

# Ocjena kvalitete video signala zasnovana na strojnom učenju

---

Ćosić, Tomislav

Master's thesis / Diplomski rad

2023

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:200:170956>

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

*Download date / Datum preuzimanja:* **2025-03-28**

*Repository / Repozitorij:*

[Faculty of Electrical Engineering, Computer Science and Information Technology Osijek](#)



**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU  
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I  
INFORMACIJSKIH TEHNOLOGIJA OSIJEK**

**Sveučilišni studij**

**Ocjena kvalitete video signala zasnovana na strojnom  
učenju**

**Diplomski rad**

**Tomislav Ćosić**

**Osijek, 2023.**

**FERIT**FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA  
I INFORMACIJSKIH TEHNOLOGIJA **OSIJEK****Obrazac D1: Obrazac za imenovanje Povjerenstva za diplomski ispit**

Osijek, 11.09.2023.

Odboru za završne i diplomske ispite

**Imenovanje Povjerenstva za diplomski ispit**

<b>Ime i prezime Pristupnika:</b>	Tomislav Ćosić
<b>Studij, smjer:</b>	Diplomski sveučilišni studij Elektrotehnika, smjer Komunikacije i informatika
<b>Mat. br. Pristupnika, godina upisa:</b>	D-1363, 07.10.2021.
<b>OIB studenta:</b>	81647902895
<b>Mentor:</b>	prof. dr. sc. Mario Vranješ
<b>Sumentor:</b>	,
<b>Sumentor iz tvrtke:</b>	Zvonimir Kaprocki
<b>Predsjednik Povjerenstva:</b>	prof. dr. sc. Marijan Herceg
<b>Član Povjerenstva 1:</b>	prof. dr. sc. Mario Vranješ
<b>Član Povjerenstva 2:</b>	izv. prof. dr. sc. Ratko Grbić
<b>Naslov diplomskog rada:</b>	Ocjena kvalitete video signala zasnovana na strojnom učenju
<b>Znanstvena grana diplomskog rada:</b>	<b>Telekomunikacije i informatika (zn. polje elektrotehnika)</b>
<b>Zadatak diplomskog rada:</b>	Praćenje kvalitete video signala od iznimne je važnosti za brojne operatere koji se bave uslugama zasnovanim na isporuci videa krajnjim korisnicima. Operaterima je stoga bitno u svakom trenutku imati informaciju o kvaliteti videa isporučenog svakom korisniku, kako bi bio upoznat s time što treba činiti da bi mu osigurao minimalno garantiranu kvalitetu usluge. U sklopu ovog rada potrebno je razviti metodu za ocjene kvalitete video signala zasnovanu na odabranom modelu strojnog učenja, gdje bi ulaz u model bili okviri videa čija se kvaliteta želi mjeriti, a izlaz iz mreže bi bila ocjena kvalitete tog videa. Razvijeni model je na kraju potrebno vrednovati na definiranim javno dostupnim skupovima podataka i usporediti njegove performanse s najboljim javno
<b>Prijedlog ocjene pismenog dijela ispita (diplomskog rada):</b>	Izvrstan (5)
<b>Kratko obrazloženje ocjene prema Kriterijima za ocjenjivanje završnih i diplomskih radova:</b>	Primjena znanja stečenih na fakultetu: 3 bod/boda Postignuti rezultati u odnosu na složenost zadatka: 3 bod/boda Jasnoća pismenog izražavanja: 2 bod/boda Razina samostalnosti: 2 razina
<b>Datum prijedloga ocjene od strane mentora:</b>	11.09.2023.
Potvrda mentora o predaji konačne verzije rada:	<i>Mentor elektronički potpisao predaju konačne verzije.</i>
	Datum:

**FERIT**FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA  
I INFORMACIJSKIH TEHNOLOGIJA OSIJEK**IZJAVA O ORIGINALNOSTI RADA**

Osijek, 26.09.2023.

**Ime i prezime studenta:**

Tomislav Ćosić

**Studij:**

Diplomski sveučilišni studij Elektrotehnika, smjer Komunikacije i informatika'

**Mat. br. studenta, godina upisa:**

D-1363, 07.10.2021.

**Turnitin podudaranje [%]:**

8

Ovom izjavom izjavljujem da je rad pod nazivom: **Ocjena kvalitete video signala zasnovana na strojnom učenju**

izrađen pod vodstvom mentora prof. dr. sc. Mario Vranješ

i sumentora ,

moj vlastiti rad i prema mom najboljem znanju ne sadrži prethodno objavljene ili neobjavljene pisane materijale drugih osoba, osim onih koji su izričito priznati navođenjem literature i drugih izvora informacija. Izjavljujem da je intelektualni sadržaj navedenog rada proizvod mog vlastitog rada, osim u onom dijelu za koji mi je bila potrebna pomoć mentora, sumentora i drugih osoba, a što je izričito navedeno u radu.

Potpis studenta:

# SADRŽAJ

<b>1. UVOD .....</b>	<b>1</b>
<b>2. PROBLEM OCJENJIVANJA KVALITETE VIDEO SIGNALA .....</b>	<b>2</b>
<b>2.1. Objektivne metode za ocjenu kvalitete video signala .....</b>	<b>4</b>
<b>2.2. Objektivne metrike za ocjenjivanje video signala zasnovane na egzaktnim računalnim algoritmima .....</b>	<b>6</b>
2.2.1. PSNR .....	6
2.2.1. SSIM.....	6
2.2.2. MS-SSIM.....	8
2.2.3. NIQE .....	9
2.2.4. VIIDEO .....	10
<b>2.3. Objektivne metrike za ocjenjivanje video signala zasnovane na algoritmima strojnog učenja .....</b>	<b>11</b>
2.3.1. BRISQUE .....	11
2.3.2. Video BLIINDS.....	13
2.3.3. VSFA.....	14
<b>3. RAZVOJ VLASTITOG ALGORITMA ZA OCJENJIVANJE VIDEO SIGNALA ZASNOVANOG NA STROJNOM UČENJU.....</b>	<b>16</b>
<b>3.1. Opis korištenih alata.....</b>	<b>16</b>
<b>3.2. Opis baza podataka za treniranje modela .....</b>	<b>17</b>
<b>3.3. Treniranje algoritma zasnovanog na strojnom učenju .....</b>	<b>19</b>
3.3.1. Priprema podataka .....	19
3.3.2. Model za izvlačenje značajki.....	19
3.3.3. Arhitektura odabranih dubokih neuronskih mreža za ocjenjivanje video signala.....	20
3.3.4. Upute za pokretanje .....	22
3.3.5. Proces treniranja predloženih dubokih neuronskih mreža za ocjenu kvalitete video signala .....	23
<b>4. EVALUACIJA RADA PREDLOŽENIH MODELA ZA OCJENJIVANJE VIDEO SIGNALA ZASNOVANOG NA STROJNOM UČENJU .....</b>	<b>49</b>
<b>4.1. Evaluacija algoritma uz prvi način podjele baze podataka .....</b>	<b>50</b>
<b>4.2. Evaluacija algoritma uz drugi način podjele baze podataka .....</b>	<b>62</b>
<b>4.3. Analiza rezultata .....</b>	<b>74</b>
<b>5. ZAKLJUČAK.....</b>	<b>76</b>

<b>LITERATURA .....</b>	<b>78</b>
<b>SAŽETAK.....</b>	<b>80</b>
<b>ABSTRACT .....</b>	<b>81</b>
<b>ŽIVOTOPIS.....</b>	<b>82</b>
<b>PRILOZI.....</b>	<b>83</b>

# 1. UVOD

Procjena kvalitete videa ključno je polje multimedijske tehnologije koje se fokusira na procjenu i kvantificiranje opažene kvalitete videa. Budući da konzumacija video sadržaja i dalje brzo raste na različitim platformama i uređajima, osiguravanje isporuke kvalitetnog video sadržaja korisnicima postalo je najvažnije. Tehnike procjene kvalitete videa daju vrijedan uvid u opaženu kvalitetu videa, pomažući u optimizaciji kodiranja videa, prijenosa i sustava reprodukcije. Subjektivna procjena kvalitete videa, koja se oslanja na ljudsku prosudbu, najbolji je način za procjenu kvalitete videa. Međutim, subjektivne procjene su dugotrajne, skupe i nepraktične za velike studije. Stoga su objektivne metode procjene kvalitete videa posljednjih godina dobile značajnu pozornost i važnost. Objektivna procjena kvalitete videozapisa ima za cilj automatsku procjenu kvalitete videozapisa analizom različitih vizualnih značajki, kao što su razlučivost, šum, artefakti kompresije i vremenska glatkoća. Ove metode koriste sofisticirane algoritme i matematičke modele za kvantificiranje percipirane kvalitete na temelju analize ovih značajki. Danas je poseban fokus na ocjenjivanju video signala kod kojih ne postoji referentni video signal, jer je većina sadržaja koji se gleda takvog tipa.

Zadatak ovog diplomskog rada je napraviti algoritam zasnovan na strojnom učenju koji ocjenjuje video signale različitog tipa izobličenja i različite razine izobličenja. Potrebno je dobivene rezultate usporediti s rezultatima drugih objektivnih metrika zasnovanim na egzaktnim računalnim algoritmima. U drugom poglavlju objašnjena je problematika ocjenjivanja video signala, što je to subjektivno ocjenjivanje video signala, na koji način se dobivaju ocjene i koje vrste testova postoje. Nadalje, opisano je što je to objektivno ocjenjivanje video signala i podjela metoda za objektivno ocjenjivanje video signala. Osim toga, dan je detaljniji opis nekoliko metrika za ocjenjivanje video signala zasnovanih na egzaktnim računalnim algoritmima i nekih postojećih rješenja za ocjenjivanje video signala zasnovanih na strojnom učenju. Treće poglavlje odnosi se na opis algoritama predloženih u ovom radu, a zasnovanih na modelima strojnog učenja. Predložena su dva modela. Prvi model zasnovan je na sloju pažnje (engl. *Attention*), dok je drugi model zasnovan na dvjema LSTM (engl. *Long short-term memory*) neuronskim mrežama. Osim toga, opisani su korišteni alati i baze video signala koje su korištene u svrhu treniranja, validacije i testiranja predloženih modela i načini provođenja treniranja. Četvrto poglavlje odnosi se na načine na koje se provodilo testiranje, prikazivanje rezultata testiranja i osvrt na prikazane rezultate. Zadnje poglavlje donosi zaključak na temelju cjelokupnog rada.

## 2. PROBLEM OCJENJIVANJA KVALITETE VIDEO SIGNALA

Zbog sve veće konzumacije video sadržaja na internetu, bilo to preko servisa za prijenos uživo ili servisa koji nude gledanje filmova ili serija, korisniku se želi osigurati što veća kvaliteta iskustva (engl. *QoE - Quality of Experience*). Da bi video servis ostvario dobar QoE, prilikom prijenosa video signala potrebno je pratiti kvalitetu usluge mrežnog prijenosa (engl. *QoS - Quality of Service*) kroz mrežne QoS parametre, kao što su kašnjenje paketa i postotak gubitka paketa [1]. Osim mrežnih parametara potrebno je pratiti i QoS parametre povezane s video signalom, uključujući kašnjenje pokretanja videa i njegovu kvalitetu, što u konačnici povećava korisnikov QoE. Kvaliteta videa može se koristiti u mjerenju izvedbe različitih komponenti prijenosnog sustava video signala, kao što su kompresija, obrada i prijenos [1]. Odgovarajuće nadziranje QoS parametara pojedinačnih komponenti prijenosnog sustava važno je za postizanje većeg QoE korisnika. Kvaliteta videa izravno utječe na raspoloženje gledatelja, primjerice zamućena slika ili pojava različitih artefakata na slici mogu znatno smetati gledatelju. Zbog toga je bitno imati pouzdanu informaciju o kvaliteti video signala, kako bi se osiguralo visoko zadovoljstvo korisnika.

Nepostojanje referentne slike ili relevantnih informacija o njoj predstavlja značajne izazove algoritmima za procjenu kvalitete videa, ograničavajući njihovu primjenjivost u scenarijima stvarnog svijeta. U praksi, procjena kvalitete često se odvija u čvorovima gdje podaci o referentnom videu nisu dostupni. Stoga, metrike koje se oslanjaju na ocjenjivanje video signala s referentnim video signalom nisu prave mjere kvalitete za ovakve slučajeve upotrebe. Takve metrike se mogu svrstati u metode kod kojih je referentni video u potpunosti dostupan (engl. *FR - Full reference*) i metode kod kojih su dostupne samo neke informacije o referentnom videu (engl. *Reduce reference*) [2]. Metode za procjenu video signala bit će dodatno objašnjenje u nastavku rada

Kvaliteta video signala može se dobiti na dva načina: subjektivnim i objektivnim ocjenjivanjem. Kod subjektivnog ocjenjivanja čovjek daje svoju ocjenu o kvaliteti videa, dok će s druge strane neki računalni ili matematički model dati objektivnu ocjenu. Kako bi objektivna ocjena bila što preciznija, računalni model mora dati procjenu kvalitete što bližu kvaliteti gledatelja. Računalni model mora se fokusirati na oštećenja slike koja kod čovjeka utječu na doživljaj kvaliteta videa, a to mogu biti: zamućenje, efekt bloka, šum i vremenske nedosljednosti. Oštećenja kod gledatelja izazivaju frustraciju i pada zadovoljstvo korisnika. Kako bi model dao ocjene što bliže ljudima, znanstvenici pripremaju baze podataka slika ili videa i angažiraju ljude kako bi dobili subjektivne ocjene za te baze, koje kasnije statistički obrađuju. Na temelju



subjektivnih ocjena provjeravaju uspješnost objektivne metrike zasnovane na računalnim modelima ili strojnom učenju. Osim što objektivno ocjenjivanje mora biti točno, isto tako i njegovo vrijeme predikcije mora biti kratko da bi algoritam bio što učinkovitiji.

Kako bi se razvio dobar algoritam, bitno je imati dovoljno veliku bazu video podataka i ocjene kvalitete gledatelja za svaki video iz baze. Jedna od metoda kako se pribavljaju subjektivne ocjene kvalitete videa je *crowdsourcing* i to na način da se izradi internetska platforma na kojoj se kreira test prema ITU-T P.910 standardu [3], prema kojem se angažiraju ljudi iz svih dijelova svijeta da bi ocijenili kvalitetu video signala, a rezultati se potom statistički obrađuju. Kod ovakvog načina prikupljanja ocjena nema potrebe za organizacijom testova u laboratoriju gdje se svakom ocjenjivaču mora osigurati jednaka oprema i jednom napravljena platforma može se koristiti za više testova što ga čini efikasnim i jeftinim. Kod *crowdsourcing*-a ocjene mogu varirati od ocjenjivača do ocjenjivača zbog toga što se ne nalaze svi u istom okruženju i uređaji na kojim se ocjenjuje su drugačiji, a to sve utječe na kvalitetu prikaza sadržaja koji se procjenjuje. Video signali se procjenjivačima mogu prikazivati jedan po jedan (engl. *single stimulus*) ili dva po dva (engl. *double stimulus*), gdje se uspoređuju međusobno. Metode koje se koriste za ocjenjivane video signala gdje video signal dolaze jedna po jedna su ACR (engl. *Absolute Category Rating*), ACR-HR (engl. *Absolute Category Rating with Hidden Reference*), SSCQE (engl. *Single Stimulus Continuous Quality Rating*) [3]. Kod ACR metode ocjenjivač gleda video signal 10 sekundi, zatim ga ocjenjuje na skali od 1 do 5, gdje 5 predstavlja odličnu kvalitetu, a 1 lošu kvalitetu i na kraju testiranja ocjene za svaki video se statistički obrađuju, te se dobije MOS (engl. *Mean Opinion Score*) ocjena. U ACR-HR metodu se osim videa s izobličenjima dodaje i referentni video signal i na temelju ocjene koju mu je ocjenjivač dao računa se DMOS (engl. *Differential Mean Opinion Score*) prema formuli (2-1). gdje je  $V(PVS)$  ocjena koju je ocjenjivač dao za izobličeni video, a  $V(REF)$  ocjena koju je ocjenjivač dao za referentni video (ne znajući pritom da ocjenjuje referentni video), a broj 5 služi kako bi se ocjene mapirale na skali od 1 do 5 [3].

$$DV(PVS) = V(PVS) - V(REF) + 5 \quad (2-1)$$

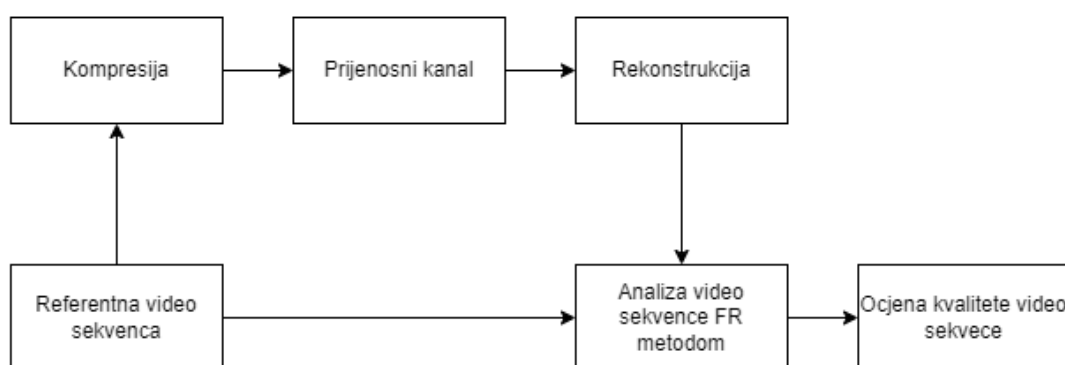
DCR (engl. *Degradation Category Rating*) jedna je od metoda gdje se video signali prikazuju u paru, tj. prvo se prikazuje referentni video, zatim izobličeni. Svaki se video prikazuje 10 sekundi. Koristi se skala od 1 do 5, gdje je 1 jako iritantno, a 5 neprimjetno i to se odnosi na oštećenje izobličenog video signala u odnosu na referentni video signal. PC (engl. *Pair Comparison*) metoda podrazumijeva da se video signal istog tipa prikazuju u parovima i to na  $n(n-1)$  načina. Ako postoje A, B, C video signali, ocjenjivaču će biti prikazani parovi AB, AC, BA, BC, CA, CB. Koristi se ista skala kao i kod DCR metode [3].

Prema [4] objektivne metode predstavljaju računalni model za procjenu kvalitete videa ili slike koje su često automatizirane, smatraju se vremenski učinkovitije i koriste se za aplikacije u stvarnom vremenu. Za objektivnu metodu ocjenjivanja videa, osnovni cilj je procijeniti kvalitetu videa što bliže odgovarajućoj ljudskoj prosudbu. Objektivne metrike kvalitete videa mogu se klasificirati prema dostupnosti referentnog video signala prilikom ocjenjivanja kvalitete, koji se smatra kao signal bez izobličenja ili savršene kvalitete te se može koristiti kao referenca za ocjenu video signala [5].

## 2.1. Objektivne metode za ocjenu kvalitete video signala

Prema dostupnosti referentnog video signala, objektivne metode se dijele na: metode kod kojih je referentni video u potpunosti dostupan (engl. FR - *Full Reference*), metode kod kojih su dostupne samo neke informacije o referentnom videu, no ne i cijeli referentni video signal (engl. RR - *Reduced Reference*) i metode kod kojih nikakve informacije o referentnom videu nisu dostupne (engl. NR - *No Reference*).

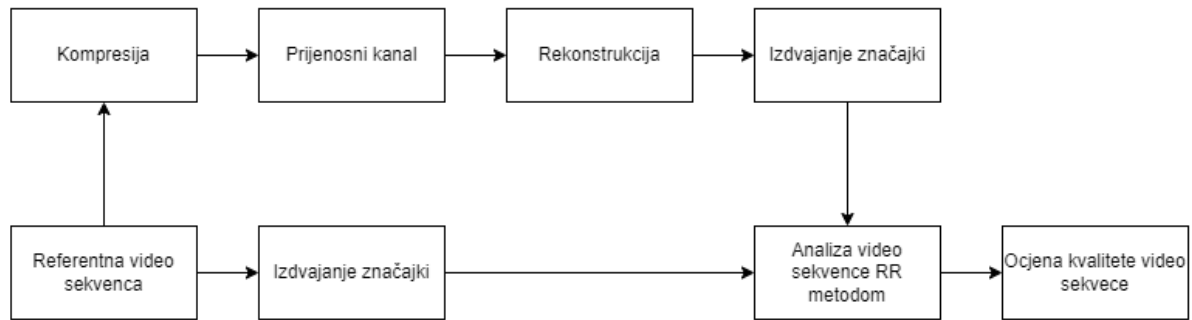
FR metode uspoređuju originalni i izobličeni video signal. One su prikladnije u izvan-mrežnoj obradi video signala, kao što je lokalno testiranje alata za kodiranje ili softvera za testiranje performansi i usporedbe. Također, FR metode se smatraju računalno složenim, jer se u većini slučajeva mora izvesti obrada na razini piksela za oba videozapisa. Aplikacije koje zahtijevaju procjenu kvalitete videa u stvarnom vremenu nemaju mogućnost pristupa referentnom video signalu i stoga se ne mogu koristiti FR metode [4]. Dijagram sustava kojim se video signal ocjenjuje FR metodom prikazan je na slici 2.1.



**Slika 2.1.** Dijagram algoritma za ocjenjivanje video signala FR metodom [6]

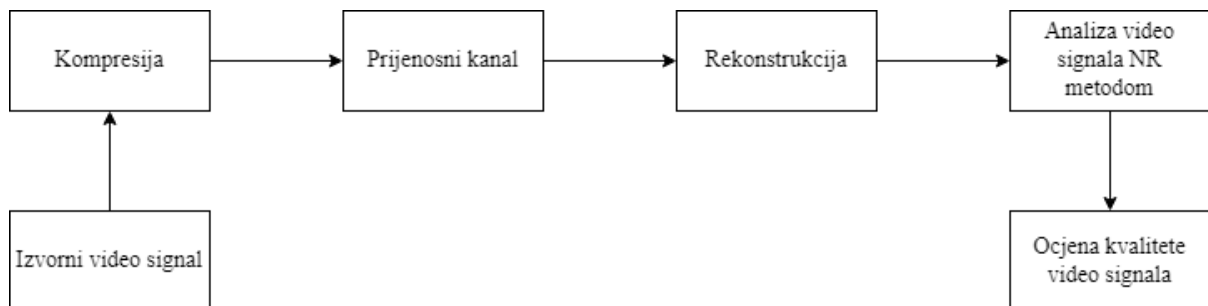
U RR metodama nije potrebno imati pristup cijelom referentnom video signalu za potrebe procjene kvalitete. RR metode obično izdvajaju određene značajke ili karakteristike iz referentnog videa,

kao što su karakteristične informacije o pikselima, koeficijentima određene transformacije ili drugim dominantnim značajkama, i uspoređuju ih s odgovarajućim značajkama izdvojenim iz izobličeneog video signala. Dijagram sustava nalazi se na slici 2.2. Razlika ili sličnost između ovih izdvojenih značajki služi kao pokazatelj degradacije kvalitete. Te su metode često dizajnirane da budu računalno učinkovite i da daju prihvatljivu aproksimaciju kvalitete videa bez potrebe za cijelim referentnim video signalom [7].



**Slika 2.2.** Dijagram algoritma za ocjenjivanje video signala RR metodom [6]

Procjena kvalitete slike ili videa bez reference doista je izazovan zadatak. U metodama koje se temelje na NR-u, procjena se provodi isključivo na temelju izdvojenih značajki iz video signala koji se ocjenjuje, jer ne postoji referentni video signal dostupna za izravnu usporedbu (slika 2.3). Procjena kvalitete slike i videa temeljena na NR metodi često se oslanja na analizu intrinzičnih svojstava i statističkih karakteristika sadržaja, kako bi se ocijenila njegova kvaliteta. Mogu se izdvojiti različite značajke, poput oštine, distribucije boja, kontrasta, teksture i vremenske koherencije. Međutim, važno je napomenuti da procjena temeljena na NR ima svojstvena ograničenja, budući da izdvojene značajke mogu pružiti samo ograničene informacije o pravoj kvaliteti. Nedostatak reference za usporedbu otežava utvrđivanje apsolutne mjere kvalitete. Osim toga, različita iskrivljenja mogu utjecati na istu značajku na različite načine, što dovodi do dvosmislenosti u procjeni kvalitete.



**Slika 2.3.** Dijagram algoritma za ocjenjivanje video signala NR metodom [6]

## 2.2. Objektivne metrike za ocjenjivanje video signala zasnovane na egzaktnim računalnim algoritmima

### 2.2.1. PSNR

PSNR (engl. *Peak signal to noise ratio*) se obično koristi u različitim aplikacijama za obradu slika i videa, kao što su kompresija slike i strujanje videa (engl. *streaming*). Omogućuje kvantitativnu mjeru kvalitete rekonstruiranog signala i osobito je koristan za usporedbu različitih algoritama kompresije ili postavki parametara. PSNR se mjeri na logaritamskoj skali i ovisi o srednjoj kvadratnoj pogrešci (engl. *MSE – Mean square error*) između izvorne i oštećene slike ili video okvira, u odnosu na  $(2^n - 1)^2$ , kvadrat najveće moguće vrijednosti signala na slici, gdje je  $n$  broj bitova po uzorku slike [8]. MSE se računa kao prosječna kvadratna razlika između odgovarajućih piksela u izvornom i rekonstruiranom signalu.

$$PSNR = 10 \log_{10} * \frac{(2^n - 1)^2}{MSE} \quad (2-2)$$

PSNR se može lako i brzo izračunati i stoga je vrlo popularna mjera kvalitete, široko se koristi za usporedbu kvalitete komprimiranih i dekomprimiranih video slika. Za danu sliku ili video, visoka vrijednost PSNR predstavlja bolju kvalitetu, dok manja predstavlja lošiju. Na slici 2.4. a) predstavljena je originalna slika, b) je zamućena slika originala s PSNR vrijednošću od 30.6 dB, a slika c) ima zamućenu pozadinu s PSNR vrijednošću od 27.7 dB. Ako se slike bolje pogledaju, može se primijetiti da slika c) ljudskom oku izgleda puno bolje u odnosu na sliku b) koja ima veći PSNR. U ovom slučaju ljudski promatrač pridaje veću važnost području lica, pa tako i je posebno osjetljiv na distorziju u ovom području [8]. Važno je napomenuti da, iako se PSNR široko koristi, ima određena ograničenja. PSNR vrlo često loše korelira sa subjektivnom ljudskom percepcijom kvalitete slike. To je čisto matematička metrika koja ne uzima u obzir faktore kao što su ljudska vizualna percepcija, osjetljivost na kontrast ili strukturna sličnost signala. Dakle, dvije slike s istom PSNR vrijednošću mogu imati potpuno različitu vizualnu kvalitetu.

### 2.2.1. SSIM

Indeks strukturne sličnosti (engl. *SSIM - Structural similarity index measure*) naširoko je korištena metrika kvalitete za procjenu sličnosti između dviju slika, na temelju čega onda procjenjuje kvalitetu izobličene slike u odnosu na referentnu. Obično se koristi u obradi slika i aplikacijama računalnog vida za kvantificiranje perceptivne sličnosti ili različitosti između referentne slike i iskrivljene slike. SSIM uzima u obzir tri glavne komponente kvalitete slike: svjetlinu, kontrast i informacije o strukturi.

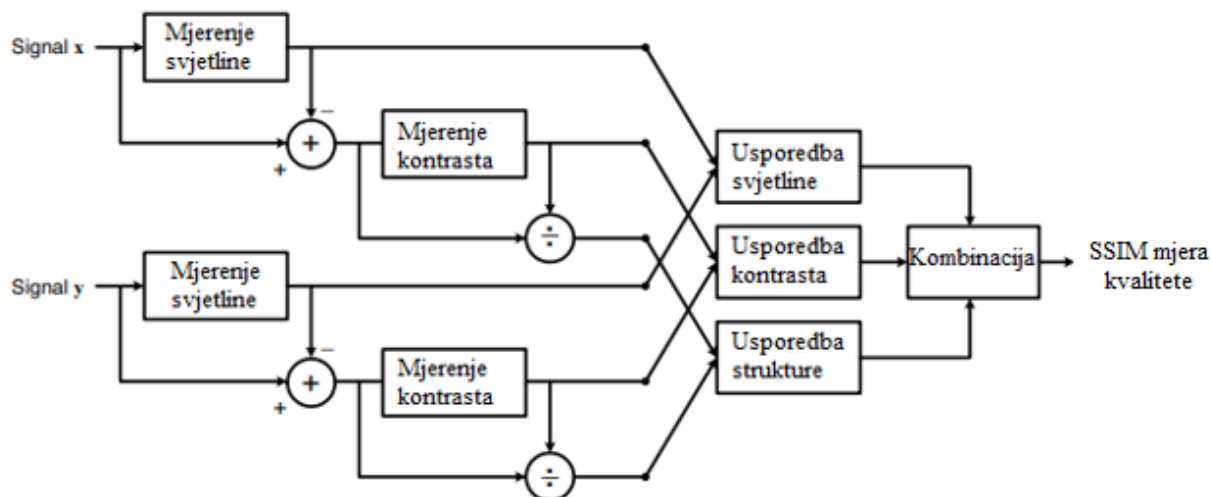


**Slika 2.4.** a) Originalna slika, b) zamućena slika sa PSNR = 30.6 dB, c) original kojem je zamućena pozadina sa PSNR=27.7 dB [8]

Ove komponente pokušavaju odraziti osjetljivost ljudskog vizualnog sustava na promjene u svjetlini, kontrastu i strukturnim detaljima. Uključujući ove čimbenike, SSIM pruža sveobuhvatniju procjenu kvalitete slike od tradicionalnih metrika kao što su MSE ili PSNR, koji se fokusiraju isključivo na razlike u pikselima. SSIM indeks mjeri sličnost između dviju slika uspoređujući njihovu svjetlinu, kontrast i strukturne sličnosti, što se može vidjeti na slici 2.5. [9]. Algoritam izdvaja svjetlinu i kontrast od danog mu signala, dobivajući strukturnu komponentu signala i na kraju uspoređuje te komponente s istim komponentama drugog signala dobivajući mjeru sličnosti. Formula (2-3) opisuje izračunavanja SSIM indeksa između dvaju signala, gdje je  $l(x,y)$  sličnost luminantne komponente između signala  $x$  i  $y$ ,  $c(x,y)$  je sličnost kontrasta  $x$  i  $y$  signala i  $s(x,y)$  sličnost strukturne komponente obaju signala, dok su  $\alpha$ ,  $\beta$  i  $\gamma$  parametri za definiranje relativne važnosti triju komponenata.

$$SSIM(x, y) = [l(x, y)]^\alpha * [c(x, y)]^\beta * [s(x, y)]^\gamma \quad (2-3)$$

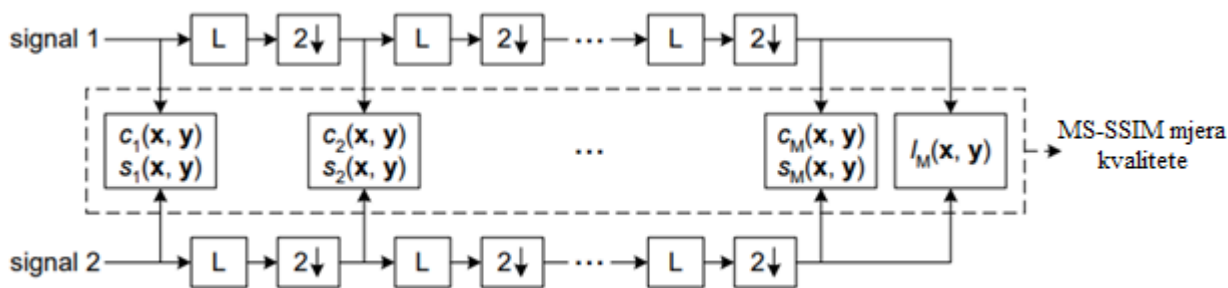
Usporedba svjetline uzima u obzir srednji intenzitet piksela i bilježi promjene u svjetlini. Usporedba kontrasta procjenjuje standardnu devijaciju intenziteta piksela, predstavljajući razlike u kontrastu između slika. Strukturna usporedba procjenjuje korelaciju između prostornog rasporeda piksela i uzima u obzir varijacije u strukturnim detaljima [9]. Rezultat SSIM je vrijednost između 0 i 1, gdje je 1 maksimalna korelacija (signali se smatraju identičnima), a 0 označava da nema korelacije između video signala.



**Slika 2.5.** Dijagram sustava za dobivanje ocjene kvalitete videa SSIM metrikom [9]

### 2.2.2. MS-SSIM

Multi-scale SSIM je proširenje tradicionalne metrike SSIM. Dok se SSIM usredotočuje na mjerenje strukturne sličnosti u jednoj skali, MS-SSIM uzima u obzir više skala ili razlučivosti slike ili videa kako bi pružio sveobuhvatniju procjenu kvalitete. MS-SSIM naširoko se koristi u području obrade slike i videa, budući da se bavi nekim ograničenjima SSIM-a i pruža točniji prikaz ljudske vizualne percepcije. Uzima u obzir činjenicu da ljudi različito percipiraju slike i videozapise na različitim razinama i to uzima u obzir pri ocjenjivanju njihove kvalitete [10]. MS-SSIM metrika funkcionira dekompozicijom slike ili videa na više ljestvica pomoću tehnike koja se naziva Gaussova piramidalna dekompozicija [11]. Svaka ljestvica predstavlja različitu razinu detalja, u rasponu od grube do fine. Na svakoj skali, SSIM indeks izračunava se između odgovarajućih blokova piksela, slično tradicionalnoj metrici SSIM. Dobivene vrijednosti sličnosti na svakoj ljestvici zatim se kombiniraju kako bi se dobila ukupna mjera kvalitete. Sustav mjerenja kvalitete je vidljiv na slici 2.6. Za svaku skalu se uspoređuju kontrast i strukturalna komponenta sve do M-tog indeksa, gdje se uzima luminantna komponenta i izračunava se mjera sličnosti. Svaka sljedeća skala se dobiva tako što se signali provode kroz nisko propusni filter i zatim se smanjuje rezolucija za 2 puta po širini i 2 puta po visini. Nadalje, ovaj sustav se dodatno može opisati formulom (2-4). gdje je jasno vidljivo da se krajnja ocjena kvalitete dobiva množenjem luminantne komponente M-tog indeksa i umnoškom komponente strukture i kontrast M puta. U slučaju procjene kvalitete videa, MS-SSIM proširuje svoju primjenu i na vremenske informacije. Izračunava sličnost između odgovarajućih okvira video signala u više rezolucija, uzimajući u obzir i prostorne i vremenske ovisnosti.



**Slika 2.6.** Sustav mjerenja kvalitete videa MS-SSIM algoritmom, gdje je L: nisko propusno filtriranje, a ↓: smanjivanje rezolucije [10]

To MS-SSIM čini posebno prikladnim za procjenu kvalitete algoritama video kompresije. Slično SSIM-u, MS-SSIM indeks kreće se od 0 do 1. Vrijednosti bliže 1 ukazuju na bolju kvalitetu, dok vrijednosti bliže 0 ukazuju na značajnu razliku između uspoređivanih slika ili okvira.

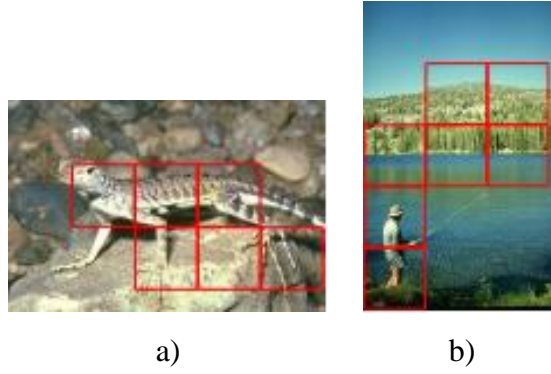
$$MSSSIM(x, y) = [l_M(x, y)]^{\alpha_M} * \prod_j^M [c_j(x, y)]^{\beta_j} * [s_j(x, y)]^{\gamma_j} \quad (2-4)$$

### 2.2.3. NIQE

NIQE (engl. *Naturalness Image Quality Evaluator*) je metrika procjene kvalitete za slike koja mjeri percipiranu kvalitetu slike bez korištenja referentne slike, tj. spada u NR metode za razliku od PSNR, SSIM i MS-SSIM. NIQE se temelji na zapažanju da prirodne slike posjeduju određena statistička svojstva koja ih razlikuju od neprirodnih ili iskrivljenih slika. To znanje koristi za kvantificiranje kvalitete slike procjenom odstupanja njezinih statističkih svojstava od onih prirodnih slika. Prema [12], IQA (engl. *Image Quality Assessment*) model temelji se na perceptualno relevantnim NSS (engl. *Natural scene statistics*) značajkama prostorne domene izdvojenim iz lokalnih regija slike koje učinkovito bilježe bitnu statistiku prirodnih slika. Koriste se klasični prostorni NSS iz [13] pri pretprocesiranju slike, uklanjajući srednju lokalnu vrijednost piksela (engl. *Local mean removal*) i djeliteljna normalizacija (engl. *Divisive Normalization*), računajući nove koeficijente slike prema formuli (2-5), gdje su  $i \in \{1, 2, \dots, M\}$  i  $j \in \{1, 2, \dots, N\}$  koordinate piksela na slici gdje se primjenjuje transformacija, a  $M$  i  $N$  su dimenzije slike [12].

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + 1} \quad (2-5)$$

Kada se izračunaju koeficijenti prema formuli (2-5), tada se izvršava selektiranje regija (engl. *Patch selection*), gdje se slika dijeli na regije  $P \times P$  piksela i odabiru se samo one čija je oštrina veća od praga  $T$ , tj. one regije slike s najviše informacija. Na slici 2.7 može se vidjeti biranje regija  $P \times P$  na temelju njihove oštine.



**Slika 2.7.** Označeni blokovi na slikama a) i b) prikazuju primjere regija slike odabrane pomoću lokalne mjere oštine[12]

Autor iz [14] spominje da se odabiru samo najoštrije regije slike iz razloga što ljudsko oko daje veću pažnju oštini pri ocjenjivanju kvalitete. Konstruirana se skup značajki koje utječu na kvalitetu (engl. *quality aware*) izvedenih iz jednostavnog modela statistike prirodne scene. Ove značajke hvataju statistička svojstva prirodnih slika koja ukazuju na kvalitetu slike. NSS modelom se izvlači 36 statističkih značajki slika koje se prosljeđuju u multivarijantni Gaussov (engl. *MVG - Multivariate Gaussian*) model. MVG model koristi značajke svjesnih kvaliteta izdvojenih iz skupa prirodnih slika i predstavlja statističku distribuciju ovih značajki u skupu podataka prirodne slike. MVG model koristi se za procjenu statistike prirodne slike i procjenu kvalitete slike. NIQE indeks se tada izražava kao udaljenost između značajki izdvojenih iz NSS modela i značajki izdvojenih iz izobličene slike prosljeđene MVG modelu, što je opisano formulom (2-6), gdje su  $v_1$ ,  $v_2$ ,  $\Sigma_1$ ,  $\Sigma_2$  vektori srednje vrijednosti i matrice kovarijance značajki NSS modela i značajki dobivenih iz izobličene slike, a  $D$  predstavlja ocjenu kvalitete slike. Za video loše kvalitete ocjena će biti veća, dok za video bolje kvalitete ocjena će biti manja.

