,04%), on značajno smanjuje varijaciju performansi među grupama za 59,45% u poređenju sa originalnim modelom i 64,70% u poređenju sa jednako oversamplovanim modelom. Jasno je da su koristi od smanjenja varijacije među grupama veće od manjeg povećanja ukupnog MAE.

Zanimljivo je da su sva četiri najbolja modela u pogledu smanjenja varijacije uključivali neko smanjenje uzoraka azijske grupe tokom obuke, što pokazuje da modeli bolje funkcionišu kako se više azijskih uzoraka uklanja. Ovaj pristup je takođe bio efikasan na APPA REAL skupu podataka, zauzimajući treće mesto u pogledu varijacije među grupama sa standardnom devijacijom od 0,11.

Suprotno tome, modeli koji su uklonili sve uzorke iz bele grupe tokom obuke pokazali su najveću varijaciju (1,36), što je odražavalo nalaze na APPA REAL skupu podataka. Najniže rangirane varijacije bile su slične onima na APPA REAL skupu podataka, sa potpunim uklanjanjem uzoraka crne grupe koji je doveo do četvrte najgore varijacije na UTK skupu podataka, što je slično petom mestu na APPA REAL.

Tokom naših testova na UTK skupu podataka, pojavio se dosledan trend: bela grupa je obično pokazivala najviši MAE (~5), što je bilo blizu crnoj grupi (~5), koja je bila nešto niža od bele grupe. Iako je imala manji broj uzoraka na početku, azijska grupa je

dosledno imala najbolje performanse sa MAE od oko 3. Međutim, degradacija performansi ove grupe na kraju je dovela do pravičnijih performansi među svim grupama.

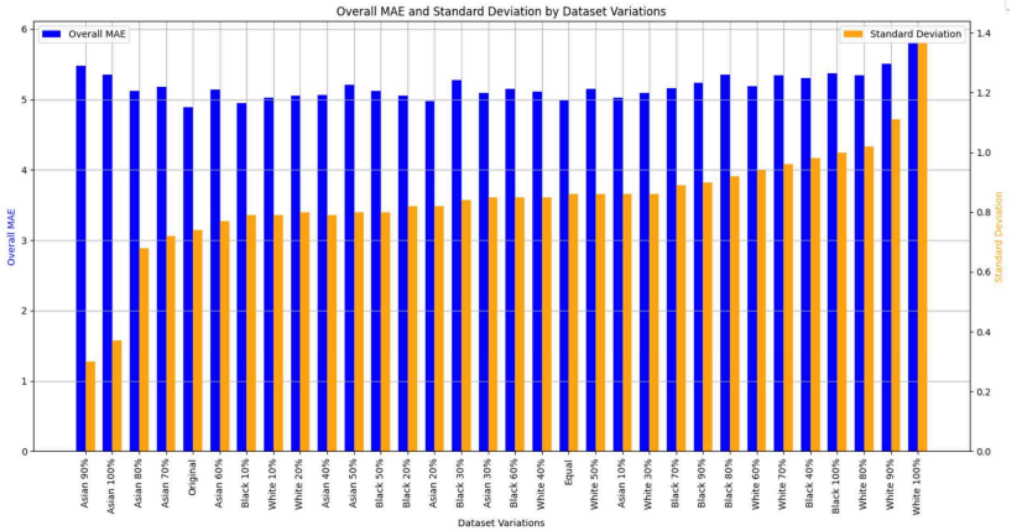
Za detaljan pregled rezultata eksperimenata na UTK Face skupu podataka, poređanih prema standardnoj devijaciji, pogledajte Tabelu 8.

Tabela 8. Poređenje svih varijacija na UTKFace skupu podataka, sortirano prema najnižoj standardnoj devijaciji između performansi grupa.

Varijacije seta podataka	Ukupni MAE	MAE Bele grupe	MAE Crne Grupe	MAE Azijske grupe	Standardna devijacija	DI Crne grupe	EoO Crne grupe	DI Azijske grupe	EoO Azijske grupe
Azijska 90%	5.4808	5.7351	5.4583	5.0036	0.3015	0.9517	0.9518	0.8724	0.8725
Azijska 100%	5.3499	5.6983	5.2564	4.7976	0.3677	0.9225	0.9226	0.8417	0.8418
Azijska 80%	5.1197	5.5972	5.2153	4.0006	0.6807	0.9318	0.9319	0.7147	0.7148
Azijska 70%	5.178	5.5635	5.2437	3.908	0.7170	0.9425	0.9426	0.7025	0.7026
Original	4.8925	5.3772	5.0445	3.6674	0.7401	0.9381	0.9382	0.6820	0.6821
Azijska 60%	5.1361	5.7023	5.2071	3.877	0.7707	0.9132	0.9133	0.6799	0.6800
Crna 10%	4.9504	5.4966	5.072	3.6505	0.7894	0.9227	0.9228	0.6643	0.6644
Bela 10%	5.028	5.5884	5.1332	3.7261	0.7926	0.9185	0.9186	0.6670	0.6671
Azijska 40%	5.0608	5.6095	5.1833	3.7543	0.7934	0.9240	0.9241	0.6683	0.6684
Bela 20%	5.0549	5.6	5.1896	3.7361	0.7996	0.9267	0.9268	0.6672	0.6673
Azijska 50%	5.2064	5.8149	5.2504	3.9063	0.8005	0.9030	0.9031	0.6717	0.6718
Crna 50%	5.1155	5.5867	5.3529	3.7797	0.8024	0.9581	0.9582	0.6764	0.6765
Azijska 20%	4.9849	5.5227	5.1505	3.6308	0.8183	0.9325	0.9326	0.6575	0.6576
Crna 30%	5.2779	5.821	5.4611	3.8847	0.8408	0.9381	0.9382	0.6678	0.6679
Azijska 30%	5.088	5.6537	5.2456	3.6904	0.8458	0.9277	0.9278	0.6528	0.6529
Crna 60%	5.1542	5.6783	5.3728	3.7417	0.8501	0.9462	0.9463	0.6591	0.6592
Bela 40%	5.1123	5.7115	5.2279	3.7152	0.8503	0.9154	0.9155	0.6505	0.6506
Jednaki	4.9879	5.5465	5.1663	3.5711	0.8557	0.9315	0.9316	0.6439	0.6440
Bela 50%	5.152	5.7823	5.2317	3.7499	0.8582	0.9047	0.9048	0.6484	0.6485
Azijska 10%	5.0262	5.6409	5.1295	3.6179	0.8588	0.9094	0.9095	0.6417	0.6418
Bela 30%	5.0891	5.6989	5.2045	3.6709	0.8634	0.9132	0.9133	0.6444	0.6445
Crna 70%	5.1646	5.5988	5.5342	3.6896	0.8851	0.9885	0.9886	0.6581	0.6582
Crna 20%	5.0512	5.6157	5.2732	3.5516	0.9031	0.9389	0.9390	0.6323	0.6324

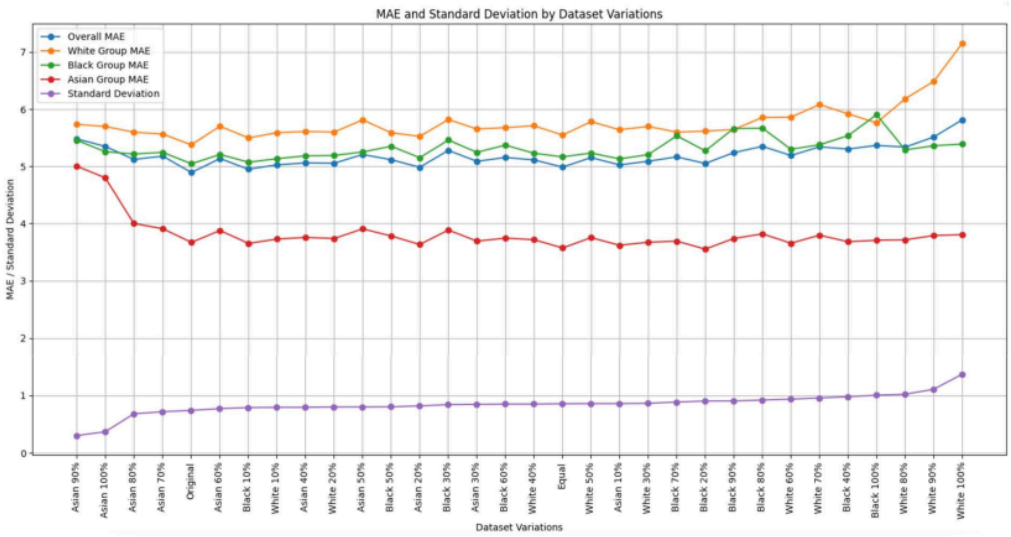
Crna 90%	5.2389	5.6447	5.6612	3.7357	0.9038	1.0029	1.0030	0.6618	0.6619
Crna 80%	5.3513	5.8564	5.6689	3.8171	0.9203	0.9679	0.9680	0.6528	0.6529
Bela 60%	5.1887	5.8597	5.2998	3.6538	0.9362	0.9043	0.9044	0.6234	0.6235
Bela 70%	5.3421	6.0802	5.3782	3.7933	0.9565	0.8846	0.8847	0.6239	0.6240
Crna 40%	5.3027	5.92	5.535	3.6802	0.9778	0.9350	0.9351	0.6215	0.6216
Crna 100%	5.3681	5.7606	5.9055	3.7052	1.0048	1.0252	1.0253	0.6432	0.6433
Bela 80%	5.3377	6.1819	5.2886	3.7123	1.0209	0.8555	0.8556	0.6005	0.6006
Bela 90%	5.5117	6.4859	5.3614	3.7876	1.1066	0.8266	0.8267	0.5841	0.5842
Bela 100%	5.8152	7.1516	5.3912	3.803	1.3676	0.7536	0.7537	0.5316	0.5317

Radi bolje jasnoće promena u performansama modela, uključili smo Sliku 22, koja prikazuje ukupni MAE i standardnu devijaciju za različite varijacije skupa podataka.



Slika 22. Ukupni MAE i standardna devijacija za varijacije skupa podataka UTK Face

Takođe Slika 23. Prikazuje detalje trendova MAE i standardne devijacije za različite grupe i ukupno.



Slika 23. MAE i standardna devijacija prema varijacijama UTK Face skupa podataka.

Za UTK Face dataset, varijacija sa 90% uklonjenih azijskih uzoraka pokazuje vrednosti DI od 0,9517 za crnu grupu i 0,8724 za azijsku grupu, što ukazuje na to da ova varijacija postiže skoro pravičan tretman ovih grupa. Slično tome, varijacija sa 100% azijskih uzoraka uklonjenih ima DI vrednosti od 0,9225 za crnu grupu i 0,8417 za azijsku grupu, što ukazuje na minimalan nejednaki uticaj i promoviše pravičnost.

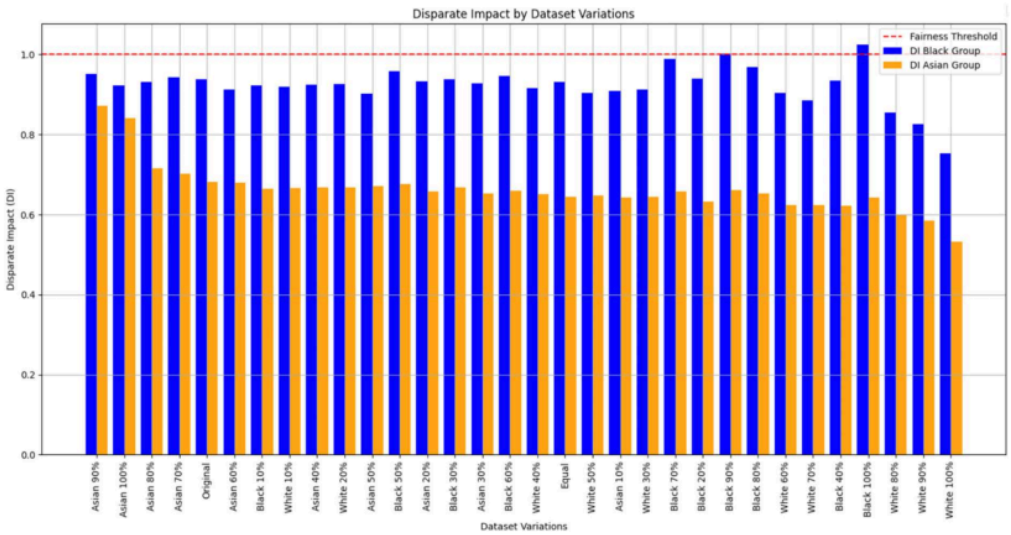
Nasuprot tome, varijacija sa jednakim datasetom, uprkos izjednačavanju veličina uzoraka, ima DI vrednosti od 0,9315 za crnu grupu i 0,6439 za azijsku grupu, što odražava prisustvo pristrasnosti. Najlošije rangirani model po pitanju DI vrednosti je varijacija sa 100% belih uzoraka uklonjenih, sa DI vrednostima od 0,7536 za crnu grupu i 0,5316 za azijsku grupu, što ukazuje na značajnu pristrasnost. Slika 9 prikazuje nejednaki uticaj za crnu i azijsku grupu kroz različite varijacije datasetova.

Jednakost prilike meri da li različite grupe imaju jednake šanse za postizanje povoljnog ishoda. Kao i DI, vrednost EoO blizu 1 ukazuje na ravnopravne performanse među grupama.

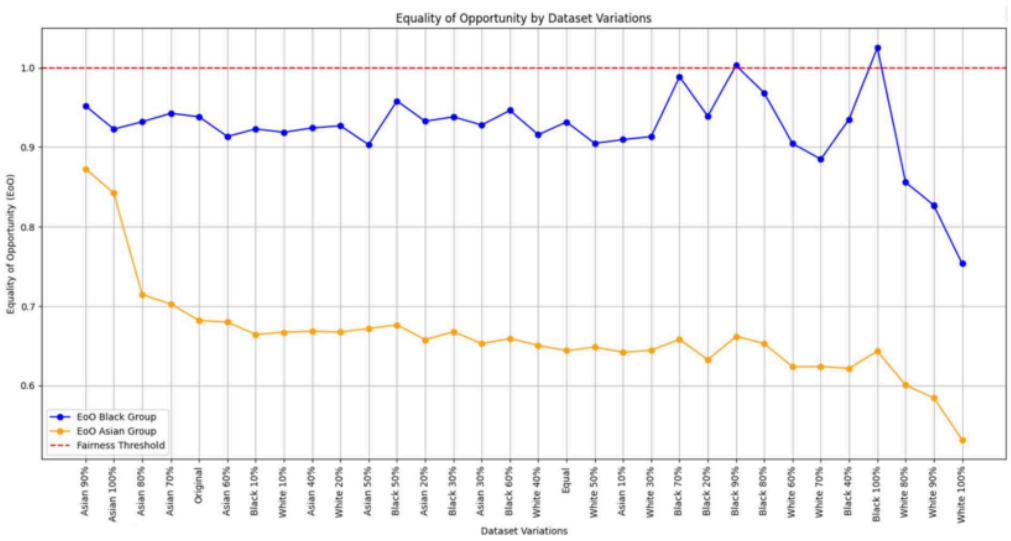
Varijacija sa 90% azijskih uzoraka uklonjenih pokazuje EoO vrednosti od 0,9518 za crnu grupu i 0,8725 za azijsku grupu, što ukazuje na to da ovaj model pruža skoro jednake prilike ovim grupama. Varijacija sa 100% azijskih uzoraka uklonjenih ima EoO vrednosti od 0,9226 za crnu grupu i 0,8418 za azijsku grupu, takođe ukazujući na dobru jednakost prilike.