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{\left( (v_1 - v_2)^T * \left( \frac{\Sigma_1 + \Sigma_2}{2} \right)^T * (v_1 - v_2) \right)} \quad (2-6)$$

#### 2.2.4. VIIDEO

VIIDEO (engl. *Video Intrinsic Integrity and Distortion Evaluation Oracle*) je NR model za ocjenjivanje kvalitete videa koji se temelji na skupu statističkih modela koji za procjenu kvalitete videa uzimaju u obzir bitne vremenske značajke dobivene iz razlike video okvira [15]. Slično NIQE algoritmu, ovaj algoritam koristi NSS model, ali umjesto da računa lokalnu statistiku svakog okvira, računa se lokalna statistika razlike video okvira opisana formulom (2-7), gdje je  $\Delta F^t$  razlika video okvira,  $F^{2t+1}$  okvir ispred trenutnog, a  $F^{2t}$  trenutni okvir. Dobivena razlika okvira više ne prati generaliziranu Gausovu distribuciju (engl. *GGD - generalized Gaussian*



*distribution*) kao kod NSS, stoga autori iz [15] primjenjuju model asimetrična generalizirane Gausove distribucije (engl. *AGGD - Asimetric generalized Gaussian distribution*) za dobivanje vektora značajki.

$$\Delta F^t = F^{2t+1} - F^{2t} \quad \forall t \in \{0, 1, 2, \dots, \frac{T-1}{2}\} \quad (2-7)$$

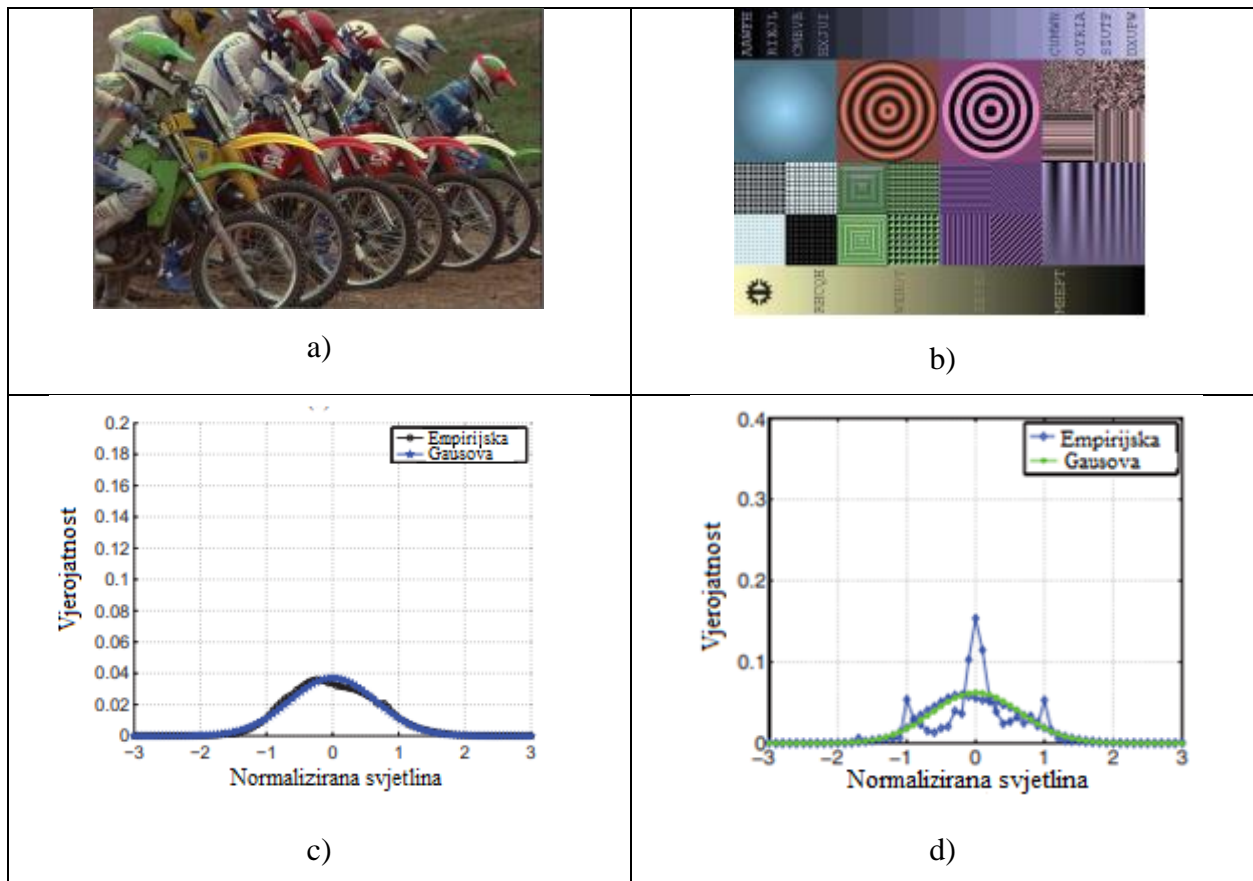
Prije samog računanja vektora značajki, dobiveni okvir iz formule 2-6 se dijeli na regije veličine  $P \times Q$  piksela, kako bi bila moguća lokalna analiza izobličenja u tim okvirima. Varijacije u transformiranim koeficijentima razlike okvira na različitim skalama tijekom vremena mogu biti indikativne ili za prirodno kretanje ili za vremensko iskrivljenje. Kada je prirodno kretanje prisutno, ovi koeficijenti pokazuju jaku korelaciju između finih i grubih ljestvica tijekom vremena. S druge strane, vremenska distorzija dovodi do prolaznih statističkih anomalija. Kako bi uhvatio te razlike, model VIIDEO fokusiran je na mjerenje vremenskih varijacija korelacija značajki statističkog modela između različitih skala koeficijenata transformirane slike na lokaliziran način. Analizirajući te korelacije, model može razlikovati prirodno kretanje i vremensko izobličenje, što omogućuje točniju procjenu vizualne kvalitete [15].

## **2.3. Objektivne metrike za ocjenjivanje video signala zasnovane na algoritmima strojnog učenja**

### **2.3.1. BRISQUE**

BRISQUE (engl. *Blind/Referenceless Image Spatial Quality Evaluator*) je jedna od metrika za ocjenjivanje kvalitete slika i video signala koja se temelji na SVM (engl. *Support Vector Machine*). SVM spada u kategoriju nadzirnog učenja i popularan je algoritam strojnog učenja koji se koristi za zadatke klasifikacije i regresije. SVM-ovi su osobito korisni za analizu bioloških podataka zbog svoje sposobnosti rukovanja visoko-dimenzionalnim skupovima podataka s ograničenim brojem uzoraka [16]. SVM-ovi rade tako da identificiraju podskup uzoraka za treniranje, koji se nazivaju vektori podrške, koji su ključni za određivanje granice odluke. Maksimiziranjem razlike između vektora podrške i ravnine koja dijeli podatke, SVM-ovi imaju za cilj postići bolju generalizaciju i robusnost. Funkcije jezgre (engl. *Kernel*) omogućuju SVM-ovima rukovanje nelinearno odvojivim podacima implicitnim mapiranjem podataka u višedimenzionalni prostor. Autor iz [17] ističe da je odabir odgovarajuće kernel funkcije ključan za postizanje točnih rezultata klasifikacije. BRISQUE se, slično kao i NIQE, temelji na principu da prirodna slika, tj. slika koja je uhvaćena optičkom kamerom i nije procesirana nikakvim računalnim programom, ima određena statistička svojstva koja prikazuju određene uzorke i karakteristike. Međutim, kada

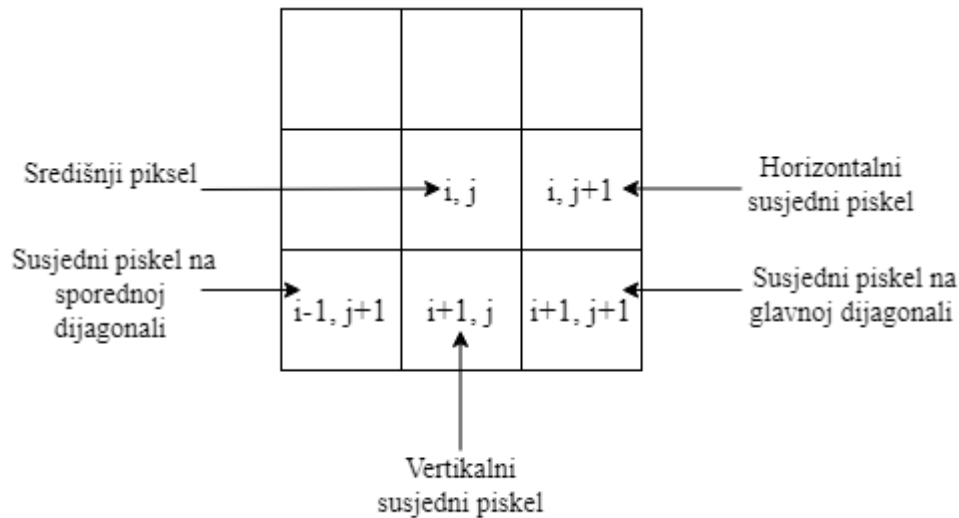
se prirodna slika modificira s izobličenjem ta se statistička svojstva mijenjaju, odnosno dolazi do određenog odstupanja od očekivanih statističkih pravilnosti. Normalizirani koeficijenti svjetline prirodne slike prate Gaussovu distribuciju, kao što je prikazano na slici 2.8. c), dok isto ne vrijedi za empirijsku distribuciju umjetno generirane slike prikazane na slici 2.8. d). BRISQUE primjenjuje NSS model u prostornoj domeni, računajući lokalnu normaliziranu svjetlinu slike pomoću formule (2-5), nazivajući ga MSCN (engl. *Mean subtracted contrast normalized*) koeficijent.



**Slika 2.8.** Primjer: a) prirodne slike i b) umjetne slike, c) Gausova i empirijska razdioba normalizirane svjetline prirodne slike, d) Gausova i empirijska razdioba normalizirane svjetline umjetne slike [17]

Nadalje, koriste se modeli generalizirane Gausove distribucije i asimetrična generalizirane Gausove distribucije za učinkovito hvatanje šireg spektra statistike iskrivljene slike. Izvlačenje značajki (engl. *Feature extraction*) vrši se na originalnoj rezoluciji i na rezoluciji smanjenoj za 2 puta i za svaku se dobije 18 značajki, što je ukupno 36 za svaku sliku. GGD se koristi za izvlačenje značajki MSCN koeficijenta, dok se AGGD iskorištava kako bi se uzele značajke susjednih piksela, parni produkti izračunavaju se duž četiri orijentacije: vodoravna, okomita, glavna dijagonala i sekundarna dijagonala na udaljenosti od 1 piksela, što je prikazano na slici 2.9. Kako

bi se dobila objektivna ocjena video signala bez reference, koristi se regresijski SVM modul, koji se trenira na dobivenim značajkama [17].

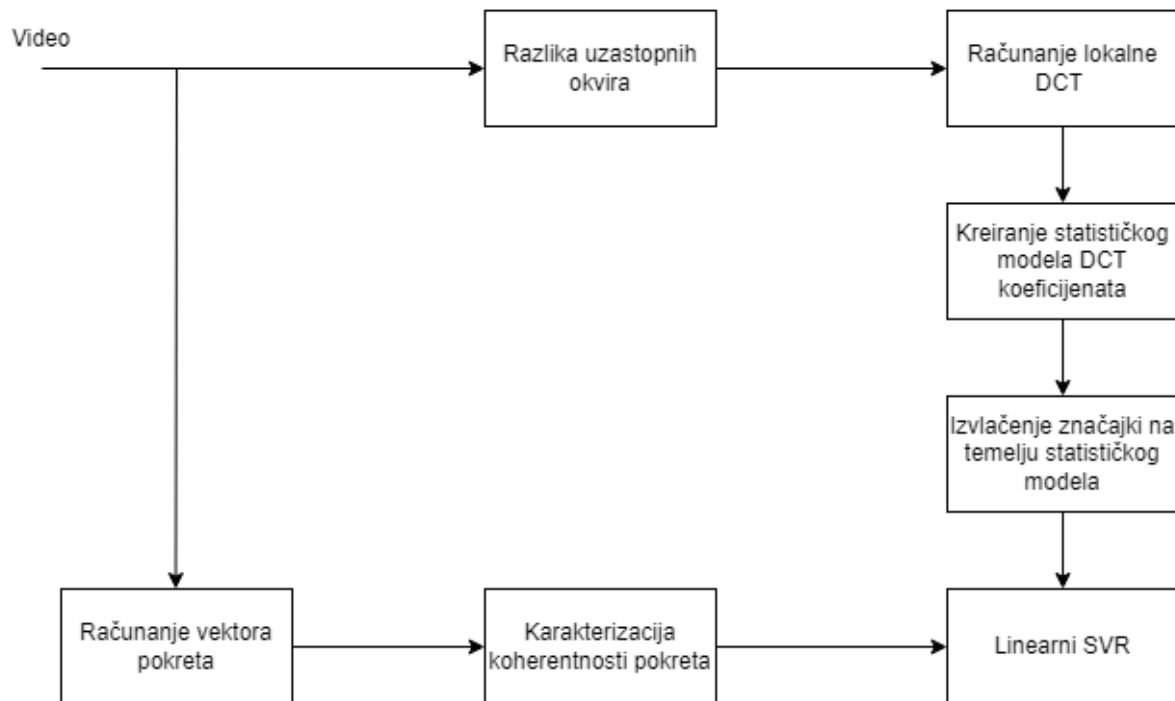


**Slika 2.9.** Upareni proizvodi susjednih MSCN koeficijenti duž četiri orijentacije [17]

### 2.3.2. Video BLIINDS

Video BLIINDS (engl. *Blind Image Integrity notator using DCT Statistics*) predstavlja NR metodu za ocjenjivanje video signala koja je zasnovana na strojnom učenju. Koristi prostorno-vremenski model video scene u domeni diskretne kosinusove transformacije i model kretanja za predviđanje kvalitete videa koji je istreniran SVM regresijskom metodom. Kao osnova prostorno-vremenskog modela koristi se model prirodne video scene (engl. *NVS – Natural video scenes*). Slično kao NIQE i BRISQUE metrike, koje koriste NSS jer se radi o metrikama za ocjenjivanje slika. Video BLIINDS upotrebljava vremenske i prostorne značajke kako bi izvukao informacije iz video signala. Sažeti prikaz kako model radi nalazi se na slici 2.10. Uključuje primjenu lokalne 2-dimenzionalne prostorne diskretne kosinusne transformacije na blokove veličine  $n \times n$  dobivene razlikom uzastopnih okvira. Ovaj pristup hvata i prostorne i vremenske lokalne frekvencije. Frekvencije se zatim modeliraju korištenjem specifične obitelji funkcija gustoće vjerojatnosti, s različitim parametrima za netaknute i iskrivljene videozapise. Model također razmatra interakciju između kretanja i prostorno-vremenskih promjena u videu. Vrsta kretanja prisutna u videu, koja je rezultat kretanja objekta i kamere, karakterizirana je pomoću mjere koherentnosti. Ova se mjera koristi u kombinaciji s parametrima izvedenim iz prostorno-vremenskog NVS modela DCT koeficijenata. Značajke izvučene na temelju prostorno-vremenskog NVS modela, zajedno s mjerom koherentnosti kretanja, koriste se kao ulazi u SVR (engl. *Support Vector Regressor*) sa

linearnom regresijskom funkcijom, koji je osposobljen za predviđanje vizualne kvalitete videozapisa [18].

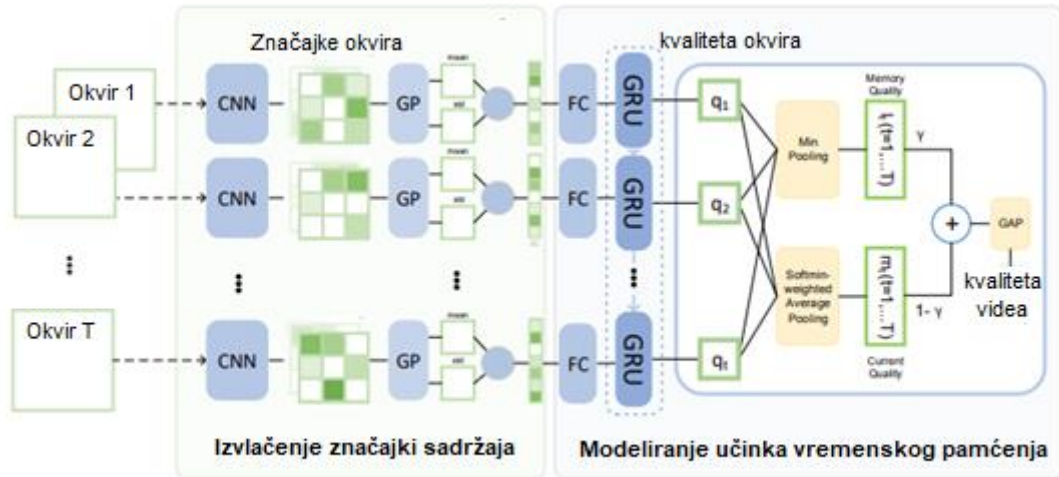


**Slika 2.10.** Dijagram Video BLIIDS algoritma [18]

### 2.3.3. VSFA

Osim problema s ocjenjivanjem video signala komprimiranih video koderima u skladu s normama poput H.264, H.265, MPEG2 bez referentnog signala, javlja se i problem s ocjenjivanjem korisnički generiranih sadržaja (engl. *UGC - User generated content*). Budući da za njih ne postoji referentni signal, trenutno najbolji pristup za rješavanje tog problema su duboke neuronske mreže. VSFA (engl. *Video Quality Assessment of in-the-wild Videos using Spatial and Frequency Analysis*) je predstavljen u [19]. Kako bi dobio što točniju kvalitetu u odnosu na DMOS ocjenu, autor koristi učinke ovisnosti o sadržaju i vremensko pamćenje. Za ovisnost o sadržaju izdvajaju se značajke iz prethodno istrenirane neuronske mreže za klasifikaciju slika ResNet50. Za učinke vremenskog pamćenja, dugoročne ovisnosti i posebne vremenske histereze integrirane su u mrežu sa zatvorenom rekurentnom jedinicom i subjektivno inspiriranim slojem vremenskog udruživanja. Slika 2.11 prikazuje model VSFA algoritma koji je podijeljen na dvije jedinice: jedinica za izdvajanje značajki sadržaja s konvolucijskom neuronskom mrežom i jedinica za vremensko pamćenje, koja koristi GRU (engl. *Gated Recurrent Unit*). GRU je rekurentna neuronska mreža s kontrolom ulaza koja je sposobna i integrirati značajke i naučiti dugoročne ovisnosti. Budući da GRU bolje uči na značajkama manjih dimenzija, a značajke koje prima iz konvolucijske neuronske mreže su veličine 1x4096, bilo je potrebno staviti jedan potpuno

povezani sloj (engl. *Fully Connected*) kako bi smanjili dimenziju na  $1 \times 128$ . Tako dobivene značajke odlaze u GRU, nakon čega odlaze do subjektivno inspiriranog sloja vremenskog udruživanja koji daje ocjenu kvalitete video signala [19].



**Slika 2.11.** Arhitektura VSFA modela [19]

### 3. RAZVOJ VLASTITOG ALGORITMA ZA OCJENJIVANJE VIDEO SIGNALA ZASNOVANOG NA STROJNOM UČENJU

U ovom poglavlju bit će opisani korišteni alati za izradu algoritma za ocjenjivanje kvalitete video signala zasnovanog na strojnom učenju i računalno okruženje na kojem je algoritam treniran. Predložena je jedna mreža koja sadrži jedan sloj pažnje (engl. *Attention layer*) i druga mreža se zasniva na dvjema LSTM neuronskim mrežama. Cijela implementacija rađena je u *Jupyter Notebook* okruženju koja koristi *Python* programski jezik. U potpoglavlju 3.2 se nalazi opis baza korištenih za treniranje, validaciju i testiranja predloženih rješenja. Za razvoj ovakvog rješenja može se koristiti više različitih pristupa: prvi pristup obuhvaća formiranje podataka tako da skupovi za trening, validaciju i test sadrži iste sadržaje, ali ne baš iste izobličene signale. Kod drugog se pristupa skupovi formiraju prema sadržaju, što znači da će trening, test i validacijski skup imati različite sadržaje. Prvi način može služiti kako bi se vidjelo hoće li model istreniran na svim sadržajima i određenim izobličnim signalima imati dobru preciznost prilikom ocjenjivanja drugačije razine izobličnih signala. S druge strane, drugi način daje nešto realniju sliku kako algoritam zapravo radi i može li se primjenjivati u realnim sustavima, jer se algoritam trenira i testira na različitim sadržajima. U sklopu diplomskog rada upravo su korištena ta dva pristupa i više detalja o tome bit će biti dano u nastavku rada. Cilj je svakog rješenja za objektivnu ocjenu kvalitete da dobro radi na video signalima koji nisu korišteni prilikom njegovog razvoja.

#### 3.1. Opis korištenih alata

Prilikom izrade zadatka diplomskog rada korišten je programski jezik *Python* koji omogućuje jednostavno implementiranje biblioteka namijenjenih za strojno učenje. Biblioteka *TensorFlow 2.6.0* omogućuje korištenje *Keras 2.6.0* API-ja koji je pristupačan i predstavlja visokoproduktivno sučelje za rješavanje problema strojnog učenja, s fokusom na moderno duboko učenje [20]. Korištena je i web aplikacija *Jupyter Notebook 6.5.2* koja omogućuje stvaranje i dijeljenje dokumenata koji sadrže kod, jednadžbe, vizualizacije i narativni tekst. Upotreba uključuje čišćenje i transformaciju podataka, numeričku simulaciju, statističko modeliranje, vizualizaciju podataka i strojno učenje [21]. Kako bi se moglo pristupiti *Jupyter Notebook* aplikaciji, bilo je potrebno instalirati *Anaconda 2.4.2 Python* distribuciju koja omogućuje kreiranje vlastitog okruženja za razne *Python* aplikacije i usmjerena je na projekte vođene podacima [22]. Radi bržeg treniranja, validacije i testiranja korištena je NVIDIA GeForce GTX 970 grafička kartica. Radi lakšeg učitavanja videa u *.yuv* video formatu korištena je biblioteka *scikit-video.io 1.1.11* [23]. Nadalje, za oblikovanje pojedinih slika video signala korišten je *OpenCV 4.7.0.72*

(engl. *Open Source Computer Vision Library*) [24] i za učitavanje subjektivnih ocjena kvalitete videa korištena je biblioteka *Csv 0.0.13* (engl. *Comma Separated Values*) [25]. Zatim je korištena MSU VQMT 14.1 (engl. *Video Quality Measurement Tool*) aplikacija koja služi za komparativnu objektivnu analizu kvalitete videa. Glavna funkcionalnost ovog softver je izračunati objektivne metrike kvalitete za video ili sliku pomoću reference ili bez reference [26].

### 3.2. Opis baza podataka za treniranje modela

U sklopu diplomskog rada potrebno je bilo pronaći relevantne baze video signala sa subjektivnim ocjenama kako bi se definirani model za ocjenjivanje kvalitete video signala mogao istrenirati. Baze koje su korištene za treniranje, testiranje i validaciju su: FERIT-RTRK1[27], FERIT-RTRK2[27], EVVQ [28], ECVQ [29], LIVE-VQA [30],[31] i LIVE-MOBILE [32].

FERIT-RTRK1 sadrži 90 video signala i 6 različitih sadržaja: *bea, bos, hon, joc, rea* i *yac*. Svaki sadržaj je komprimiran s 5 razina H.264, H.265 i MPEG2 kompresijom, što ukupno daje 15 video signala po sadržaju. Dodatno, svi su video signali u mp4 formatu rezolucije 1920x1080 piksela. Trajanje svakog video signala je 5 sekundi, uz 60 okvira po sekundi, što daje 300 okvira po video signalu. Subjektivne ocjene za video signale izražene su u obliku DMOS-a od 0-10 i dobivene su ACR-HR metodom za subjektivno ocjenjivanje video signala. Za svaku kategoriju izobličenja FERIT-RTRK1 sadrži po 30 video signala.

FERIT-RTRK2 baza sadrži 486 video signala za 6 sadržaja, istih kao i kod FERIT-RTRK1. Svaki video je komprimiran je s po 3 razine H.264, H.265 i MPEG2 koderima. Osim toga, svakoj razini kompresije svakog sadržaja dodana je određena stopa gubitka pakete, dvije razine smrzavanja (engl. *freeze*) i kombinacija ta 2 mrežna artefakta. Nadalje, svaki video ova baze traje po 5 sekundi uz 60 okvira po sekundi. Subjektivne ocjene za video signale izražene su u DMOS-u od 0-10 i dobivene su ACR-HR metodom za subjektivno ocjenjivanje video signala. FERIT-RTRK2 baza sadrži po 162 video signala komprimirana s H.264, H.265, MPEG2 koderima, no dio tih signala uz tip kompresije sadrže i dodatne artefakte, tj. smrzavanje (engl. *freezing*) i određeni postotak gubitka paketa. Od ukupno 486 video signala, 324 video signala sadrže određenu razinu smrzavanja, dok 108 video signala sadrže određenu stopu gubitka paketa.

ECVQ baza podataka sadrži 90 video signala 8 različitih sadržaja: *container\_ship, flower\_garden, football, foreman, hall, mobile, news, silent*. ECVQ ukupno sadrži 43 video signala komprimiranih s H.264 koderom i 47 video signala komprimiranih s MPEG4 koderom. Svaki video ima 25 okvira u sekundi, CIF rezoluciju od 352x288 piksela i nalazi se u 4:2:0 YUV planar

formatu. Za subjektivnu ocjenu kvalitete video signal koristila se metoda evaluacije SAMVIQ (engl. *Subjective Assessment Method for Video Quality*) sa skalom ocjenjivanja od 0-100.

EVVQ baza podataka sadrži 90 video signala 8 različitih sadržaja: *cartoon, cheerleaders, football, discussion, flower\_garden, mobile, town\_plan, weather*. EVVQ ukupno sadrži 42 video signala komprimiranih s H.264 kodekom i 48 video signala komprimiranih s MPEG4 kodekom. Svaki video ima 25 okvira u sekundi, rezolucije 640x480 piksela i nalazi se u 4:2:0 YUV planar formatu. Za subjektivnu ocjenu kvalitete video signal koristila se metoda evaluacije SAMVIQ (engl. *Subjective Assessment Method for Video Quality*) sa skalom ocjenjivanja od 0-100.

LIVE VQA video baza ima ukupno 150 video signala komprimiranih s H.264 i MPEG2 koderima, te uz različite postotke gubitka paketa preko IP mreže i bežične mreže. Ova baza ima 10 različitih sadržaja: *Blue sky, Mobile and Calendar, Pedestrian area, Park run, Riverbed, Rushhour, Sunflower, Shields, Tractor*. Video signali su u YUV 4:2:0 planar formatu, različitog broja okvira po sekundi i rezolucije 768x432 piksela. Za subjektivne ocjene kvalitete video signala koristila se ACR-HR evaluacijska metoda sa skalom od 0-100. LIVE VQA ima po 40 video signala komprimiranih s H.264 i MPEG2 kodekom, 30 video signala s određenom stopom gubitka paketa preko IP mreže i 40 video signala s gubitkom paketa preko bežične mreže

LIVE Mobile baza podataka sadrži 200 video signala, od kojih je samo 120 korišteno u sklopu diplomskog rada, jer su samo video signali s određenim tipovima izobličenja bili potrebni za diplomski rad. Video signali koji su korišteni izobličeni su: H.264 koderom, mrežnim artefaktima zamrzavanja i određenom stopom gubitka paketa. Ukupno ima 40 video signala po tipu izobličenja. Video signali koji nisu korišteni su oni koji sadrže vremensku dinamiku (engl. *Temporal Dynamics*) i video signali koji sadrže prilagodbu brzine (engl. *Rate adaptation*). Ona sadrži ukupno 10 sadržaja: *Friend Drinking Coke, Two Swan Dunking, Runners Skinny Guy, Students Looming Across Street, Bulldozer With Fence, Panning Under Oak, Landing Airplane, Barton Springs Pool Diving, Trail Pink Kid, Harmonicat*. Video signali su u YUV planar 4:2:0 formatu, prostorne rezolucije 1280x720 piksela i trajanja 690 okvira. Za subjektivne ocjene kvalitete video signala koristila se ACR-HR evaluacijska metoda i ocjene su izražene u DMOS-u od 1-5.



### 3.3. Treniranje algoritma zasnovanog na strojnom učenju

#### 3.3.1. Priprema podataka

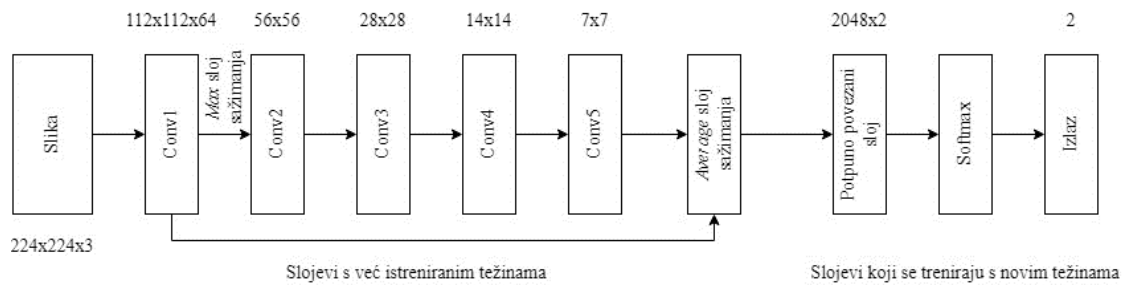
Prvi korak u nadzornom učenju je priprema podataka za treniranje, validaciju i testiranje. Budući da je zadatak diplomskog rada regresijski problem, uz dostupne baze video signala nalaze se i meta-podaci o prosječnoj DMOS ocjeni za svaki video signal. Dostupne podatke o DMOS ocjenama za neke baze potrebno je skalirati tako da su sve u istom rasponu. Odabrani raspon je od 0-10.

#### 3.3.2. Model za izvlačenje značajki

Nadalje, prije samog treniranja modela bilo je potrebno odraditi postupak izvlačenja značajki iz okvira videa. Taj proces odrađen je pomoću već dostupne neuronske mreže ResNet50 istrenirane na *ImageNet* bazi, čija je arhitektura prikazana na slici 3.1. Ona se sastoji od niza konvolucijskih slojeva koji iz slike željene rezolucije izvlače značajke i gornjeg sloja koji se sastoji od potpuno povezanih slojeva i *softmax* aktivacijske funkcije i često se gornji sloj modificira ovisno o zadatku na kojem se koristi ResNet50 neuronska mreža. Budući da je zadatak diplomskog regresijskog tipa, gornji sloj nije bio potreban, stoga je izbačen. Slojeve konvolucijske neuronske mreže bilo je potrebno zamrznuti, jer se u praksi na taj način ResNet50 koristi za izvlačenje značajki. Poslije zadnjeg konvolucijskog sloja dodan je *GlobalAveragePooling2D* sloj sažimanja, koji za predani mu izlaz iz konvolucijskog sloja računa prosječnu vrijednost piksela dajući izlaz od 2048 značajki. Zbog toga što baze opisane u prošlom potpoglavlju sadrže video signale različitih rezolucija, bilo je potrebno postaviti fiksnu rezoluciju. Postavljena rezolucija je 224x224 piksela, čime je promijenjen omjer slike (engl. *aspect ratio*) svakog videa na 1:1, što znatno mijenja vizualnu kvalitetu video signala. Promjena rezolucije donosi veću brzinu izvlačenja značajki, ali uz gubitak detalja pojedinog okvira video signala. Zbog različitosti u broju okvira po sekundi, bilo je potrebno sve video signale svesti na isti broj okvira, a odabran je broj 300. Za video signale koji su imali više od 300 okvira odabrano je prvih 300 okvira. Ovdje svakako treba naglasiti da je broj okvira za izvlačenja značajki morao biti fiksna zbog načina funkcioniranja modela, što može imati posljedice na krajnji rezultat cijelog algoritma ocjene kvalitete videa. Naime, ocjena gledatelja zasnovana je na svim okvirima videa, a ocjene modela bit će zasnovane samo na kvaliteti prvih 300 okvira, što može dovesti do znatnog odstupanja od subjektivnih ocjena ako je dio videa iza 300. okvira najviše utjecao na ocjenu gledatelja.

Prije samog procesa izvlačenja značajki, osim promjene rezolucije bilo je potrebno i sliku pretvoriti iz BGR formata boja u RGB format, jer *skvideo.io.vread* naredba za čitanje video signala

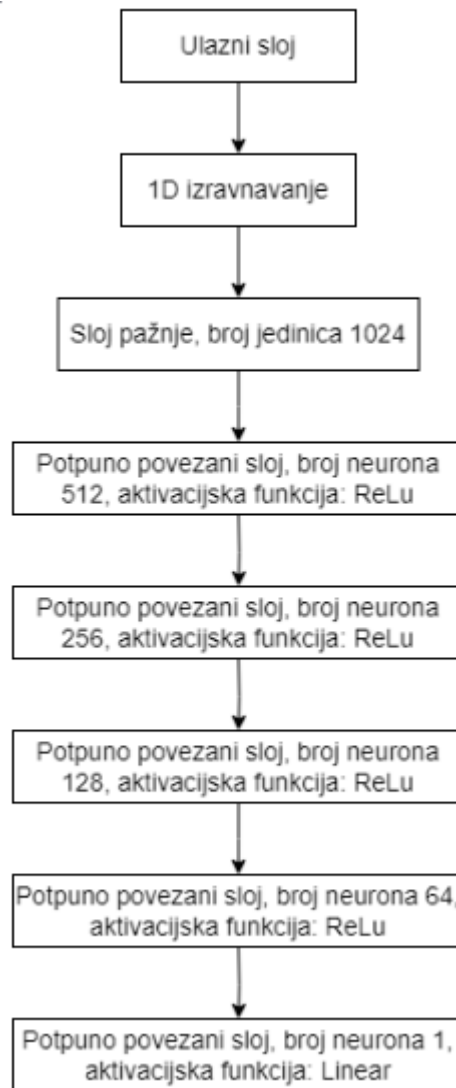
učitava video u BGR formatu. Na taj način je slika svakog video signala pripremljena je za *resnet50.preprocess\_input* naredbu, koja osigurava da su podaci koji ulaze u model za izvlačenje značajki kompatibilni s podacima na kojima je model treniran. Krajnji rezultat algoritma za izvlačenje značajki je trodimenzionalno *numpy* polje koji je oblika (broj video signala po izobličenju, broj okvira video signala, broj značajki po okviru). Primjerice za MPEG2 trening skup *numpy* polje izgledalo bi: (60, 300, 2048).



**Slika 3.1.** Arhitektura ResNet50 konvolucijske neuronske mreže [33]

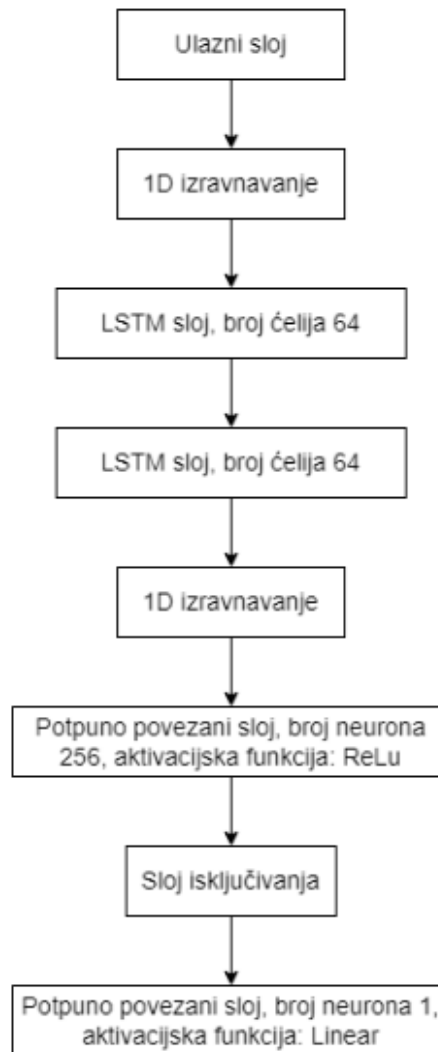
### 3.3.3. Arhitektura odabranih dubokih neuronskih mreža za ocjenjivanje video signala

U svrhu ocjenjivanja video signala zasnovanog na strojnom učenju, odabrane su dvije različite arhitekture. Prva arhitektura (slika 3.2.) koristi sloj pažnje (engl. *Attention*) poslije ulaza, koji koristi *bahdanau* mehanizmu ocjenjivanja. *Bahdanau* mehanizma izračunava rezultate pažnje koji se normaliziraju pomoću funkcije *softmax* u vrijednosti težine između 0 i 1. Rezultat pažnje koristi se za stvaranje vektora konteksta direktno iz značajki dobivenih iz ResNet50 konvolucijske mreže. Kroz operaciju težinskog zbrajanja računa se završni vektor konteksta. Zatim se vektor konteksta prosljeđuje u potpuno povezani sloj s ReLU aktivacijskom funkcijom i 512 neurona, nakon kojeg ide niz potpuno povezanih slojeva. I na kraju se nalazi potpuno povezani sloj s jednim neuronom i linearnom aktivacijskom funkcijom, koji predstavlja izlaz. Nadalje će se u tekstu ovaj model zvati Model 1.



**Slika 3.2.** Arhitektura duboke neuronske mreže sa slojem pažnje korištene za ocjenjivanje video signala u sklopu ovog diplomskog rada (Model 1)

Druga arhitektura koja je prikazana na slici 3.3. zasnovana je na dvjema LSTM povratnim neuronskim mrežama, koristi trodimenzionalni ulazni podatak iz modela za izvlačenje značajki. LSTM se obično koristi u slučaju kada postoji niz podataka kod kojih je bitna prijašnja i buduća informacija. Na taj način se ostvaruje da neuronska mreža, osim prostornih informacija, prati i vremenske, što je izrazito bitno za analiziranje video signala. Budući da ulazni podatak predstavlja sekvencu na početku duboke neuronske mreže, dodan je *TimeDistributed* sloj. Slijede dvije LSTM mreže. Nakon druge LSTM mreže dolazi sloj izravnavanja (engl. *Flatten*), koji 3D podatke sažima u jedno 1D polje. Takvo 1D polje prosljeđuje se potpuno povezanom sloju sa 256 neurona i *ReLU* aktivacijskom funkcijom. Nadalje, stavljen je sloj isključivanja (engl. *Dropout*) sa stopom isključivanja od 20%. Izlaz iz duboke neuronske mreže je potpuno povezani sloj s jednim neuronom i linearnom aktivacijskom funkcijom. Nadalje će se u tekstu ovaj model zvati Model 2.



**Slika 3.3.** Arhitektura duboke neuronske mreže s dvjema LSTM mrežama korištena za ocjenjivanje video signala u sklopu ovog diplomskog rada (Model 2)

### 3.3.4. Upute za pokretanje

Prvi korak za pokretanje rješenja je instalirati *Anaconda* okruženje, unutar kojeg treba stvoriti virtualno programsko okruženje. Za stvaranje virtualnog programskog okruženja koristi se datoteka *requirements.txt*, koja sadrži biblioteke potrebne za pokretanje algoritma i dostupna je na priloženom DVD-u u direktoriju P.3.1. Ono se stvara s naredbom:

```
conda create -n <ime_okruženja> --file <putanja_do_requirements.txt>
```

Stvoreno okruženje potrebno je aktivirati s naredbom:

```
conda activate <ime_okruženja>
```

Nakon toga potrebno je pripremiti *Excel* datoteku za *feature\_extraction.py*. Ona se sastoji od 6 stupaca: naziv video signala, DMOS ocjena, putanja do direktorija gdje se nalazi video signal, format video signala, rezolucija video signala i prikazana je na slici 3.4.