Nasuprot tome, varijacija sa jednakim datasetom pokazuje EoO vrednosti od 0,9316 za crnu grupu i 0,6440 za azijsku grupu, što odražava određenu nejednakost. Varijacija sa 100% belih uzoraka ima najlošije EoO vrednosti od 0,7537 za crnu grupu i 0,5317 za azijsku grupu, što ukazuje na značajnu nejednakost. Slika 25 prikazuje vrednosti jednakosti prilike za crnu i azijsku grupu kroz različite kompozicije datasetova.

Usmeravajući pažnju na performanse svake grupe pojedinačno, počinjemo sa belom grupom. Kao što je prikazano u Tabeli 9, uočava se primetan trend gde smanjenje uzoraka za 10% ne dovodi nužno do pada performansi grupe za 10%. Umesto toga, primećujemo promene u performansama koje se obično kreću između 1% i 3%. Najznačajnije povećanje MAE, od približno 10,26%, javlja se tek kada se broj uzoraka bele grupe smanji sa 90% na potpuno izostavljanje.



Slika 24. Nejednaki uticaj prema varijacijama UTK Face skupa podataka.



Slika 25. Jednakost mogućnosti prema varijacijama UTK Face skupa podataka.

Poredeći ove rezultate sa jednakim i originalnim skupovima podataka, slično nalazima sa APPA REAL skupa podataka, bela grupa pokazuje blažu reakciju na smanjenje uzoraka u jednakom skupu podataka u poređenju sa originalnim. Najveće povećanje MAE, u poređenju sa jednakim skupom podataka, dostiže 28,94%, dok u poređenju sa originalnim skupom podataka dostiže 33%.

Tabela 9. Promene u performansama nakon smanjenja uzorka bele grupe na UTKFace skupu podataka.

Varijacije seta podataka	MAE Ukupni	MAE Bele grupe	MAE Crne grupe	MAE Azijske grupe	Standardna devijacija	MAE Promena u odnosu na prethodnu vrednost	MAE u odnosu na jednak set podataka	MAE u odnosu na originalan set podataka	Black Group Compared to Original	MAE Crne grupe u odnosu na jednak set podataka	MAE Azijske grupe u odnosu na originalan set podataka	MAE Azijske grupe u odnosu na jednak set podataka
Original	4.8925	5.3772	5.0445	3.6674	0.7401							
Jednaki	4.9879	5.5465	5.1663	3.5711	0.8557	3.15%		3.15%	2.41%		−2.63%	
10%	5.028	5.5884	5.1332	3.7261	0.7926	0.76%	0.76%	3.93%	1.76%	−0.64%	1.60%	4.34%
20%	5.0549	5.6	5.1896	3.7361	0.7996	0.21%	0.96%	4.14%	2.88%	0.45%	1.87%	4.62%
30%	5.0891	5.6989	5.2045	3.6709	0.8634	1.77%	2.75%	5.98%	3.17%	0.74%	0.10%	2.79%
40%	5.1123	5.7115	5.2279	3.7152	0.8503	0.22%	2.97%	6.22%	3.64%	1.19%	1.30%	4.04%

50%	5.152	5.7823	5.2317	3.7499	0.8582	1.24%	4.25%	7.53%	3.71%	1.27%	2.25%	5.01%
60%	5.1887	5.8597	5.2998	3.6538	0.9362	1.34%	5.65%	8.97%	5.06%	2.58%	−0.37%	2.32%
70%	5.3421	6.0802	5.3782	3.7933	0.9565	3.76%	9.62%	13.07%	6.62%	4.10%	3.43%	6.22%
80%	5.3377	6.1819	5.2886	3.7123	1.0209	1.67%	11.46%	14.97%	4.84%	2.37%	1.22%	3.95%
90%	5.5117	6.4859	5.3614	3.7876	1.1066	4.92%	16.94%	20.62%	6.28%	3.78%	3.28%	6.06%
100%	5.8152	7.1516	5.3912	3.803	1.3676	10.26%	28.94%	33.00%	6.87%	4.35%	3.70%	6.49%

Zanimljivo, slično zapažanjima u APPA REAL skupu podataka, bela grupa pokazuje značajnu osetljivost na smanjenje veličine uzorka tokom obuke. S druge strane, druge grupe pokazuju manje izražene reakcije na promene u sastavu uzorka bele grupe. Performanse crne grupe pogoršavaju se za oko 1% do 4%, bez obzira na procenat uklonjenih belih uzoraka. Slično tome, performanse azijske grupe dosledno opadaju za oko 6%, bez obzira na to da li je 20% ili 90% uzoraka bele grupe uklonjeno. Ovaj obrazac sugeriše, slično nalazima u APPA REAL skupu podataka, da azijska grupa ostaje relativno nepogođena takvim promenama.

Međutim, za razliku od APPA-REAL skupa podataka, u ovoj analizi opažamo slabiju korelaciju u performansama između bele i crne grupe.

Ispitivanjem performansi crne grupe, detaljno prikazano u Tabeli 10, otkriva se jasna tendencija gde povećanje uklanjanja uzoraka korelira sa višim MAE. Najveći porast MAE javlja se kada su svi uzorci crne grupe isključeni iz obuke, što rezultira povećanjem od 17,07% u poređenju sa originalnim performansama. Zanimljivo je da je ovaj porast nešto manji, 14,31%, kada se uporedi sa performansama jednakog modela, što ukazuje na to da oversampling koristi crnoj grupi stabilizovanjem njenih performansi.

Tabela 10. Promene u performansama prilikom smanjenja uzorka crne grupe na UTKFace skupu podataka.

Varijacije seta podataka	Ukupni MAE	MAE Bele grupe	MAE Crne grupe	MAE Azijske grupe	Standardna devijacija	MAE promena u odnosu na prethodnu vrednost	MAE u odnosu na jednak set podataka	MAE u odnosu na originalan set podataka	MAE Bele grupe u odnosu na original set podataka	MAE Bele grupe u odnosu na jednak set podataka	MAE Azijske grupe u odnosu na originalan set podataka	MAE Azijske grupe u odnosu na jednak set podataka
Original	4.8925	5.3772	5.0445	3.6674	0.7401							
Jednaki	4.9879	5.5465	5.1663	3.5711	0.8557	2.41%		2.41%	3.15%		−2.63%	
10%	4.9504	5.4966	5.072	3.6505	0.7894	−1.83%	−1.83%	0.55%	2.22%	−0.90%	−0.46%	2.22%
20%	5.0512	5.6157	5.2732	3.5516	0.9031	3.97%	2.07%	4.53%	4.44%	1.25%	−3.16%	−0.55%
30%	5.2779	5.821	5.4611	3.8847	0.8408	3.56%	5.71%	8.26%	8.25%	4.95%	5.93%	8.78%
40%	5.3027	5.92	5.535	3.6802	0.9778	1.35%	7.14%	9.72%	10.09%	6.73%	0.35%	3.06%
50%	5.1155	5.5867	5.3529	3.7797	0.8024	−3.29%	3.61%	6.11%	3.90%	0.72%	3.06%	5.84%
60%	5.1542	5.6783	5.3728	3.7417	0.8501	0.37%	4.00%	6.51%	5.60%	2.38%	2.03%	4.78%
70%	5.1646	5.5988	5.5342	3.6896	0.8851	3.00%	7.12%	9.71%	4.12%	0.94%	0.61%	3.32%
80%	5.3513	5.8564	5.6689	3.8171	0.9203	2.43%	9.73%	12.38%	8.91%	5.59%	4.08%	6.89%
90%	5.2389	5.6447	5.6612	3.7357	0.9038	−0.14%	9.58%	12.23%	4.97%	1.77%	1.86%	4.61%
100%	5.3681	5.7606	5.9055	3.7052	1.0048	4.32%	14.31%	17.07%	7.13%	3.86%	1.03%	3.76%

Slično posmatranjima sa belom grupom, smanjenje uzorka za 10% nije linearno povezano sa povećanjem MAE od 10% za crnu grupu. Umesto toga, primećujemo visoko jednodižno povećanje, sa porastom od 14,31% prilikom prelaska sa smanjenja od 90% na potpuno uklanjanje uzoraka crne grupe.

U kontrastu, druge grupe pokazuju minimalnu reakciju na promene u veličini uzorka crne grupe. MAE bele grupe povećava se za otprilike 1% do 6%, bez obzira na procenat uklonjenih uzoraka crne grupe. Slično tome, azijska grupa pokazuje varijacije u povećanju MAE od 2% do 8%, pokazujući dosledan obrazac sa rezultatima testova na APPA REAL datasetu za crnu grupu. Okrećući se azijskoj grupi, kao što je prikazano u Tabeli 11, primećujemo obrazac sličan onom viđenom u rezultatima sa APPA-REAL

datasetom. Smanjenja veličine uzorka uglavnom rezultiraju manjim fluktuacijama u MAE, obično u opsegu od blagih smanjenja do povećanja od 1–4%. Jedna izuzetnost javlja se pri smanjenju uzorka sa 80% na 90%, gde primetimo značajan skok u MAE od 25,07%. Iznenađujuće, promena u performansama između 90% uklonjenih uzoraka i 100% zapravo je poboljšanje performansi od 4,12%, što ukazuje na složen odnos između veličine uzorka i performansi za azijsku grupu.

Tabela 11. Promene performansi usled smanjenja uzorka azijske grupe na UTKFace datasetu.

Varijacije seta podataka	MAE Ukupni	MAE Bele grupe	MAE Crne grupe	MAE Azijske grupe	Standardna devijacija	MAE Promena u odnosu na prethodnu vrednost	MAE u odnosu na jednak set podataka	MAE u odnosu na originalan set podataka	MAE Bele grupe u odnosu na original set podataka	MAE Bele grupe u odnosu na jednak set podataka	MAE Crne grupe u odnosu na originalan set podataka	MAE Crne grupe u odnosu na jednak set podataka
Original	4.8925	5.3772	5.0445	3.6674	0.7401							
Jednaki	4.9879	5.5465	5.1663	3.5711	0.8557	−2.63%		−2.63%	3.15%		2.41%	
10%	5.0262	5.6409	5.1295	3.6179	0.8588	1.31%	1.31%	−1.35%	4.90%	1.70%	1.69%	−0.71%
20%	4.9849	5.5227	5.1505	3.6308	0.8183	0.36%	1.67%	−1.00%	2.71%	−0.43%	2.10%	−0.31%
30%	5.088	5.6537	5.2456	3.6904	0.8458	1.64%	3.34%	0.63%	5.14%	1.93%	3.99%	1.53%
40%	5.0608	5.6095	5.1833	3.7543	0.7934	1.73%	5.13%	2.37%	4.32%	1.14%	2.75%	0.33%
50%	5.2064	5.8149	5.2504	3.9063	0.8005	4.05%	9.39%	6.51%	8.14%	4.84%	4.08%	1.63%
60%	5.1361	5.7023	5.2071	3.877	0.7707	−0.75%	8.57%	5.72%	6.05%	2.81%	3.22%	0.79%
70%	5.178	5.5635	5.2437	3.908	0.7170	0.80%	9.43%	6.56%	3.46%	0.31%	3.95%	1.50%
80%	5.1197	5.5972	5.2153	4.0006	0.6807	2.37%	12.03%	9.09%	4.09%	0.91%	3.39%	0.95%
90%	5.4808	5.7351	5.4583	5.0036	0.3015	25.07%	40.11%	36.43%	6.66%	3.40%	8.20%	5.65%
100%	5.3499	5.6983	5.2564	4.7976	0.3677	−4.12%	34.35%	30.82%	5.97%	2.74%	4.20%	1.74%

Upoređujući ove rezultate sa performansama originalnog i jednako raspodeljenog skupa podataka, azijska grupa pokazuje obrazac reakcije sličan onom kod bele grupe. Uklanjanje 100% uzoraka azijske grupe rezultira povećanjem MAE od 34,35% u poređenju sa jednakim skupom podataka i povećanjem od 30,82% u poređenju sa originalnim skupom podataka. Ova reakcija je znatno izraženija u poređenju sa rezultatima dobijenim sa APPA REAL datasetom.

Što se tiče reakcija drugih grupa na smanjenje broja uzoraka azijske grupe, primećujemo konstantan porast MAE u opsegu od 1% do 8%, bez obzira na obim smanjenja uzoraka azijske grupe. Ovo se donekle razlikuje od nalaza sa APPA REAL skupa podataka, gde je crna grupa pokazala značajnije povećanje MAE koje nije bilo striktno povezano sa smanjenjem uzoraka azijske grupe. Najveća reakcija na APPA REAL datasetu bila je povećanje MAE od 24,75% za crnu grupu.

5.4. Uporedna analiza: naš model naspram sličnih radova

Kako bismo pružili sveobuhvatnu evaluaciju performansi našeg modela, upoređujemo naše rezultate sa onima iz drugih sličnih radova, fokusirajući se na ukupni MAE, tačnost prema rasama i standardnu devijaciju. Iako su prethodne studije o proceni starosti izvele različite tačnosti za različite etničke grupe, nijedna eksplicitno nije adresirala razlike u performansama između različitih demografskih grupa. Tabela 12 u nastavku prikazuje rezime naših nalaza zajedno sa sličnim radovima.

Kao što je primećeno, naš model postiže znatno niži ukupni MAE i standardnu devijaciju na oba skupa podataka, UTKFace i APPA-REAL, u poređenju sa prethodnim studijama. Ovo ukazuje na uravnoteženije performanse među različitim etničkim grupama, naglašavajući efikasnost našeg pristupa u rešavanju razlika u performansama.

Tabela 12. Upoređivanje naših rezultata sa sličnim radovima

Rad	Dataset	Ukupni MAE	Grupa			Standardna Devijacija
			Bela	Crna	Azijska	
[11]	UTKFace		9.79	7.71	9.56	0.931
[11]	APPA-REAL		7.79	8.23	7.85	0.1948
[18]	APPA-REAL	7.356	7.40	7.73	6.59	0.4789
[16]	APPA-REAL	13.5774	13.6386	14.1607	12.4729	0.7055
Naši rezultati	UTKFace	5.4808	5.7351	5.4583	5.0036	0.3015
Naši rezultati	APPA-REAL	7.1603	7.1676	7.1663	7.0735	0.044

6. Zaključak

6.1. Ključni ciljevi i metodološki pristup

U ovom istraživanju, ispitivali smo uticaj sastava skupa podataka na performanse modela za procenu starosti, fokusirajući se na ublažavanje pristrasnosti među različitim etničkim grupama. Koristili smo pristup transfernog učenja, koristeći unapred obučene CNN modele (VGG16, VGG19, ResNet50 i MobileNetV2) i prilagodili ih na UTKFace i APPA-REAL skupovima podataka, koji su izabrani zbog svoje demografske raznovrsnosti i uključivanja relevantnih oznaka, kao što su stvarna starost i etnička pripadnost.