Vid_name	DMOS	Vid_dir	Format	Height	Width
sh13_50fps	37.0212	H:/Downloads/VQA	YUV	432	768
sh14_50fps	44.0813	H:/Downloads/VQA	YUV	432	768
sh15_50fps	57.5757	H:/Downloads/VQA	YUV	432	768
sh16_50fps	62.0745	H:/Downloads/VQA	YUV	432	768
joc_MPEG2_1	8.733333	H:/Downloads/FERIT_RTRK/			
joc_MPEG2_1	8.727273	H:/Downloads/FERIT_RTRK2/			
joc_MPEG2_2	7.363636	H:/Downloads/FERIT_RTRK2/			
joc_MPEG2_3	5.181818	H:/Downloads/FERIT_RTRK2/			

**Slika 3.4.** Izgled *Excel* datoteke

Nakon toga potrebno je otvoriti *feature\_extraction.py*, koja se nalazi na DVD-u u direktoriju P.3.2, unutar *Jupyter Notebook* okruženja i promijeniti varijablu *dataset* s direktorijem gdje se nalazi *Excel* datoteka, te promijeniti u funkciji *feature\_extraction* nazive *numpy* polja koja se spremaju. Prvo *numpy* polje sadrži značajke svih video signala za pojedini tip izobličenja, dok drugo sadrži DMOS ocjene.

Nakon što se završi proces izvlačenja značajki, potrebno je otvoriti *python* datoteku *train.py*, koja se nalazi na DVD-u u direktoriju P.3.2, unutar koje se mora promijeniti ime datoteke koja se učitava s naredbom *np.load()* za *X\_train*, *y\_train*, *X\_val* i *y\_val* varijable. Pokrenuti sve ćelije unutar *Jupiter Notebook-a*, nakon čega će početi treniranje modela. Nadalje, otvoriti *test.py*, koja se nalazi na DVD-u u direktoriju P.3.2, gdje se mora promijeniti ime datoteke koja se učitava s *np.load()* za *X\_test* i *y\_test* varijable. Pokrenuti sve ćelije u *test.py* poslije čega će se provesti testiranje naredbom *predict()* za dani testni skup. Poslije toga će se prikazati graf odnosa subjektivnih ocjena i ocjena predviđenih modelom, te PLCC i SROCC vrijednosti između tih dvaju skupova ocjena.

### **3.3.5. Proces treniranja predloženih dubokih neuronskih mreža za ocjenu kvalitete video signala**

Treniranje predloženih neuronskih mreža zahtjeva stvaranje baze podataka na temelju dostupnih baza video signala s prisutnim oštećenjima nastalim dodavanjem mrežnih artefakata i različitim algoritmima za kompresiju video signala. Pripremanje baze odrađivalo se na 2 načina. Kod prvog načina skupovi za treniranje, validaciju i testiranje sadrže sve sadržaje dostupne po tipu

izobličenja, dok je drugi način koristio podjelu prema sadržaju gdje skupovi za treniranje, validaciju i testiranje sadrže različite sadržaje. Kod drugog načina sadržaji su se dijelili približno (tj. što je moguće bliže) u odnosu 70:15:15, gdje 70% sadržaja ide u skup za treniranje, a po 15% u skupove za testiranje i validaciju, točniji odnos će biti napisan pri predstavljanju rezultata treniranja. Kod prvog pristupa se zapravo svaki sadržaj nalazio u svakom od triju skupova, dok se kod drugog pristupa nijedan sadržaj nije nalazio istovremeno u dvama različitim skupovima.

Za svaki način podjele baze podataka izvršeno je 9 odvojenih treniranja, validacija i testiranja. Od tih 9 odvojenih treniranja, njih 7 se izvršavalo za različite tipove izobličenja: H.264, H.265, MPEG2, MPEG4, zamrzavanje, % gubitka paketa pri prijenosu IP mrežom i % gubitka paketa pri prijenosu bežičnom mrežom. Potom su zadnja 2 treniranja bila odrađena na način da su se grupirali svi signali s izobličenjima nastalim isključivo zbog kompresije te svi signali s izobličenjima nastalim zbog prijenosa mrežom. U tablici 3.1 nalazi se primjer validacijskog skupa za prvi način treniranja modela na H.264 skupu, a u tablici 3.2 validacijski skup za drugi način treniranja istog tipa izobličenja. S obzirom da FERIT-RTRK1 i FERIT-RTRK2 baze imaju jednake sadržaje, za prvi način treniranja za validacijski skup uzeti su video signali iz FERIT-RTRK1 baze.

**Tablica 3.1.** Validacijski skup za prvi način treniranja modela za H.264 tip izobličenja

Ime baze:	Validacijski video signali:
FERIT-RTRK1	bea_H264_3, bos_H264_4, hon_H264_4, joc_H264_5, rea_H264_4, yac_H264_1
EVVQ	cartoon_6, cheerleaders_3, discussion_4, flower_garden_1, football_3, mobile_1, town_plan_1, weather_4
ECVQ	container_ship_5, flower_garden_2, football_3, foreman_6, hall_2, mobile_4, news_6, silent_1
LIVE-VQA	bs12_25fps, mc9_50fps, pa9_25fps, pr11_50fps, rb12_25fps, rh10_25fps, sf11_25fps, st12_25fps, tr9_25fps
MOBILE-LIVE	bf_r4, dv_r3, hc_r1, la_r2, po_r3, sd_r1, ss_r4, tk_r2

**Tablica 3.2.** Validacijski skup za drugi način treniranja modela za H.264 tip izobličenja

Ime baze:	Validacijski video signali:
FERIT-RTRK1	bea_H264_1, bea_H264_2, bea_H264_3, bea_H264_4, bea_H264_5

FERIT-RTRK2	bea_H264_1, bea_H264_2, bea_H264_3
EVVQ	discussion_1, discussion_2, discussion_3, discussion_4, discussion_5, discussion_6
ECVQ	silent_1, silent_2, silent_3, silent_4, silent_6, silent_6,
LIVE-VQA	mc11_50fps, mc12_50fps, mc10_50fps, mc9_50fps
MOBILE-LIVE	bf_r1, bf_r2, bf_r3, bf_r4

Kako bi se mogla pratiti uspješnost treniranja i validacije modela, potrebno je odabrati mjere za gubitak i točnost. One se implementiraju unutar optimizatora koji se definira na kraju modela, prije stvaranje objekta modela. Za optimizator odabran je ADAM (engl. *Adaptive Moment Estimation*) [34], koji dinamički prilagođava brzinu učenja tijekom treninga na temelju procjene gradijenta. Gradijent mjeri promjenu greške iz iteracije u iteraciju i s obzirom na iznos promjene mijenjaju se težine (engl. *weights*). ADAM pomaže u postizanju brže konvergencije ka minimalnom gubitku ili ka maksimalnoj točnosti. Za gubitak je odabrana mjera MSE i ona računa srednju kvadratnu pogrešku između dviju vrijednosti (formula (3-1)), točnije za potrebe diplomskog je računala srednju kvadratnu pogrešku između subjektivne ocjene videa i modelom predviđene ocjene za isti video signal:

$$MSE = \frac{1}{n} * \sum (y_{DMOS} - y_{PRED})^2 \quad (3-1)$$

Za mjeru točnosti korišten je Pearson-ov linearni koeficijent korelacije (engl. *PLCC – Pearson linear correlation coefficient*), koji prikazuje na određeni način koliko su ocjene predviđene modelom slične subjektivnim ocjenama. Računa se prema formuli (3-2), gdje je  $y_{DMOS}$  subjektivna ocjena video signala,  $y_{PRED}$  predviđena ocjena video signala, a  $n$  broj parova video signala. Rezultat PLCC može biti između -1 i 1, gdje je -1 maksimalna negativna korelacija, što znači da kada se jedna varijabla mijenja druga se mijenja u suprotnom smjeru. PLCC jednak 1 označava maksimalnu pozitivnu korelaciju. Ako je PLCC jednak 0, to znači da ne postoji nikakva korelacija između varijabli [35].

$$PLCC = \frac{n * \sum y_{DMOS} y_{PRED} - \sum y_{DMOS} * \sum y_{PRED}}{\sqrt{(n * \sum y_{DMOS}^2 - (\sum y_{DMOS})^2) * (n * \sum y_{PRED}^2 - (\sum y_{PRED})^2)}} \quad (3-2)$$

Za mjeru monotonosti korišten je Spearman-ov koeficijent korelacije ranga (engl. *SROCC – Spearman's rank correlation coefficient*). SROCC je statistička mjera koja kvantificira jačinu i smjer međusobne monotonosti dvaju nizova. Vrijednost SROCC može biti između -1 i 1, gdje je -1 savršeni negativni monotoni odnos, a 1 savršeni pozitivni monotoni odnos. Ako je vrijednost 0,

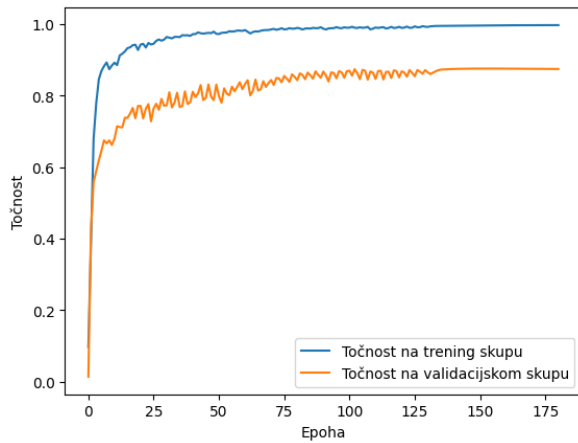
tada nema nikakvog monotonog odnosa. SROCC je ne parametarska mjera što znači da ne pretpostavlja nikakvu specifičnu distribuciju podataka [36].

Implementirane su funkcije za ranije zaustavljanje treniranja, za automatsko smanjivanje brzine učenja i za spremanje modela kada se postigne manji gubitak na validacijskom skupu podataka u odnosu na dotad najbolji model. Funkcija za ranije zaustavljanje treniranja će zaustaviti treniranje u slučaju da se validacijski gubitak nije smanjio unutar 80 uzastopnih epoha. Validacijski gubitak računa se svaku epohu i uvijek je različit i u slučaju da se 80 epoha za redom on nije smanjio od trenutno najmanjeg, tada se zaustavlja treniranje. Funkcija za automatsko smanjivanje brzine učenja će smanjiti brzinu učenja za 50% trenutne brzine učenja svakih 30 epoha ako validacijski gubitak bude veći od trenutnog najmanjeg validacijskog gubitka. Početna brzina učenja postavljena je na 0.001.

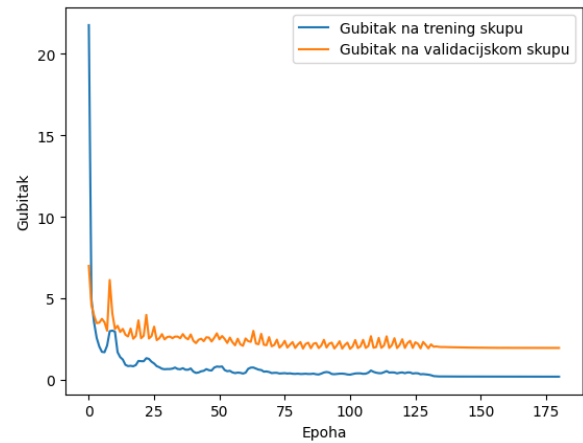
### **3.3.5.1. Treniranja modela za ocjenjivanje video signala uz prvi način podjele baze podataka**

Prvi načina treniranja obuhvaća podjelu video signala na način da se za svako izobličenje u skupove za treniranje, testiranje i validaciju stavljaju svi video sadržaji, podjela se nalazi na DVD-u u direktoriju P.3.3. Na taj način se može provjeriti hoće li model uspješno ocjenjivati video signale različite razine izobličenja ako je istim sadržajima imao pristup u skupu za treniranje. Na slici 3.5 se nalaze grafovi za točnost i gubitak za trening i validacijske skupove u odnosu na broj proteklih epoha za H.264 tip izobličenja obaju predložena modela. Prema grafovima priloženim na slici 3.5 vidi se da Model 2 ima veće skokove za vrijednost MSE i PLCC nego Model 1. Može se vidjeti da je gubitak trening skupa niži za Model 1 u odnosu na Model 2, dok je za validacijski skup gubitak niži kod Modela 2. To se može potvrditi i tablicom 3.5 gdje su prikazane najmanje vrijednosti MSE za trening i validacijske skupove. Nadalje, točnost obaju modela na trening skupu je jednaka, dok je za validacijski skup točnost nešto veća kod Modela 2, što se može potvrditi u tablicama 3.3 i 3.4. Rezultati iz grafova na slici 3.5 i tablicama 3.3 i 3.4 pokazuju da oba predložena modela uspješno predviđaju ocjenu kvalitete na obama skupovima podataka. Isto tako, ako se pogledaju tablice 3.3 i 3.4 za H.264 tip izobličenja može se vidjeti da predloženi modeli uspješnije predviđaju ocjene nego objektivne metrike zasnovane na egzaktnim računalnim algoritmima s kojima su uspoređivani. Odnos podjele skupova podataka za ovaj tip izobličenja je 64:18:18.

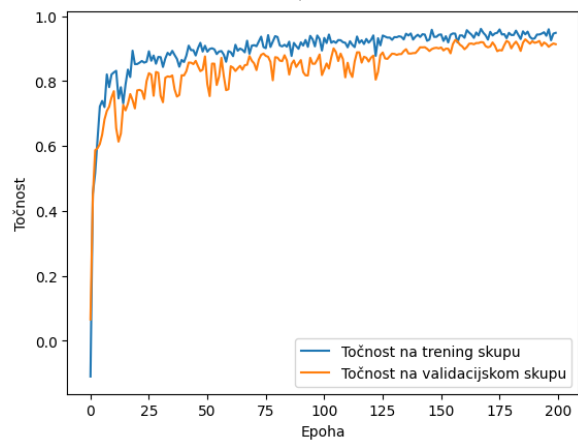




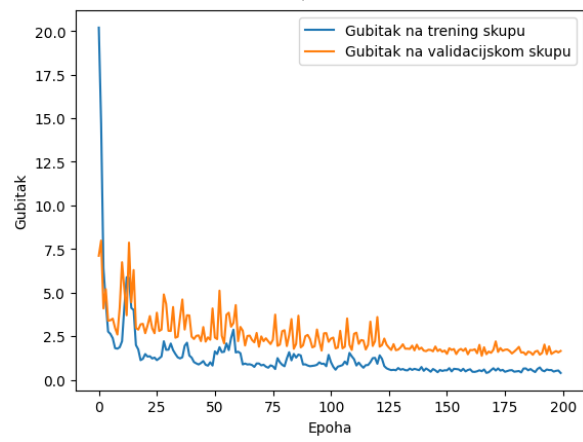
a)



b)



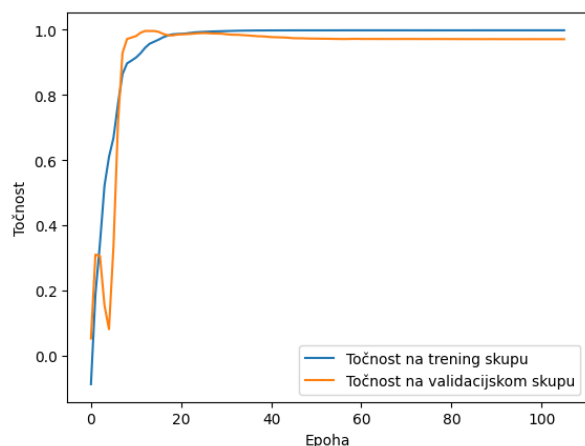
c)



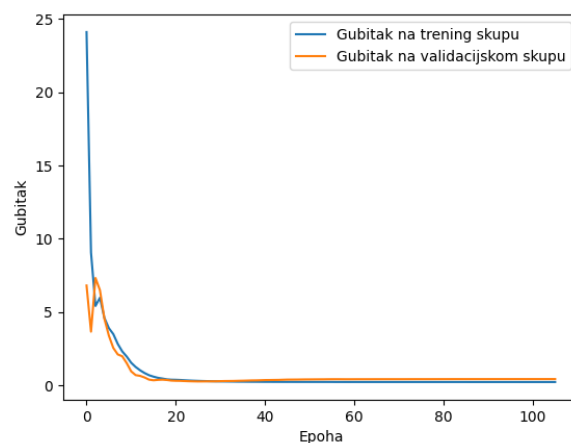
d)

**Slika 3.5.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za H.264 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

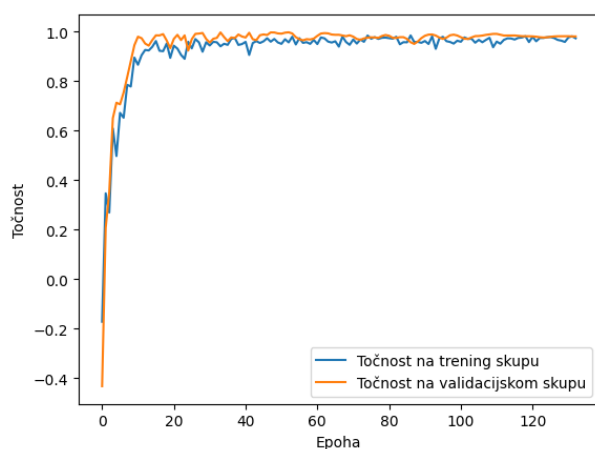
Slika 3.6 prikazuje rezultate treniranja i validacije za H.265 tip izobličenja za Model 1 i Model 2. Točnost i gubitak za oba modela postižu približno jednake vrijednosti na trening skupu, što se može vidjeti u tablicama 3.3 i 3.5. Model 2 postiže manji gubitak i nešto veću točnost na validacijskom skupu nego Model 1. Model 2 ima nešto više skokova na grafovima za točnost i gubitak prikazanim slikama 3.6 c) i d) za oba skupa, dok za Model 1 grafovi izgledaju glatko (slika 3.6 a) i b)). U odnosu na objektivne metrike zasnovane na egzaktnim računalnim algoritmima, oba modela postižu bolje rezultate za PLCC i SROCC za oba skupa. Podjela skupova podataka za H.265 tip izobličenja je 75:12.5:12.5.



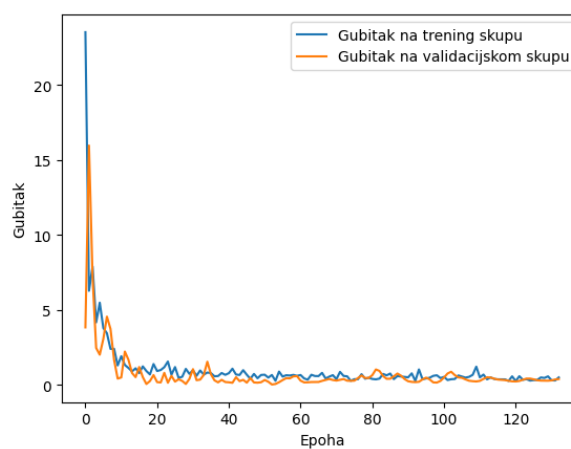
a)



b)



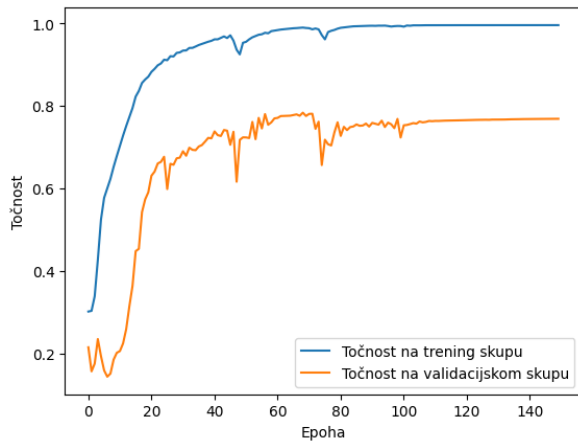
c)



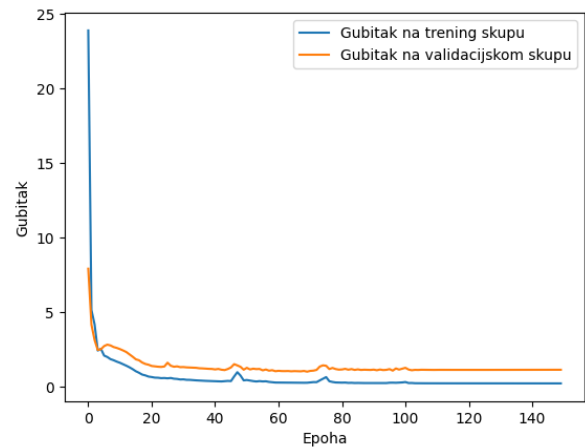
d)

**Slika 3.6.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za H.265 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

Rezultati treniranja za MPEG2 tip izobličenja prikazani na slici 3.7 pokazuju da je Model 1 ostvario bolje rezultate za točnost i gubitak za trening skup nego Model 2. Dok je Model 2 ostvario nešto bolje rezultate za gubitak i točnost na validacijskom skupu, što se može vidjeti u tablici 3.5. Isto kao i za H.264, prilikom treniranja Model 2 ima puno više skokova PLCC i MSE prikazano na slikama 3.7 c) i d), u odnosu na Model 1 gdje nema toliko skokova vrijednosti. Ako se pogledaju tablice 3.3 i 3.4 može se vidjeti da su PLCC i SROCC za trening i validacijski skup puno veći nego PLCC i SROCC dobiven između subjektivnih ocjena i objektivnih metrika s kojima su uspoređivani. Podjela skupova podataka za MPEG2 tip izobličenja je 68.2:15.9:15.9.



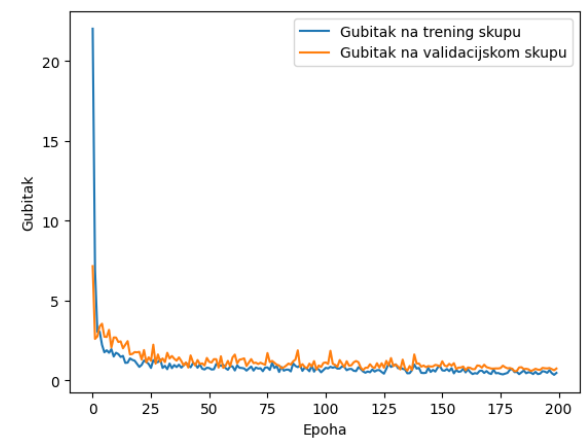
a)



b)



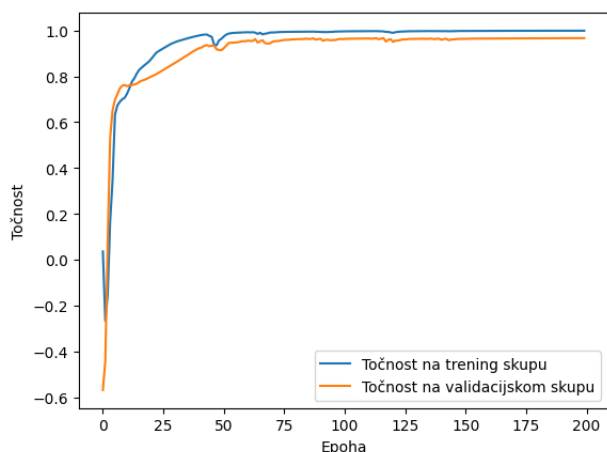
c)



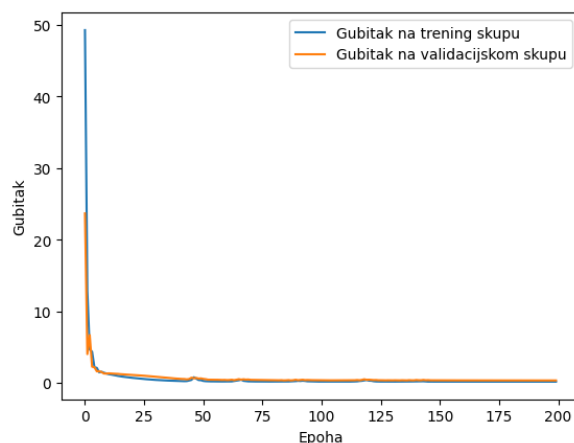
d)

**Slika 3.7.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za MPEG2 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

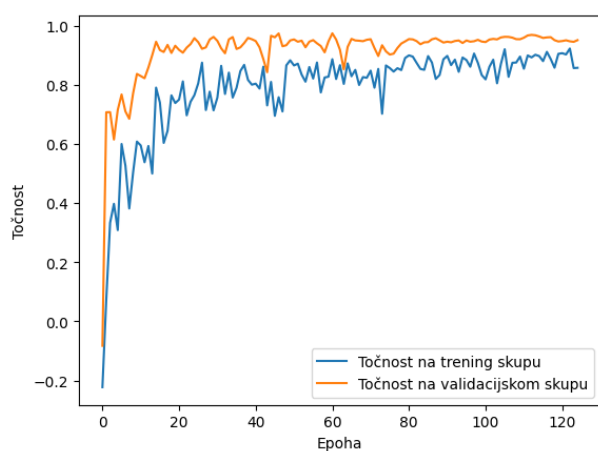
Na slici 3.8 prikazani su rezultati treniranja i validacije za MPEG4 tip izobličenja, iz čega se može vidjeti da su grafovi za Model 1 puno glađi nego za Model 2, na kojem ima puno skokova vrijednosti za točnost i gubitak. Iz tablice 3.3. može se vidjeti da Model 1 postiže nešto veću vrijednost PLCC nego Model 2 za trening skup, dok za validacijski skup oba modela imaju približno jednaku vrijednost PLCC. SROCC za validacijski skup veći je za Model 1, što se može potvrditi tablicom 3.4. Predloženi modeli uspješno predviđaju ocjene za oba skupa podataka. Prema tablici 3.5. može se vidjeti da Model 2 postiže manju vrijednost MSE nego Model 1 na validacijskom skupu. Predloženi modeli postigli su bolje vrijednosti za PLCC i SROCC u odnosu na PSNR, SSIM, MS-SSIM i NIQE metrike za oba skupa. Podjela skupova podataka za ovaj tip izobličenja je 64.4:16.8:16.8.



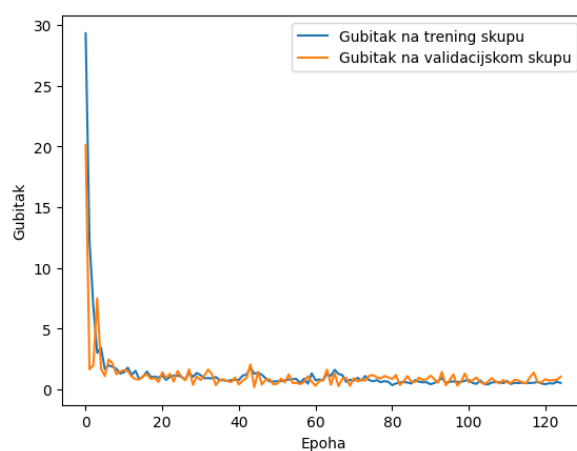
a)



b)



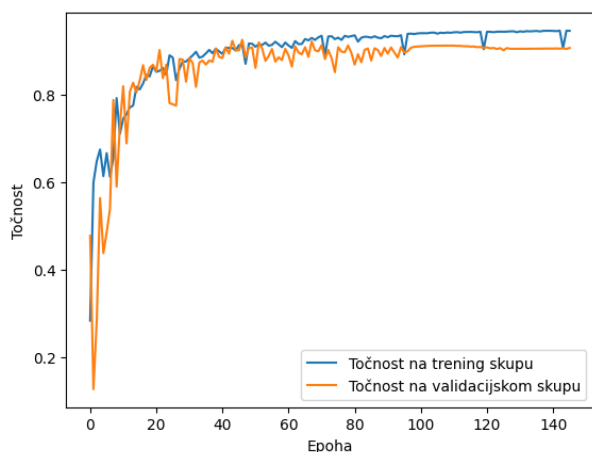
c)



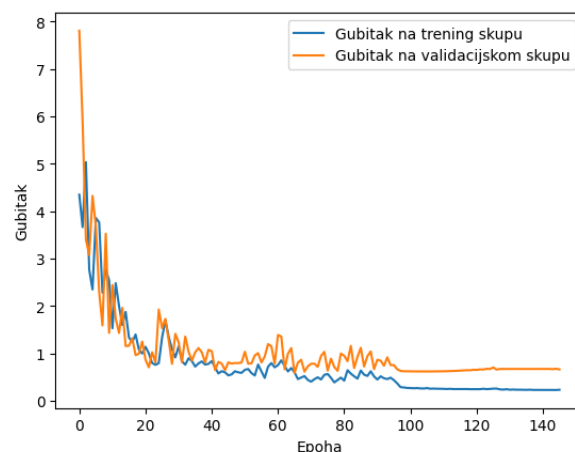
d)

**Slika 3.8.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za MPEG4 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

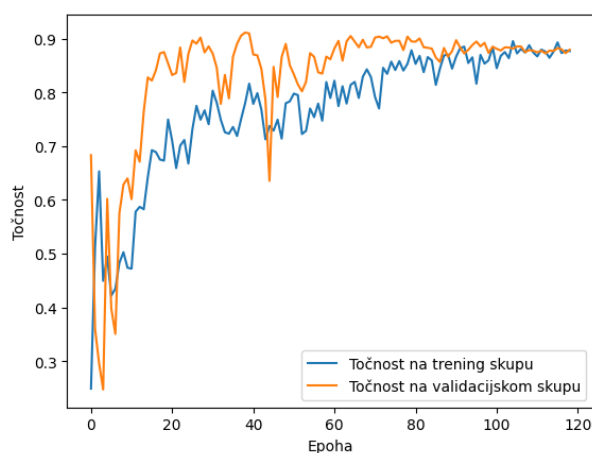
Na slici 3.9 prikazani su rezultati treniranja za izobličenje nastalo smrzavanjem, a Model 1 postiže bolje rezultate nego Model 2. Na slici 3.9 d) može vidjeti da dolazi do jako velikih skokova vrijednosti MSE prilikom treniranja na oba skupovima. U tablici 3.5. vidljivo je da Model 1 postiže manji gubitak nego Model 2 za trening skup, dok Model 2 postiže manji gubitak i veću točnost za validacijski skup. U tablicama 3.3 i 3.4 vidljivi su rezultati PLCC i SROCC za oba skupa koji to potvrđuju. Isto tako vidljivo je da predloženi modeli postižu puno bolje rezultate nego objektivne metrike zasnovane na egzaktnim računalnim algoritmima s kojima su uspoređivani. Podjela skupova podataka za ovaj tip izobličenja je 69.8:15.1:15.1.



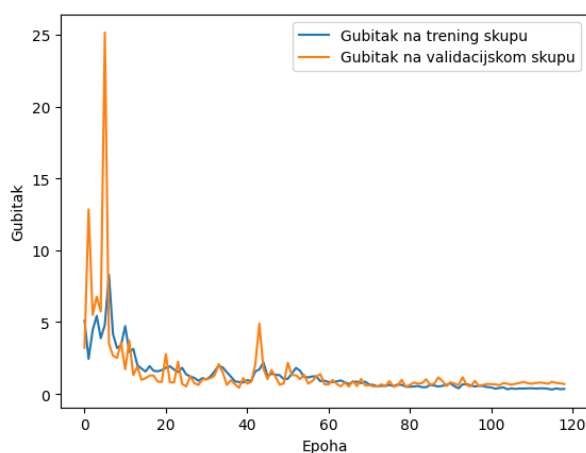
a)



b)



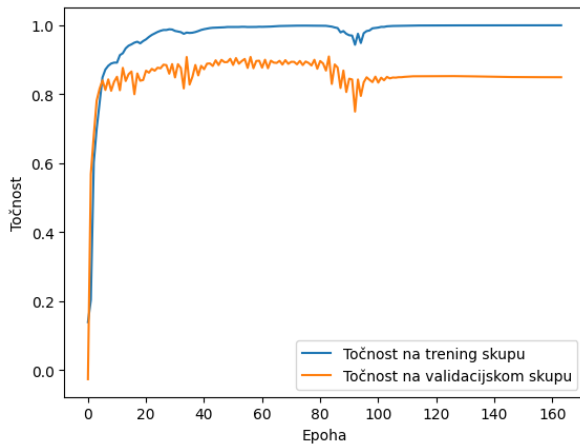
c)



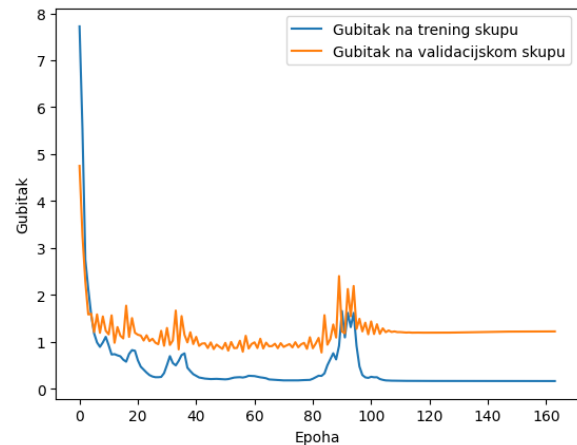
d)

**Slika 3.9.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za izobličenje nastalo smrzavanjem: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

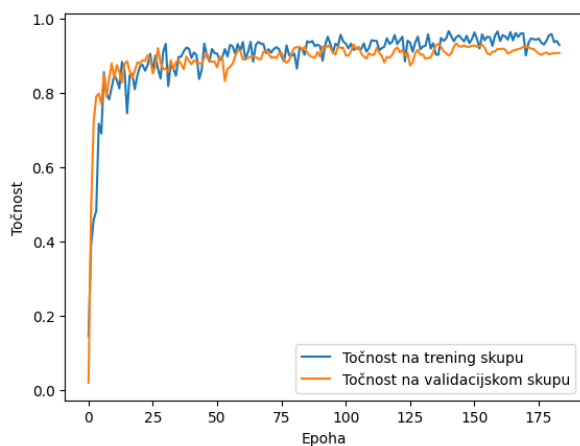
Iz rezultata treniranja za izobličenja nastala određenim % gubitka paketa pri prijenosu IP mrežom prikazanih na slici 3.10, vidljivo je kako Model 1 postiže veću vrijednost PLCC i manju vrijednost MSE nego Model 2 za trening skup. S druge strane Model 2 postiže bolje rezultate za točnost i gubitak na validacijskom skupu nego Model 1, što se može vidjeti i u tablicama 3.4 i 3.5. Grafovi za točnost i gubitak malo su pravilniji za Model 2, tj. ima manje skokova vrijednost PLCC-a i MSE-a. Iz tablice 3.3 i 3.4 prikazano je da oba predložena modela postižu bolje rezultate nego objektivne metrike zasnovane na egzaktnim računalnim algoritmima s kojima su uspoređeni. Podjela skupova podataka za ovaj tip izobličenja je 75.4:12.3:12.3.



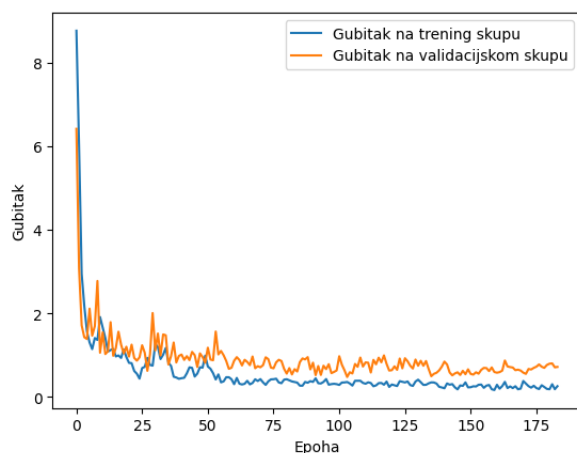
a)



b)



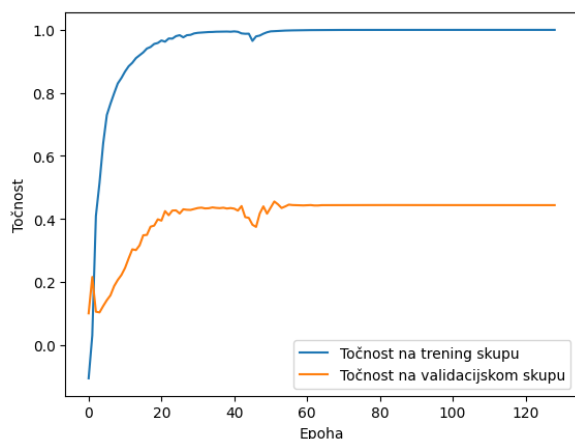
c)



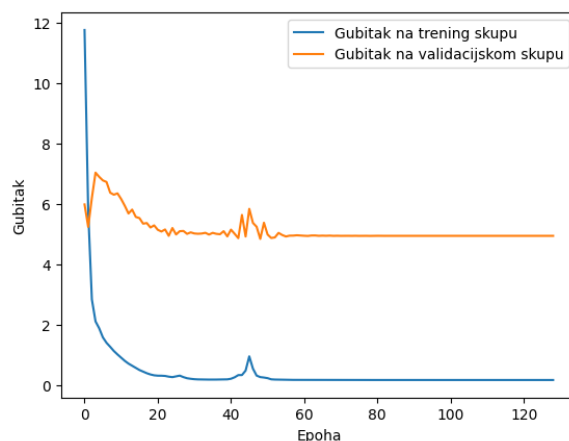
d)

**Slika 3.10.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za izobličenje nastalo određenim % gubitka paketa IP mreže: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražen preko PLCC, d) gubitak za Model 2 izražena preko MSE

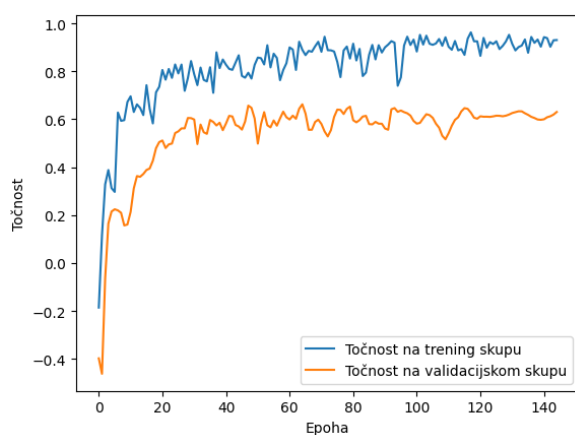
Iz rezultata treniranja i validacije za izobličenja nastala određenim % gubitka paketa pri prijenosu bežičnom mrežom prikazanih na slici 3.11, vidljivo je kako Model 2 postiže veću vrijednost PLCC i manju vrijednost MSE nego Model 1. Grafovi na slikama 3.11 c) i d) prikazuju veliki broj skokova vrijednosti PLCC i MSE tijekom treniranja za Model 2 u odnosu na Model 1. Isto tako, ostvarena vrijednost točnosti može se potvrditi ako se pogledaju tablice 3.3 i tablice 3.4. Nadalje, Model 2 je ostvario bolje PLCC i SROCC za oba skupa u odnosu na objektivne metrike zasnovane na egzaktnim računalnim algoritmima s kojima je uspoređivan. Model 1 se nije pokazao dobrim za ovaj tip izobličenja i to se može vidjeti iz tablica 3.3, 3.4 i 3.5. Podjela skupova podataka za ovaj tip izobličenja je 62.4:18.8:18.8.