Naša metodologija je uključivala sistematsku manipulaciju sastavom skupa podataka oversampling-om manjinskih grupa kako bi se izjednačili sa većinskom grupom, a zatim postepeno smanjenje veličine uzorka svake grupe. Ovo nam je omogućilo da analiziramo odnos između sastava skupa podataka i performansi modela, kako ukupno, tako i za specifične etničke grupe. Kao primarne metrike za ocenu koristili smo Srednju apsolutnu grešku (MAE) i standardnu devijaciju.

6.2. Efekti manipulacije podacima na pravičnost

Naši nalazi otkrivaju da samo balansiranje skupa podataka oversampling-om manjinskih grupa ne vodi nužno ka pravičnim performansama među etničkim grupama. Ovo je u skladu sa zapažanjima Puc-a i saradnika (2020) [8], koji su uočili razlike u performansama među različitim rasnim grupama u modelima za procenu starosti, ali nisu aktivno manipulisali skupovima podataka da bi ublažili ove pristrasnosti. Nasuprot tome, naše istraživanje je pokazalo da variranje broja uzoraka iz većinske grupe (bela) može dovesti do mešanih rezultata. U nekim slučajevima, to je dovelo do izjednačenijih performansi među etničkim grupama, što je označeno nižom standardnom devijacijom MAE, dok u drugim slučajevima nije značajno poboljšalo pravičnost ili ukupne performanse. Ovo sugeriše da oversampling i undersampling možda nisu uvek efikasne strategije same po sebi za ublažavanje pristrasnosti i da je potreban suptilniji pristup.

6.3. Rezultati modela na UTKFace i APPA-REAL skupovima

Za UTKFace skup podataka, model obučen na originalnom sastavu skupa podataka postigao je ukupni MAE od 4.89 sa standardnom devijacijom od 0.74.

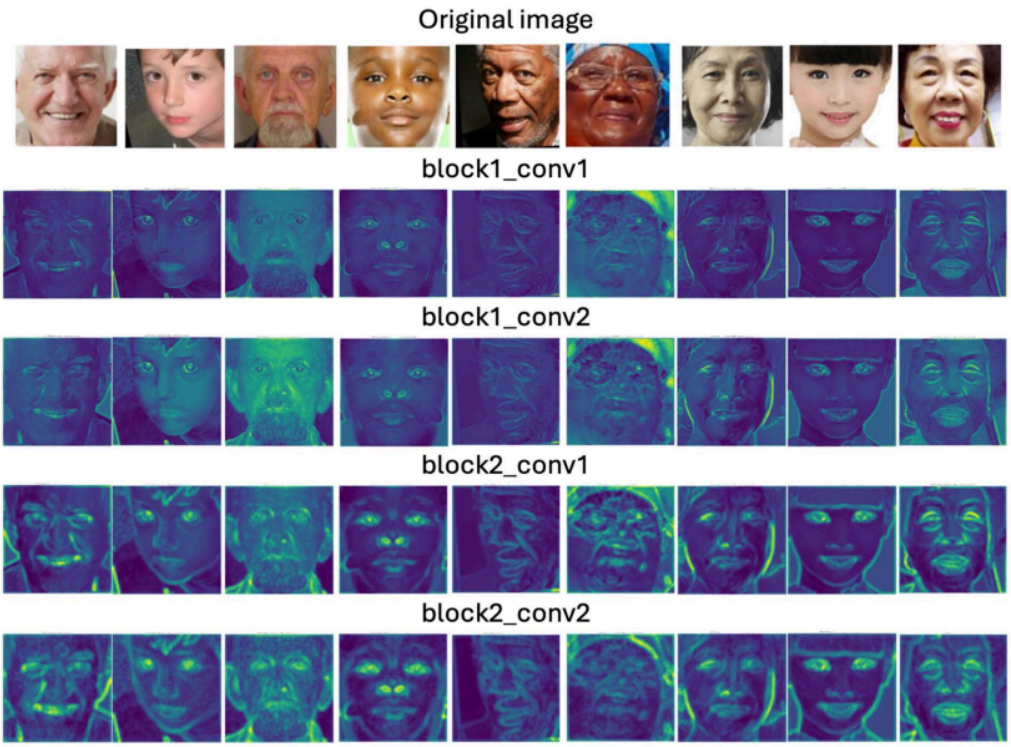
Smanjenje veličine uzorka Azijske grupe za 90% dovelo je do najmanje standardne devijacije (0.30) i ukupnog MAE od 5.48. Ovo smanjenje je poboljšalo doslednost performansi u poređenju sa originalnim skupom podataka. S druge strane, smanjenje uzoraka drugih grupa nije uvek dovelo do boljih rezultata, što naglašava složenost postizanja izjednačenih performansi među svim demografijama.

APPA-REAL skup podataka, s druge strane, pokazao je različite trendove. Originalni sastav skupa podataka doveo je do ukupne MAE od 6.45 i standardne devijacije od 0.20. Smanjenje Bele grupe na 20% postiglo je najbalansiranije performanse, sa standardnom devijacijom od 0.04 i ukupnom MAE od 7.16. Ovo pokazuje da smanjenje zastupljenosti većinske grupe može poboljšati pravičnost među različitim etničkim grupama. Model sa najlošijim performansama izostavio je u potpunosti belu grupu, što je dovelo do ukupnog MAE od 8.89 i standardne devijacije od 0.94, što pokazuje da potpuno isključivanje dominantne grupe negativno utiče na performanse modela.

Ovi nalazi ističu sposobnost našeg pristupa da sistematski otkrije i adresira pristrasnosti u modelima za procenu starosti manipulacijom sastava skupa podataka. Naša metoda pruža jasan okvir za analizu i poboljšanje pravičnosti modela.

6.4. Analiza mapa karakteristika i razlike u ponašanju modela

Bili smo zaintrigirani različitim performansama Azijske grupe unutar UTK Face skupa podataka u poređenju sa drugim grupama, koja je dosledno pokazivala manju MAE za otprilike 2 poena. Da bismo ovo dodatno istražili, sproveli smo analizu koristeći naše obučene modele, ocenjujući primerke slika svih grupa unutar oba skupa podataka, UTK i APPA REAL. Naši nalazi, prikazani na Slici 26, otkrivaju interesantno zapažanje.



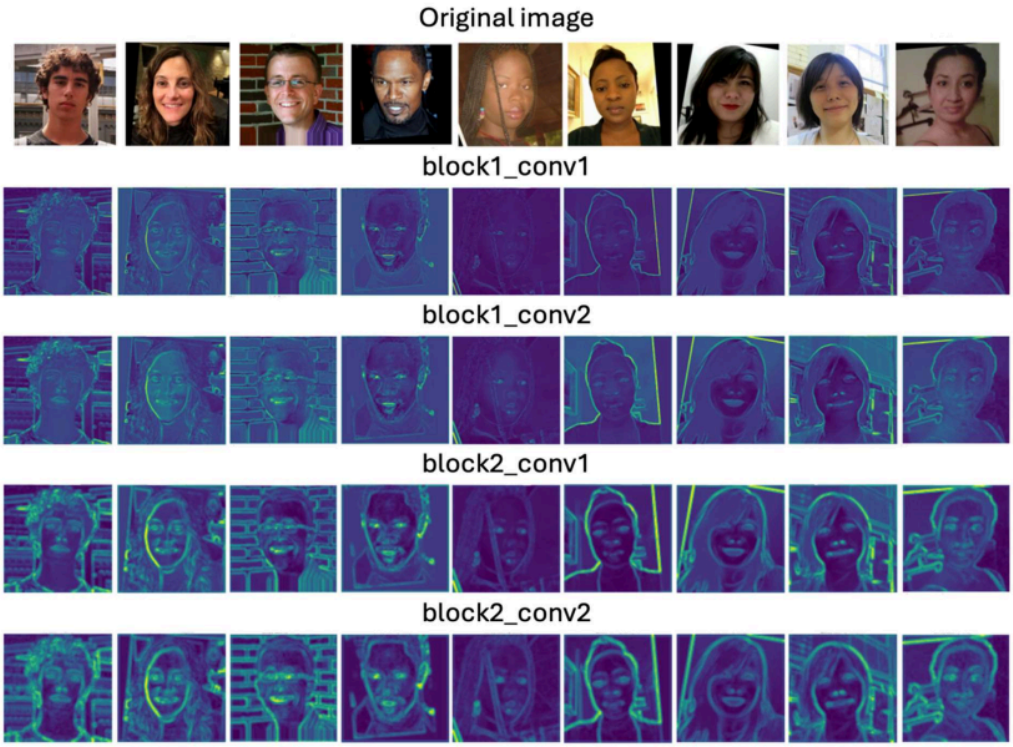
Slika 26. Primeri slika iz UTKFace skupa podataka i njihove mape karakteristika kroz slojeve.