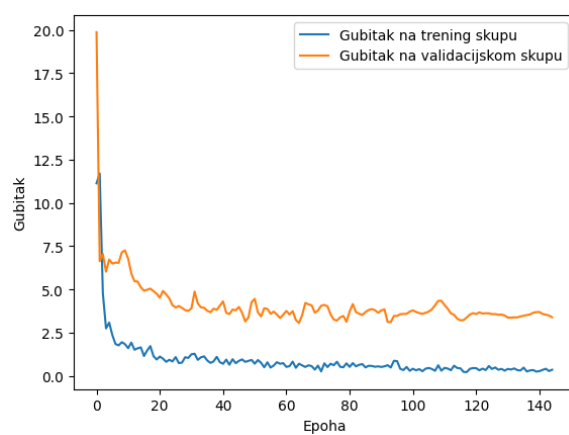
a)



b)



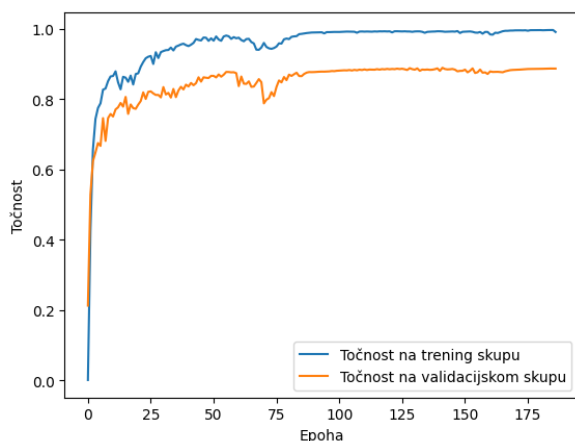
c)



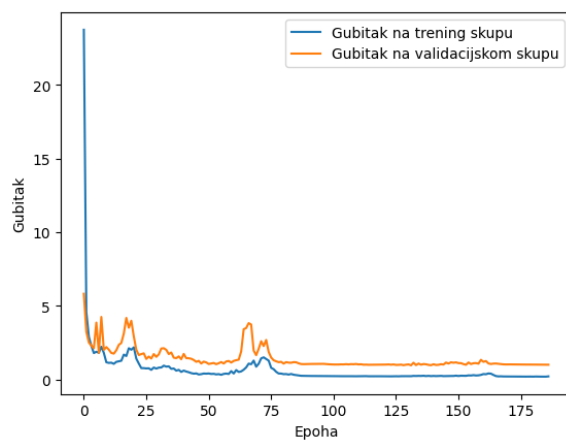
d)

**Slika 3.11.** Grafovi dobiveni nakon završetka treninga za trening i validacijski skup za izobličenje nastalo određenim % gubitka paketa bežične mreže: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

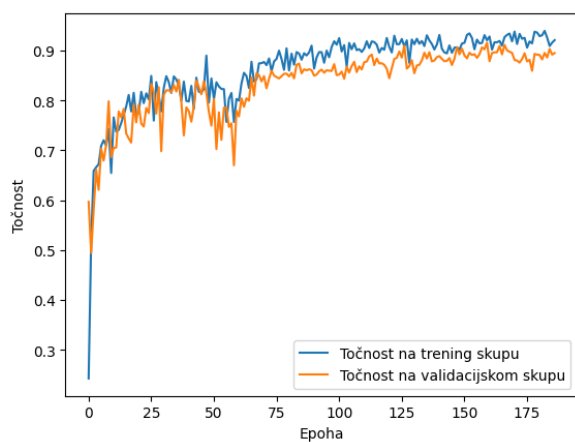
Na slici 3.12 mogu se vidjeti rezultati treniranja za izobličenje nastalo kompresijom za predložene modele. Prema rezultatima u tablicama 3.3 i 3.5 i grafovima 3.12 a) i c) može se vidjeti da je Model 1 ostvario bolje rezultate na skupu za treniranje u odnosu na Model 2, dok je Model 2 ostvario bolje rezultate za točnost i gubitak za validacijski skup, što je vidljivo u tablicama 3.4 i 3.5. Predloženi modeli ostvarili su puno bolje rezultate za PLCC i SROCC u odnosu na objektivne metrike s kojima su uspoređivani. Podjela skupova podataka za ovaj tip izobličenja je 66.4:16.8:16.8.



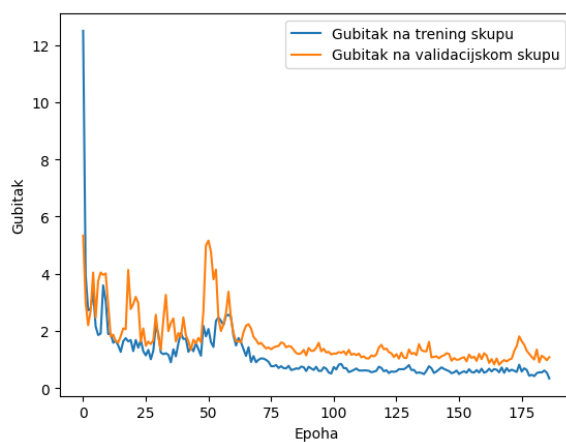
a)



b)



c)

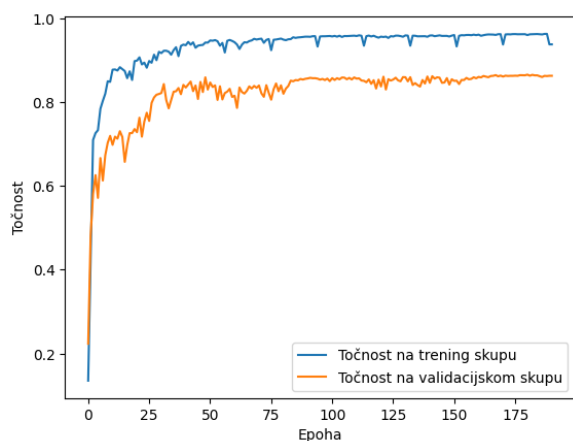


d)

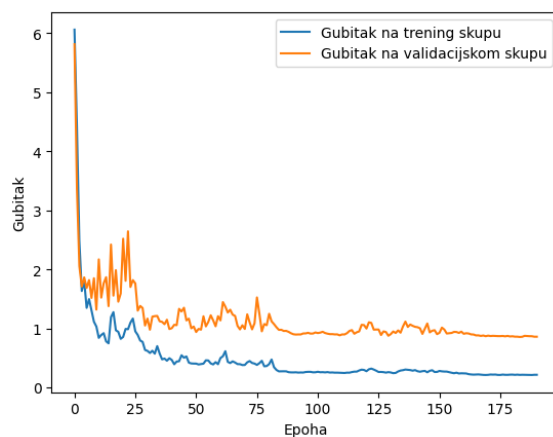
**Slika 3.12.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za izobličenja uzrokovana kompresijom: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

Na slici 3.13 mogu se vidjeti rezultati treniranja za Model 1 i Model 2 za izobličenja nastala pri prijenosu mrežom. Prema rezultatima u tablicama 3.3 i 3.5 i grafovima 3.13 a) i c) može se vidjeti da je Model 1 ostvario bolje rezultate na skupu za treniranje u odnosu na Model 2, dok je Model 2 ostvario bolje rezultate za točnost i gubitak za validacijski skup, što je vidljivo u tablicama 3.4 i 3.5. Predloženi modeli ostvarili su puno bolje rezultate za PLCC i SROCC u odnosu na objektivne metrike s kojima su uspoređivani. Podjela skupova podataka za ovaj tip izobličenja je 70.2:14.9:14.9. U tablicama 3.3 i 3.4 predstavljeni su rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na trening i validacijskim skupovima za prvi način podjele podataka, gdje su istaknuti najbolji rezultati postignuti za pojedini tip izobličenja. Tablica 3.5 predstavlja rezultate gubitka izražene u MSE predloženih modela na trening i validacijskim skupovima za pojedine tipove izobličenja.

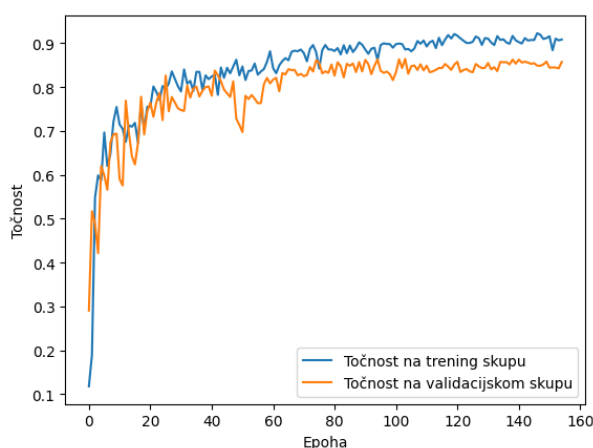




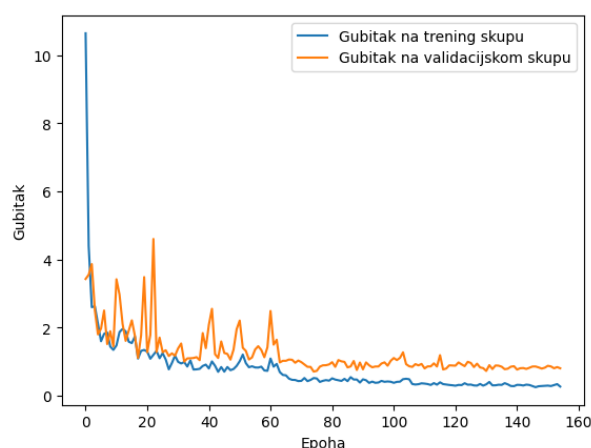
a)



b)



c)



d)

**Slika 3.13.** Grafovi dobiveni nakon završetka treninga za trening i validacijski skup za izobličenja nastala pri prijenosu mrežom: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

**Tablica 3.3.** Rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na trening skupovima podataka za prvi način podjele baze podataka

Tip izobličenja		PSNR	SSIM	MS-SSIM	NIQE	Model1	Model2
H264	PLCC	0.511	0.549	0.563	0.225	0.996	<b>0.991</b>
	SROCC	0.493	0.556	0.642	0.097	<b>0.994</b>	0.987
H265	PLCC	0.723	0.730	0.749	-0.710	<b>0.991</b>	0.989
	SROCC	0.700	0.657	0.673	-0.729	<b>0.987</b>	0.985
MPEG2	PLCC	0.486	0.703	0.739	-0.394	<b>0.994</b>	0.940
	SROCC	0.492	0.678	0.714	-0.412	<b>0.994</b>	0.942
MPEG4	PLCC	0.667	0.829	0.829	-0.274	<b>0.998</b>	0.971
	SROCC	0.674	0.810	0.772	-0.256	<b>0.992</b>	0.951
Smrzavanje	PLCC	0.707	0.651	0.577	0.101	<b>0.987</b>	0.929
	SROCC	0.732	0.664	0.607	0.233	<b>0.988</b>	0.951
	PLCC	0.514	0.438	0.318	0.369	<b>0.993</b>	0.979

% gubitka paketa pri prijenosu IP mrežom	SROCC	0.655	0.476	0.408	0.468	<b>0.994</b>	0.984
% gubitka paketa pri prijenosu bežičnom mrežom	PLCC	0.275	0.104	0.020	0.124	0.967	<b>0.991</b>
	SROCC	0.352	0.187	0.094	0.118	<b>0.960</b>	0.984
Izobličenja uzrokovana kompresijom	PLCC	0.257	0.619	0.681	0.151	<b>0.957</b>	0.934
	SROCC	0.244	0.662	0.727	0.058	<b>0.945</b>	0.922
Izobličenja nastala pri prijenosu mrežom	PLCC	0.577	0.396	0.328	0.260	<b>0.963</b>	0.934
	SROCC	0.628	0.420	0.374	0.345	<b>0.953</b>	0.928

**Tablica 3.4.** Rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na validacijskim skupovima podataka za prvi način podjele baze podataka

Tip izobličenja		PSNR	SSIM	MS-SSIM	NIQE	Model1	Model2
H264	PLCC	0.525	0.535	0.583	0.367	0.840	<b>0.902</b>
	SROCC	0.510	0.607	0.683	0.321	0.824	<b>0.881</b>
H265	PLCC	0.758	0.612	0.610	-0.510	<b>0.986</b>	0.994
	SROCC	0.829	0.486	0.658	-0.543	<b>0.829</b>	0.772
MPEG2	PLCC	0.591	0.817	0.741	0.001	0.782	<b>0.856</b>
	SROCC	0.771	0.882	0.806	0.359	<b>0.666</b>	0.662
MPEG4	PLCC	0.669	0.740	0.781	-0.319	<b>0.967</b>	0.963
	SROCC	0.762	0.889	0.892	-0.209	<b>0.988</b>	0.895
Smrzavanje	PLCC	0.887	0.565	0.545	-0.038	0.887	<b>0.898</b>
	SROCC	0.891	0.532	0.491	0.066	0.872	<b>0.896</b>
% gubitka paketa pri prijenosu IP mrežom	PLCC	0.668	0.482	0.452	0.657	0.885	<b>0.917</b>
	SROCC	0.686	0.409	0.417	0.765	0.744	<b>0.856</b>
% gubitka paketa pri prijenosu bežičnom mrežom	PLCC	0.445	0.391	0.402	0.226	0.418	<b>0.663</b>
	SROCC	0.586	0.524	0.399	0.536	0.300	<b>0.454</b>
Izobličenja uzrokovana kompresijom	PLCC	0.564	0.654	0.674	0.230	0.894	<b>0.929</b>
	SROCC	0.605	0.667	0.702	0.173	<b>0.895</b>	0.890
Izobličenja nastala pri prijenosu mrežom	PLCC	0.620	0.582	0.397	0.214	0.822	<b>0.851</b>
	SROCC	0.619	0.543	0.400	0.266	0.831	<b>0.843</b>

**Tablica 3.5.** Rezultati MSE postignuti predloženim modelima na trening i validacijskim skupovima podataka za prvi način podjele baze podataka

Tip izobličenja	MSE			
	Skup za treniranje		Skup za validaciju	
	Model1	Model2	Model1	Model2
H.264	0.1970	0.3940	1.9600	1.3743
H.265	0.2070	0.2060	0.9700	0.1068

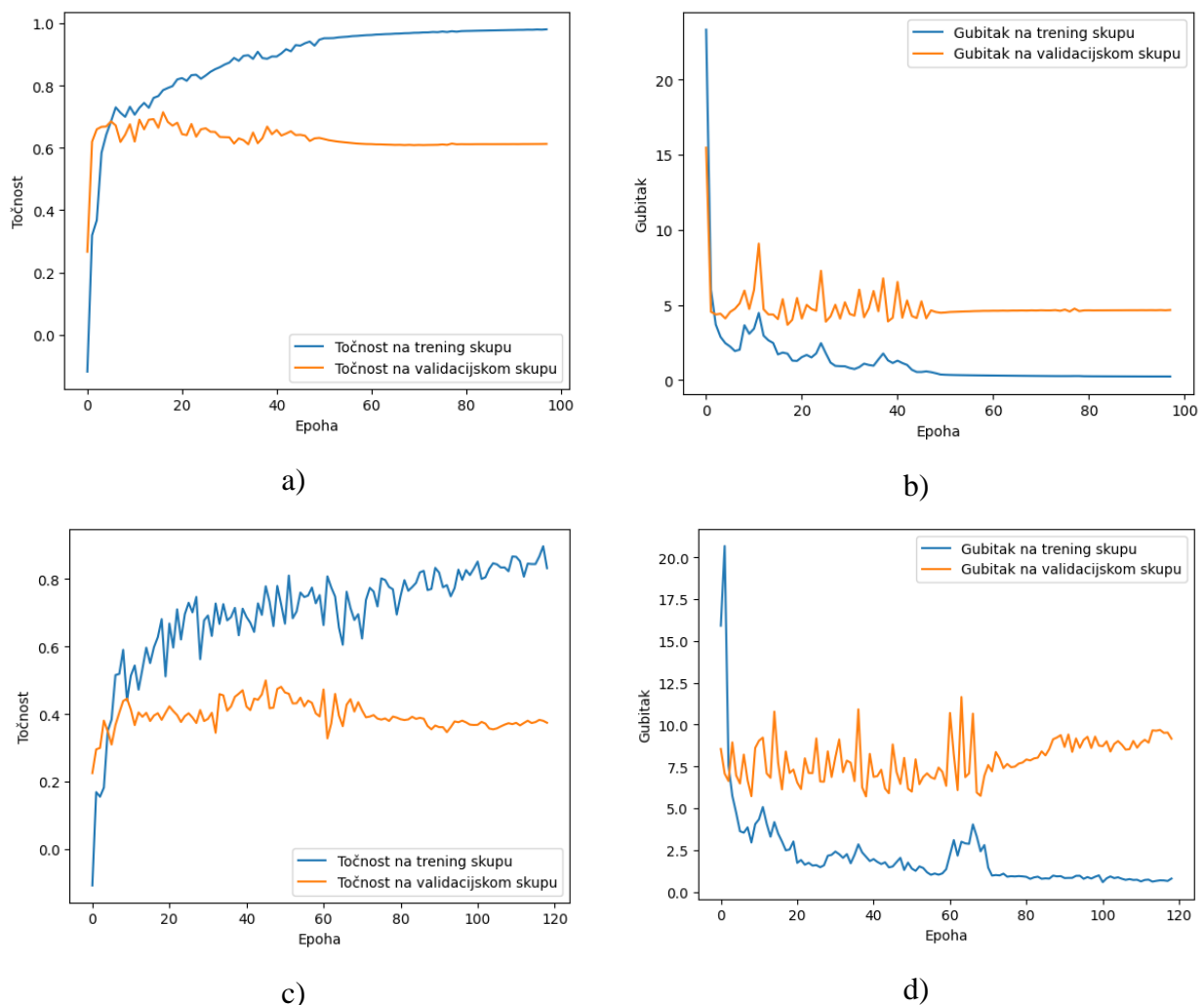
MPEG2	0.2065	0.3375	0.9938	0.5616
MPEG4	0.1678	0.3494	0.3393	0.1852
Smrzavanje	0.2265	0.3118	0.5893	0.4360
% gubitka paketa pri prijenosu IP mrežom	0.1657	0.1641	0.7741	0.4770
% gubitka paketa pri prijenosu bežičnom mrežom	0.1660	0.1560	4.8351	3.0510
Izobličenja uzrokovana kompresijom	0.1905	0.3438	0.9614	0.8230
Izobličenja nastala pri prijenosu mrežom	0.2119	0.2461	0.8556	0.7051

### 3.3.5.2. Treniranje modela za ocjenjivanje video signala uz drugi način podjele baze podataka

Drugi način treniranja modela odrađen je uz već ranije opisanu, nešto drugačiju podjelu video signala u tri skupa nego prvi način. Prema ovom načinu podjele, video signali se dijele tako da se baza podataka podijeli prema sadržaju i to na način da približno (što je bliže moguće) 70% signala bude u trening skupu, a po 15% u test i validacijskom skupu, što znači da će testni i validacijski skupovi sadržavati sadržaje koji nisu bili dostupni modelu za vrijeme treninga. Takva podjela podataka može se vidjeti u tablici 3.2. Ovim načinom hoće se provjeriti funkcionalnost istreniranog modela za različite sadržaje. Podjela video signala prema tipu izobličenja za ovaj način podjele nalazi na DVD-u u direktoriju P.3.4. U nastavku će biti napisan točan omjer skupova podataka za pojedini tip izobličenja.

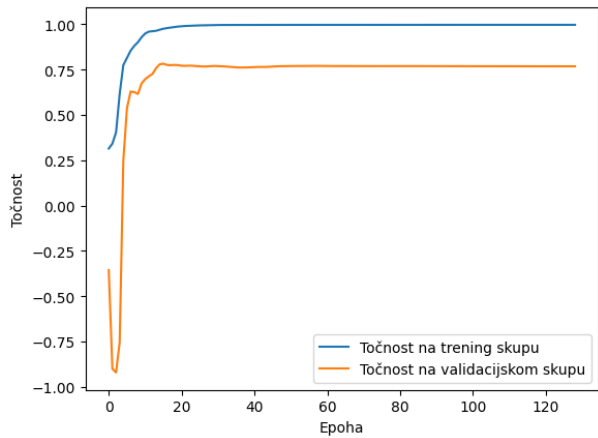
Na slici 3.14 mogu se vidjeti rezultati treniranja modela za H.264 tipa izobličenja uz drugi način podjele video signala. Model 1 poslije 20 epohe prestaje učiti, PLCC se počinje smanjivati što se može vidjeti na slici 3.14 a). To znači da se model počinje previše usklađivati na trening podatke jer točnost raste na trening skupu, a na validacijskom skupu pada. Tada se uzima model koji je bio najbolji na validacijskom skupu, bez obzira na to što postoji bolji model na trening skupu. Ipak, Model 1 postiže bolje rezultate validacije nego Model 2 i to je prikazano u tablici 3.7. Model 1 isto tako postiže manji gubitak za oba skupovima podataka u odnosu na Model 2, što je vidljivo u tablici 3.8. Ako se pogledaju tablice 3.6 i 3.7 može se vidjeti da Model 1 i Model 2 postižu puno bolje rezultate na trening skupu nego objektivne metrike s kojima su uspoređivani,

dok Model 1 postiže bolje rezultate na validacijskom skupu. Podjela skupova podataka za ovaj tip izobličenja je 84.3:7.85:7.85.

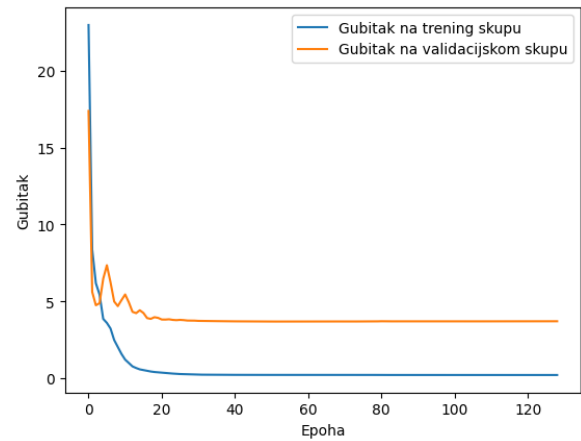


**Slika 3.14.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za H.264 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

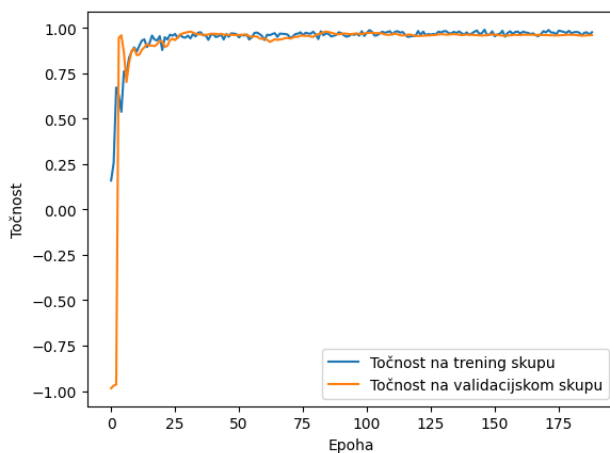
Na slici 3.15 mogu se vidjeti rezultati treniranja za H.265 tip izobličenja za predložene modele. Model 1 i Model 2 postižu približno jednake rezultate za točnost i gubitak na trening skupu, što se može potvrditi tablicama 3.6 i 3.8. S druge strane, Model 2 postiže puno bolje rezultate za točnost i gubitak na validacijskom skupu u odnosu na Model 1, vidljivo u tablici 3.7. Model 2 izrazito točno predviđa ocjene na validacijskom skupu. Predloženi modeli postižu puno bolje rezultate PLCC za trening skup u odnosu na objektivne metrike s kojima su uspoređivani, dok za validacijski skup Model 2 predviđa približno jednako kao i objektivne metrike. Podjela skupova podataka za ovaj tip izobličenja je 66.66:16.67:16.67.



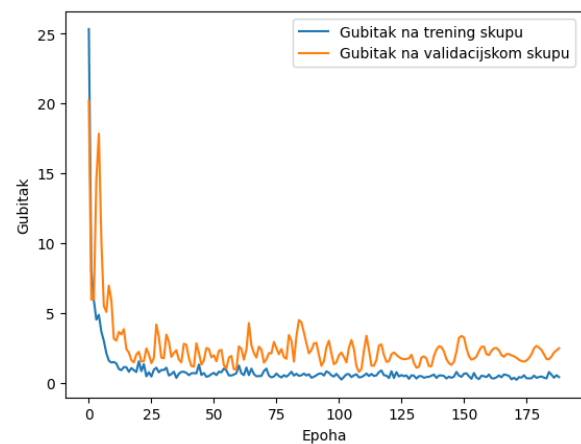
a)



b)



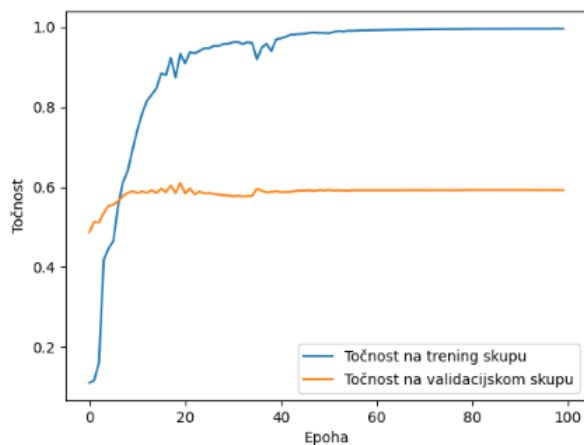
c)



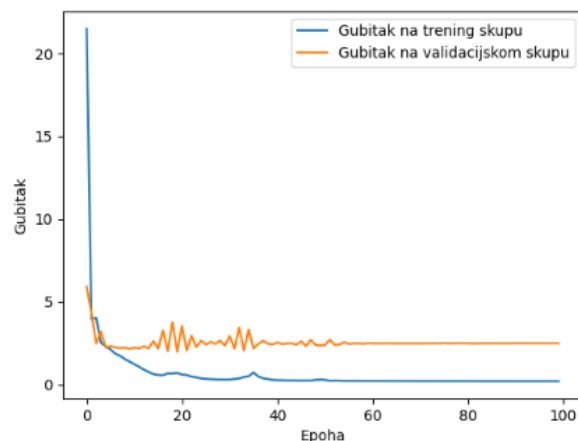
d)

**Slika 3.15.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za H.265 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

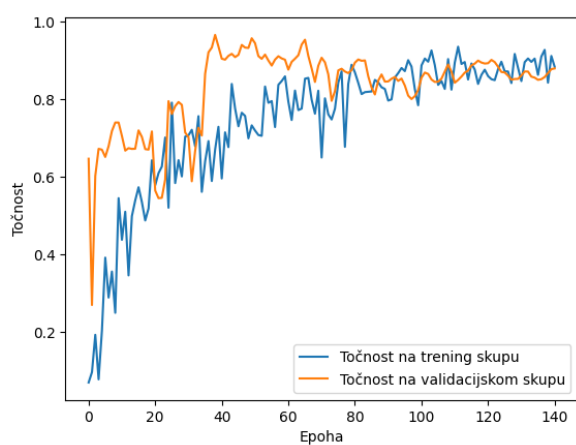
Na slici 3.16. prikazani su rezultati treniranja za MPEG2 tip izobličenja za predložene modele iz kojih se može vidjeti da je Model 1 postigao bolje rezultate za točnost i gubitak za trening skup od Modela 2, što se može potvrditi ako se pogleda u tablice 3.6 i 3.8. S druge strane, Model 2 je ostvario bolje rezultate za validacijski skup od Modela 1. Ako se usporede rezultati predviđanja ocjena za trening skup, oba su modela bolja nego objektivne metrike s kojima su uspoređivani. S druge strane za validacijski skup situacija je nešto drugačija i prikazana je tablicom 3.7. Model 1 ima najgori rezultat, dok je Model 2 prilično dobro predvidio ocjene za validacijski skup u odnosu na objektivne metrike. Podjela skupova podataka za ovaj tip izobličenja je 72.8:13.6:13.6.



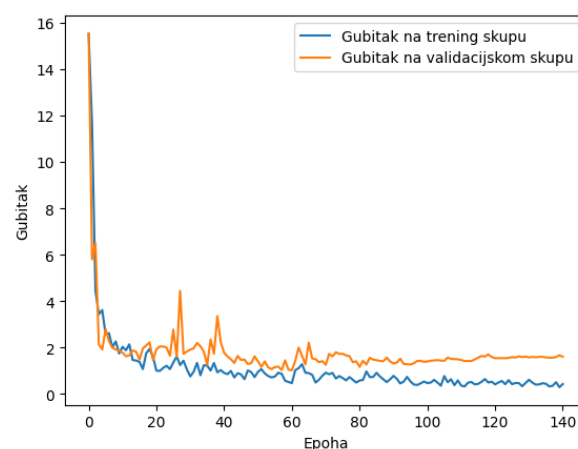
a)



b)



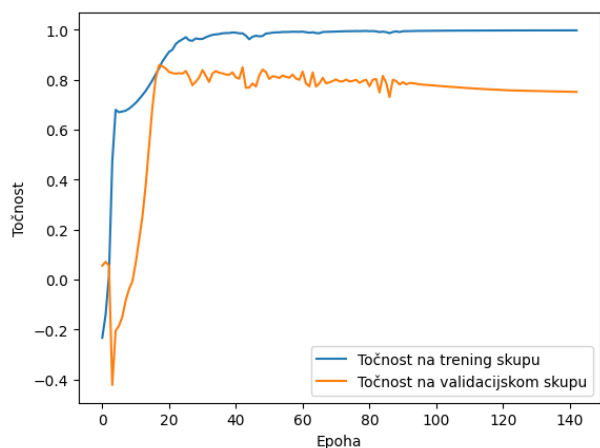
c)



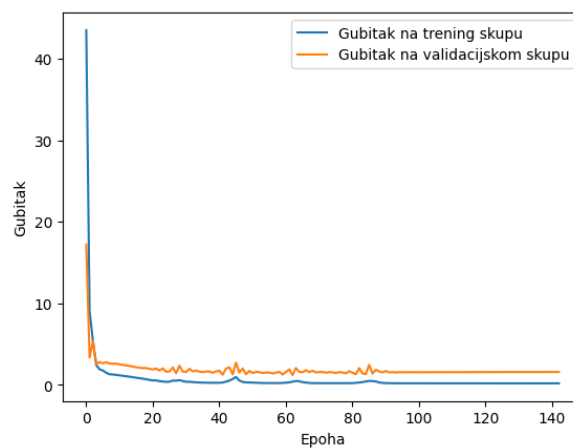
d)

**Slika 3.16.** Grafovi dobiveni nakon završetka treninga za trening i validacijski skup za MPEG2 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

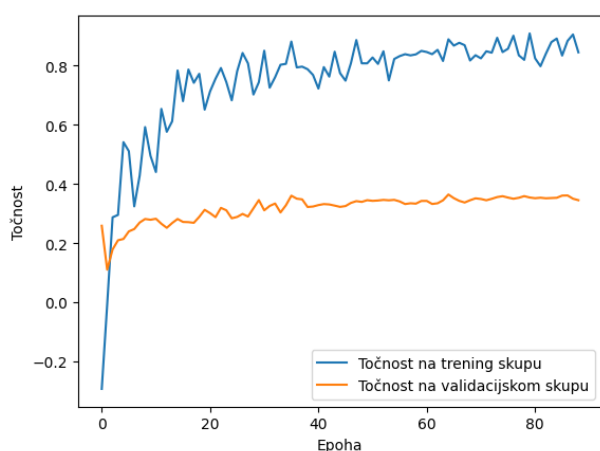
Na slici 3.17. prikazani su rezultati treniranja za MPEG4 tip izobličenja za predložene modele iz kojih se može vidjeti da je Model 1 postigao bolje rezultate za točnost i gubitak za trening skup i validacijski skup u odnosu na Model 2, što se može potvrditi ako se pogleda u tablice 3.6, 3.7 i 3.8. Ako se usporede rezultati predviđanja ocjena za trening skup, oba su modela bolja nego objektivne metrike s kojima su uspoređivani. S druge strane, za validacijski skup Model 1 ima najbolji rezultat, dok je Model 2 samo bolji od NIQE metrike. Podjela skupova podataka za ovaj tip izobličenja je 75.8:11.6:12.6.



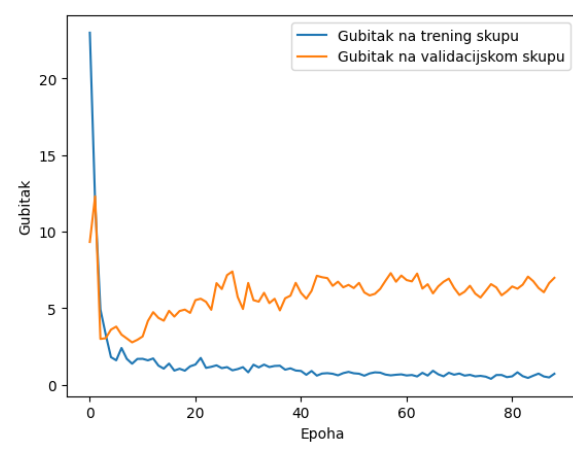
a)



b)



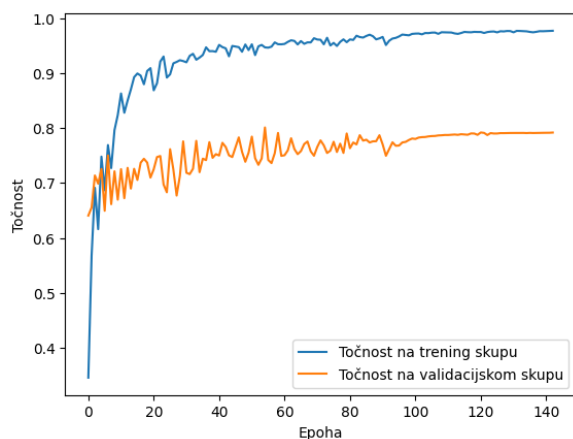
c)



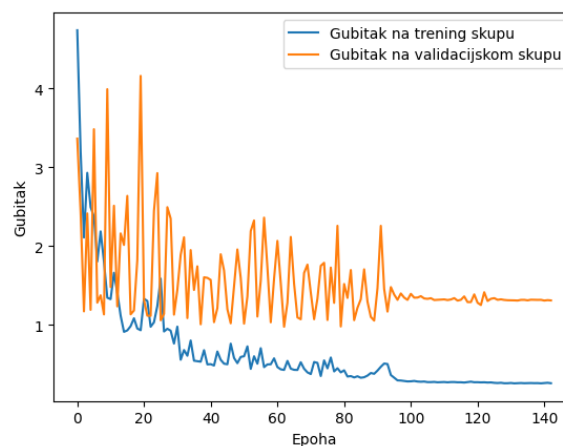
d)

**Slika 3.17.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za MPEG4 tip izobličenja: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

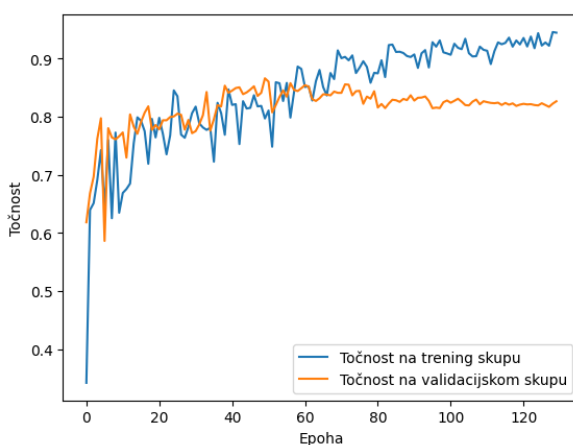
Na slici 3.18 mogu se vidjeti rezultati treniranja za izobličenja uzrokovana smrzavanjem za predložene modele. Model 1 i Model 2 postižu približno jednake rezultate za točnost i gubitak na trening skupu, što se može potvrditi tablicama 3.6 i 3.8. S druge strane, Model 2 postiže puno bolje rezultate za točnost i gubitak na validacijskom skupu u odnosu na Model 1, vidljivo u tablici 3.7. Model 2 točnije predviđa ocjene na validacijskom skupu nego Model 1. Predloženi modeli postižu puno bolje rezultate PLCC za trening skup u odnosu na objektivne metrike s kojima su uspoređivani, dok za validacijski skup Model 2 daje točnije predviđanje ocjena nego ostale objektivne metrike. Model 1 je bolji od SSIM i MS-SSIM objektivnih metrika. Podjela skupova podataka za ovaj tip izobličenja je 68.2:15.9:15.9.



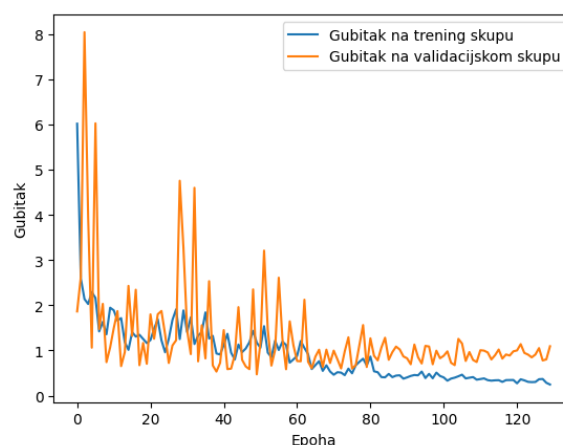
a)



b)



c)

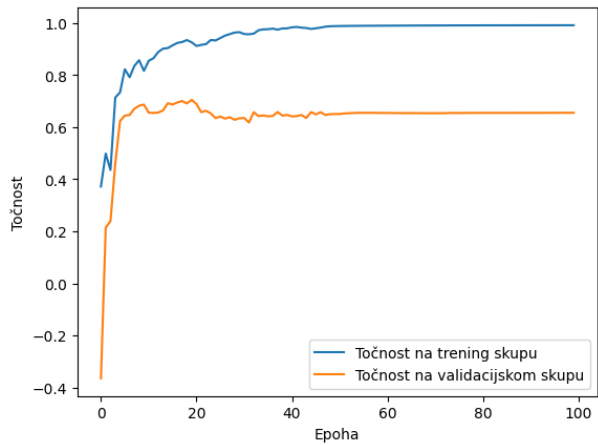


d)

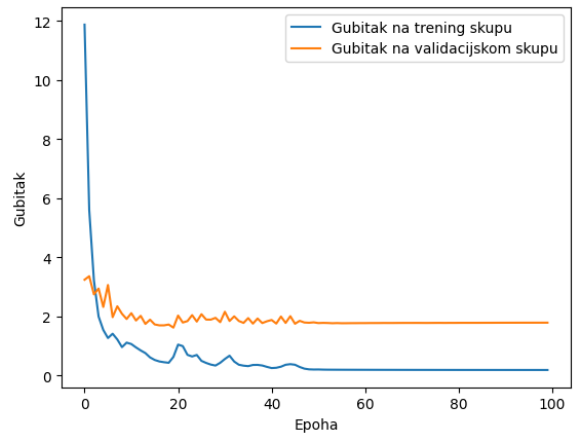
**Slika 3.18.** Grafovi dobiveni nakon završetka treninga za trening i validacijski skup za izobličenje nastalo smrzavanjem: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

Na slici 3.19 mogu se vidjeti rezultati treniranja za izobličenja uzrokovana određenim % gubitka paketa pri prijenosu IP mrežom za predložene modele. Model 1 postiže bolje rezultate za točnost i gubitak na trening skupu, što se može potvrditi tablicama 3.6 i 3.8. S druge strane, Model 2 postiže puno bolje rezultate za točnost i gubitak na validacijskom skupu u odnosu na Model 1, vidljivo u tablici 3.7. Model 2 točnije predviđa ocjene na validacijskom skupu nego Model 1. Predloženi modeli postižu puno bolje rezultate PLCC za trening skup u odnosu na objektivne metrike s kojima su uspoređivani. Za validacijski skup Model 2 daje točnije predviđanje ocjena nego ostale objektivne metrike, dok je Model 1 bolji od NIQE metrika. Podjela skupova podataka za ovaj tip izobličenja je 69.6:15.2:15.2.