Prilikom ispitivanja primera iz UTKFace skupa podataka, model pokazuje preferenciju prema određenim facijalnim karakteristikama kod različitih grupa. Za belu i crnu grupu primećuje se izražen fokus na oči, nos i usta, sa jačim aktivacijama u tim regijama. Pored toga, određeni naglasak je stavljen na obrubljivanje ivica lica. Nasuprot tome, aktivacije za azijsku grupu uglavnom su koncentrisane na ivice lica, posebno oko oblika obraza, uz dodatni fokus na usta, nos i oči.

Iako model analizira slične facijalne karakteristike kod svih grupa, intenzitet i raspodela ovih aktivacija se razlikuju. Ova razlika u obrascima aktivacija može pomoći u objašnjenju zašto azijska grupa pokazuje drugačije performanse u poređenju sa belom i crnom grupom unutar UTK skupa podataka. Nasuprot tome, performanse bele i crne grupe deluju usklađenije kroz različite testove.

Sada, hajde da istražimo zašto ne primećujemo isti odnos u rezultatima APPA REAL skupa podataka. Ovo se može objasniti ispitivanjem mapa karakteristika. Kao što je prikazano na slici 27, mape karakteristika APPA REAL skupa podataka otkrivaju da model pokazuje dosledan obrazac aktivacija kod svih etničkih grupa. Značajne aktivacije

su primećene unutar lica, posebno ističući obraze, osmeh, oči i nos, zajedno sa nekim detaljima pozadine. Ovaj uniformni obrazac ekstrakcije karakteristika među različitim etničkim grupama verovatno doprinosi relativno ujednačenim rezultatima performansi koji su primećeni kod ovih grupa u APPA REAL skupu podataka.



Slika 27. Primeri slika iz APPA-REAL skupa podataka i njihove mape karakteristika kroz slojeve.

6.5. Tumačenje i složenost pravičnih performansi

Ova razlika u detekciji karakteristika i performansama modela može se pripisati nekoliko faktora. Jedna od mogućnosti je da prisustvo vidljivije pozadine u slikama APPA REAL skupa podataka vodi model ka drugačijem pristupu u prepoznavanju slika, što dovodi do njihovog ujednačenijeg tretmana. Analizom mapa karakteristika stičemo dragocene uvide u način na koji model obrađuje slike iz različitih skupova podataka i zašto mogu postojati varijacije u performansama među različitim etničkim grupama.

Analiza mapa karakteristika pružila je dodatne uvide u ponašanje modela. Za UTKFace skup podataka, primećeni su specifični obrasci aktivacija za azijsku grupu u poređenju sa belom i crnom grupom, što može objasniti razlike u performansama. Ovo

otkriće se poklapa sa radom Abdolrashidi i saradnika (2020) [15], koji su istakli izazove u predikciji godina zbog varijacija unutar iste klase. Nasuprot tome, APPA-REAL skup podataka pokazao je konzistentne obrasce aktivacija kod svih etničkih grupa, što doprinosi uravnoteženijim performansama primećenim u ovom skupu podataka.

Ovi rezultati ističu složenost postizanja jednakih performansi među različitim demografskim grupama. Jasno je da postizanje jednakih performansi među grupama ili klasama predstavlja veoma složen problem bez jednostavnog rešenja. Prekomerno uzorkovanje skupa podataka kako bi distribucije bile ravnomerne ne bi trebalo da bude jedini korak; to je samo uvod u detaljniju obradu koja mora biti sprovedena. Iako smo primetili određene zajedničke obrasce između dva skupa podataka i način na koji performanse variraju među grupama u različitim scenarijima, reakcije nisu identične. Jedan sastav skupa podataka koji funkcioniše za jedan skup možda neće funkcionisati za drugi zbog različitih faktora, poput osvetljenja i kontrasta.

6.6. Strategije za poboljšanje pravičnosti

Naši nalazi sugerišu da uravnotežen pristup, koji uključuje kombinaciju poduzorkovanja većinske klase i preuzorkovanja manjinskih grupa, može biti efikasniji u smanjenju pristrasnosti nego oslanjanje isključivo na preuzorkovanje. Manjinske klase u našim skupovima podataka imaju značajno manje uzoraka u poređenju sa većinskom grupom, što zahteva preuzorkovanje radi uravnoteženja skupa podataka. Ova razlika predstavlja izazove, i uprkos korišćenju augmentacije podataka za smanjenje prenaučavanja, veći skupovi podataka sa jednakim veličinama grupa od samog početka bili bi mnogo korisniji. To bi eliminisalo potrebu za generisanjem sintetičkih podataka, čime bi se izbeglo unošenje potencijalnih šumova i poboljšala pouzdanost modela. Iako smo pokazali da jednako uzorkovani skupovi podataka ne dovode automatski do jednakih performansi, rad sa jedinstvenim uzorcima iz uravnoteženog skupa podataka poboljšao bi i pravednost i pouzdanost modela.

6.7. Ograničenja i moguća poboljšanja

Nedostatak naše metode, kao i nekih drugih predloženih metoda, jeste to što je računski zahtevna. Zahteva opsežno testiranje različitih kombinacija. Moguće poboljšanje moglo bi uključivati upotrebu reprezentativnog uzorka skupa podataka za

pronalaženje odgovarajuće kombinacije odnosa grupa. Iako se u ovoj fazi ne može očekivati visoka tačnost, odnos između grupa i uticaj prilagođavanja skupa podataka na ravnopravnost među njima trebalo bi da budu vidljivi, dok se vreme testiranja smanjuje zbog manje količine podataka. Nakon toga, moglo bi se nastaviti treniranje sa ansambl-modelima na prilagođenim odnosima grupa, ali sada koristeći ceo skup podataka.

6.8. Poređenje sa prethodnim istraživanjima

Naši nalazi su u skladu sa prethodnim istraživanjima, poput obimnog pregleda koji su sproveli Hasib i saradnici (2020) [19], koji su se bavili neuravnoteženošću klasa u skupovima podataka. Oni su metodologije klasifikovali u metode na nivou podataka, metode na nivou algoritma, metode ansambla i hibridne metode. Tehnike poput SMOTE i ADASYN, koje generišu sintetičke podatke za uravnotežavanje distribucije klasa, kao i tehnike poduzorkovanja poput RUS i T-Link, istaknute su zbog svoje efikasnosti. Međutim, povećani troškovi obrade i potencijalni gubitak informacija tokom poduzorkovanja identifikovani su kao značajni izazovi. Naš pristup, koji sistematski smanjuje uzorke iz previše zastupljenih grupa i posmatra uticaj na performanse modela, otkrio je da, iako rebalansiranje skupa podataka može smanjiti standardne devijacije i poboljšati konzistentnost performansi među etničkim grupama, to ne garantuje uvek bolje ukupne performanse. Ovaj zaključak je u skladu sa zaključcima Ramyachitre i Manikandana (2014) [21], koji su primetili da, iako su tehnike poput SMOTE-a i učenja osetljivog na troškove efikasne, izbor metode treba prilagoditi specifičnim karakteristikama skupa podataka.