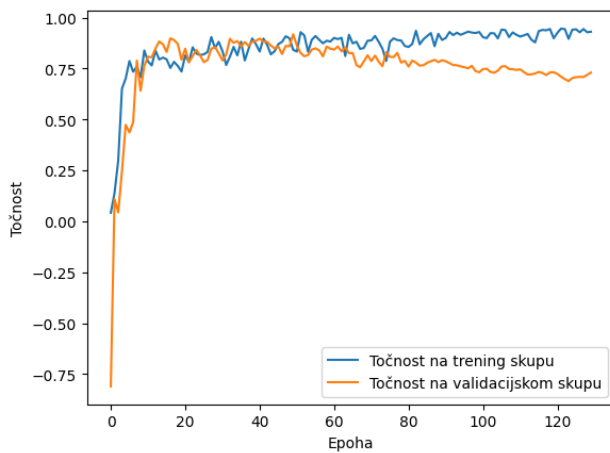




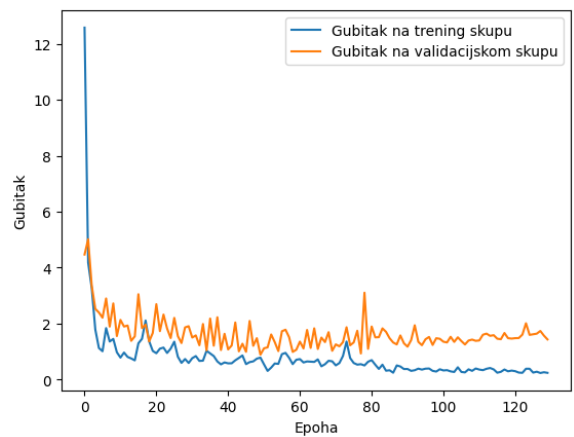
a)



b)



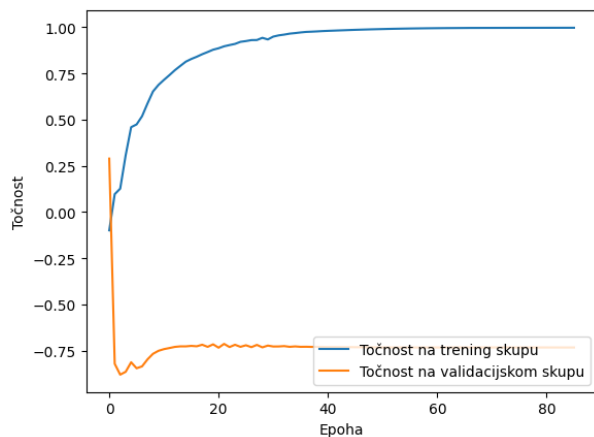
c)



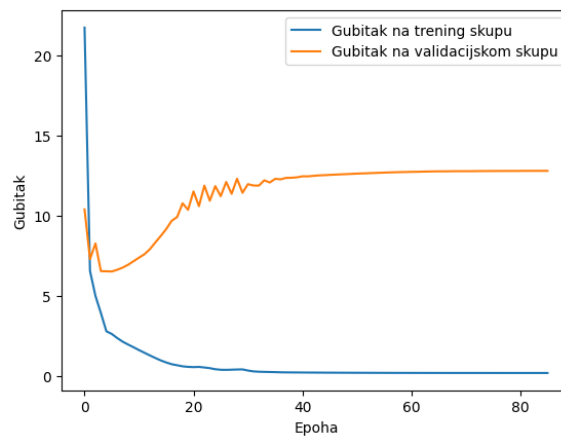
d)

**Slika 3.19.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za izobličenje nastalo određenim % gubitka paketa pri prijenosu IP mrežom: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražen preko PLCC, d) gubitak za Model 2 izražena preko MSE

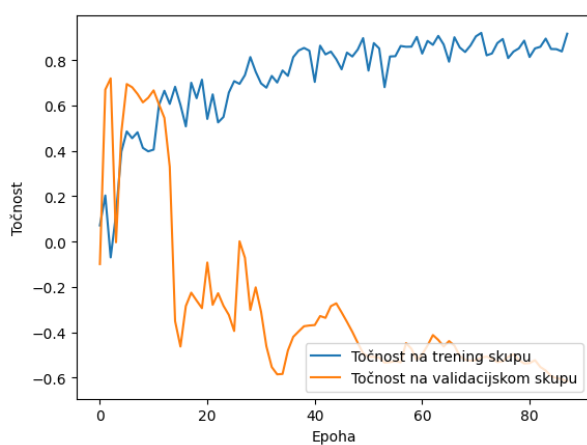
Na slici 3.20. prikazani su rezultati treniranja za izobličenja uzrokovana određenim % gubitka paketa pri prijenosu bežičnom mrežom za predložene modele iz kojih se može vidjeti da je Model 1 postigao bolje rezultate za točnost i gubitak za trening skup od Modela 2, što se može potvrditi ako se pogleda u tablice 3.6 i 3.8. S druge strane, Model 2 je ostvario bolje rezultate za validacijski skup od Modela 1. Ako se pogleda graf na slici 3.20 c), može se vidjeti kako Model 2 prestaje učiti na epohi broj 5 gdje postiže najveću validacijsku točnost i zatim ona počinje padati jer ima premalo podataka. Ako se usporede rezultati predviđanja ocjena za trening skup, oba su modela bolja nego objektivne metrike s kojima su uspoređivani. S druge strane, za validacijski skup situacija je nešto drugačija i prikazana je tablicom 3.7. Model 2 predviđa ocjene puno bolje od Modela 1, ali u odnosu na objektivne metrike s kojima je uspoređivan daje bolje rezultate samo od MS-SSIM i NIQE. Podjela skupova podataka za ovaj tip izobličenja je 80:10:10.



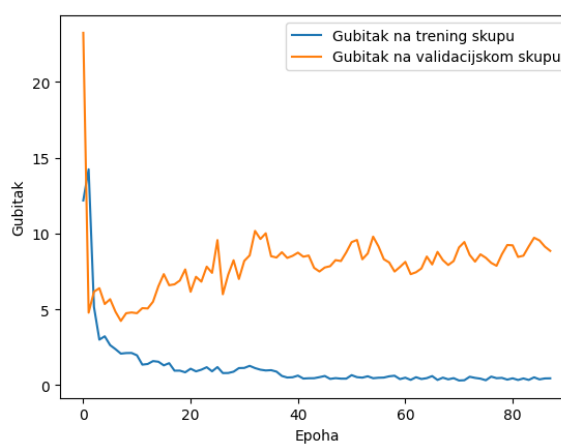
a)



b)



c)

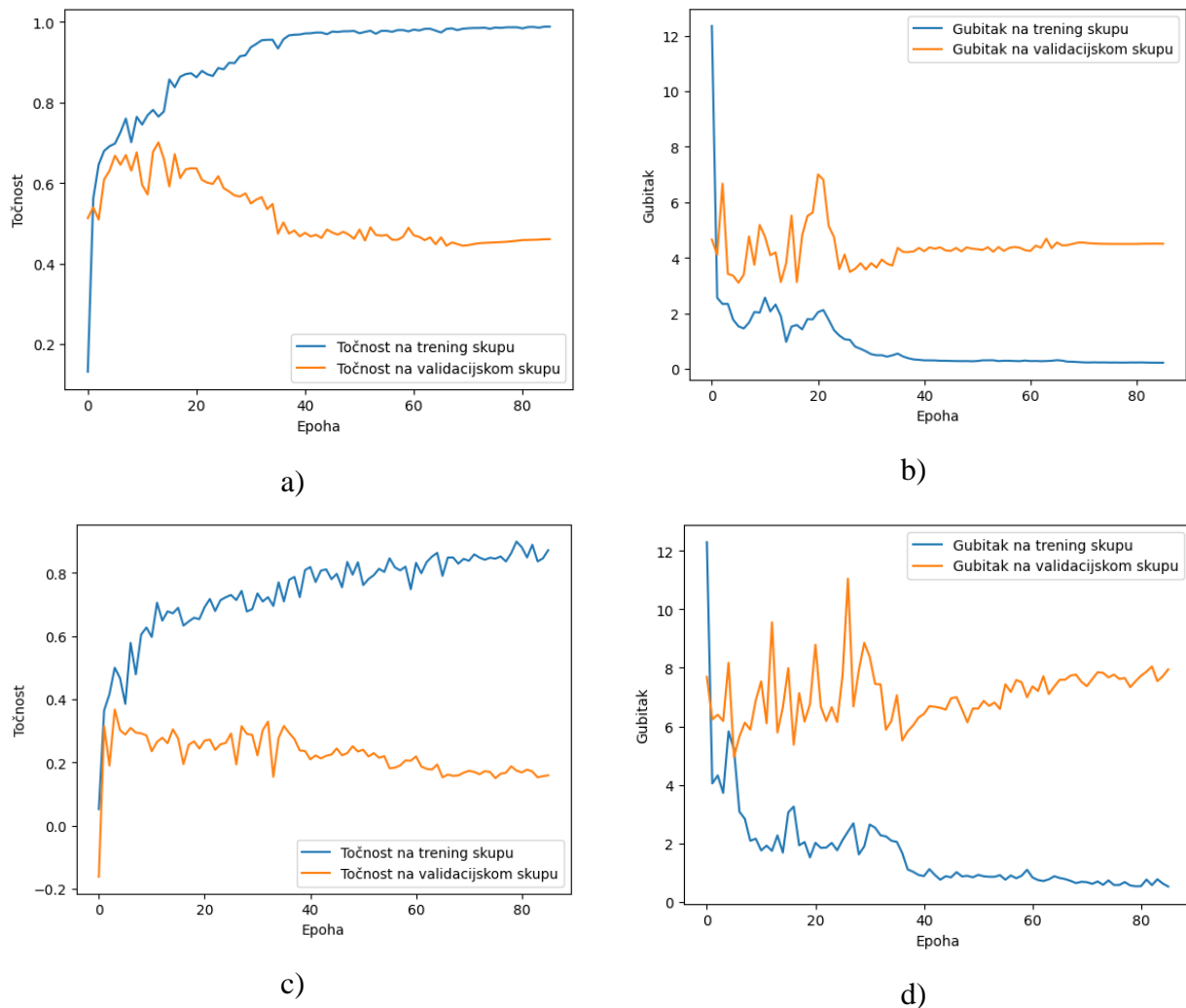


d)

**Slika 3.20.** Grafovi dobiveni nakon završetka treniranja za trening i validacijski skup za izobličenje nastalo određenim % gubitka paketa pri prijenosu bežičnom mrežom: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

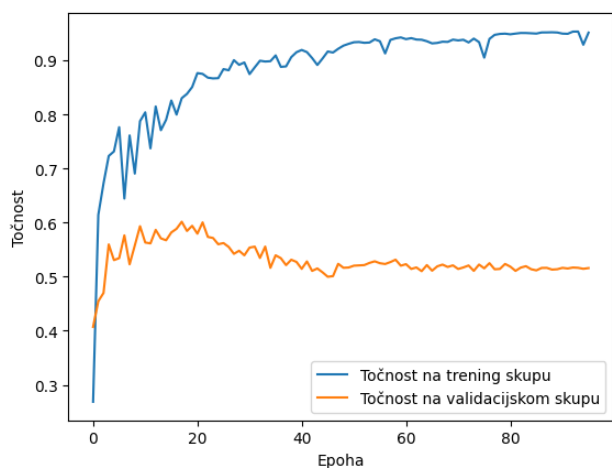
Slika 3.21 prikazuje rezultate treniranja za izobličenja uzrokovana kompresijom i može se vidjeti kako Model 1 ostvaruje puno bolje rezultate treniranja za oba skupa podataka. Ako se pogleda tablica 3.6, oba modela ostvaruju bolje rezultate točnosti za trening skup nego objektivne metrike zasnovane na egzaktnim računalnim algoritmima s kojima su uspoređivani. Validacijski gubitak za predložene modele dosta je visok u odnosu na gubitak na trening skupu, što se može vidjeti u tablici 3.8. Model 1 ostvaruje bolje rezultate predviđanja ocjena u odnosu na Model 2 i ostale objektivne metrike s kojima je uspoređivan, dok Model 2 ostvaruje bolje rezultate samo od NIQE metrike. Podjela skupova podataka za ovaj tip izobličenja je 73.8:13.1:13.1. Slika 3.22 prikazuje rezultate treniranja za izobličenja nastala pri prijenosu mrežom i može se vidjeti kako Model 1 ostvaruje puno bolje rezultate treniranja za oba skupa podataka. Ako se pogleda tablica 3.6, oba

modela ostvaruju bolje rezultate točnosti za trening skup nego objektivne metrike zasnovane na egzaktnim računalnim algoritmima s kojima su uspoređivani.

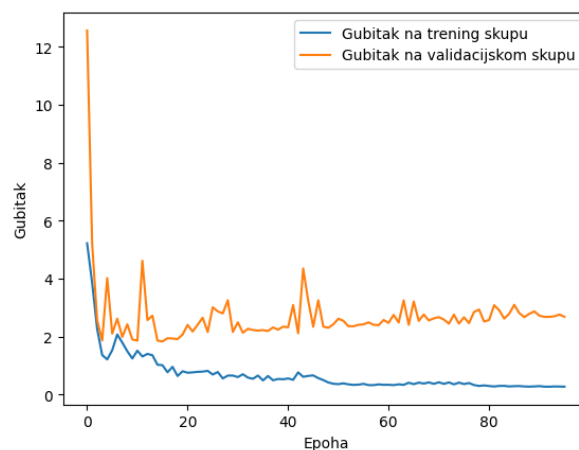


**Slika 3.21.** Grafovi dobiveni nakon završetka treninga za trening i validacijski skup za izobličenja uzrokovana kompresijom: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

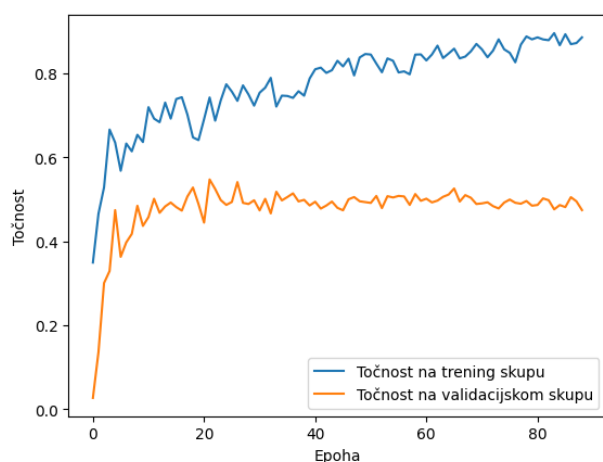
Validacijski gubitak veći je kod Modela 2 nego Modela 1, što se može vidjeti u tablici 3.8. Najbolji rezultati predviđanja ostvareni su PSNR metrikom, dok su predloženi modeli bolji od NIQE metrike. Podjela skupova podataka za ovaj tip izobličenja je 70.2:14.9:14.9. U tablicama 3.6 i 3.7 predstavljeni su rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na trening i validacijskim skupovima za drugi način podjele podataka, gdje su istaknuti rezultati najbolji rezultati postignuti predviđanja za pojedini tip izobličenja.



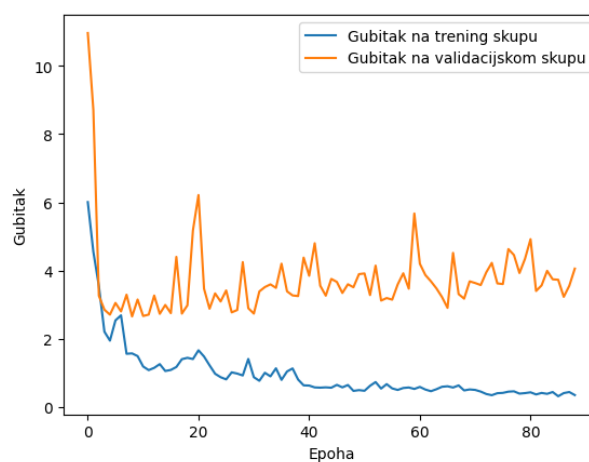
a)



b)



c)



d)

**Slika 3.22.** Grafovi dobiveni nakon završetka treninga za trening i validacijski skup za izobličenja nastala pri prijenosu mrežom: a) točnost za Model 1 izražena preko PLCC, b) gubitak za Model 1 izražen preko MSE, c) točnost za Model 2 izražena preko PLCC, d) gubitak za Model 2 izražen preko MSE

Tablica 3.8 predstavlja rezultate gubitka izražene u MSE predloženih modela na trening i validacijskim skupovima za pojedine tipove izobličenja.

**Tablica 3.6.** Rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na trening skupu podataka za drugi način podjele baze podataka

Tip izobličenja		PSNR	SSIM	MS-SSIM	NIQE	Model1	Model2
H264	PLCC	0.399	0.556	0.542	0.292	0.936	<b>0.943</b>
	SROCC	0.402	0.604	0.676	0.224	0.925	<b>0.943</b>
H265	PLCC	0.710	0.750	0.749	-0.639	<b>0.997</b>	0.981
	SROCC	0.684	0.699	0.753	-0.732	<b>0.989</b>	0.979
MPEG2	PLCC	0.399	0.676	0.590	-0.366	<b>0.958</b>	0.954
	SROCC	0.405	0.696	0.570	-0.315	<b>0.951</b>	0.938
MPEG4	PLCC	0.736	0.865	0.851	-0.392	<b>0.963</b>	0.829

	SROCC	0.754	0.866	0.833	-0.339	<b>0.946</b>	0.833
Smrzavanje	PLCC	0.687	0.658	0.541	-0.026	<b>0.972</b>	0.956
	SROCC	0.695	0.668	0.561	0.153	<b>0.978</b>	0.965
% gubitka paketa pri prijenosu IP mrežom	PLCC	0.365	0.368	0.306	0.348	<b>0.978</b>	0.949
	SROCC	0.511	0.461	0.435	0.483	<b>0.983</b>	0.953
% gubitka paketa pri prijenosu bežičnom mrežom	PLCC	0.391	0.146	0.220	0.043	<b>0.995</b>	0.899
	SROCC	0.406	0.247	0.356	0.022	<b>0.976</b>	0.836
Izobličenja uzrokovana kompresijom	PLCC	0.301	0.445	0.513	0.184	<b>0.922</b>	0.830
	SROCC	0.373	0.418	0.641	0.134	<b>0.911</b>	0.834
Izobličenja nastala pri prijenosu mrežom	PLCC	0.501	0.431	0.343	0.185	<b>0.927</b>	0.854
	SROCC	0.559	0.488	0.426	0.294	<b>0.934</b>	0.861

**Tablica 3.7.** Rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na validacijskom skupu podataka za drugi način podjele baze podataka

Tip izobličenja		PSNR	SSIM	MS-SSIM	NIQE	Model1	Model2
H264	PLCC	0.391	0.535	0.495	0.200	<b>0.684</b>	0.461
	SROCC	0.375	0.585	0.591	-0.256	<b>0.706</b>	0.615
H265	PLCC	0.971	0.983	<b>0.988</b>	-0.976	0.770	0.964
	SROCC	0.936	0.941	<b>0.946</b>	<b>-0.946</b>	0.873	0.946
MPEG2	PLCC	0.893	0.882	<b>0.912</b>	-0.694	0.610	0.877
	SROCC	0.861	0.861	<b>0.912</b>	-0.808	0.706	0.846
MPEG4	PLCC	0.405	0.775	0.755	-0.245	<b>0.789</b>	0.282
	SROCC	0.455	0.719	0.719	-0.419	<b>0.855</b>	0.510
Smrzavanje	PLCC	0.863	0.627	0.663	0.231	0.816	<b>0.876</b>
	SROCC	<b>0.887</b>	0.758	0.702	0.196	0.807	0.875
% gubitka paketa pri prijenosu IP mrežom	PLCC	0.882	0.863	0.798	0.646	0.692	<b>0.919</b>
	SROCC	<b>0.858</b>	0.876	0.779	0.530	0.598	0.833
% gubitka paketa pri prijenosu bežičnom mrežom	PLCC	<b>0.871</b>	0.740	0.542	0.532	-0.845	0.652
	SROCC	<b>0.810</b>	0.715	0.667	0.334	-0.834	0.572
Izobličenja uzrokovana kompresijom	PLCC	0.473	0.508	0.534	0.128	<b>0.645</b>	0.258
	SROCC	0.527	0.563	0.556	-0.072	<b>0.723</b>	0.255
Izobličenja nastala pri prijenosu mrežom	PLCC	<b>0.739</b>	0.606	0.552	0.649	0.528	0.399
	SROCC	<b>0.777</b>	0.675	0.556	0.675	0.548	0.343

**Tablica 3.8.** Rezultati MSE postignuti predloženim modelima na trening i validacijskim skupovima podataka za drugi način podjele baze podataka

Tip izobličenja	MSE			
	Skup za treniranje		Skup za validaciju	
	Model1	Model2	Model1	Model2
H.264	0.2427	0.5736	3.6815	5.7006

H.265	0.2107	0.2105	3.6877	0.7887
MPEG2	0.2177	0.2876	1.9982	1.0164
MPEG4	0.1739	0.3890	1.1905	2.7691
Smrzavanje	0.2628	0.2460	0.9760	0.4708
% gubitka paketa pri prijenosu IP mrežom	0.1821	0.2316	1.6156	0.8848
% gubitka paketa pri prijenosu bežičnom mrežom	0.1727	0.3093	6.5074	4.2344
Izobličenja uzrokovana kompresijom	0.2173	0.526	3.1032	4.9542
Izobličenja nastala pri prijenosu mrežom	0.2704	0.3184	1.8332	2.6567

## 4. EVALUACIJA RADA PREDLOŽENIH MODELA ZA OCJENJIVANJE VIDEO SIGNALA ZASNOVANOG NA STROJNOM UČENJU

U ovom poglavlju bit će opisani rezultati testiranja modela za ocjenjivanje video signala predloženih u ovom radu i bit će dana usporedba tih rezultata s rezultatima nekih postojećih objektivnih metrika. Rezultati predloženih modela i ostalih objektivnih metrika (PSNR, SSIM, MSSIM, NIQE) uspoređivat će se sa subjektivnim ocjenama za oba načina treniranja, tj. za oba načina podjele signala u trening, validacijski i testni skup. Kako bi se rezultati mogli uspoređivati, korištene su dvije metrike: PLCC i SROCC. Prije samog predstavljanja rezultata bitno je naglasiti da su se rezultati prilikom računanja PLCC i SROCC mjera između DMOS-a i objektivnih metrika zasnovanih na egzaktnim računalnim algoritmima predstavljenih u ovo radu za sve tipove izobličenja skalirali između 0 i 1, zbog primjene nelinearne regresije u [30]. Nelinearna regresija je tehnika statističkog modeliranja koja se koristi za procjenu odnosa između zavisnih i nezavisnih varijabli kada odnos nije linearan. Za razliku od linearne regresije, koja pretpostavlja linearni odnos između varijabli, nelinearna regresija omogućuje složenije i fleksibilnije modeliranje. Za računanje novih vrijednosti objektivnih metrika u odnosu na subjektivne ocjene koristila su logistička funkcija opisana formulom 4-1, gdje su  $\beta_1, \beta_2, \beta_3, \beta_4$  parametri koji se određenim brojem iteracija računaju kako bi se dobila najbolja procjena odnosa između DMOS i objektivne metrike.

$$Q' = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-\frac{(Q - \beta_3)}{|\beta_4|}}} \quad (4-1)$$

Potrebno je bilo i odrediti početne vrijednosti tih parametara, a one su prema [40] bile:

- $\beta_1 = \max(DMOS)$
- $\beta_2 = \min(DMOS)$
- $\beta_3 = \bar{Q}$ , gdje je  $\bar{Q}$  srednja vrijednost ocjena dobivenih objektivnom metrikom
- $\beta_4 = 1$

U idućim potpoglavljima bit će predstavljeni rezultati testiranja modela za 7 različitih tipova izobličenja: H.264, H.265, MPEG2, MPEG4, smrzavanje, % gubitka paketa u IP mreži i % gubitka paketa u bežičnoj mreži. Nakon toga će biti predstavljeni rezultati izobličenja uzrokovani zbog kompresije i izobličenja nastalih pri prijenosu mrežom.

## 4.1. Evaluacija algoritma uz prvi način podjele baze podataka

U ovom potpoglavlju bit će predstavljeni rezultati testiranja predloženih modela za ocjenjivanje kvalitete video signala. Rezultati modela uspoređivani su s trima FR metrikama: PSNR, SSIM i MS-SSIM te NIQE metrikom, koja je NR. Osim numeričkih rezultata, grafički je prikazan odnos ocjena dobiven za svaku metriku u odnosu na subjektivne ocjene za pojedini tip izobličenja. U tablici 4.1 predstavljeni su rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na testnom skupu za prvi način podjele podataka, gdje su istaknuti rezultati najbolji rezultati postignuti predviđanjem za pojedini tip izobličenja.

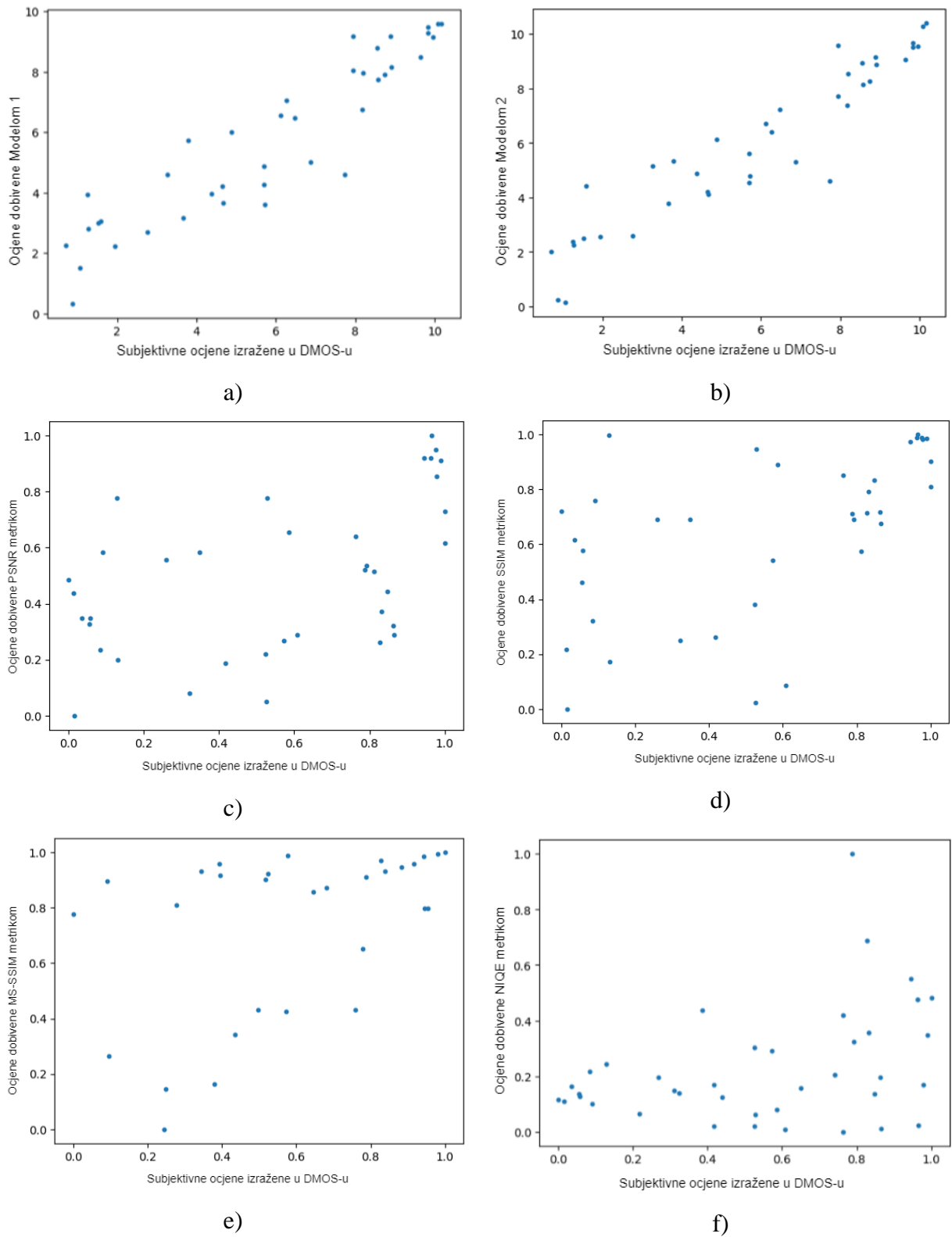
**Tablica 4.1.** Rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na testnom skupu podataka za prvi način podjele baze podataka

Tip izobličenja		PSNR	SSIM	MS-SSIM	NIQE	Model 1	Model 2
H264	PLCC	0.499	0.554	0.568	0.367	0.924	<b>0.940</b>
	SROCC	0.515	0.606	0.670	0.320	0.934	<b>0.945</b>
H265	PLCC	0.636	0.339	0.606	-0.509	<b>0.996</b>	0.896
	SROCC	0.771	0.428	0.257	-0.542	<b>1.000</b>	0.943
MPEG2	PLCC	0.445	0.591	0.516	0.000	<b>0.910</b>	0.824
	SROCC	0.410	0.586	0.509	0.358	<b>0.855</b>	0.622
MPEG4	PLCC	0.846	0.916	0.831	-0.318	<b>0.927</b>	0.910
	SROCC	0.803	0.879	0.818	-0.208	0.885	<b>0.915</b>
Smrzavanje	PLCC	0.684	0.671	0.500	-0.038	<b>0.882</b>	0.862
	SROCC	0.697	0.708	0.506	0.066	<b>0.868</b>	0.846
% gubitka paketa pri prijenosu IP mrežom	PLCC	0.722	0.602	0.640	0.657	0.866	<b>0.910</b>
	SROCC	0.635	0.583	0.630	0.765	0.865	<b>0.912</b>
% gubitka paketa pri prijenosu bežičnom mrežom	PLCC	0.137	0.172	0.640	0.226	0.466	<b>0.747</b>
	SROCC	0.090	0.296	<b>0.640</b>	0.536	0.504	0.611
Izobličenja uzrokovana kompresijom	PLCC	0.514	0.540	0.580	0.419	0.955	<b>0.959</b>
	SROCC	0.540	0.591	0.622	0.400	<b>0.951</b>	<b>0.951</b>
Izobličenja nastala pri prijenosu mrežom	PLCC	0.615	0.602	0.530	0.277	0.874	<b>0.900</b>
	SROCC	0.630	0.608	0.519	0.285	0.846	<b>0.866</b>

Prema rezultatima za H.264 tip izobličenja, može se vidjeti da su rezultati ocjena obaju predloženih modela daleko bolji od FR i NR objektivnih metrika koje se temelje na egzaktnim računalnim algoritmima. To je vidljivo i na slici 4.1, gdje je raspodjela točkica na grafu za a) i b) slike puno linearnija u odnosu na ostale. Dobiveni rezultati za objektivne FR metode su očekivani, budući da su SSIM i MS-SSIM kompleksniji algoritmi, stoga postižu veću korelaciju sa subjektivnim ocjenama u odnosu na PSNR. S druge strane, predloženi su modeli postigli gotovo jednake

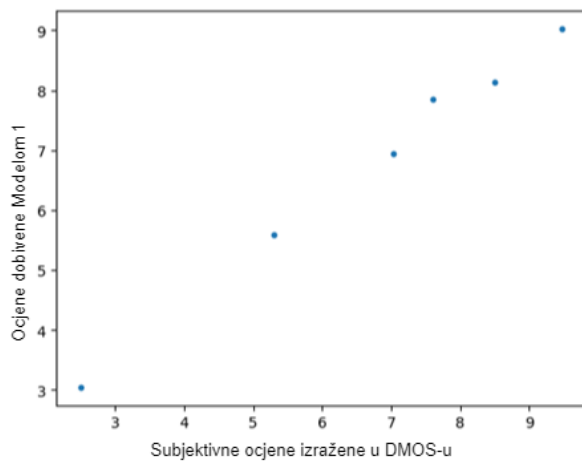


rezultate za ovaj tip izobličenja. Za H.264 tip izobličenja bilo je ukupno 39 video signal u testnom skupu.

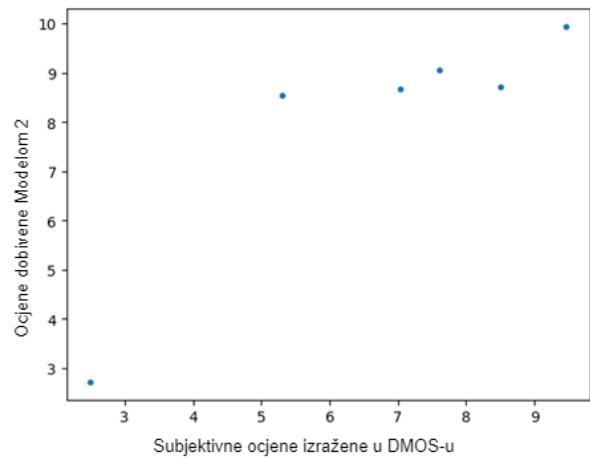


**Slika 4.1.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike d) SSIM metrike e) MS-SSIM metrike i f) NIQE metrike za H.264 tip izobličenja

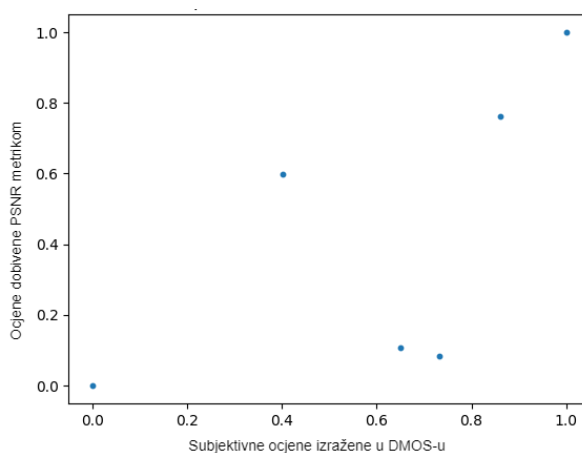
Iz rezultata testiranja predloženih modela za H.265 tip izobličenja, prikazanih u tablici 4.1, može se vidjeti da je Model 1 ostvario puno bolje rezultate nego Model 2. Oba modela su postigla bolje rezultate nego objektivne metrike koje su korištene u usporedbi. Nadalje, PSNR je ostvario bolje rezultate nego SSIM, što je pomalo neobično s obzirom na razlike u kompleksnosti algoritama. Od svih objektivnih metrika najbolje rezultate postigao je MS-SSIM, no oni su daleko lošiji nego rezultate predloženih modela, što se može vidjeti i na slici 4.2, gdje je vidljiva raspodjela ocjena pojedinih algoritama u odnosu na subjektivne ocjene. Na slici 4.2 c) može se vidjeti kako PSNR lošije ocjenjuje video signale veće kvalitete, kao i SSIM. Ukupan broj signala u testnom skupu za ovaj tip izobličenja bio je 6, te se u biti zbog toga ovi rezultati ne mogu smatrati pouzdanima i trebalo bi osigurati testni skup s većim brojem video signala.