Slično tome, Rahman i Davis (2013) [20] su istraživali performanse tehnika preuzorkovanja i poduzorkovanja radi balansiranja podataka o kardiovaskularnim bolestima. Njihovi nalazi su istakli da, iako je SMOTE pokazao dobre rezultate u klasifikaciji, modifikovana metoda poduzorkovanja zasnovana na klasterima nadmašila je tradicionalne metode. Ovo naglašava efikasnost hibridnih metoda koje integrišu uzorkovanje podataka i poboljšanje algoritama kako bi se rešila neuravnoteženost klasa. Naši nalazi podržavaju ovaj pristup, pokazujući da kombinacija tehnika može obezbediti uravnoteženije performanse među različitim etničkim grupama.

Kotsiantis, Kanellopoulos i Pintelas (2006) [22] su takođe naglasili potrebu za prilagođenim pristupom zasnovanim na karakteristikama skupa podataka. Oni su diskutovali o korišćenju metoda ansambla, kao što su boosting i bagging, za poboljšanje rezultata klasifikacije kombinovanjem više modela. Naša studija podržava ovaj pristup, pokazujući da sistematski pristup prilagođavanju zastupljenosti grupa može dopuniti postojeće tehnike i ponuditi nove perspektive u postizanju pravednosti.

6.9. Preporuke za praktične sisteme i buduća istraživanja

Za primenu nalaza našeg eksperimenta na praktične sisteme za prepoznavanje lica, može se preduzeti nekoliko ključnih koraka. Prvo, naš pristup balansiranju skupova podataka kroz kombinaciju preuzorkovanja manjinskih grupa i poduzorkovanja većinske klase treba implementirati. Ovo obezbeđuje pravičnije performanse među različitim etničkim grupama. Pored toga, uvide stečene analizom mapa karakteristika, koja je identifikovala varijacije u obrascima aktivacija među etničkim grupama, treba iskoristiti za fino podešavanje arhitektura modela i poboljšanje njihove osetljivosti na raznovrsne facijalne karakteristike.

Aplikacije iz stvarnog sveta treba da daju prioritet kreiranju i korišćenju većih i uravnoteženijih skupova podataka od samog početka kako bi se izbegla potreba za sintetičkim podacima i smanjili potencijalni šumovi. Ovo će poboljšati pouzdanost i pravičnost modela. Takođe, korišćenje reprezentativnih uzoraka skupova podataka za određivanje optimalnih odnosa grupa može pomoći u pojednostavljivanju procesa, smanjujući troškove obrade i vreme testiranja. Kombinovanje ovih prilagođavanja na nivou podataka sa naprednim algoritamskim tehnikama, poput metoda ansambla, može dodatno poboljšati performanse modela i osigurati pravičniju primenu tehnologije prepoznavanja lica.

Moguća poboljšanja uključuju kombinovanje naše tehnike rebalansiranja skupa podataka sa naprednim metodama za smanjenje pristrasnosti, kao što su protivnička obuka (adversarial training) ili ograničenja za pravičnost (fairness constraints). Pored toga, istraživanje raznovrsnijih i sveobuhvatnijih skupova podataka moglo bi dodatno unaprediti generalizaciju i pravičnost modela. Buduća istraživanja takođe treba da

razmotre integraciju socio-demografskih faktora kako bi razvili nijansiranije modele koji bolje obuhvataju složenost ljudskih lica.

6.10. Završna razmatranja

Zaključno, naša studija pokazuje da, iako jednostavne tehnike preuzorkovanja i poduzorkovanja mogu do određene mere smanjiti pristrasnost, postizanje istinske pravičnosti zahteva kombinaciju preciznih prilagođavanja skupa podataka i sofisticiranih algoritamskih metoda. Sistematskom analizom uticaja različitih kompozicija skupova podataka pružamo dragocene uvide za razvoj pravičnijih i tačnijih modela za prepoznavanje lica.

Literatura

[1] Terhörst, P.; Kolf, J.N.; Huber, M.; Kirchbuchner, F.; Damer, N.; Moreno, A.M.; Fierrez, J.; Kuijper, A. A Comprehensive Study on Face Recognition Biases Beyond Demographics. *IEEE Trans. Technol. Soc.* **2022**, *3*, 16–30. [CrossRef]

[2] Voigt, P.; Bussche, A.V.D. *The EU General Data Protection Regulation (GDPR): A Practical Guide*, 1st ed.; Springer: Cham, Switzerland, 2017; pp. 141–187.

[3] Albert, A.M.; Ricanek, K.; Patterson, E. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Sci. Int.* **2007**, *172*, 1–9. [CrossRef] [PubMed]

[4] Michalski, D.; Yiu, S.Y.; Malec, C. The Impact of Age and Threshold Variation on Facial Recognition Algorithm Performance Using Images of Children. In Proceedings of the International Conference on Biometrics (ICB), Gold Coast, Australia, 20–23 February 2018; pp. 217–224.

[5] Srinivas, N.; Ricanek, K.; Michalski, D.; Bolme, D.S.; King, M. Face recognition algorithm bias: Performance differences on images of children and adults. In Proceedings of the EEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–20 June 2019.

[6] Albiero, V.; Bowyer, K.W. Is face recognition sexist? no, gendered hairstyles and biology are. *arXiv* **2020**, arXiv:2008.06989.

[7] Albiero, V.; Zhang, K.; Bowyer, K.W. How does gender balance in training data affect face recognition accuracy. In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), Houston, TX, USA, 28 September–1 October 2020; pp. 1–10.

[8] Andraz, P.; Vitomir, S.; Klemen, G. Analysis of Race and Gender Bias in Deep Age Estimation Models. In Proceedings of the 8th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 18–22 January 2021.

[9] Kimmo, K.; Jungseock, J. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age. *arXiv* **2019**, arXiv:1908.04913.

[10] Clapes, A.; Bilici, O.; Temirova, D.; Avots, E.; Anbarjafari, G.; Escalera, S. From Apparent to Real Age: Gender, Age, Ethnic, Makeup, and Expression Bias Analysis in Real Age Estimation. In Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2373–2382.

[11] Jacques, J.C.S.; Ozcinar, C.; Marjanovic, M.; Baró, X.; Anbarjafari, G.; Escalera, S. On the effect of age perception biases for real age regression. In Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition, Lille, France, 14–18 May 2019; pp. 1–8.

[12] Angulu, R.; Tapamo, J.R.; Adewumi, A.O. Age estimation via face images: A survey. *J. Image Video Proc.* **2018**, *2018*, 42. [CrossRef]

[13] Khaled, E.K.; Valliappan, R.; Patrick, T. Facial Age Estimation Using Machine Learning Techniques: An Overview. *Big Data Cogn. Comput.* **2022**, *6*, 128.

[14] Age Detection Using Facial Images: Traditional Machine Learning vs. Deep Learning, towardsdatascience.com. Dostupno online: https://towardsdatascience.com/age-detection-using-facial-images-traditional-machine-learning-vs-deep-learning2437b2feeab2 (Datum pristupa 18 Jun 2024).

[15] Abdolrashidi, A.; Minaei, M.; Azimi, E. Age and Gender Prediction From Face Images Using Attentional Convolutional Network. *arXiv* **2020**, arXiv:2010.03791.