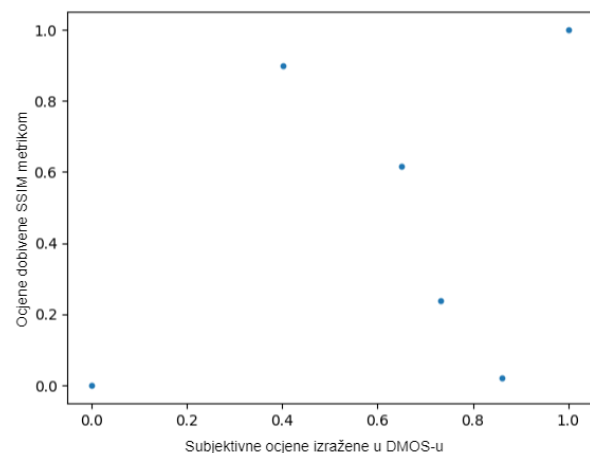
a)



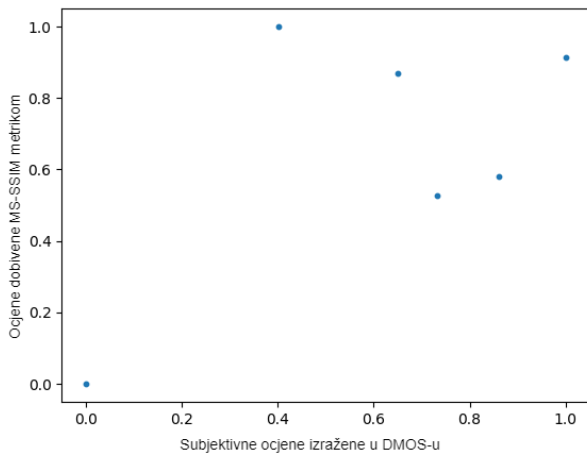
b)



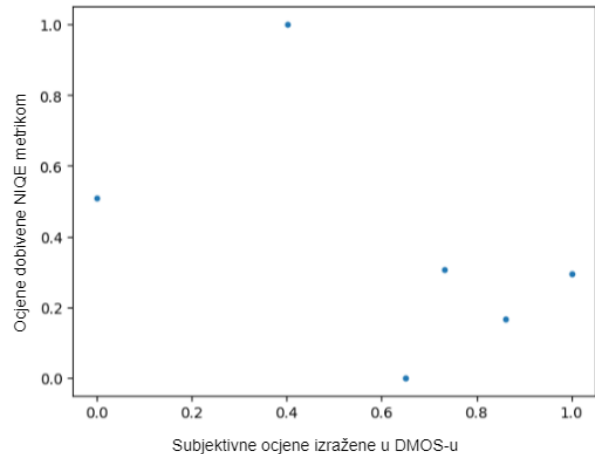
c)



d)



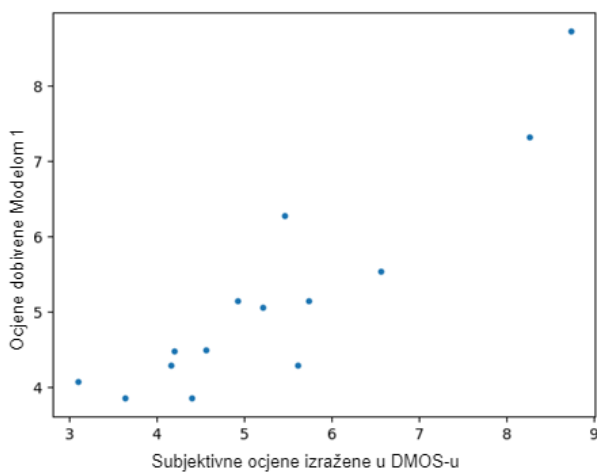
e)



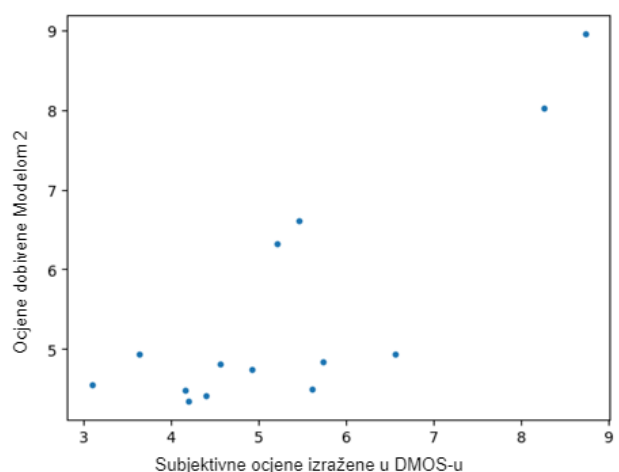
f)

**Slika 4.2.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za H.265 tip izobličenja

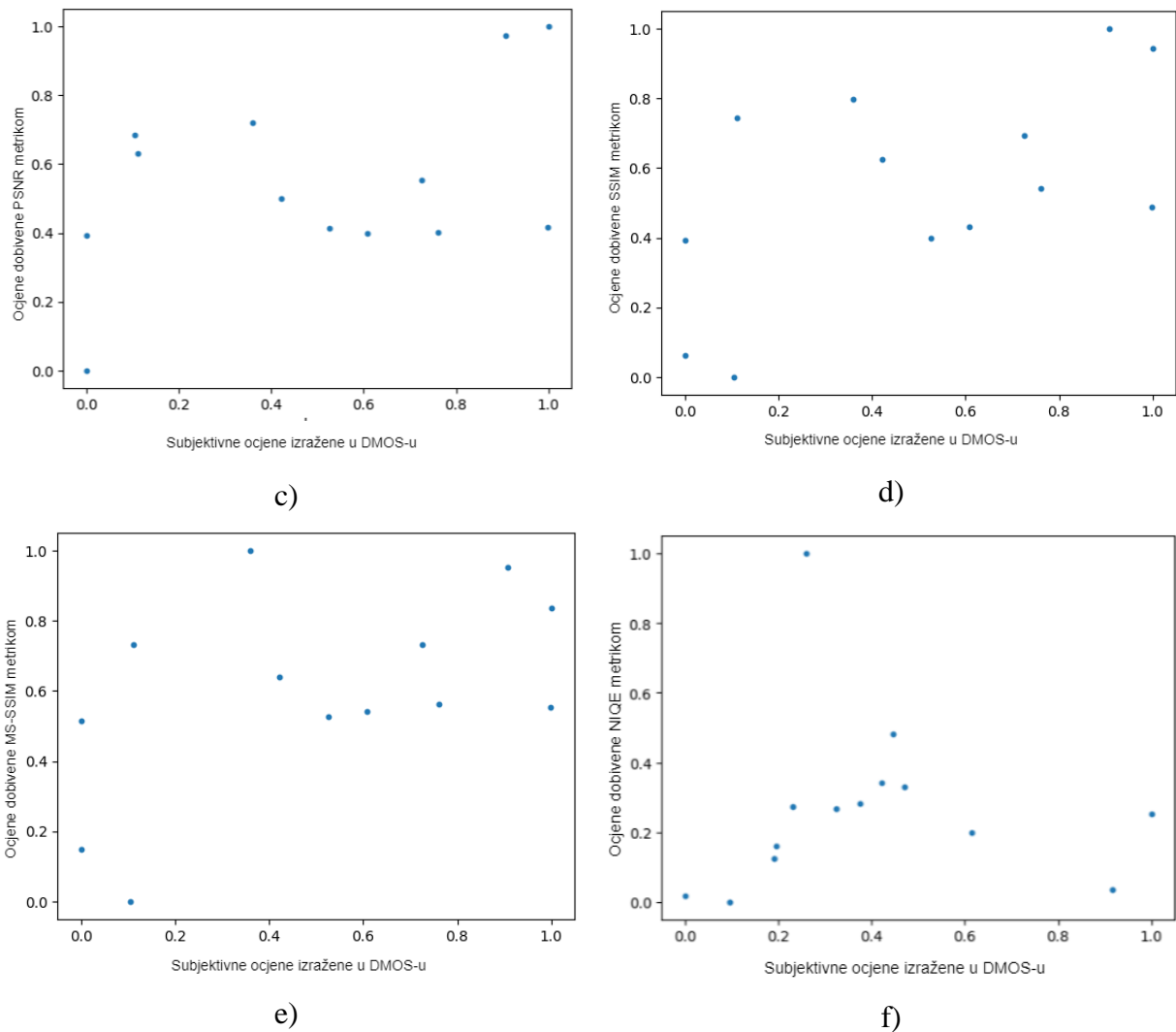
Prema prikazanim rezultatima u tablici 4.1 za MPEG2 tip izobličenja, može se vidjeti da se Model 1 pokazao boljim u odnosu na Model 2. Oba su modela postigla bolje rezultate nego objektivne metrike zasnovane na egzaktnim računalnim algoritmima u vidu PLCC, dok je monotonost ocjena izražena SROCC metrikom kod Modela 1 znatno veća nego kod ostalih metrika (slika 4.3). SSIM daje najbolje rezultate u odnosu na druge metrike koje su korištene za usporedbu s predloženim modelima. Za ovaj tip izobličenja bilo je ukupno 14 video signala u testnom skupu.



a)



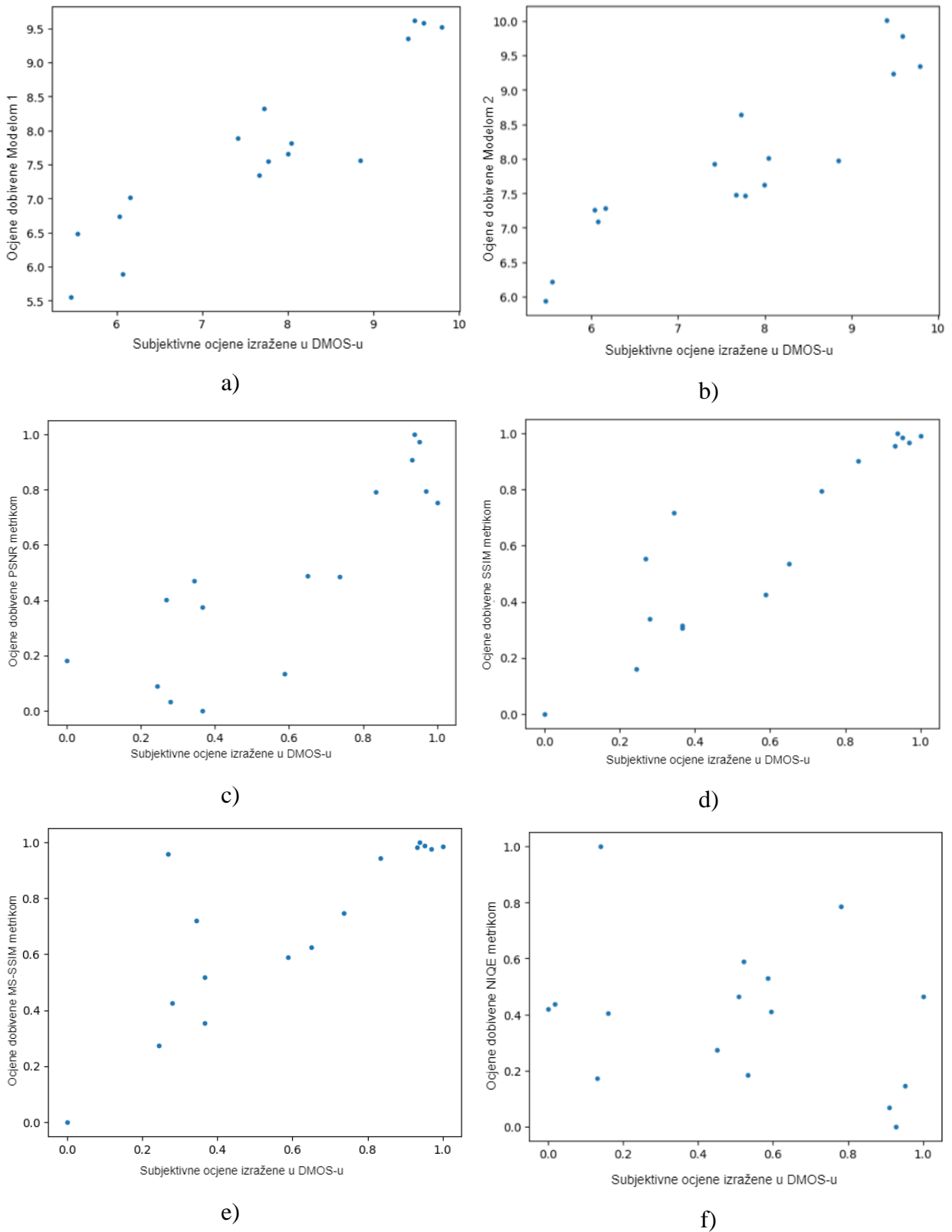
b)



**Slika 4.3.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za MPEG2 tip izobličenja

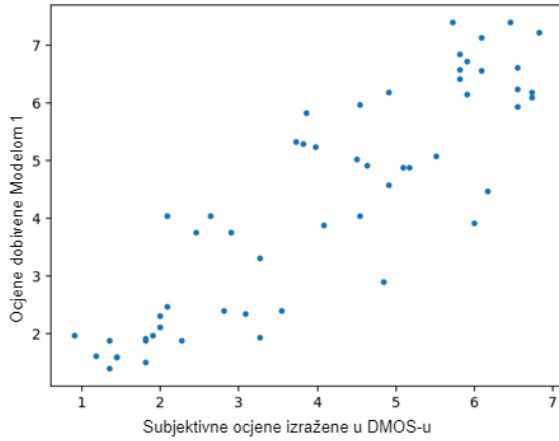
Za MPEG4 tip izobličenja sve objektivne metrike pokazale su izrazito dobre rezultate, osim NIQE metrike, što je vidljivo u tablici 4.1. Model 1 bolje predviđa ocjene u odnosu na Model 2. Najgore rezultate opet je dala NIQE metrika. Najbolji rezultat ostvario je Model 1 u vidu PLCC, dok je Model 2 ostvario najbolji rezultat SROCC-a. NIQE metrika, iako je ostvarila najlošije rezultate, dobro predviđa video signale niže kvalitete. Na slici 4.4 može se vidjeti točkasti graf koji predstavlja raspodjelu predviđenih ocjena predloženih modela i objektivnih metrika s kojima su uspoređivani. Za ovaj tip izobličenja bilo je ukupno 14 video signala u testnom skupu. Iz rezultata prikazanih tablicom 4.1. za izobličenja video signala nastala smrzavanjem okvira, može se vidjeti kako su oba modela predložena u ovom radu ostvarila bolje rezultate nego objektivne FR i NR metode s kojima su uspoređeni. U vidu PLCC oba predložena modela dali su podjednake rezultate, no SROCC kod Modela 1 nešto je bolji nego kod Modela 2. Od objektivnih metrika temeljnih na

egzaktним računalnim algoritmima najbolje su se pokazale PSNR i SSIM za procjenjivanje kvalitete video signal sa smrzavanjem okvira. Najgorom se pokazala NIQE metrika.

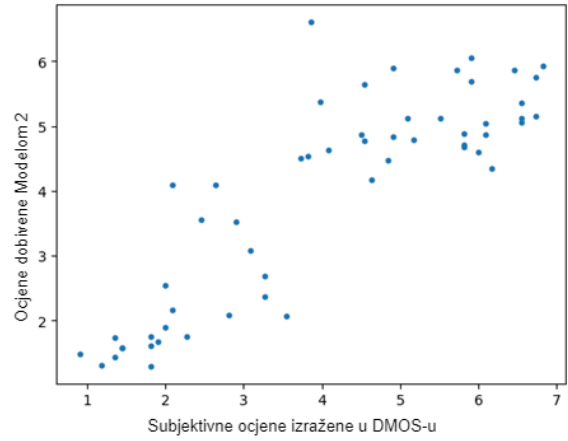


**Slika 4.4.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za MPEG4 tip izobličenja

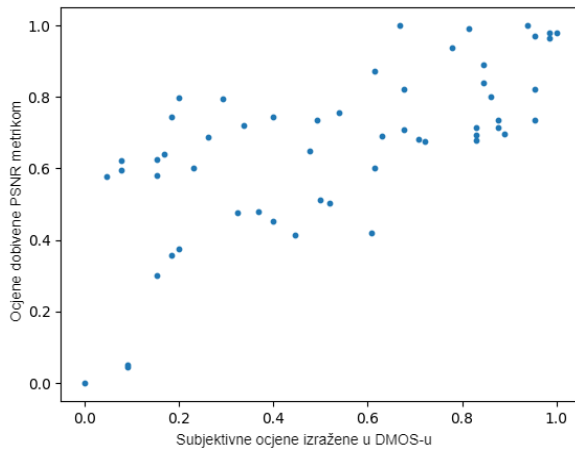
Na slici 4.5 može se vidjeti da su grafovi rezultata odnosa subjektivnih ocjena i ocjena dobivenih predloženim modelima linearniji u odnosu na metrike s kojim se uspoređuje, što potvrđuje rezultate iz tablice 4.1. Za ovaj tip izobličenja bilo je ukupno 55 video signala u testnom skupu.



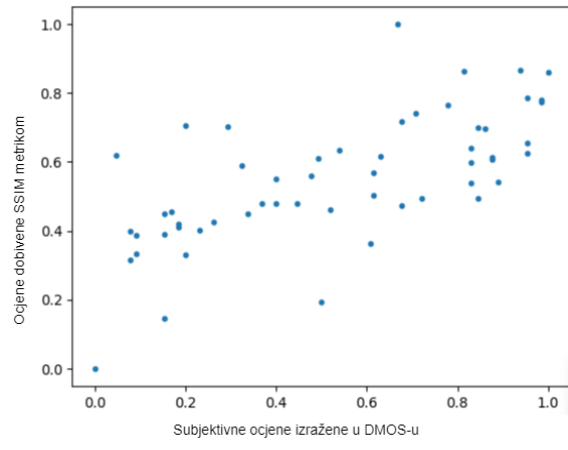
a)



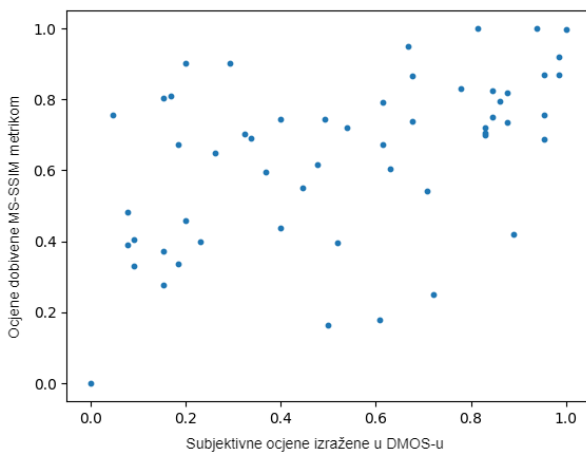
b)



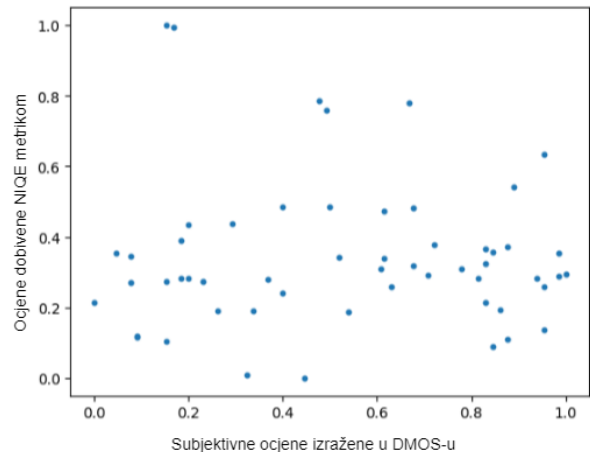
c)



d)



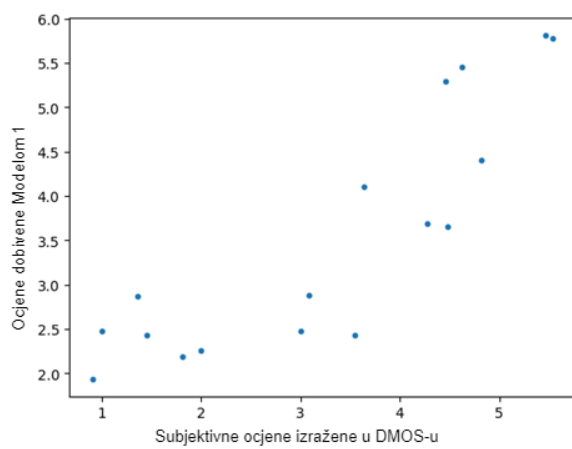
e)



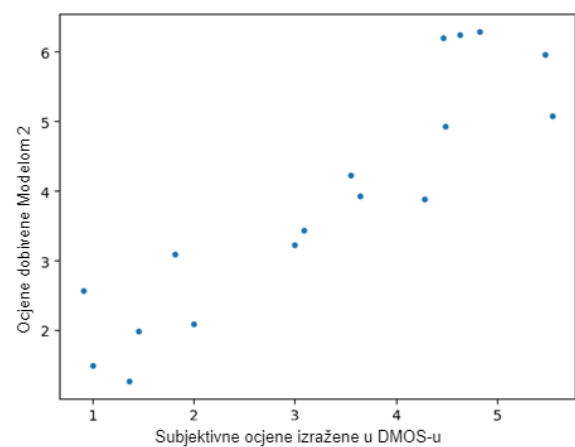
f)

**Slika 4.5.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo smrzavanje okvira

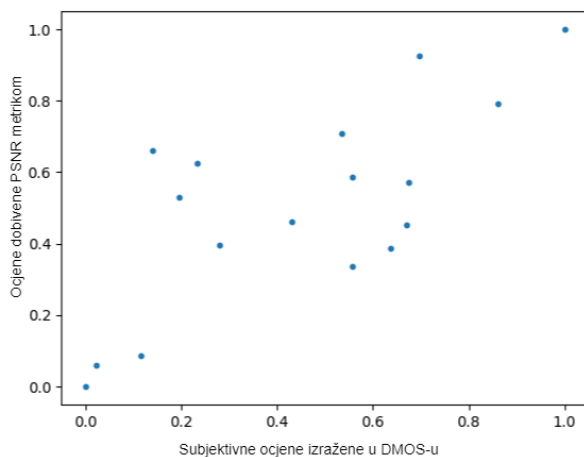
Iz rezultata ostvarenih testiranjem na skupu za izobličenje nastalo gubitkom paketa u IP mreži, vidljivo je da je Model 2 ostvario najbolje rezultate u odnosu na sve ostale metrike s kojima je uspoređivan. S druge strane NIQE metrika pokazala najlošijom za ovaj tip izobličenja u odnosu na rezultate koje je postizala kod prije spomenutih tipova izobličenja. Model 1 je ostvario nešto lošije rezultate u odnosu na Model 2, no u odnosu na ostale metrike ima puno bolje rezultate. Od objektivnih metrika, PSNR je neočekivano ostvario najbolje rezultate, dok SSIM postiže najlošije rezultate. Na slici 4.6 može se vidjeti da je odnos subjektivnih ocjena i ostalih metrika najlinearniji kod modela predloženih u ovom radu. Za ovaj tip izobličenja bilo je ukupno 17 video signala u testnom skupu.



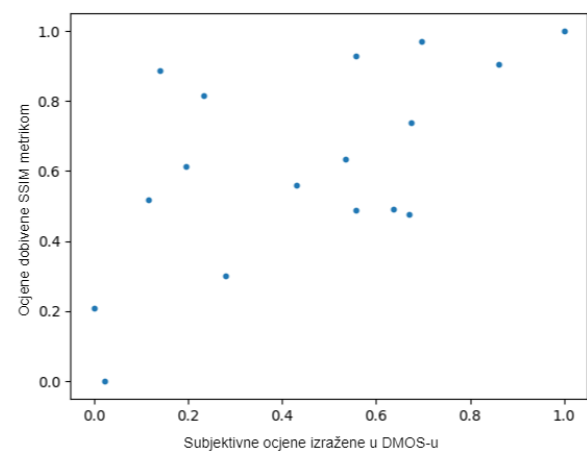
a)



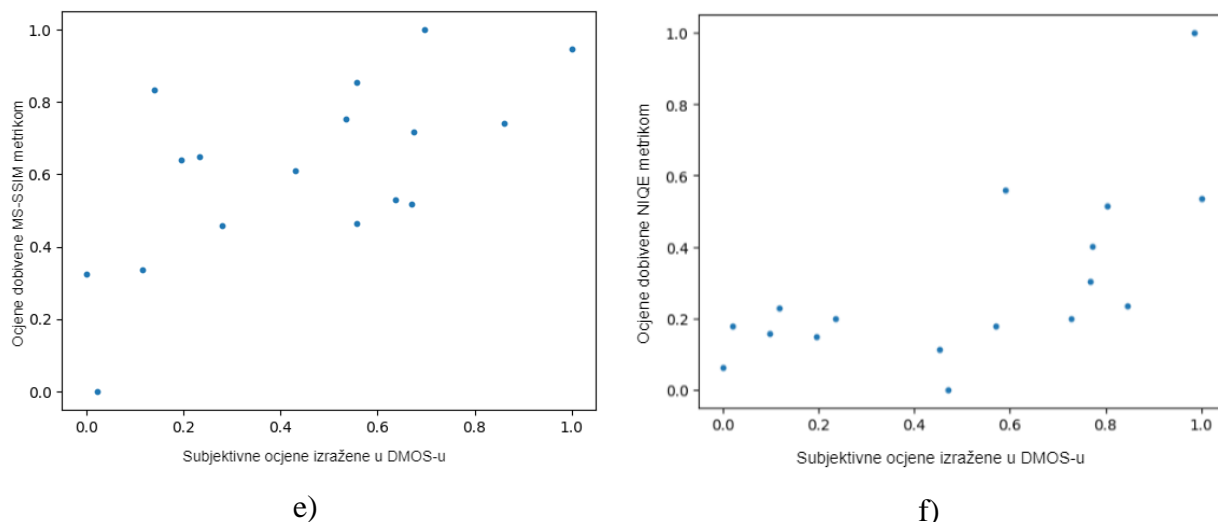
b)



c)

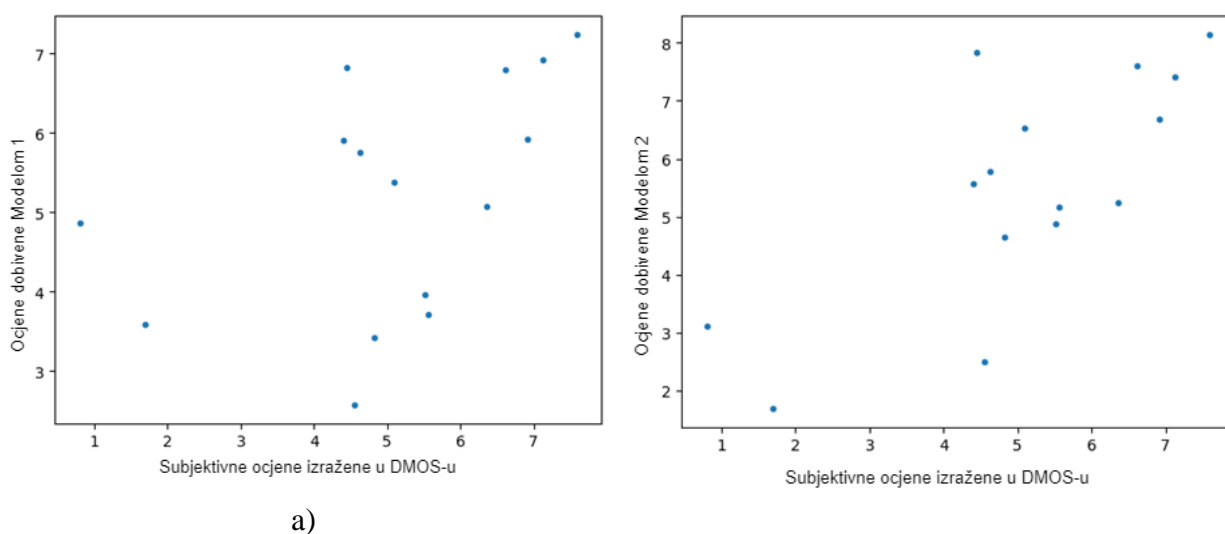


d)

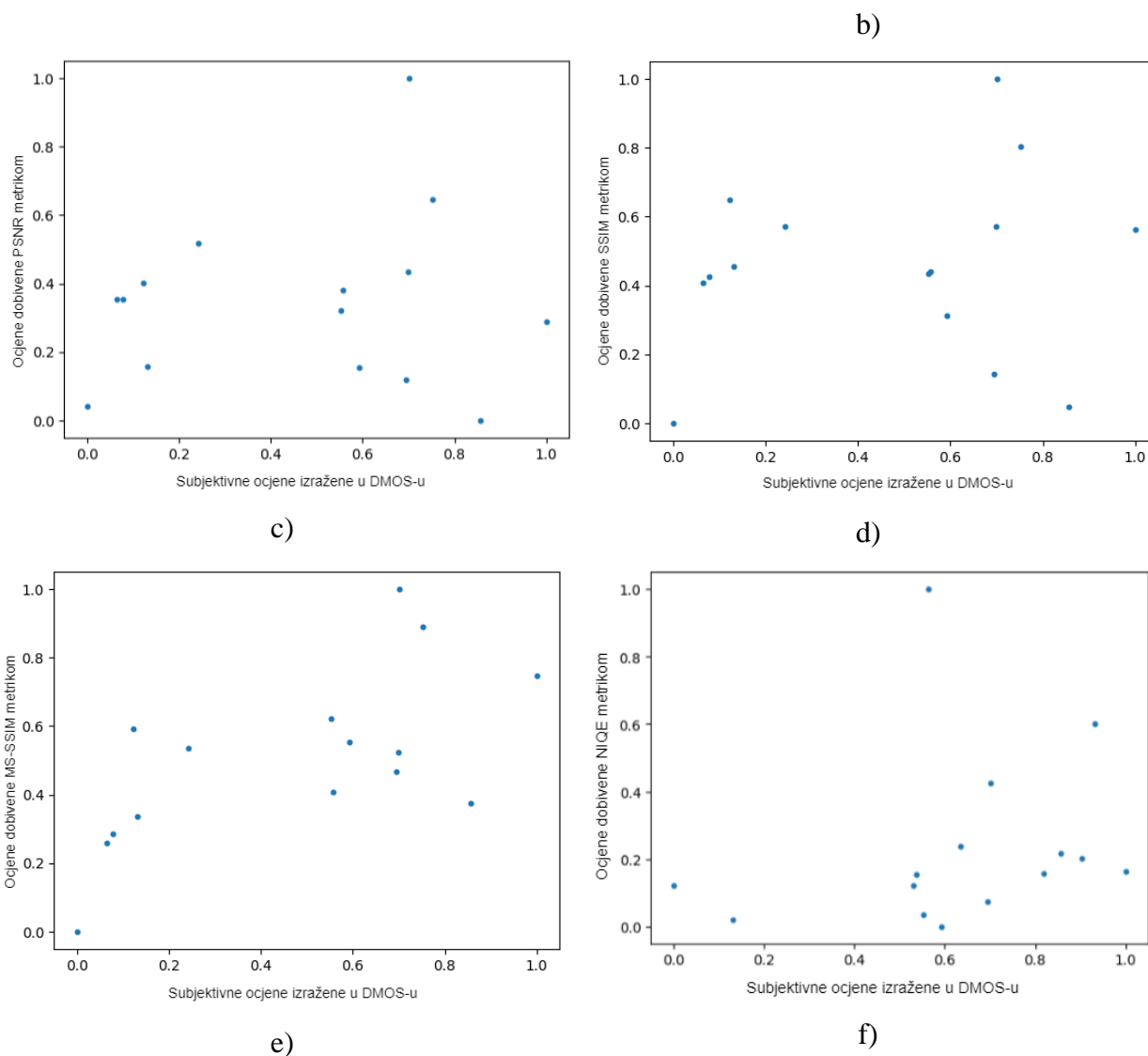


**Slika 4.6.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo % gubitka paketa pri prijenosu IP mreže

Nadalje, za izobličenja nastalo % gubitkom paketa pri prijenosu u bežičnoj mreži rezultati iz tablice 4.1 pokazuju da se daleko najbolje pokazao Model 2, dok je Model 1 dao nešto lošije rezultate. Od objektivnih metrika zasnovanih na egzaktnim računalnim algoritmima najbolje rezultate dala je MS-SSIM metrika, dok su rezultati PSNR-a, SSIM-a i NIQE-a poprilično podjednaki. Osim toga MS-SSIM ostvaruje i najveći SROCC u odnosu na sve objektivne metrike. Na slici 4.7 b) može se vidjeti da Model 1 ima problema s ocjenjivanjem video signala srednje kvalitete, dok NIQE i PSNR dobro ocjenjuju video signale niže kvalitete. Za ovaj tip izobličenja bilo je ukupno 15 video signala u testnom skupu. Nadalje, za izobličenje nastalo općenito kompresijom, gdje su se za treniranje i testiranje koristili zajedno H.264, H.265, MPEG2 i MPEG4 tipovi izobličenja, rezultati iz tablice 4.1 prikazuju da su najbolje ocjene dali predloženi modeli i pokazuju veliku korelaciju sa subjektivnim ocjenama, što se može vidjeti na slikama 4.8 a) i b).

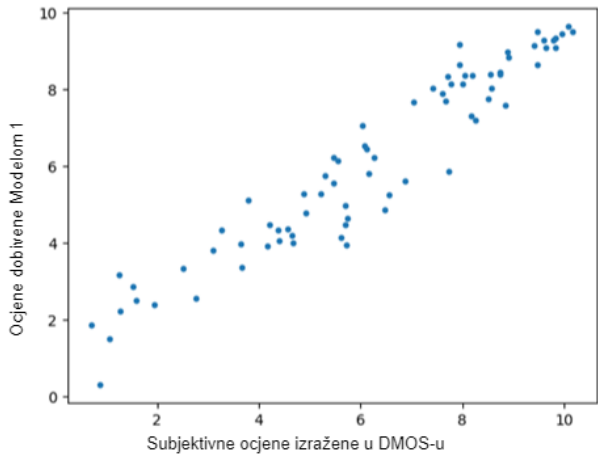




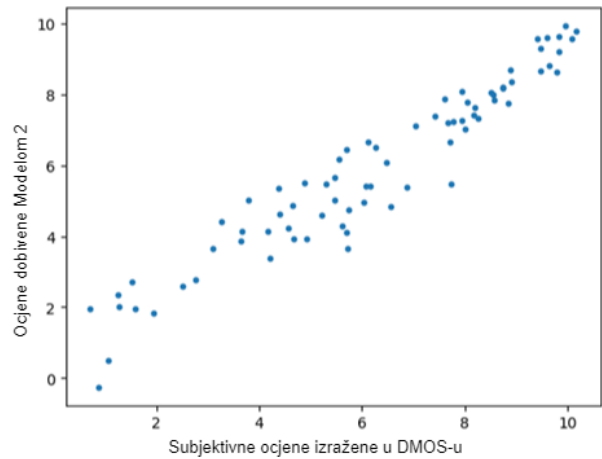


**Slika 4.7.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo % gubitka paketa pri prijenosu bežične mreže

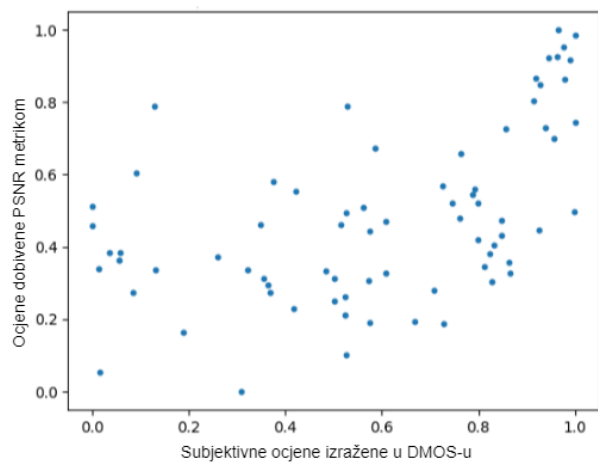
U odnosu na ostale objektivne metrike, oba predložena modela pokazali su jako dobre rezultate. Najgore rezultate ostvarila je NIQE metrika, dok je najbolje rezultate od objektivnih metrika zasnovanih na egzaktnim računalnim algoritmima ostvarila MS-SSIM metrika. Osim toga, iz slike 4.8 c), d), f) vidljivo je da PSNR, SSIM i MS-SSIM metrike bolje procjenjuju kvalitetu video signale bolje kvalitete, nego one lošije kvalitete. Za ovaj tip izobličenja bilo je ukupno 75 video signala u testnom skupu. Prema tablici 4.1, najbolje rezultate na testnom skupu izobličenja nastalih mrežnim artefaktima smrzavanja i gubitka paketa bežične i IP mreže ostvario je Model 2. Oba predložena modela ostvarila su bolje rezultate nego ostale metrike s kojima su uspoređivani.



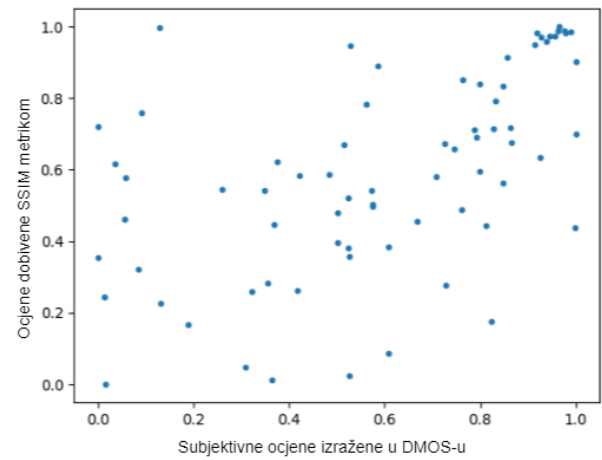
a)



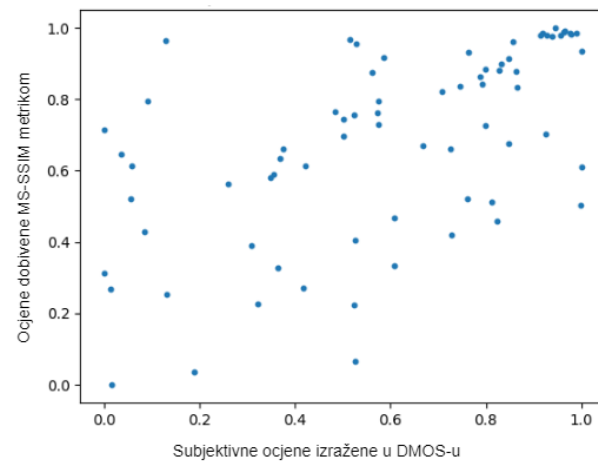
b)



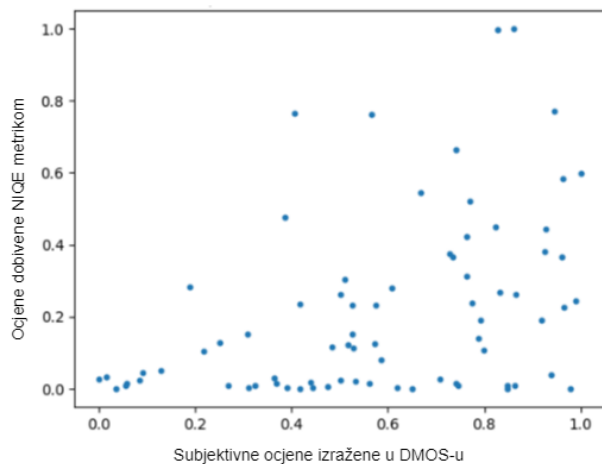
c)



d)



e)

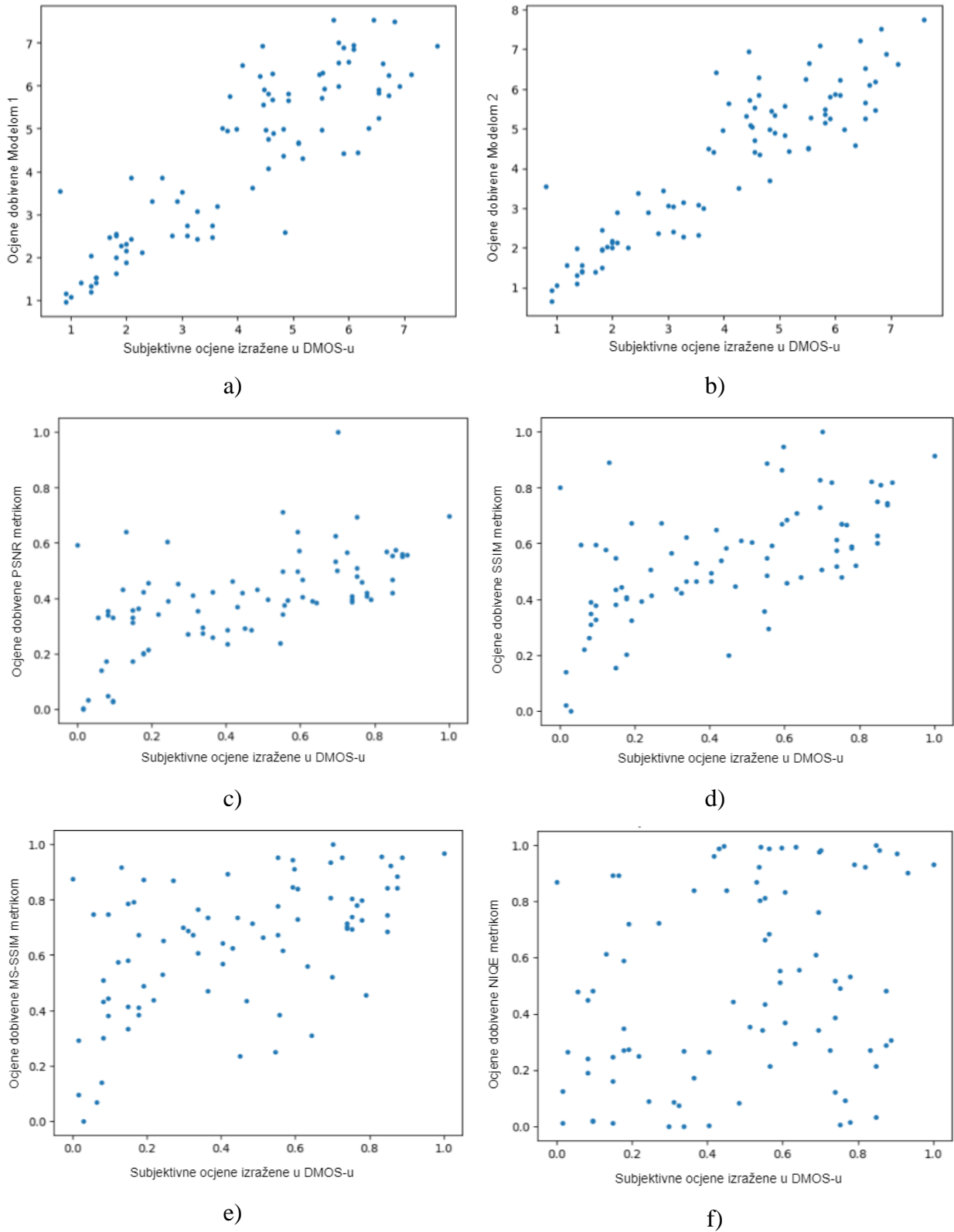


f)

**Slika 4.8.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo kompresijom

Prema slici 4.9. koja prikazuje odnose subjektivnih ocjena i ocjena koje su postigli objektivni algoritmi, najveću linearnost ima Model 2, pa Model 1, dok je najgora kod NIQE metrike. Od

ostalih objektivnih metrika najbolje rezultate postigla je PSNR metrika, što potvrđuju rezultati korelacije iz tablice 4.1. Za ovaj tip izobličenja bilo je ukupno 87 video signala u testnom skupu.



**Slika 4.9.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo pri prijenosu mrežom

## 4.2. Evaluacija algoritma uz drugi način podjele baze podataka

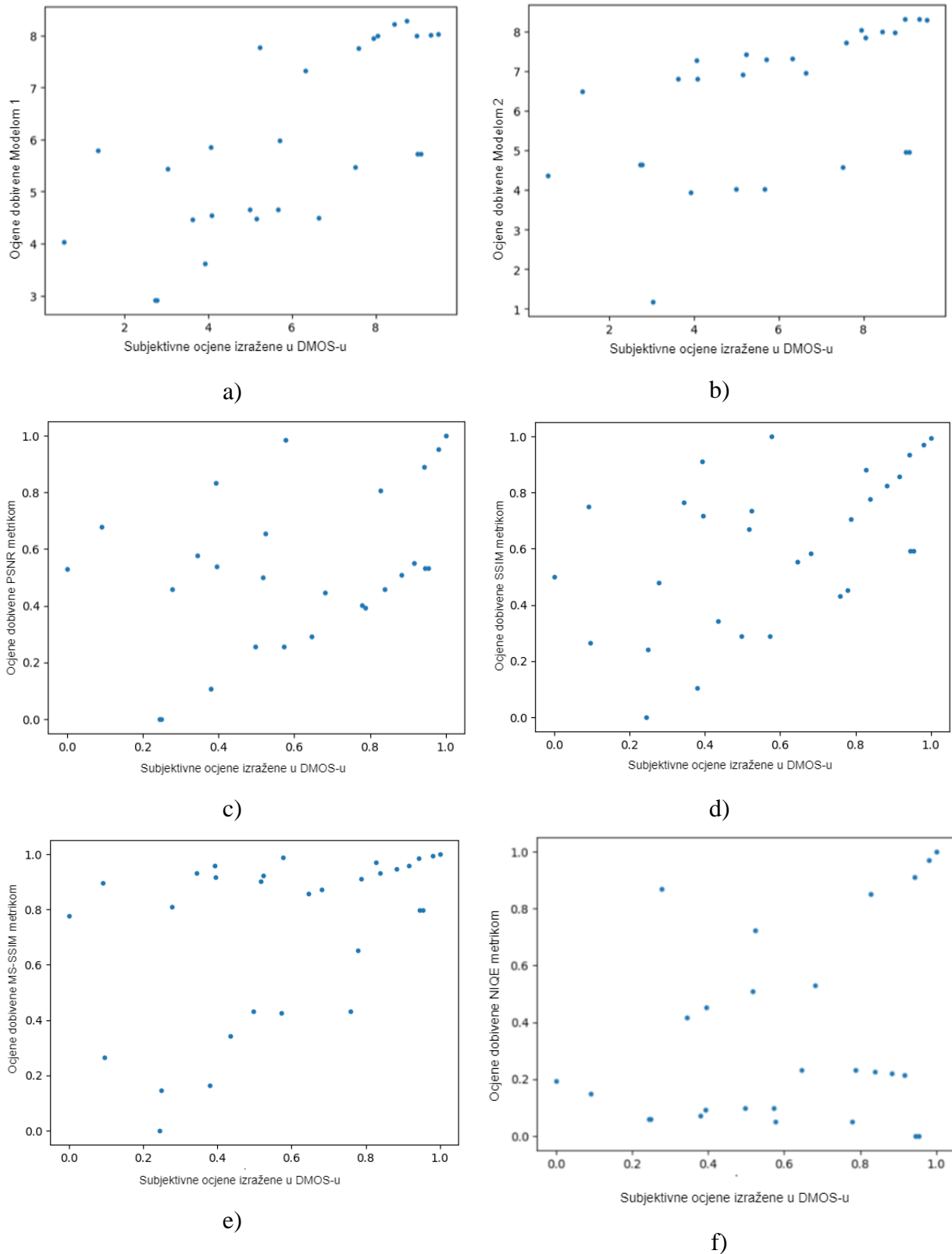
U ovom potpoglavlju prikazani su rezultati testiranja modela na skupu video signala uz drugi način podjele baze podataka. Svaki skup za testiranje (za pojedini tip izobličenja) imao je iste sadržaje iz razloga što su se tipovi izobličenja grupirali prilikom treniranja za izobličenja nastala kompresijom i izobličenja nastala prijenosom kroz mrežu i na taj način se izbjeglo preklapanje sadržaja. U tablici 4.2. prikazani su rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na testnom skupu podataka za drugi način podjele baze podataka. Rezultati metrike koja je postigla najbolje performanse za pojedini tip izobličenja istaknuti su na način da su podebljani.

**Tablica 4.2.** Rezultati postignuti predloženim modelima i ostalim testiranim objektivnim metrikama na testnom skupu podataka za drugi način podjele baze podataka

Tip izobličenja		PSNR	SSIM	MS-SSIM	NIQE	Model1	Model2
H264	PLCC	0.391	0.511	0.463	0.292	<b>0.719</b>	0.516
	SROCC	0.375	0.538	0.543	0.224	<b>0.742</b>	0.656
H265	PLCC	0.688	0.755	0.704	<b>-0.988</b>	-0.026	0.869
	SROCC	0.764	0.764	0.764	<b>-0.982</b>	-0.327	0.837
MPEG2	PLCC	0.847	<b>0.920</b>	0.917	0.605	-0.360	0.654
	SROCC	0.851	<b>0.916</b>	0.911	0.637	-0.133	0.807
MPEG4	PLCC	0.962	0.970	<b>0.993</b>	-0.222	0.600	0.364
	SROCC	0.959	0.931	<b>0.987</b>	-0.427	0.706	0.601
Smrzavanje	PLCC	0.792	<b>0.819</b>	0.605	0.264	0.393	0.281
	SROCC	0.873	<b>0.939</b>	0.753	0.300	0.655	0.528
% gubitka paketa pri prijenosu IP mrežom	PLCC	0.811	<b>0.877</b>	0.876	0.672	0.511	0.339
	SROCC	0.919	<b>0.95</b>	0.928	0.635	0.381	0.390
% gubitka paketa pri prijenosu bežičnom mrežom	PLCC	<b>0.952</b>	0.874	0.893	0.190	0.754	-0.568
	SROCC	<b>0.953</b>	0.953	0.905	0.524	0.786	-0.405
Izobličenja uzrokovana kompresijom	PLCC	0.712	0.548	0.603	0.234	<b>0.717</b>	0.418
	SROCC	0.763	0.757	<b>0.784</b>	0.079	0.706	0.404
Izobličenja nastala pri prijenosu mrežom	PLCC	<b>0.728</b>	0.710	0.654	0.490	0.245	0.661
	SROCC	0.824	<b>0.867</b>	0.739	0.487	0.341	0.652

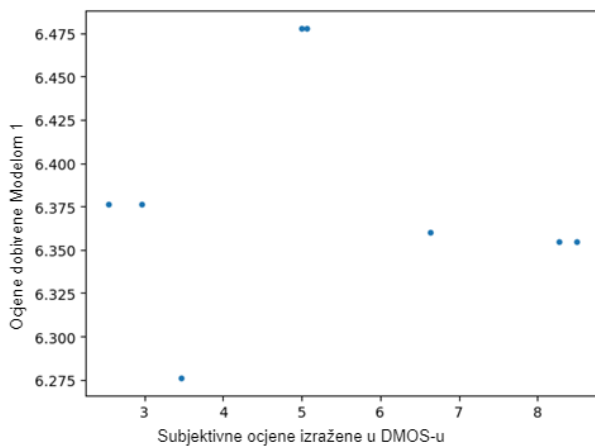
Iz rezultata prikazanih u tablici 4.2 za H.264 tip izobličenja, najbolji rezultat postigao je Model 1. Oba predložena modela su postigla bolje rezultate u odnosu na druge objektivne metrike s kojima su uspoređivani. Na slici 4.10 b) može se vidjeti da Model 2 lošije ocjenjuje video signale niže i srednje kvalitete, u odnosu na Model 1 koji lošije ocjenjuje video signale niže kvalitete. Od objektivnih metrika s kojima su predloženi modeli uspoređivani, najveću korelaciju s subjektivnim

ocjenama ima SSIM metrika, dok NIQE ima najmanju. Za ovaj tip izobličenja bilo je ukupno 27 video signala u testnom skupu.

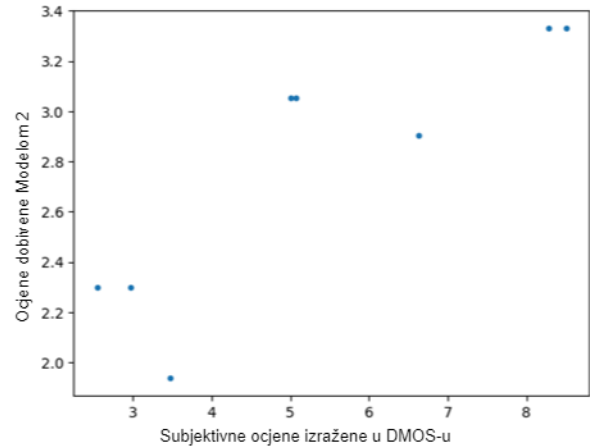


**Slika 4.10.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za H.264 tip izobličenja

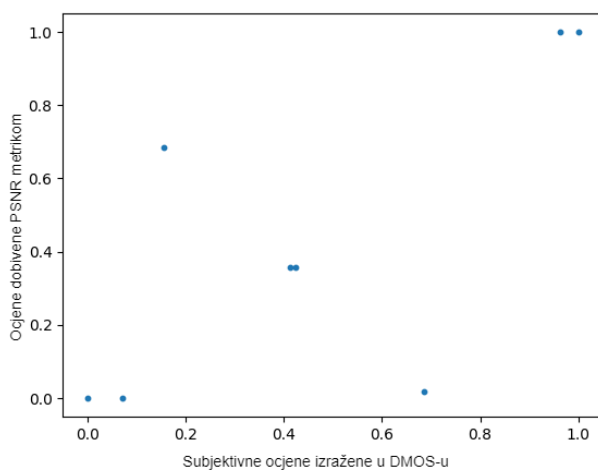
Iz tablice 4.2, gdje su prikazani rezultati testiranja za H.265 tip izobličenja, može se vidjeti kako Model 2 ima višu korelaciju sa subjektivnim ocjenama nego Model 1, no najbolji rezultati ostvareni su NIQE metrikom. Drugi najbolji rezultat ostvario je Model 2, dok Model 1 ostvaruje najlošije rezultate. Model 2 prema slici 4.11 b) lošije ocjenjuje video signale srednje kvalitete, dok Model 1 prema slici 4.11 b) izrazito loše ocjenjuje video signale, jer za sve daje slične rezultate. Model 2, iako je ostvario prilično veliku vrijednost PLCC, pokazuje raspon ocjena koji je između 1 i 3.4 dok su subjektivne ocjene u rasponu između 2 i 9. Za ovaj tip izobličenja bilo je ukupno 8 video signala u testnom skupu. Rezultati ostvareni za MPEG2 tip izobličenja jako su loši za Model 1 u odnos na objektivne metrike zasnovane na egzaktnim računalnim algoritmima, što se može vidjeti u tablici 4.2 i na slici 4.12 a). S druge strane, Model 2 daje puno bolje ocjene i bolje prepoznaje video signale različite razine izobličenja nego Model 1, no ako se uspoređi s rezultatima ostalih objektivnih metrika s kojim je uspoređivan on daje najlošije rezultate.



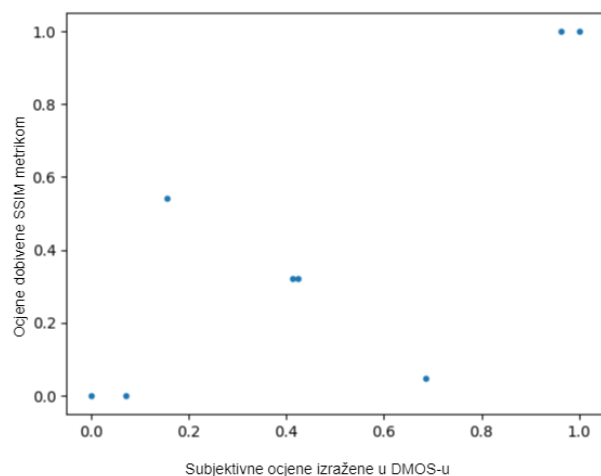
a)



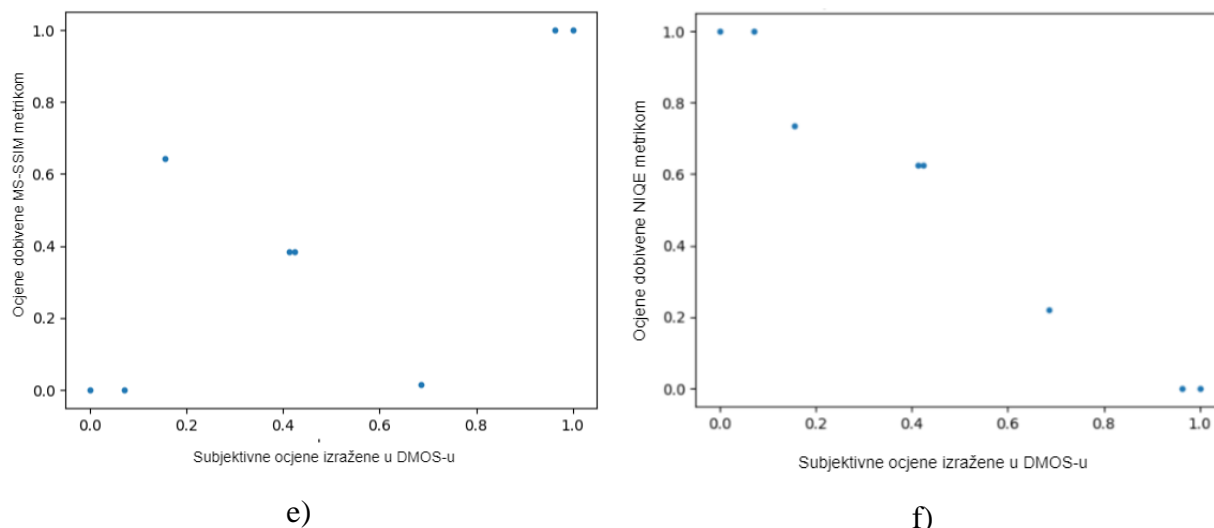
b)



c)

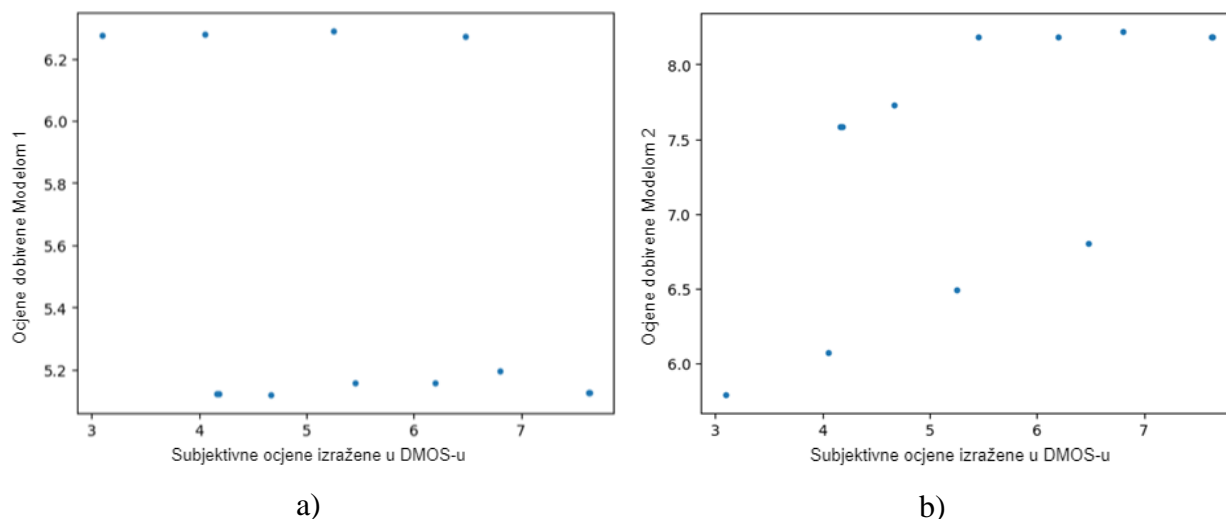


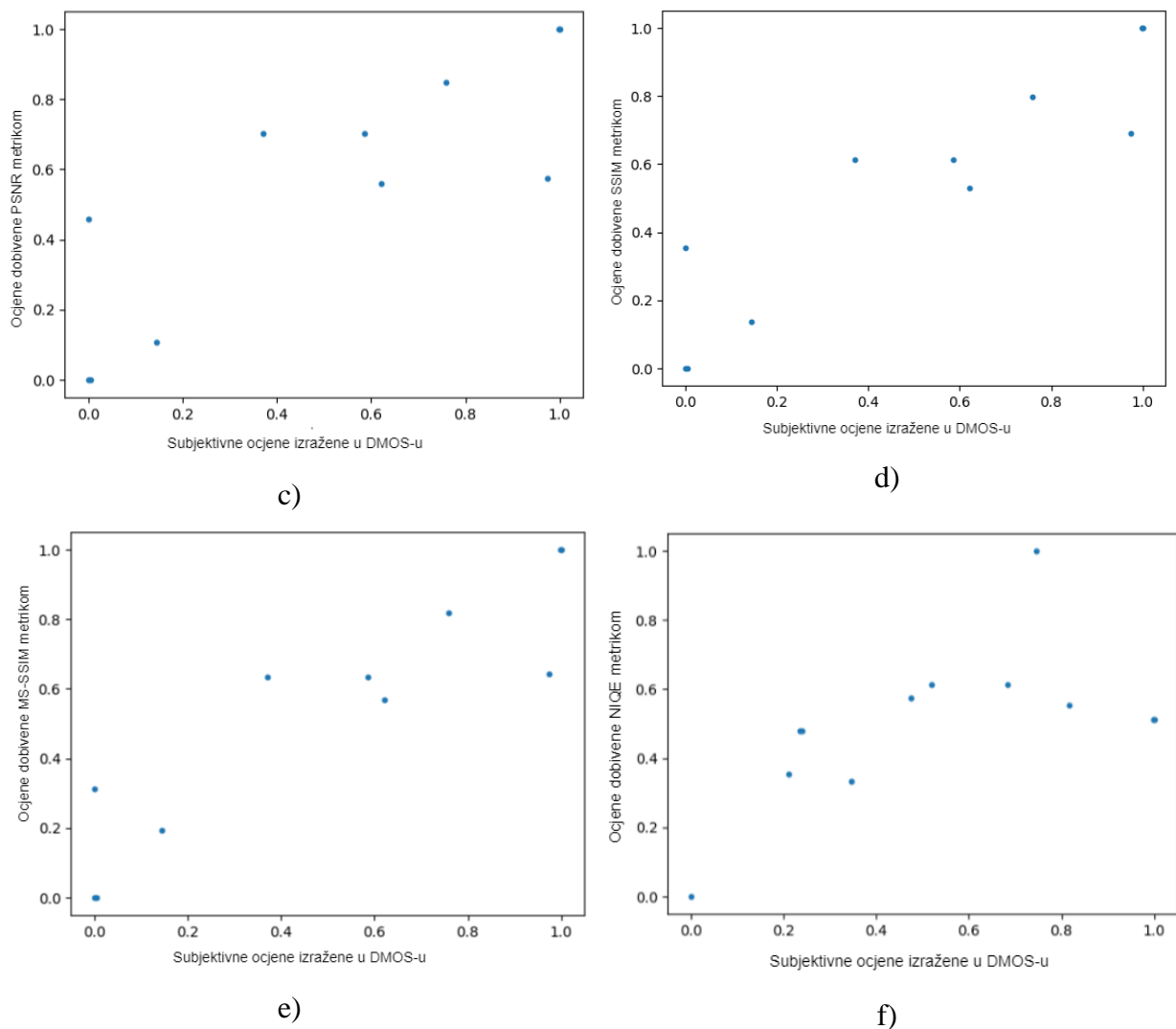
d)



**Slika 4.11.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za H.265 tip izobličenja

S druge strane, Model 2 daje puno bolje ocjene i bolje prepoznaje video signale različite razine izobličenja nego Model 1, no ako se uspoređi s rezultatima ostalih objektivnih metrika s kojim je uspoređivan on daje najlošije rezultate. Najbolje rezultate ostvareni su SSIM metrikom, no i ostale FR metrike postigle približne rezultate. NIQE je ostvario najlošije rezultate. Isto kao i kod H.265, raspon ocjena koje je dao Model 2 puno je manji u odnosu na raspon ocjena subjektivnih ocjena. Za ovaj tip izobličenja bilo je ukupno 12 video signala u testnom skupu.

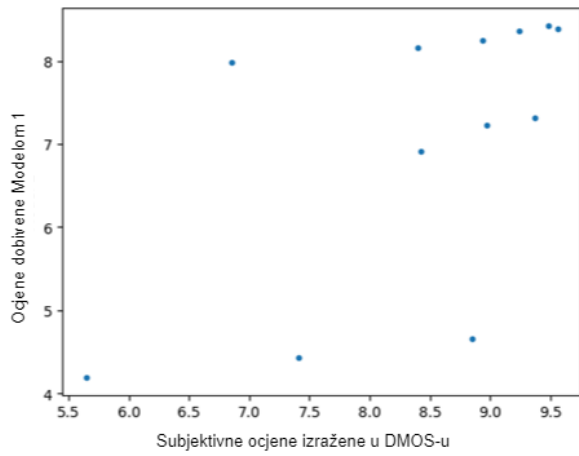




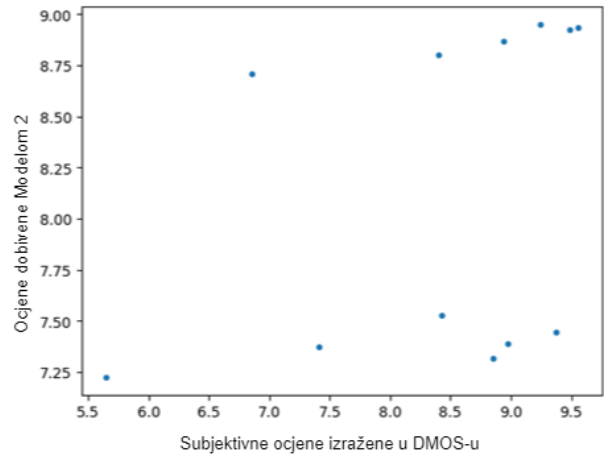
**Slika 4.12.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za MPEG2 tip izobličenja

Nadalje, rezultati za MPEG4 tip izobličenja iz tablice 4.2 prikazuju izuzetno dobru korelaciju objektivnih FR metrika i subjektivnih ocjena video signala testnog skupa. Najbolji rezultati su postignuti MS-SSIM metrikom. Od predloženih modela bolje se pokazao Model 1, čija je korelacija dosta veća nego za Model 2. Najlošije ocjene ostvarene su NIQE metrikom. Prema slici 4.13. a) može se vidjeti da Model 1 dobro ocjenjuje video signale više kvalitete. Za ovaj tip izobličenja bilo je ukupno 12 video signala u testnom skupu. Iz prikazanih rezultata u tablici 4.2, za tip izobličenja nastao smrzavanjem okvira vidljivo je da oba predložena modela imaju nisku linearnu korelaciju sa subjektivnim ocjenama testnog skupa, što je vidljivo na slikama 4.14 a) i b). Model 1 ostvaruje bolje rezultate u odnosu na Model 2, no oni su lošiji u odnosu na rezultate koji su dobiveni FR objektivnim metrikama.

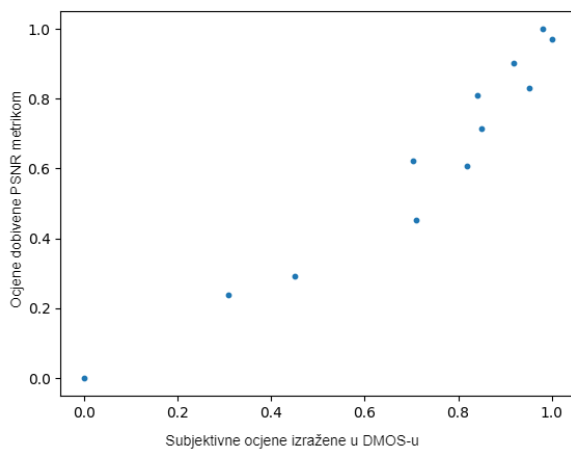




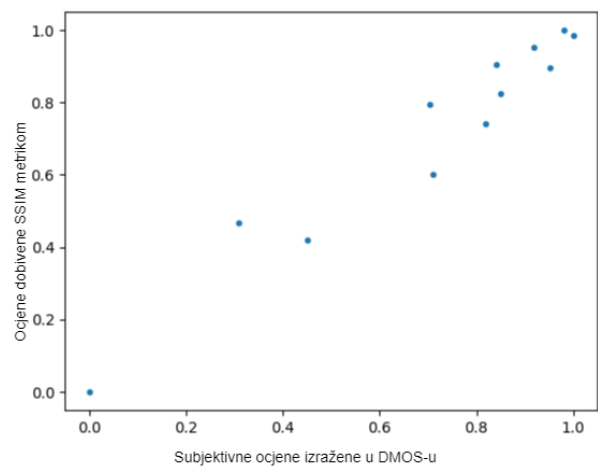
a)



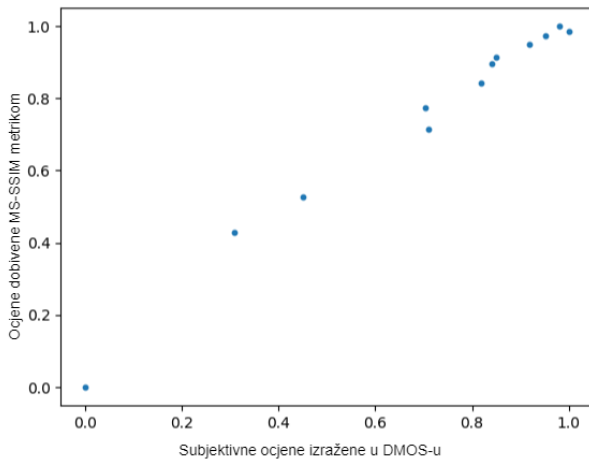
b)



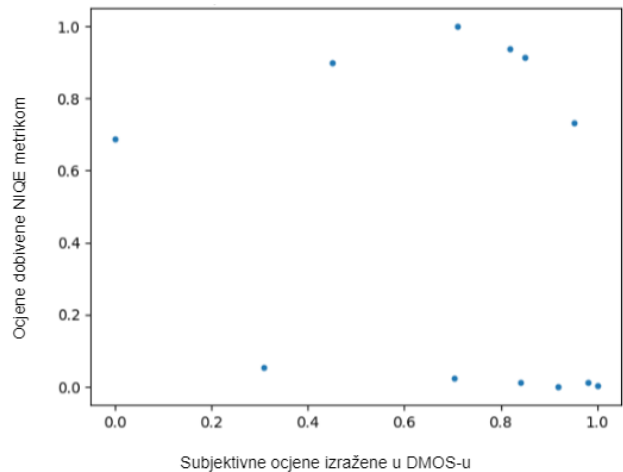
c)



d)



e)

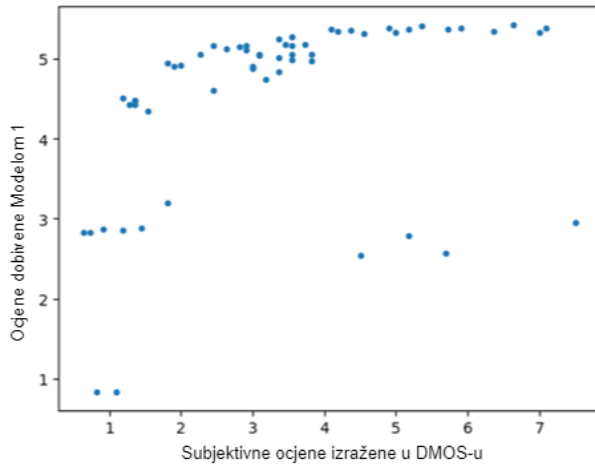


f)

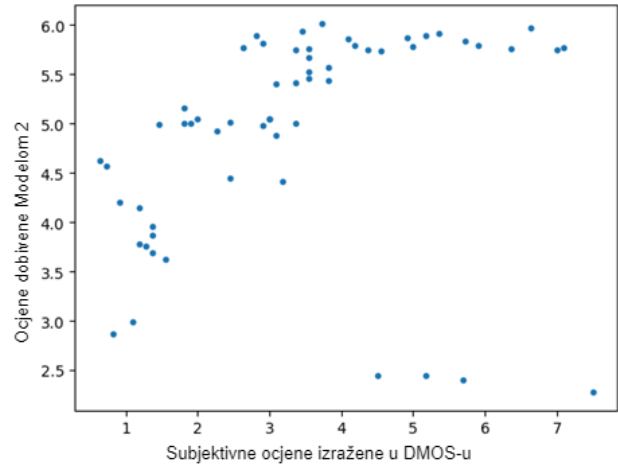
**Slika 4.13.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za MPEG4 tip izobličenja

U odnosu na NIQE koja je NR metrika, oba predložena modela ostvaruju bolje rezultate, dok je najveći korelacija dobivena SSIM metrikom. Za ovaj tip izobličenja bilo je ukupno 58 video

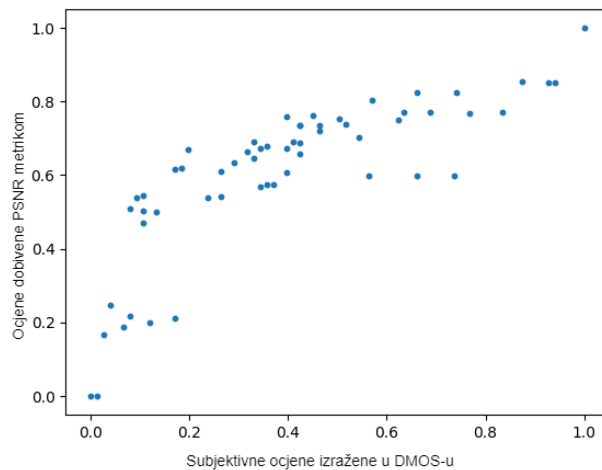
signala u testnom skupu. Nadalje, za tip izobličenja nastao određenim % gubitkom paketa u prijenosu IP mreži predloženi modeli ostvaruju bolje rezultate samo od NIQE metrike. Model 1 ostvaruje nešto bolje rezultate od Modela 2, dok najbolju korelaciju ima SSIM metrika. Prema slici 4.15 a) i b) može se vidjeti da predloženi modeli postižu dobre ocjene za video signale srednje kvalitete.



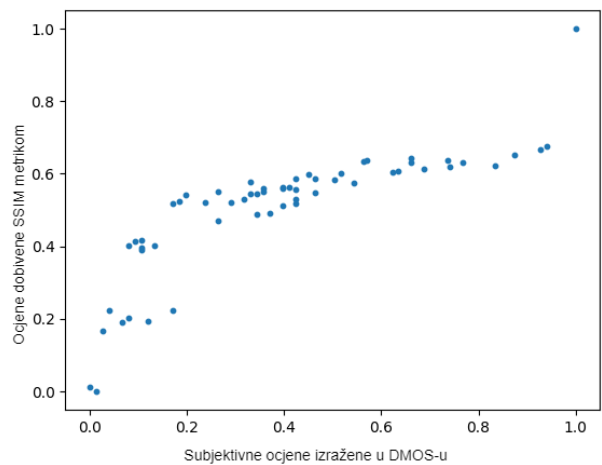
a)



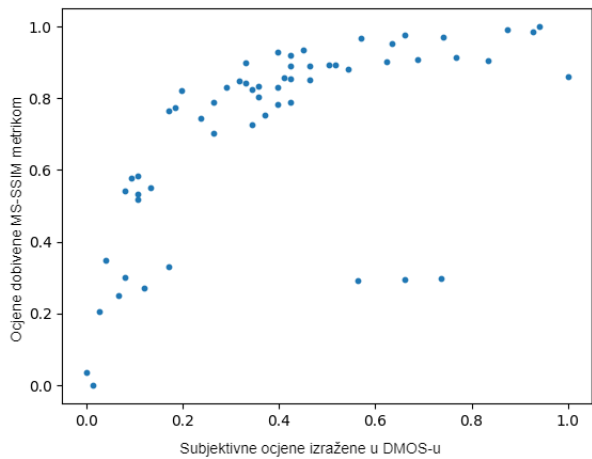
b)



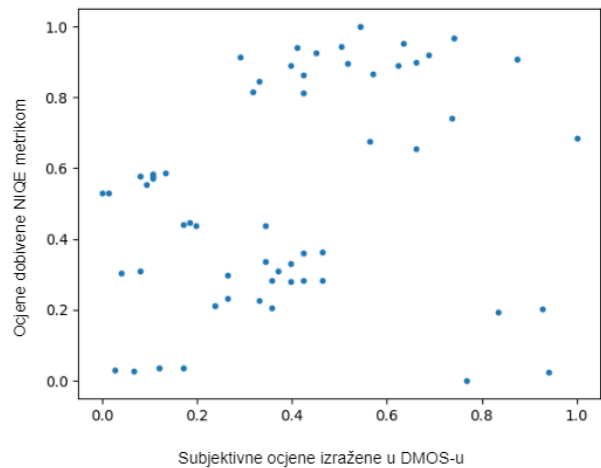
c)



d)



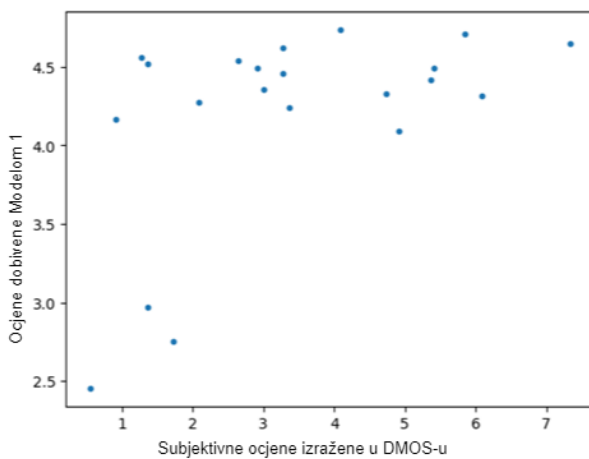
e)



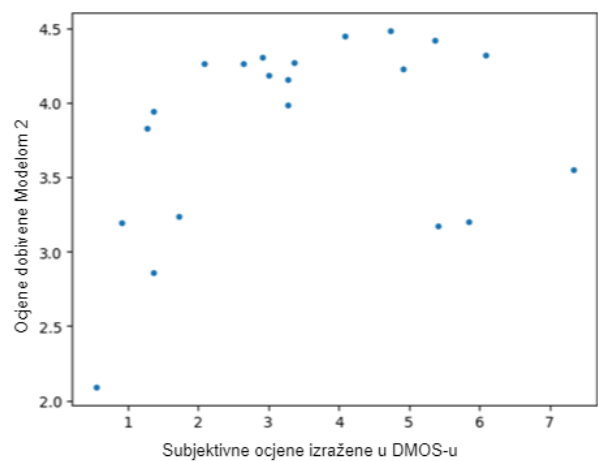
f)

**Slika 4.14.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo smrzavanje okvira

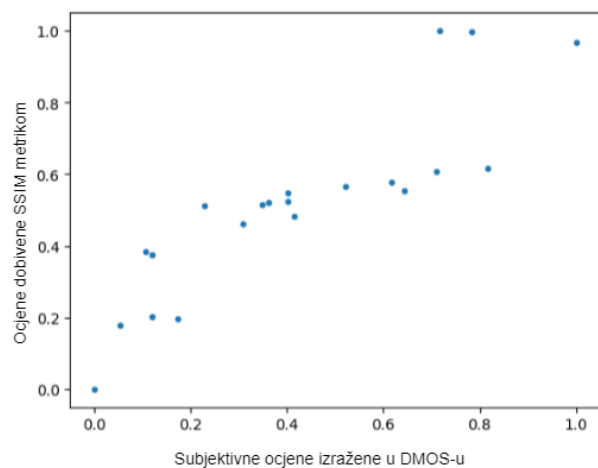
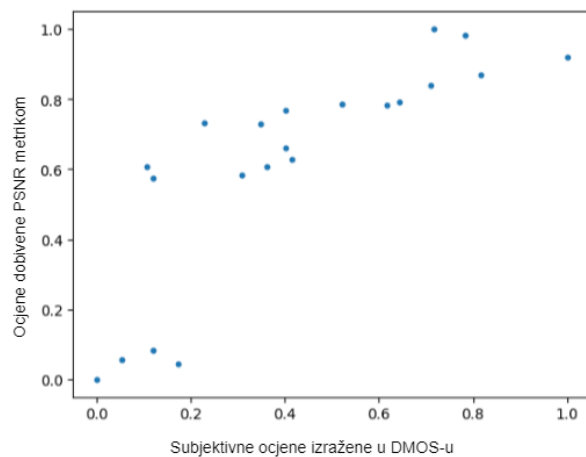
Najgori rezultat ostvaruje NIQE metrika. Za ovaj tip izobličenja bilo je ukupno 21 video signala u testnom skupu.

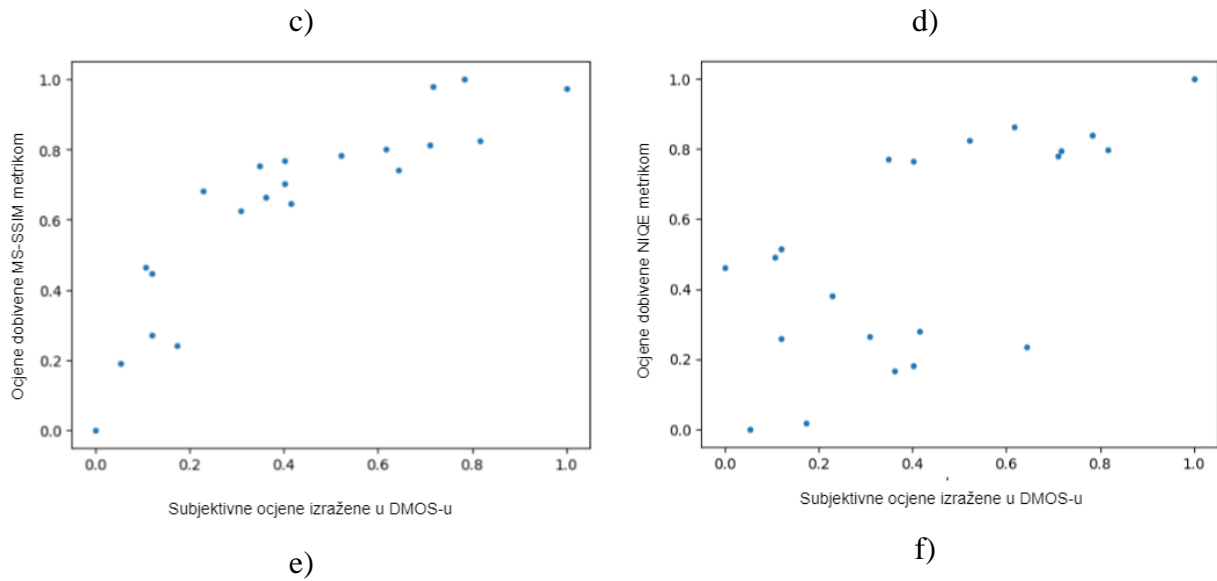


a)



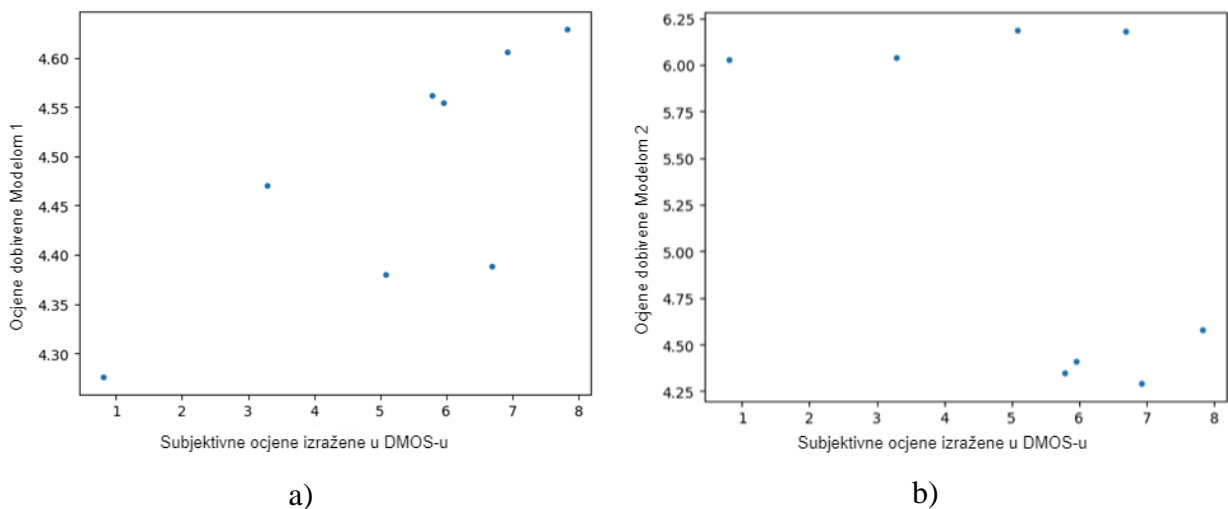
b)

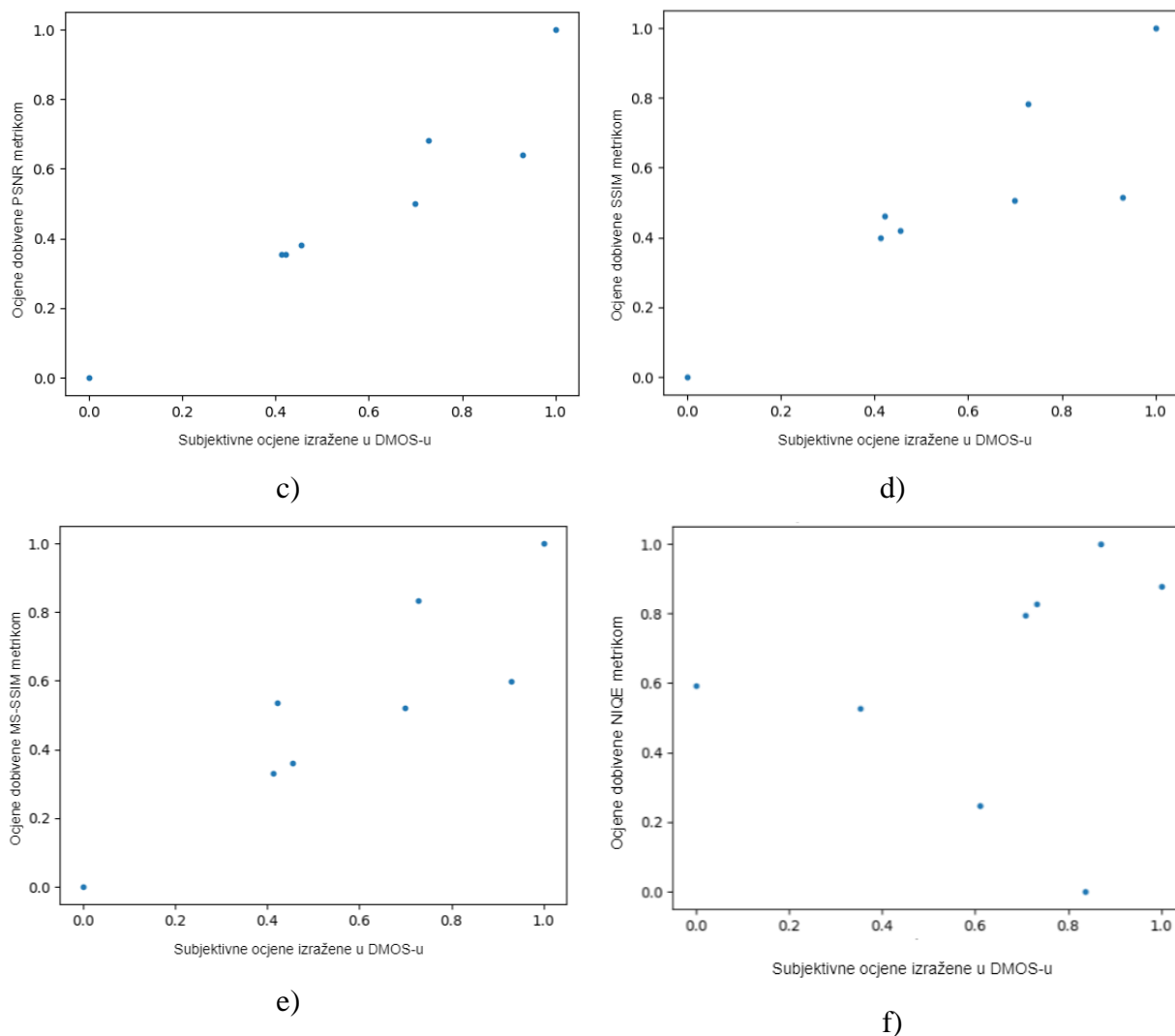




**Slika 4.15.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo određenim % gubitka paketa pri prijenosu IP mrežom

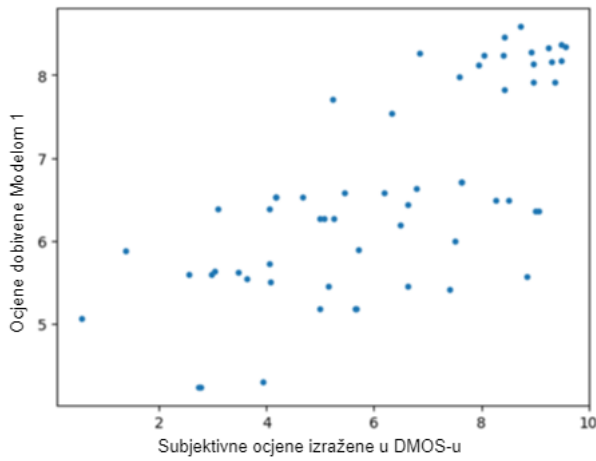
Model 1 za tip izobličenja nastao određenim % gubitka paketa pri prijenosu u bežičnoj mreži bolje rezultate u odnosu na izobličenje nastalo % gubitka paketa pri prijenosu IP mrežom. Isto tako Model 1 ostvaruje puno bolje rezultate nego Model 2. Dobiveni rezultati u tablici 4.2 za Model 2 prikazuju negativnu korelaciju sa subjektivnim ocjenama testnog skupa. Najbolji rezultat ostvaruje PSNR metrika i to ima veliku korelaciju sa subjektivnim ocjena, dok Model 1 ima najveći iznos monotonosti izražen SROCC-om. Odnosi sa subjektivnim ocjena za sve metrike vidljivi su na slici 4.16. Za ovaj tip izobličenja bilo je ukupno 10 video signala u testnom skupu.



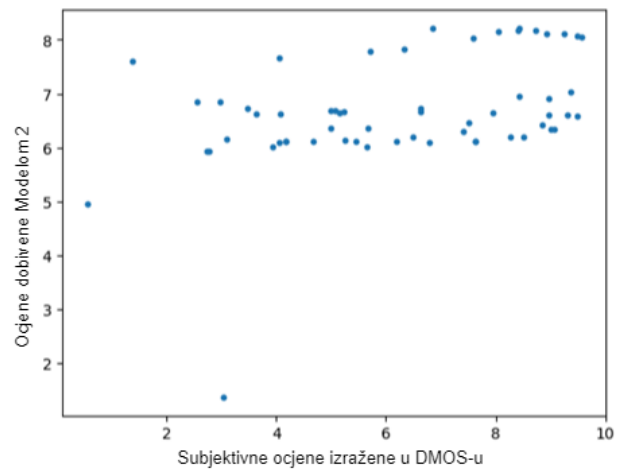


**Slika 4.16.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo određenim % gubitka paketa pri prijenosu bežične mreže

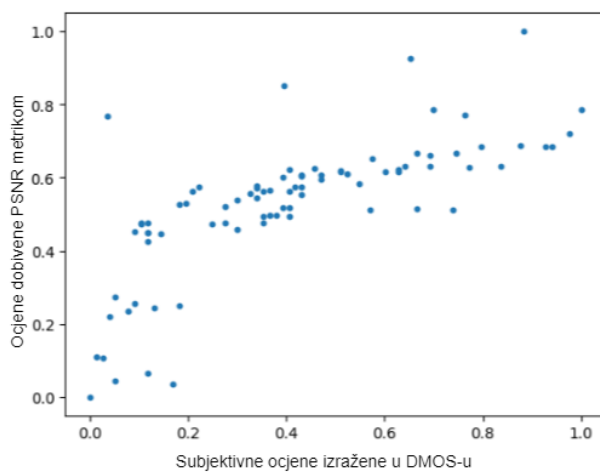
Za kombinaciju izobličenja uzrokovanih kompresijom Model 1 daje najbolje rezultate u vidu PLCC vrijednosti, dok najveći iznos SROCC-a daje MS-SSIM metrika. Model 1 ostvaruje bolje rezultate nego Model 2. Prema slici 4.17 a) može se vidjeti da Model 2 lošije ocjenjuje video signale niže i srednje kvalitete, dok Model 1 lošije ocjenjuje video signale niže kvalitete. Najveću korelaciju sa subjektivnim ocjenama od objektivnih metrika zasnovanih na egzaktnim računalnim algoritmima daje PSNR metrika. Najlošiji rezultat daje NIQE metrika. Za ovaj tip izobličenja bilo je ukupno 59 video signala u testnom skupu.



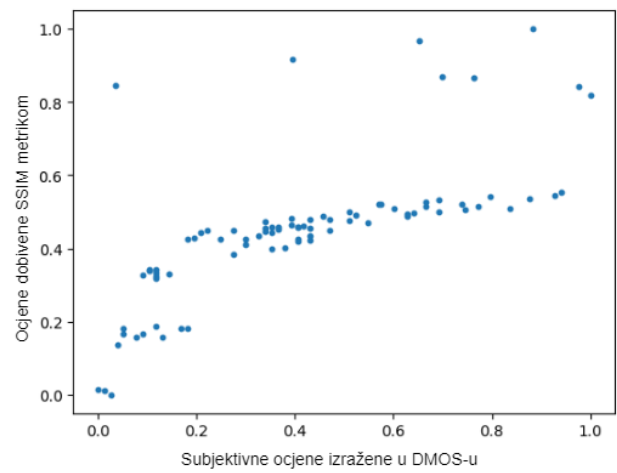
a)



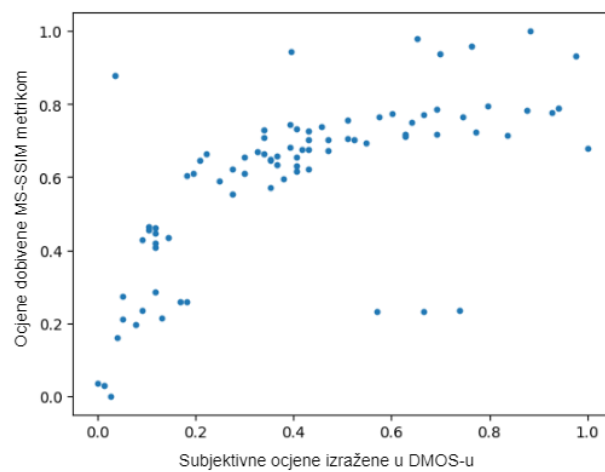
b)



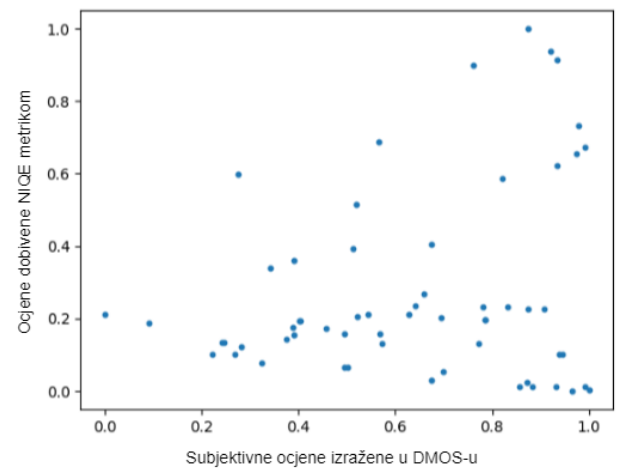
c)



d)



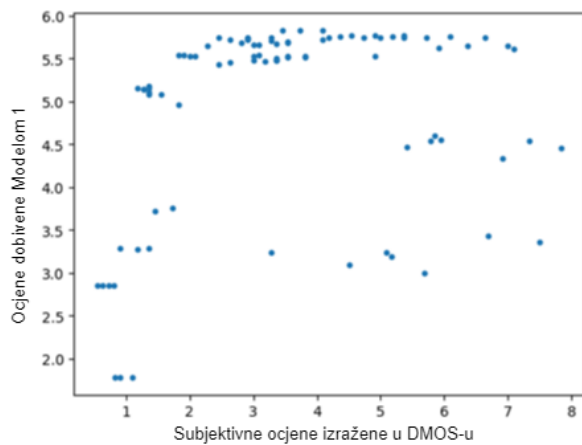
e)



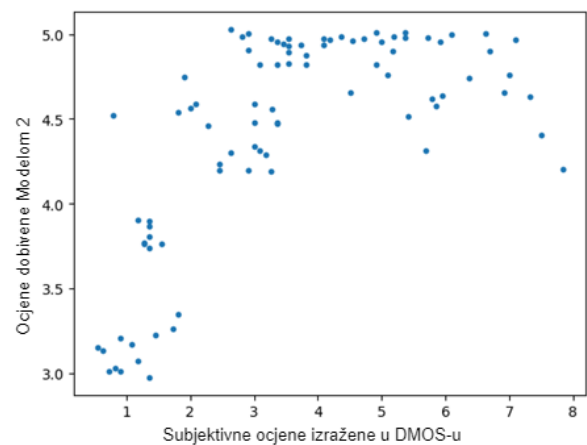
f)

**Slika 4.17.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metričke, d) SSIM metričke, e) MS-SSIM metričke i f) NIQE metričke za izobličenje nastalo kompresijom

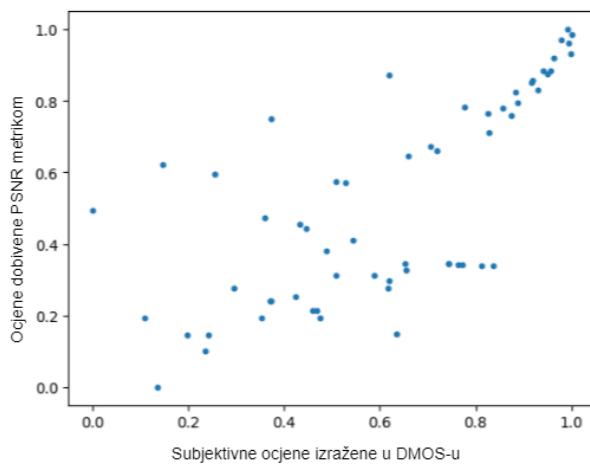
Nadalje, iz rezultata prikazanih u tablici 4.2 izobličenje nastalo pri prijenosu mrežom vidljivo je kako je Model 2 postigao veću korelaciju od MS-SSIM i NIQE metrike, dok je Model 1 bolje rezultate postigao jedino od NIQE metrike. Za Model 1 distribucija predviđenih ocjena je nelinearna, što se može vidjeti na slikama 4.18 a). Najveću linearnost ostvaruje PSNR metrika. Model 2 dobro ocjenjuje video signale srednje kvalitete. Za ovaj tip izobličenja bilo je ukupno 87 video signala u testnom skupu.



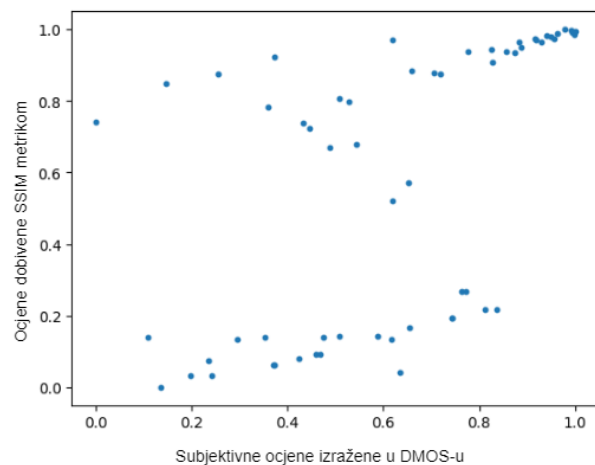
a)



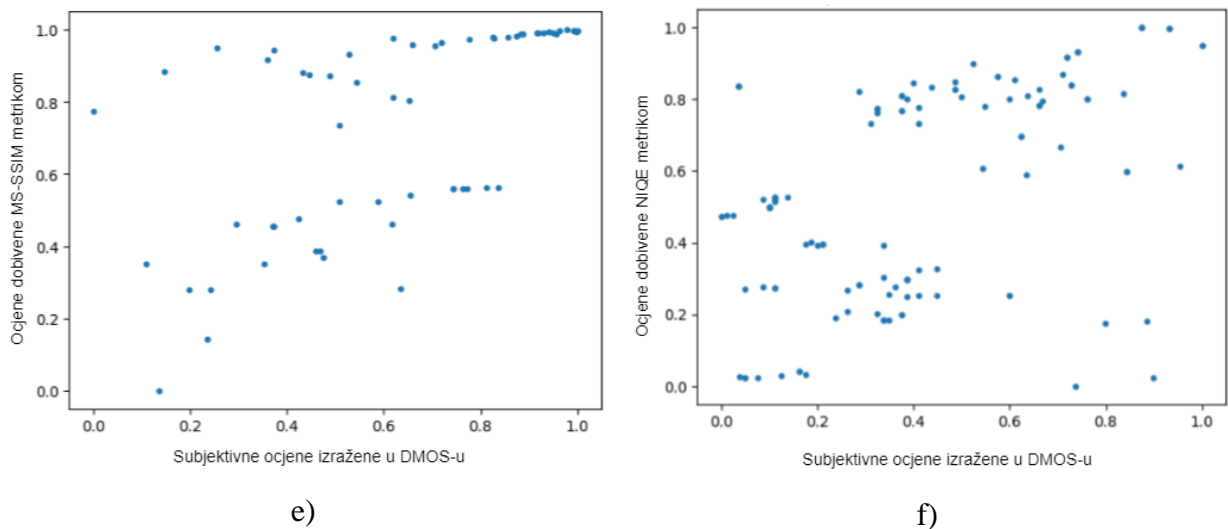
b)



c)



d)



**Slika 4.18.** Grafički prikaz odnosa subjektivnih ocjena i ocjena : a) Model 1, b) Model 2, c) PSNR metrike, d) SSIM metrike, e) MS-SSIM metrike i f) NIQE metrike za izobličenje nastalo pri prijenosu mrežom

### 4.3. Analiza rezultata

Iz prethodno opisanih rezultata testiranja za različite tipove izobličenja, za prvi način podjele podijele baze podataka može se zaključiti da predloženi modeli postižu puno bolje rezultate nego objektivne metrike zasnovane na egzaktnim računalnim algoritmima. S druge strane, za drugi način podjele baze podataka, Model 1 ostvaruje bolje rezultate za H.264 tip izobličenja i izobličenja nastala zbog kompresije nego ostale objektivne metrike s kojima je uspoređen. Model 2 ostvaruje dobre rezultate za H.265 tip izobličenja i izobličenja nastala pri prijenosu mrežom. Ako se pogledaju rezultati za FR objektivne metrike za oba načina testiranja, može se vidjeti da su korelacije sa subjektivnim ocjenama puno veće kod drugog načina podjele baze podataka. Iz toga se može zaključiti da objektivne FR metrike postižu veću korelaciju sa subjektivnim ocjenama kada se u testnom skupu nalaze različiti video signali komprimirani/izobličeni različitim razina izobličenja, ali su istog sadržaja. Puno manje koreliraju sa subjektivnim ocjenama kada se testni skup sastoji od više različitih sadržaja sličnog tipa, bez obzira na količinu izobličenja u signalima.

Model 2 se pokazao puno bolji pri ocjenjivanju kvalitete video signala kada se sadržaji iz trening skupa nalaze i u testnom skupu, no nije velika razlika u odnosu na Model 1. S druge strane Model 1 pokazao se bolji pri drugom načinu testiranja i to gotovo na svim tipovima izobličenja, osim na MPEG2. Predloženi modeli nisu se dobro pokazali za većinu tipova izobličenja na drugom načinu testiranja iz razloga što je rezolucija video signala na kojima su trenirani puno manja od originalne rezolucije za pojedine baze. Kada se rezolucija video signala smanji, izobličenja su



manje vidljiva i model ne može dobro prepoznati različite razine izobličenja. Stoga kod nekih tipova izobličenja se javlja manji raspon predviđenih ocjena, i veća vrijednost PLCC-a. Osim toga promjenom same rezolucije promijenjen je i omjer slike što znatno mijenja vizualnu kvalitetu video signala. To utječe na raspored piksela unutar video signala i s time na originalno izobličenje. Jedna prednosti smanjenja rezolucije kod ovakvog zadatka je brže procesiranje video signala i pogodnije je za sustave koje rade u stvarnom vremenu. S druge strane, objektivne metrike dobivaju video signale pune rezolucije i puno bolje vide izobličenja u video signalima. Budući da su se značajke pojedinih video signala koji su se ocjenjivali Modelom 1 i Modelom 2 izdvajane samo iz prvih 300 okvira videa (bez obzira na njegovu veću duljinu), predloženi su modeli samo na temelju tih 300 okvira dali svoje ocjene. Kada bi se analizirali svi okviri svakog video signala, ocjene predloženih modela sigurno bi bile bolje i tu svakako leži prostor za napredak. Predloženi modeli spadaju pod NR metode za ocjenjivanje video signala i kao takvi imaju znatnu prednost u sustavima gdje ne postoji pristup referentnom video signalu ni njegovim značajkama, tj. gdje se ne mogu koristiti FR i RR metrike.

## 5. ZAKLJUČAK

U ovom radu opisano je što su to subjektivna i objektivna procjena kvalitete video signala i na koji način se provode. Osim toga, opisana je problematika ocjenjivanja video signala koje sadrže različite tipove izobličenja, te na koje načine se rješava. Nadalje, opisana su neka postojeća rješenja za ocjenjivanje kvalitete video signala koja se temelje na egzaktnim računalnim algoritmima i strojnom učenju. Dan je opis postojećih baza video podataka koje sadrže H.264, H.265, MPEG2, MPEG4 izobličenja i izobličenja nastala smrzanjem, gubitkom paketa u bežičnoj i IP mreži.

U sklopu diplomskog rada razvijena su rješenja za ocjenjivanje video signala zasnovana na strojnom učenju. Predložene su dvije arhitekture duboke neuronske mreže koje su trenirane na dva različita načina. Za svaki način odrađeno je po 9 zasebnih treniranja, 7 za svaki tip izobličenja i po 1 treniranje za izobličenja nastala kompresijom i izobličenja nastala pri prijenosu mrežnom. Za prvi način treniranja video signali su podijeljeni na način da su svi sadržaji prisutni u skupovima za trening, test i validaciju, dok je drugi način obuhvaćao podjelu video signala tako da su se u testnom i validacijskom skupu nalazili sadržaji koji se nisu nalazili u trening skupu. Prema rezultatima treniranja može se zaključiti da Model 1 u većini slučajeva ostvaruje bolje rezultate nego Model 2. Osim toga, iz grafova za trening i validaciju Modela 2 vidljivi su određeni skokovi u vrijednostima za PLCC i MSE koji su uzrokovani preprilagodбом (engl. *overfitting*) modela na trening skup podataka.

Rezultati testiranja za prvi način podjele video signala pokazuju visoku korelaciju obaju predloženih modela sa subjektivnim ocjenama. Model 2 se ipak boljim pokazao za većinu izobličenja, gdje mu je PLCC za većinu izobličenja veći od 0.9, dok se Model 1 boljim pokazao za H.265 tip izobličenja, gdje je ostvario SROCC od 1. Nadalje, za drugi način podjele video signala, Model 1 postiže bolje rezultate nego Model 2, no i dalje su rezultati lošiji u odnosu na objektivne metrike. Gotovo za svaku vrstu izobličenja, kod drugog načina podjele, objektivne FR metrike zasnovane na egzaktnim računalnim algoritmima imale su bolje rezultate od predloženih modela. Model 1 ipak se pokazao bolji od drugih objektivnih metrika za H.264 tip izobličenja i izobličenja uzrokovana kompresijom. NIQE NR metrika pokazala je iznenađujuće loše rezultate pri ocjenjivanju video signala s različitim izobličenjima, ali treba uzeti u obzir da je to NR metrika i kao takve nema pristup referentnom video signalu prilikom ocjenjivanja, za razliku od PSNR, SSIM i MS-SSIM koji to imaju.

Kako bi se poboljšali rezultati za Model 1, bilo bi dobro trenirati model s podacima koji osim što daju informaciju o prostornim značajkama video signala, daju informaciju o razlici između okvira, tj. na neki način dati mu informaciju o vremenskim značajkama video signala. Osim toga, dobro bi bilo pokušati istrenirati oba predložena modela na video signalima originalne rezolucije ili rezolucije koja održava originalni omjer slike. U tom slučaju bolje bi mogli prepoznati različite tipove izobličenja i različite razine izobličenja. No takav način nije prikladan za sustave koji rade u stvarnom vremenu jer procesiranje video signala pune rezolucije zahtjeva veliku količinu memorije i sporije procesiranje video signala, posebno ako se radi o procesiranju video signala 8K ili 4K rezolucije. S obzirom da su ocjene predloženih modela zasnovane na prvih 300 okvira video signal, određeni broj okvira nekih video signala nije analiziran. Kako bi se taj problem riješio, postoji mogućnost implementacije prilagodljivog sažimanja podataka, koji će za sve podatke istog oblika dati jednaki izlazni oblik.

## LITERATURA

- [1] S. Chikkerur, V. Sundaram, M. Reisslein, i L. J. Karam, „Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison“, *IEEE Trans. Broadcast.*, sv. 57, izd. 2, str. 165–182, lip. 2011, doi: 10.1109/TBC.2011.2104671.
- [2] A. C. Bovik, „Perceptual image processing: Seeing the future“, *Proc IEEE*, sv. 98, izd. 11, str. 1799–1803, stu. 2010.
- [3] „Subjective video quality assessment methods for multimedia applications“. International Telecommunication Union, 2022.
- [4] M. Shahid, *Methods for objective and subjective video quality assessment and for speech enhancement*,. Blekinge Institute of Technology, Sweden, 2014.
- [5] Wang, Zhou, Sheikh, Hamid, i Bovik, Alan, „41 OBJECTIVE VIDEO QUALITY ASSESSMENT“, 2003.
- [6] M. A. Usman, „PERFORMANCE EVALUATION OF VIDEO QUALITY ASSESSMENT METHODS BASED ON FRAME FREEZING“, 2014, doi: 10.13140/RG.2.2.10940.85126.
- [7] S. Dost, F. Saud, M. Shabbir, M. G. Khan, M. Shahid, i B. Lovstrom, „Reduced reference image and video quality assessments: review of methods“, *EURASIP J. Image Video Process.*, sv. 2022, izd. 1, str. 1, sij. 2022, doi: 10.1186/s13640-021-00578-y.
- [8] I. Richardson, *The H.264 advanced video compression standard*. Wiley, Ujedinjeno Kraljevstvo, 2010.
- [9] Z. Wang, A. C. Bovik, H. R. Sheikh, i E. P. Simoncelli, „Image Quality Assessment: From Error Visibility to Structural Similarity“, *IEEE Trans. Image Process.*, sv. 13, izd. 4, str. 600–612, tra. 2004, doi: 10.1109/TIP.2003.819861.
- [10] Z. Wang, E. P. Simoncelli, i A. C. Bovik, „Multiscale structural similarity for image quality assessment“, u *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Pacific Grove, CA, USA: IEEE, 2003, str. 1398–1402. doi: 10.1109/ACSSC.2003.1292216.
- [11] J. Whittle, M. W. Jones, i R. Mantiuk, „Analysis of reported error in Monte Carlo rendered images“, *Vis. Comput.*, sv. 33, izd. 6–8, str. 705–713, lip. 2017, doi: 10.1007/s00371-017-1384-7.
- [12] A. Mittal, R. Soundararajan, i A. C. Bovik, „Making a “Completely Blind” Image Quality Analyzer“, *IEEE Signal Process. Lett.*, sv. 20, izd. 3, str. 209–212, ožu. 2013, doi: 10.1109/LSP.2012.2227726.
- [13] D. L. Ruderman, „The statistics of natural images“, *Netw. Comput. Neural Syst.*, sv. 5, izd. 4, str. 517–548, sij. 1994, doi: 10.1088/0954-898X\_5\_4\_006.
- [14] R. Soundararajan i A. C. Bovik, „Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing“, *IEEE Trans. Circuits Syst. Video Technol.*, sv. 23, izd. 4, str. 684–694, tra. 2013, doi: 10.1109/TCSVT.2012.2214933.
- [15] A. Mittal, M. A. Saad, i A. C. Bovik, „A Completely Blind Video Integrity Oracle“, *IEEE Trans. Image Process.*, sv. 25, izd. 1, str. 289–300, sij. 2016, doi: 10.1109/TIP.2015.2502725.
- [16] W. S. Noble, „What is a support vector machine?“, *Nat. Biotechnol.*, sv. 24, izd. 12, str. 1565–1567, pros. 2006, doi: 10.1038/nbt1206-1565.
- [17] A. Mittal, A. K. Moorthy, i A. C. Bovik, „No-Reference Image Quality Assessment in the Spatial Domain“, *IEEE Trans. Image Process.*, sv. 21, izd. 12, str. 4695–4708, pros. 2012, doi: 10.1109/TIP.2012.2214050.
- [18] M. A. Saad, A. C. Bovik, i C. Charrier, „Blind Prediction of Natural Video Quality“, *IEEE Trans. Image Process.*, sv. 23, izd. 3, str. 1352–1365, ožu. 2014, doi: 10.1109/TIP.2014.2299154.

- [19] D. Li, T. Jiang, i M. Jiang, „Quality Assessment of In-the-Wild Videos“, u *Proceedings of the 27th ACM International Conference on Multimedia*, Nice France: ACM, lis. 2019, str. 2351–2359. doi: 10.1145/3343031.3351028.
- [20] „Keras, the Python deep learning API“. [Na internetu]. Dostupno na: <https://keras.io/>
- [21] „Jupyter Notebook“. [Na internetu]. Dostupno na: <https://docs.jupyter.org/en/latest/reference/content-reference.html>
- [22] „Anaconda“. [Na internetu]. Dostupno na: <https://docs.anaconda.com/free/anaconda/reference/index.html>
- [23] „Skvideo.io, Video input and output utilities“. [Na internetu]. Dostupno na: <http://www.scikit-video.org/stable/modules/io.html#module-skvideo.io>
- [24] „OpenCV,“ OpenCV.“ [Na internetu]. Dostupno na: <https://opencv.org/>
- [25] „CSV, file reading and writing“. [Na internetu]. Dostupno na: <https://docs.python.org/3/library/csv.html>
- [26] „MSU VQMT, Video Quality Measurement Tool“. [Na internetu]. Dostupno na: <https://videoprocessing.ai/vqmt/vqmt-doc-toc/>
- [27] V. Bajcinovci, M. Vranjes, D. Babic, i B. Kovacevic, „Subjective and objective quality assessment of MPEG-2, H.264 and H.265 videos“, u *2017 International Symposium ELMAR*, Zadar: IEEE, ruj. 2017, str. 73–77. doi: 10.23919/ELMAR.2017.8124438.
- [28] S. Rimac-Drlje, M. Vranješ, i D. Žagar, „Foveated mean squared error—a novel video quality metric“, *Multimed. Tools Appl.*, sv. 49, izd. 3, str. 425–445, ruj. 2010, doi: 10.1007/s11042-009-0442-1.
- [29] M. Vranješ, S. Rimac-Drlje, i K. Grgić, „Review of objective video quality metrics and performance comparison using different databases“, *Signal Process. Image Commun.*, sv. 28, izd. 1, str. 1–19, sij. 2013, doi: 10.1016/j.image.2012.10.003.
- [30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, i L. K. Cormack, „Study of Subjective and Objective Quality Assessment of Video“, *IEEE Trans. Image Process.*, sv. 19, izd. 6, str. 1427–1441, lip. 2010, doi: 10.1109/TIP.2010.2042111.
- [31] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, i L. K. Cormack, „A subjective study to evaluate video quality assessment algorithms“, predstavljeno na IS&T/SPIE Electronic Imaging, B. E. Rogowitz i T. N. Pappas, Ur., San Jose, California, velj. 2010, str. 75270H. doi: 10.1117/12.845382.
- [32] A. K. Moorthy, L. K. Choi, A. C. Bovik, i G. De Veciana, „Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies“, *IEEE J. Sel. Top. Signal Process.*, sv. 6, izd. 6, str. 652–671, lis. 2012, doi: 10.1109/JSTSP.2012.2212417.
- [33] A. Telili, S. A. Fezza, W. Hamidouche, and H. F. Z. B. Meftah, „2BiVQA: Double Bi-LSTM based Video Quality Assessment of UGC Videos.“ arXiv, 2022. doi: 10.48550/ARXIV.2208.14774.
- [34] D. P. Kingma i J. Ba, „Adam: A Method for Stochastic Optimization“. arXiv, 2014. doi: 10.48550/ARXIV.1412.6980.
- [35] F. Zinzendoff Okwonu, B. Laro Asaju, i F. Irimisose Arunaye, „Breakdown Analysis of Pearson Correlation Coefficient and Robust Correlation Methods“, IOP Conference Series: Materials Science and Engineering, sv. 917, izd. 1. IOP Publishing, str. 012065, ruj. 01, 2020. doi: 10.1088/1757-899x/917/1/012065.
- [36] P. Sedgwick, „Spearman’s rank correlation coefficient“, *BMJ*. *BMJ*, str. g7327, stu. 28, 2014. doi: 10.1136/bmj.g7327.

## SAŽETAK

U sklopu ovog diplomskog rada izrađen je algoritam za ocjenjivanje video signala zasnovan na strojnom učenju. Opisano je što su to subjektivna i objektivna procjena kvalitete video signala i na koji način se provode. Osim toga, dan je pregled postojećih algoritama za ocjenjivanje video signala zasnovanih na egzaktnim računalnim algoritmima te nekih postojećih algoritama za ocjenjivanje video signala zasnovanih na strojnom učenju. Nadalje, opisane su korištene baze video signala koje sadrže H.264, H.265, MPEG2, MPEG4 izobličenja i izobličenja nastala smrzanjem, određenim % gubitka paketa pri prijenosu IP i bežičnim mrežama. Predložena su dva modela zasnovana na strojnom učenju koja rješavaju problem ocjenjivanja video signala bez potrebe za referentnim signalom. Za realizaciju rješenja korištena je *Tensorflow* biblioteka za *Python* programski jezik. Prikazani su rezultati testiranja za predložene modele koji su uspoređeni s rezultatima PSNR, SSIM, MS-SSIM i NIQE metrika. Predloženi modeli točnije predviđaju ocjene na testnom skupu nego objektivne metrike s kojima su uspoređeni za prvi način podjele podataka. Za drugi način podjele podataka Model 1 postiže bolje rezultate u odnosu na Model 2, no u odnosu na objektivne metrike s kojima je uspoređivan postiže bolje rezultate samo za H.264 tip izobličenja i izobličenja uzrokovana kompresijom.

**Ključne riječi:** subjektivna ocjena video kvalitete, objektivna ocjena video kvalitete, CNN, LSTM, sloj pažnje

# VIDEO QUALITY ASSESSMENT BASED ON MACHINE LEARNING

## ABSTRACT

Within this master's thesis, an algorithm for assessing video signals based on machine learning was developed. It describes the concepts of subjective and objective quality assessment of video signals and the methods used for conducting them. Additionally, an overview of existing algorithms for video signal assessment based on exact computational algorithms, as well as some existing algorithms based on machine learning, is provided. Furthermore, the used video signal databases containing H.264, H.265, MPEG2, MPEG4 distortions, and freeze distortions, as well as % of packet loss in IP and wireless networks, are described. Two machine learning-based models are proposed to address the problem of video signal assessment without the need for a reference signal. The TensorFlow library for the Python programming language was used for implementing the solutions. The testing results for the proposed models are presented and compared with the results of the PSNR, SSIM, MS-SSIM, and NIQE metrics. The proposed models achieve more accurate predictions on the test set compared to the objective metrics they were compared to for the first data partition method. For the second data partition method, Model 1 performs better than Model 2, but it achieves better results compared to the objective metrics only for H.264 distortion type and compression-induced distortions.

**Keywords:** Subjective video quality assessment, objective video quality assessment, CNN, LSTM, Attention layer

## **ŽIVOTOPIS**

Tomislav Čosić rođen je 4.8.1999. u Osijeku. Završio je elektrotehničku i prometnu školu u Osijeku s vrlo dobrim uspjehom. Nakon srednje škole upisuje Fakultet elektrotehnike računarstva i informacijskih tehnologija u Osijeku, smjer Elektrotehnika. Godine 2021. stječe akademski naziv sveučilišni prvostupnik (baccalaureus) inženjer elektrotehnike i informacijske tehnologije. Iste godine upisuje diplomski sveučilišni studij elektrotehnika, smjer mrežne tehnologije na istom fakultetu i uz to postaje stipendist grupacije TTech Auto u Osijeku što se nastavilo i sljedeće godine. Aktivno se služi engleskim jezikom u govoru i pismu.



## PRILOZI

- P.3.1. konfiguracijska datoteka potrebna za stvaranje programskog okruženja (priloženo na DVD-u uz rad)
- P.3.2. *feature\_extraction.py*, *train.py*, *test.py* datoteke (priloženo na DVD-u uz rad)
- P.3.3. Excel datoteke za pojedini tip izobličenja koje sadrže nazive videa, DMOS, direktorij videa, format videa, rezoluciju videa podijeljene u mape za prvi način podjele (priloženo na DVD-u uz rad)
- P.3.4. Excel datoteke za pojedini tip izobličenja koje sadrže nazive videa, DMOS, direktorij videa, format videa, rezoluciju videa podijeljene u mape za drugi način podjele (priloženo na DVD-u uz rad)