[16] Sathyavathi, S.; Baskaran, K.R. An Intelligent Human Age Prediction from Face Image Framework Based on Deep Learning Algorithms. *Inf. Technol. Control* **2023**, *52*, 245–257. [CrossRef]

[17] Amelia, J.S.; Wahyono. Age Estimation on Human Face Image Using Support Vector Regression and Text-Based Features. *Int. J. Adv. Comput. Sci. Appl.* **2022**. [CrossRef]

[18] Xing, J.; Li, K.; Hu, W.; Yuan, C.; Ling, H. Diagnosing deep learning models for high accuracy age estimation from a single image. *Pattern Recognit.* **2017**, *66*, 106–116. [CrossRef]

[19] Khan, H.; Iqbal, S.; Shah, F.M.; Mahmud, J.A.; Popel, M.H.; Showrow, I.H.; Ahmed, S.; Rahman, O. A Survey of Methods for Managing the Classification and Solution of Data Imbalance Problem. *arXiv* **2020**, arXiv:2012.11870. [CrossRef]

[20] Rahman, M.M.; Davis, D.N. Addressing the class imbalance problem in medical datasets. *Int. J. Mach. Learn. Comput.* **2013**, *3*, 224. [CrossRef]

[21] Ramyachitra, D.; Manikandan, P. Imbalanced dataset classification and solutions: A review. *Int. J. Comput. Bus. Res.* **2014**, *5*, 1–29.

[22] Kotsiantis, S.; Kanellopoulos, D.; Pintelas, P. Handling imbalanced datasets: A review. *GESTS Int. Trans. Comput. Sci. Eng.* **2006**, *30*, 25–36.

[23] IMDB-WIKI, data.vision.ee.ethz.ch. Dostupno online: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki (Datum pristupa 26. Avgust 2025.)

[24] FGNET dataset, datasets.activeloop.ai. Dostupno online: https://datasets.activeloop.ai/docs/ml/datasets/fgnet-dataset/ (Datum pristupa 26. Avgust 2025.)

[25] Cross-Age Celebrity Dataset, bcsiriuschen.github.io. Dostupno online: https://bcsiriuschen.github.io/CARC/ (Datum pristupa 26. Avgust 2025.)

[26] MORPH, people.uncw.edu. Dostupno online: people.uncw.edu/vetterr/morph-noncommercial-stats.pdf (Datum pristupa 26. Avgust 2025.)

[27] Adience, talhassnet.github.io. Dostupno online: https://talhassner.github.io/home/projects/Adience/Adience-data.html (Datum pristupa 26. Avgust 2025.)

[28] AgeDB, openaccess.thecvf.com. Dostupno online: openaccess.thecvf.com/content_cvpr_2017_workshops/w33/papers/Moschoglou_AgeDB_The_First_CVPR_2017_paper.pdf (Datum pristupa 26. Avgust 2025.)

[29] MegaAge, mmlab.ie.cuhk.edu.hk. Dostupno online: https://mmlab.ie.cuhk.edu.hk/projects/MegaAge/ (Datum pristupa 26. Avgust 2025.)

[30] UTKFace, github.io. Dostupno online: https://susanqq.github.io/UTKFace/ (Datum pristupa 19 Jun 2024).

[31] APPA-REAL, chalearnlap.cvc.uab.cat. Dostupno online: https://chalearnlap.cvc.uab.cat/dataset/26/description/ (Datum pristupa 19 Jun 2024).

[32] Victoria, Z. Data Challenges in Building a Facial Recognition Model and How to Mitigate Them. SOA Research Institute, 2023.

[33] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [CrossRef]

[34] Kaiming, H.; Xiangyu, Z.; Shaoqing, R.; Jian, S. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385. [CrossRef]

[35] Howard, A.G.; Zho, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Hartwig, A. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]

[36] ImageNet. Dostupno online: https://www.image-net.org (Datum pristupa 19 Jun 2024).

Spisak slika

Slika 1. Primer slika iz IMDB-WIKI skupa podataka 24

Slika 2. Prikaz distribucije godina uzoraka iz IMDB-WIKI skupa podataka 24

Slika 3. Primer slika iz FG-NET skupa podataka 25

Slika 4. Primer slika iz CACD skupa podataka 26

Slika 5. Primer slika iz MORPH II skupa podataka 27

Slika 6. Primer slika iz Adience skupa podataka 28

Slika 7. Primer slika iz AgeDB skupa podataka 29

Slika 8. Primer slika iz MegaAge-Asian skupa podataka 30

Slika 9. Primer slika iz FairFace skupa podataka 31

Slika 10. Poređenje demografske uravnoteženosti FairFace skupa sa ostalim skupovima podataka 31

Slika 11. Primeri slika iz UTKFace skupa podataka 36

Slika 12. Distribucija uzoraka UTKFace skupa podataka po grupi godina 37

Slika 13. Vizuelna reprezentacija podele uzoraka prema etničkoj pripadnosti unutar UTKFace skupa podataka 38

Slika 14. Primeri slika iz APPA-REAL skupa podataka ..... **Error! Bookmark not defined.**

Slika 15. Distribucija uzoraka APPA-REAL skupa podataka po grupi godina ..... **Error! Bookmark not defined.**

Slika 16. Vizuelna reprezentacija podele uzoraka prema etničkoj pripadnosti unutar APPA-REAL skupa podataka..... **Error! Bookmark not defined.**

Slika 17. Reprezentacija arhitekture VGG-19 modela korišćenog u ovom radu..... **Error! Bookmark not defined.**

Slika 18. Ukupni MAE i standardna devijacija prema varijacijama APPA-REAL skupa podataka 60

Slika 19. MAE i standardna devijacija prema varijacijama APPA-REAL skupa podataka.61

Slika 20. Nejednaki uticaj kroz varijacije APPA-REAL skupa podataka**Error! Bookmark not defined.**

Slika 21. Jednakost mogućnosti prema varijacijama APPA-REAL skupa podataka 63

Slika 22. Ukupni MAE i standardna devijacija za varijacije skupa podataka UTK Face... 71

Slika 23. MAE i standardna devijacija prema varijacijama UTK Face skupa podataka... 72

Slika 24. Nejednaki uticaj prema varijacijama UTK Face skupa podataka. 73

Slika 25. Jednakost mogućnosti prema varijacijama UTK Face skupa podataka 74

Slika 26. Primeri slika iz UTKFace skupa podataka i njihove mape karakteristika kroz slojeve..... 81

Slika 27. Primeri slika iz APPA-REAL skupa podataka i njihove mape karakteristika kroz slojeve..... 82

Spisak tabela

No table of figures entries found.

METOD ZA SMANJENJE DEMOGRAFSKE PRISTRASNOSTI U MODELIMA PROCENE STAROSTI PUTEM OPTIMIZOVANE SKTRUKTURE SKUPOVA PODATAKA

ORIGINALITY REPORT

2%
SIMILARITY INDEX

PRIMARY SOURCES

1	discovery.researcher.life Internet	210 words — 1%
2	nardus.mpn.gov.rs Internet	76 words — < 1%
3	www.researchgate.net Internet	50 words — < 1%
4	doaj.org Internet	14 words — < 1%
5	www.hajdeda.org.rs Internet	12 words — < 1%
6	www.mdpi.com Internet	11 words — < 1%
7	www.cris.uns.ac.rs Internet	10 words — < 1%
8	Nithyashri, J., and G. Kulanthaivel. "Classification of human age based on Neural Network using FG-NET Aging database and Wavelets", 2012 Fourth International Conference on Advanced Computing (ICoAC), 2012. Crossref	9 words — < 1%

9	singipedia.singidunum.ac.rs Internet	9 words — < 1%
10	elibrary.matf.bg.ac.rs Internet	8 words — < 1%
11	mojauprava.hr Internet	8 words — < 1%
12	pdfs.semanticscholar.org Internet	8 words — < 1%
13	www.itnetwork.rs Internet	8 words — < 1%
14	www.scribd.com Internet	8 words — < 1%
15	Andrija Popović, Andreja Todorović, Vladan Vučić. "Assessing the influence of economic growth on competitiveness and innovation within the circular economy", Ekonomski horizonti, 2025 Crossref	7 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF