

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 2387

**WEB APLIKACIJA ZA KLASIFIKACIJU GLAZBENIH
ISJEČAKA U GLAZBENA RAZDOBLJA**

Antun Modrušan

Zagreb, lipanj 2021.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 2387

**WEB APLIKACIJA ZA KLASIFIKACIJU GLAZBENIH
ISJEČAKA U GLAZBENA RAZDOBLJA**

Antun Modrušan

Zagreb, lipanj 2021.

DIPLOMSKI ZADATAK br. 2387

Pristupnik: **Antun Modrušan (0036490733)**

Studij: Računarstvo

Profil: Računarska znanost

Mentor: izv. prof. dr. sc. Alan Jović

Zadatak: **Web aplikacija za klasifikaciju glazbenih isječaka u glazbena razdoblja**

Opis zadatka:

Glazba se razvijala kroz stoljeća i većinu kompozicija istrenirano uho može smjestiti u pripadajuće glazbeno razdoblje. Pritom se tu u obzir uzima više značajki glazbe, primjerice složenost glazbene linije ili instrumenti koji ju izvode. U ovom diplomskom radu istražiti će se koliko takvih značajki postoji i koje su značajne kako bi model strojnog učenja mogao, slično stručnjaku, prepoznati kojem glazbenom razdoblju određeni glazbeni isječak pripada. Bit će razmatrani i složeniji modeli dubokog učenja, poput konvolucijskih i povratnih modela, jer se takvi modeli često koriste za kompleksnije probleme. Za učenje modela potrebno je prikupiti slobodno dostupne glazbene isječke sa snimaka iz većeg broja razdoblja s weba i ispravno ih označiti. Vrednovanje je potrebno provesti na izdvojenom i označenom testnom skupu podataka. Osim rezultata vrednovanja svih isječaka, rezultat rada bit će i web aplikacija preko koje će biti omogućen unos audio zapisa, njegova obrada te prikaz vjerojatnosti pripadanja zapisa u svako od odabranih razdoblja. Za implementaciju rješenja koristit će se programski jezik Python.

Rok za predaju rada: 28. lipnja 2021.

Zahvaljujem se svom mentoru izv. prof. dr. sc. Alanu Joviću za svu pruženu pomoć i usmjerenje tijekom izrade ovog diplomskog rada.

Sadržaj

Uvod	1
1. Glazba i glazbena razdoblja.....	2
1.1 Elementi glazbe	2
1.2 Glazbena razdoblja	5
2. Analiza glazbenog isječka	7
2.1. Značajke.....	7
2.1.1. Značajke vremenske domene	9
2.1.2. Značajke frekvencijske domene	11
2.1.3. Značajke vremensko-frekvencijske domene	14
3. Pribavljanje i priprema podataka.....	22
4. Algoritmi strojnog učenja za klasifikaciju glazbenih isječaka	25
4.1. Potpuno povezana neuronska mreža.....	27
4.2. Konvolucijska neuronska mreža	29
4.3. Povratna neuronska mreža	32
5. Rezultati klasifikacije	37
5.1. Potpuno povezana neuronska mreža.....	37
5.2. Konvolucijska neuronska mreža	39
5.3. Povratna neuronska mreža	41
5.4. Odabrani model	42
6. Web-aplikacija.....	45
Zaključak	46
Literatura	47

Uvod

Glazba postoji gotovo otkad postoje ljudi i s vremenom su ju počeli razvrstavati na razne načine, primjerice na sakralnu i svjetovnu ili na vokalnu, vokalno-instrumentalnu i instrumentalnu. Kroz stoljeća se glazba razvijala u raznim smjerovima, koji su međusobno često bili kontrastni, i tako se određene periode glazbe u povijesti naziva glazbenim razdobljima. Iskusnom poznavatelju klasične glazbe na temelju odslušane kompozicije brzo je poznato o kojem se glazbenom razdoblju radi i, ovisno o tome koliko poznaje kompozitore toga razdoblja, moguće je da čak ima i ideju o kojem se kompozitoru radi. Ima puno razloga zbog kojih bi se neka kompozicija svrstala u neko razdoblje, primjerice izvođački sastav, kompleksnost glazbenih linija, apstraktnost glazbe, harmonijska podloga i dr. Cilj ovog rada je proanalizirati glazbene kompozicije kroz glazbena razdoblja i pronaći značajke koje bi dobro mogle raspoznati razliku između pojedinih razdoblja. Te značajke će biti iskorištene pri izgradnji više modela dubokog učenja kako bi se pronašao model koji bi što točnije mogao predložiti o kojem se razdoblju radi. U prvom poglavlju opisane su značajke glazbe i razdoblja u koja će biti provedena klasifikacija. U drugom poglavlju istražene su značajke glazbe u vremenskoj, frekvencijskoj i vremensko-frekvencijskoj domeni. Treće poglavlje opisuje način prikupljanja audio zapisa za učenje modela i tehnike koje su provedene nad zapisima kako bi ih se pripremilo za ulaz u neuronsku mrežu. Također je prikazano koji kompozitori su odabrani za pojedina razdoblja i koliko njihovih isječaka se koristi za učenje modela. U četvrtom poglavlju dan je pregled neuronskih mreža i njihovih hiperparametara. Za svaku od odabranih tipova mreže prikazan je tip modela koji će se koristiti tijekom učenja i hiperparametri koji će biti optimirani. Peto poglavlje bavi se analizom rezultata učenja za različite mreže s različitim parametrima i obrazlaže odabir modela koji je korišten u aplikaciji. U šestom poglavlju opisana je izrađena web aplikacija za klasifikaciju glazbenih isječaka u glazbena razdoblja.

1. Glazba i glazbena razdoblja

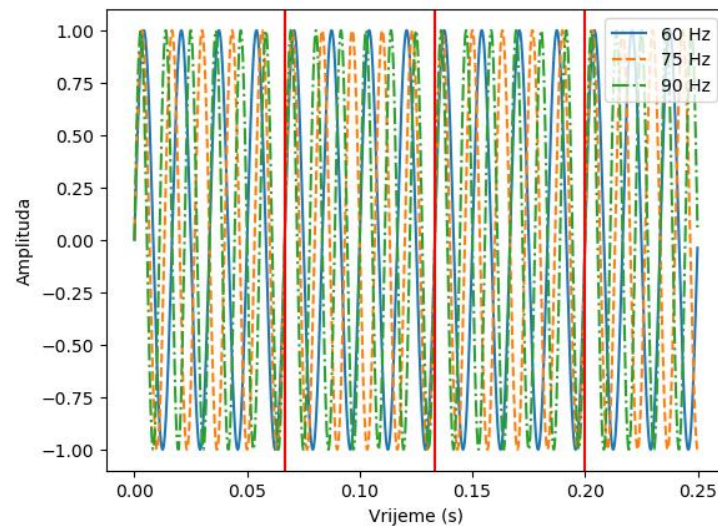
Glazba postoji već tisućama godina i o njenim počecima imamo informacije samo kroz istraživanja arheologa [1]. Neka od prvih mjesta o kojima postoje podaci o glazbi su Babilon, Egipat i Kina [1]. Povijest instrumentalne glazbe moguće je procijeniti prema pronađenim instrumentima, primjerice u egipatskim grobnicama pronađeni su razni puhački instrumenti slični flauti [1]. Najstarija glazba je vokalna glazba, no ona se do početka 10. stoljeća prenosila usmenom predajom pa ne postoje materijalni dokazi o ranijoj vokalnoj glazbi [2].

1.1 Elementi glazbe

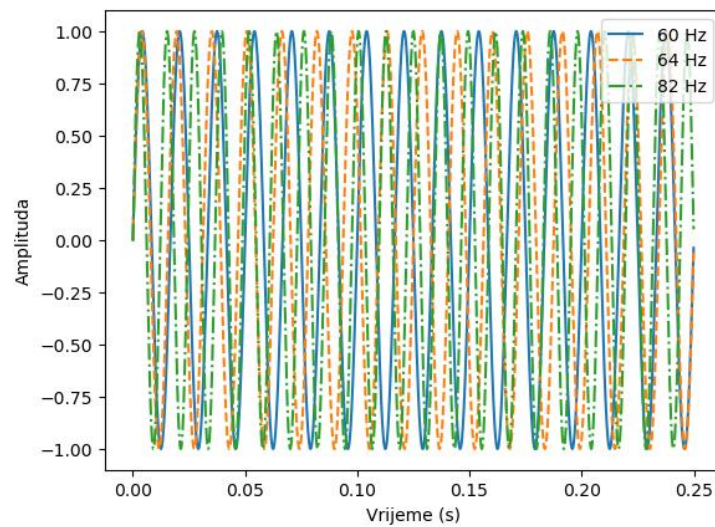
Glazba se u osnovi sastoji od dva elementa, zvuka i ritma, no vrlo je teško opisati što je ona zapravo [3]. Oba prethodno navedena elementa glazbe su neopipljiva i nevidljiva, za razliku od elemenata od kojih se sastoje ostale umjetnosti poput arhitekture, kiparstva i slikarstva. Zvuk je jedan od najosnovnijih i najmisterioznijih fizičkih fenomena [3]. Kada vibracije izvođačkih tijela, primjerice glasa, drva ili metala, pokreću zrak, naše uši omogućavaju nam da te vibracije tumačimo kao zvuk određene frekvencije [3]. Bitno je napomenuti kako zvuk zapravo nije jedna značajka, već se sastoji od dva faktora, frekvencije i dinamike.

Frekvencija, odnosno broj titraja u sekundi, određuje visinu tona. Ton prema kojem se orkestar usklađuje da zvuči sinkrono, odnosno uštima, je a_1 i njegova frekvencija je najčešće 440 Hz [4]. Na temelju tog tona instrument se uštima tako da bi ton za oktavu, ili 12 polutona, viši trebao biti duplo veće frekvencije, odnosno 880 Hz [5]. Na suprotan način dobiva se ton koji je za oktavu niži i njegova frekvencija će biti 220 Hz. Pitagora je za uštima ostalih polutona odredio omjere koji bi davali idealne frekvencije tih tonova, primjerice kvinta, odnosno razmak od 7 polutona, bi imala omjer frekvencija 3:2, a kvarta, odnosno razmak od 5 polutona, omjer 4:3 [5]. Za glazbala ugođena s takvim frekvencijama kaže se da su ugođena u prirodnoj intonaciji [5]. Problem s takvom intonacijom je da se transpozicijom kompozicija, odnosno pomicanjem kompozicije određeni broj polutona više ili niže, gubi njena vjernost i ona zvuči drugačije [5]. Taj problem rješava se jednoliko ugođenom skalom koju preporučuje antički muzički teoretičar Aristoksen [5]. Zbog toga što se uzima da je frekvencija više oktave duplo viša, poluton iznad osnovnog može se izračunati tako da se osnovni množi s $\sqrt[12]{2}$ [5]. Uzastopnim množenjem polutona s tom vrijednosti dobivaju se svi polutoni. Više frekvencija koje

se emitiraju istovremeno, čije valne duljine se često poklapaju u beskonačno puno točaka, zvučat će ugodno i takvo poklapanje se naziva konsonanca [6]. Suprotno tome, disonanca je istodobno zvučanje frekvencija kada se one poklapaju jako rijetko [6]. Na slici 1 crvenim linijama označena su poklapanja sve tri frekvencije jednog od glavnih akorda, durskog kvintakorda. Na slici 2 prikazano je gotovo nepostojeće poklapanje frekvencija disonantnog akorda.



Slika 1 Podudaranje frekvencija konsonantnog akorda



Slika 2 Podudaranje frekvencija disonantnog akorda

Dinamika označava glasnoću izvođenja kompozicije i dvije glavne dinamike su glasno (tal. *forte*) i tiho (tal. *piano*) [7]. Koristeći te dvije osnovne dinamike i oznaku umjerenosti (tal. *mezzo*) sastavljen je cijeli spektar glasnoće, od jedva čujne glazbe do pune glasnoće simfonijskog orkestra [7].

Ritam je uzastopno ponavljanje dvaju ili više različitih, i često kontrastnih, elemenata [8]. Uzmimo za primjer poznati motiv iz 1. stavka Beethovenove 5. simfonije koji se sastoji od 3 kratke note i jedne duge note. Te tri kratke note služe za pripremu ove duge, koja slijedi nakon njih, i daju joj na važnosti. Kada bi sve četiri bile jednake duljine, ova tema vjerojatno bi prošla nezamijećeno u povijesti. Naravno, frekvencija tih tonova je dobrim dijelom zaslužna za to što je motiv svjetski poznat, no iz ovoga se jasno vidi da je ritam jednako bitna značajka neke kompozicije kao i zvuk.

Osim ova 3 glavna elementa glazbe, frekvencije, dinamike i ritma, postoji još cijeli niz sporednih elemenata koji će ukratko biti spomenuti. Tempo označava brzinu kojom se kompozicija izvodi s obzirom na trajanje osnovne metričke jedinice, odnosno dobe [8]. Istu kompoziciju moguće je izvesti polako, umjereno ili brzo i to će utjecati na stil i kvalitetu izvedbe. Artikulacija opisuje na koji način će ton biti izveden, odnosno međusobno povezivanje i odjeljivanje tonova u interpretaciji [8]. Za različite skupine instrumenata postoje specifični tipovi artikulacije, a neki osnovni tipovi su *legato*, odnosno povezano sviranje tonova, i *staccato*, odnosno kraće i odvojeno sviranje tonova [8]. Vibrato je zvučni efekt koji se postiže veoma brzim, višekratnim oscilacijama u visini tona [8]. Koristi se za povećanje izražajnosti tona, no prevelika upotreba smatra se lošom praksom [8]. Na zvuk kompozicije također utječe i boja tona, odnosno vrsta zvuka koji melodiju proizvodi. Istu kompoziciju moguće je izvesti na više instrumenata te će ona na nekima zvučati prikladnije od drugih. Zbog razlika u tipovima instrumenata, primjerice puhačkih i gudačkih, melodija izvedena na instrumentu za koji je namijenjena najbolje će opisati srž, odnosno poruku, koja je u djelu sadržana. Tonalnost glazbe definira koji je osnovni ton prema kojem se odabiru ostali tonovi prema svojim funkcijama unutar tonaliteta [8]. S vremenom se tonalnost počela primjenjivati na manje dijelove kompozicije sve do potpunog gubitka tonalnosti u kasnijim razdobljima [8]. Kontrapunkt je horizontalna značajka glazbe i bavi se proučavanjem različitih melodijskih linija i njihovim kombiniranjem bez gubitka njihovog melodijskog značaja [9]. Ako u kompoziciji postoji jedna glavna melodija, a ostale simultane melodije služe samo kao podloga toj melodiji, riječ je o homofoniji [8]. Suprotno tome, polifonija je simultano izvođenje više samostalnih melodijskih linija [8]. Te samostalne melodijske linije započinju i završavaju u različitim trenucima i potrebno je veliko znanje kako bi se one kvalitetno kombinirale. Harmonija je vertikalna značajka glazbe i definira se kao simultana kombinacija zvukova, odnosno frekvencija [9]. Oslabljivanjem tonalnosti i učestalijim izmjenama tonaliteta dolazi do razvoja novih, apstraktnijih obrazaca uporabe harmonije [9].

1.2 Glazbena razdoblja

Kako se glazba razvijala kroz vrijeme, prethodno navedene značajke glazbe su se mijenjale i periodi u povijesti glazbe, okarakterizirani specifičnim značajkama, nazvani su glazbenim razdobljima. Glazbena razdoblja često se preklapaju zbog velike brojnosti kompozitora i utjecaja raznih kultura, no neke aproksimativne godine su definirane kao granice razdoblja.

Od srednjeg vijeka, cca. od 1150. g. do 1400. g., postoji više očuvanih glazbenih zapisa te će to razdoblje biti prvo glazbeno razdoblje koje će biti promatrano. Gregorijanski korali i jednoglasne pjesme postupno su se razvile u organum, odnosno simultano izvođenje više melodijskih linija, te bi se to moglo smatrati početkom harmonije [10]. Neki od značajnijih kompozitora tog doba su Philippe de Vitry, Guillaume de Machaut i Francesco Landini.

Tijekom renesanse, od cca. 1400. g. do 1600. g., glazba je imala ulogu ostvarivanja ekspresivne komunikacije, jasne projekcije teksta i poticanja emocija slušatelja [12]. Pojavljuje se imitacija, odnosno oponašanje melodije u drugom glasu ili instrumentu, koja služi kao jedna od osnova polifone glazbe [8]. Neki od instrumenata koji su korišteni su viola, lutnja i kornet, a značajniji kompozitori su John Dunstable, Giovanni Pierluigi da Palestrina i Tomás Luis de Victoria [11].

Nakon renesanse slijedi razdoblje baroka, od cca. 1580. g. do 1750. g., koje se smatra razdobljem ekstaze, dinamičkih tenzija i žudnje te je time kontrastno uvjerenosti renesanse [9]. U baroku se razvija pojam *basso continuo* koji označava sustav notacije u kojem bi kompozitor napisao melodiju i liniju basa, no harmonizaciju kompozicije prepustio izvođačima [8]. Ti izvođači nisu bili specificirani, no često su to bili čembalo, orgulje, lutnja ili bas viola [12]. Značajniji kompozitori ovog razdoblja su Johann Sebastian Bach, Antonio Vivaldi i Georg Friedrich Händel.

Klasicizam, 1750. g. - 1820. g., je razdoblje koje obilježava gotovo opsesija oko strukturalne jasnoće kompozicije [10]. Tada je nastao sonatni oblik koji se sastoji od ekspozicije, provedbe i reprize [10]. U ekspoziciji se iznosi muzička ideja, u provedbi se ta tema razrađuje, a u reprizi se ponavlja ekspozicija uz određene promjene [9]. S razvojem tog oblika došlo je do razvoja modernog koncerta, simfonije i sonate [10]. Najznačajniji kompozitori klasicizma su Wolfgang Amadeus Mozart, Joseph Haydn i Ludwig van Beethoven.

U romantizmu, 1830. g. - 1920. g., naglasak se stavlja na emocije i subjektivnost, povezanost umjetnosti te na nacionalni identitet [10]. Dolazi do proširenja harmonijskih obrazaca te opsega i broja instrumenata koji sačinjavaju orkestar [8]. Osim programne glazbe, koja opisuje izvanglazbenu temu i često nudi program koji tu temu objašnjava, razvija se apsolutna glazba koja glazbu smatra apstraktnom i autonomnom [8]. Najznačajniji kompozitori ovog razdoblja su Richard Wagner, Giuseppe Verdi i Hector Berlioz.

2. Analiza glazbenog isječka

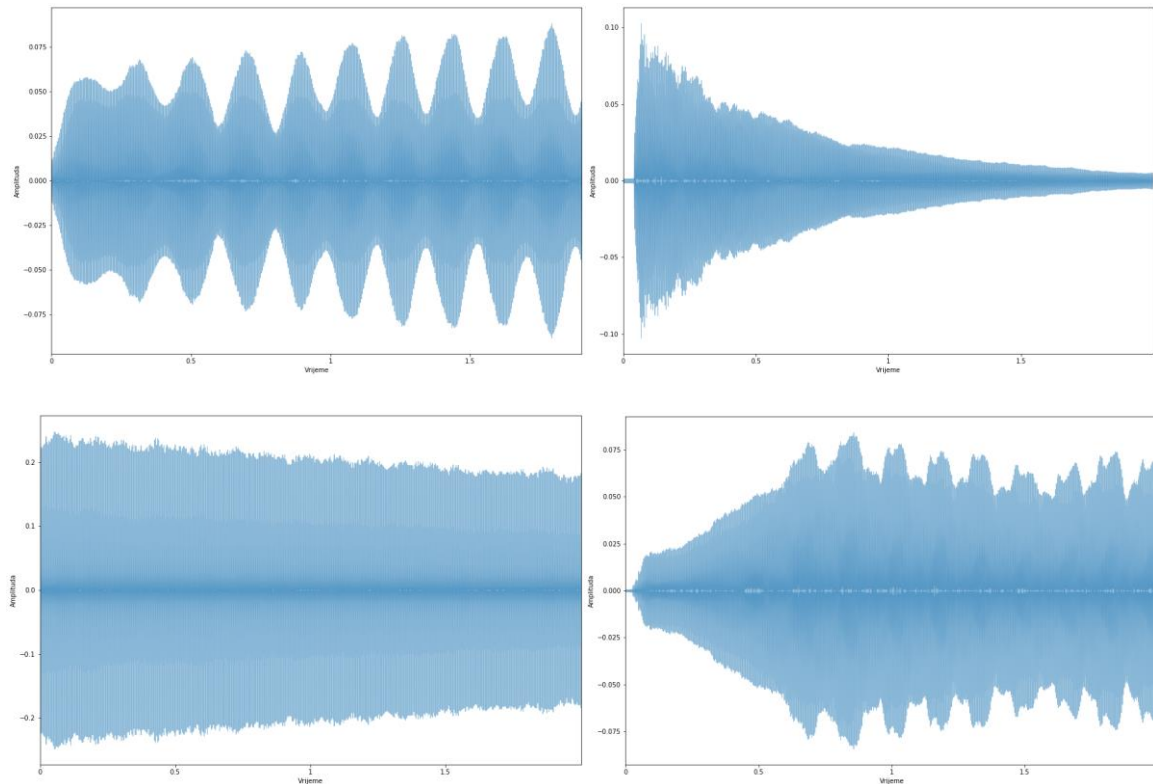
U prethodnom poglavlju opisane su značajke glazbe i glazbenih razdoblja, no dobar dio tih značajki, primjerice ritam, artikulaciju ili polifonost, bit će teško izvući iz kompozicije. Ovaj rad fokusira se na učenje modela strojnog i dubokog učenja i potrebno je pronaći primjerenu analizu glazbenog isječka koja će pružiti najviše bitnih informacija sadržanih u glazbi. Kako bi detaljnija analiza značajki bila napravljena, pripremljeno je nekoliko zvučnih zapisa nad kojima će se provoditi testovi. Sa stranice Sveučilišta Columbia u New Yorku [12] preuzeti su zvučni primjeri izvođenja tona c1, čija je frekvencija otprilike 261 Hz, na flauti, klaviru, trubi i violini. Ti zapisi su usklađeni da su im trajanje i glasnoća jednaki kako to ne bi bili faktori prilikom analize. Također im je uklonjena istosmjerna pristranost kako ne bi bila aktivna frekvencija 0 Hz. Od tih jednostavnih slučajeva napravljena je kombinacija jednog izvođenja tona c1 i to je peti testni slučaj. Posljednji, najkompleksniji, primjer je kratki isječak iz Beethovenove 5. simfonije u c-molu gdje se glavna tema izvodi na gotovo svim instrumentima u orkestru. Zvučni primjeri:

- izvođenje tona c1 na flauti
- izvođenje tona c1 na klaviru
- izvođenje tona c1 na trubi
- izvođenje tona c1 na violini
- zajedničko izvođenje tona c1 na flauti, klaviru, trubi i violini
- isječak iz Beethovenove 5. simfonije

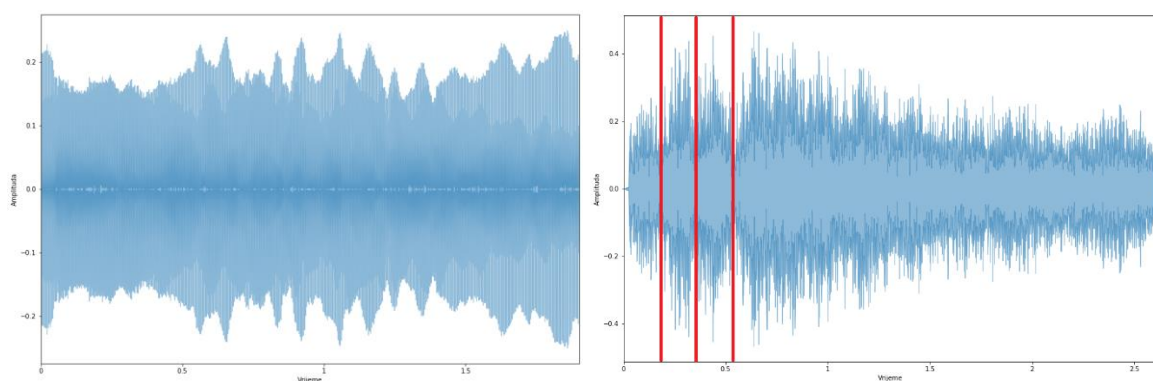
2.1. Značajke

Na početku analize zvuka prva ideja je gledati reprezentaciju zvuka kao graf s vremenom na x-osi i amplitudom vala na y-osi. Na slici 3 prikazani su ti grafovi za 4 jednostavna slučaja analize i jasno je da ovakva reprezentacija neće biti od velike koristi. Na flauti, trubi i violini se kontinuirano može kontrolirati jačina zvuka pa ovakva reprezentacije ne daje konkretne rezultate. Mogu se primijetiti valovi kod flaute i violine zbog malih oscilacija u frekvenciji koje nastaju uporabom vibrata. Na klaviru je drugačija situacija jer se na jačinu odsviranog tona ne može utjecati te se on polako stišava. Također, bitno je napomenuti da bi uz različitu artikulaciju ovi zapisi izgledali drugačije te je to još jedan razlog zbog čega od ovakve analize neće biti konkretnih rezultata. Prethodno iskazani nedostaci samo se potvrđuju na

složenijim primjerima sa slike 4. Kombinacijom istog tona uočena je djelomična sličnost zbog valova koji su bili prisutni kod violine i flaute, no za nepoznate ulaze bilo bi nemoguće dešifrirati o kojim se instrumentima i tonovima radi. U posljednjem primjeru bi se jedino moglo razabrati označene tri kratke note i jednu dugu glavne teme jer je poznata. Kada ta tema ne bi bila poznata, ovakva analiza ne bi bila od koristi.



Slika 3 Prikaz zvučnih valova za izvođenje tona c1 na flauti (gore lijevo), klaviru (gore desno), trubi (dolje lijevo) i violini (gore desno)



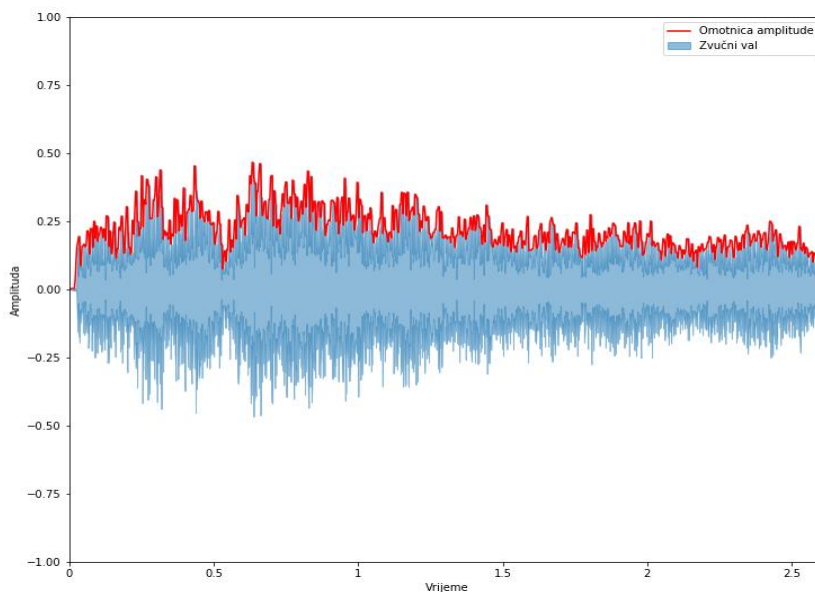
Slika 4 Prikaz zvučnih valova za kombinirano izvođenje tona c1 (lijevo) i za isječak iz Beethovenove 5. simfonije (desno)

Ovi prikazi neće biti korišteni za klasifikaciju, no iz tih zvučnih signala radit će se obrade amplituda kako bi se dobile značajke u vremenskoj i frekvencijskoj domeni.

2.1.1. Značajke vremenske domene

Omotnica amplitude (engl. *Amplitude Envelope, AE*) sadrži maksimalne amplitude svih uzoraka u vremenskim okvirima [14]. U jednom glazbenom isječku ima primjerice 22050 uzoraka u sekundi, a *AE* traži maksimalnu amplitudu unutar vremenskog okvira od K uzoraka. Ovime će se reducirati količina podataka koja je sadržana u signalu samo na maksimalne vrijednosti. Vrijednost *AE* u trenutku t računa se prema izrazu (1) i odgovara maksimalnoj amplitudi signala unutar K uzoraka u rasponu $[t*K, (t+1)*K-1]$, gdje je K proizvoljna veličina vremenskog okvira. Nakon izračuna prve vrijednosti, t se povećava za 1 i tako se računa cijela omotnica amplitude. Za bolje rezultate moguće je povećavati t za neku manju vrijednost, primjerice 0.5, i tako postići da dolazi do djelomičnog preklapanja vremenskih prozora. Na slici 5 prikazana je *AE* za isječak iz Beethovenove 5. simfonije.

$$AE_t = \max_{k = t * K}^{(t + 1) * K - 1} A(k) \quad (1)$$

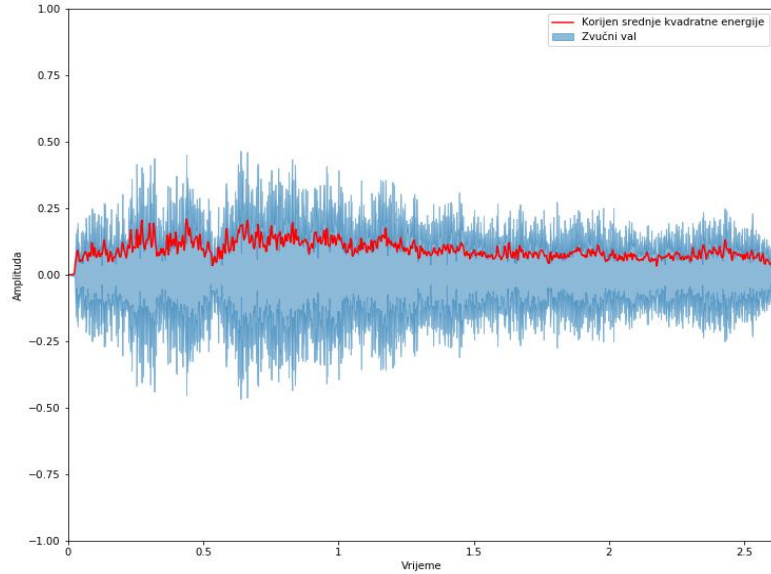


Slika 5 Zvučni val i omotnica amplitude za isječak iz Beethovenove 5. simfonije

Korijen srednje kvadratne energije (engl. *Root Mean Square Energy, RMSE*) također daje informaciju o glasnoći, no manje je osjetljiv na ekstremne vrijednosti od omotnice amplitude. Zbog toga što se ova značajka puno mijenja od jedne glazbene cjeline do druge, dobra je u pronalaženju novih cjelina. Vrijednost *RMSE* u trenutku t računa se izrazom (2) i odgovara korijenu prosjeka zbroja kvadriranih amplituda u rasponu $[t*K, (t+1)*K-1]$, gdje je K proizvoljna veličina vremenskog okvira. Kao i kod *AE*, nakon izračuna prve vrijednosti,

t se povećava za vrijednost u intervalu $[0, 1]$ i tako se računaju sve vrijednosti za signal. Na slici 6 prikazan je $RMSE$ za isječak iz Beethovenove 5. simfonije.

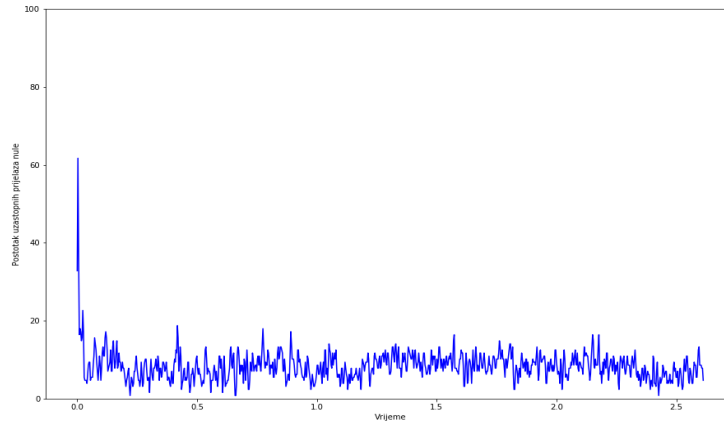
$$RMSE_t = \sqrt{\frac{1}{K} * \sum_{k=t*K}^{(t+1)*K-1} A(k)^2} \quad (2)$$



Slika 6 Zvučni val i korijen srednje kvadratne energije ($RMSE$) za isječak iz Beethovenove 5. simfonije

Učestalost prelaska nule (engl. *Zero Crossing Rate, ZCR*) opisuje koliko puta signal prelazi horizontalnu os. Ova značajka koristi se u razlikovanju govora od glazbe, detekciji govora te klasifikaciji glazbe u žanrove [15]. Vrijednost ZCR u trenutku t računa se izrazom (3) i odgovara koliko prosječno puta u rasponu $[t*K, (t+1)*K-1]$, gdje je K veličina vremenskog okvira, signal prelazi iz pozitivne u negativnu vrijednost, ili obrnuto. Za izračun se koristi funkcija $sgn(x)$, koja za pozitivne vrijednosti vraća 1, a za negativne -1. Nakon izračuna prve vrijednosti, t se povećava za vrijednost u intervalu $[0, 1]$ i dalje se računaju sve vrijednosti za signal. Na slici 7 prikazano je koliko posto vrijednosti u vremenskom prozoru je različitog predznaka od prethodnih vrijednosti za isječak iz Beethovenove 5. simfonije. Vrijednost ZCR je u intervalu $[0,1]$, no prikazana je kao postotak radi veće jasnoće.

$$ZCR_t = \frac{1}{2K} * \sum_{k=t*K}^{(t+1)*K-1} |sgn(A(k)) - sgn(A(k+1))| \quad (3)$$



Slika 7 Postotak uzastopnih prijelaza nule za isječak iz Beethovenove 5. simfonije

2.1.2. Značajke frekvencijske domene

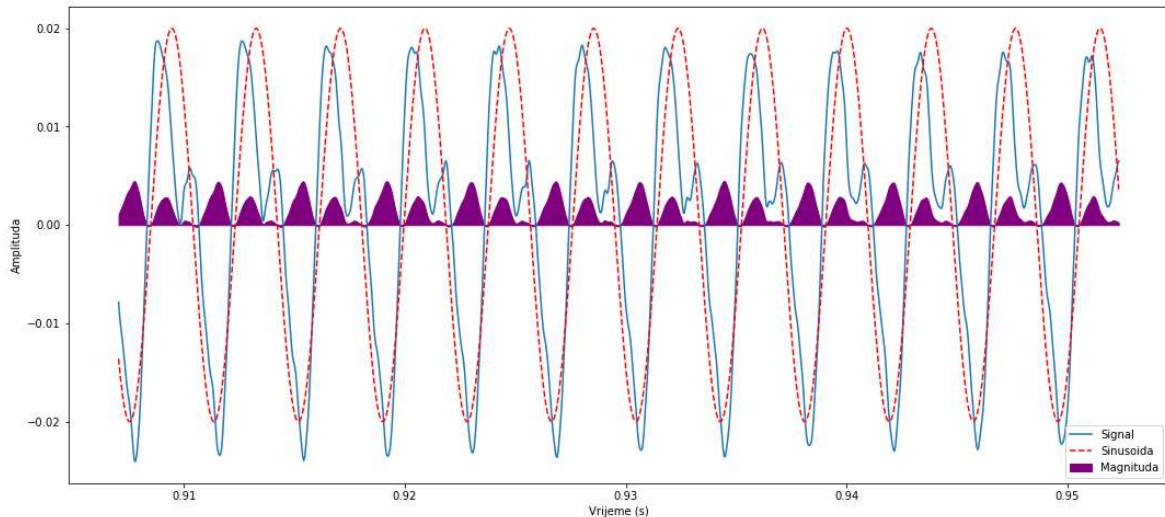
Iz vremenske domene zvučnog zapisa mogu se iščitati informacije poput amplitude i učestalosti prelaska nule, no za dublju analizu potrebno je skupiti informacije o frekvencijama koje su aktivne u tom zvučnom zapisu. Za pretvorbu iz vremenske u frekvencijsku domenu koristi se **Fourierova transformacija**. Vremenski kontinuirana Fourierova transformacija računa se prema izrazu (4).

$$X(2\pi jf) = \int_{-\infty}^{+\infty} x(t) * e^{-2\pi jft} dt \quad (4)$$

Njome se uspoređuju signali različitih frekvencija i faznih pomaka s frekvencijama signala nad kojim se obavlja transformacija i kao rezultat se dobiva spektar frekvencija s označenim stupnjem podudaranja s ulaznim signalom. Izrazom (5) prikazano je računanje stupnja podudaranja kao traženja kuta za koji je površina ispod krivulje, koju se dobiva umnoškom ulaznog signala $s(t)$ i sinusoide koja ovisi o frekvenciji i kutu, najveća.

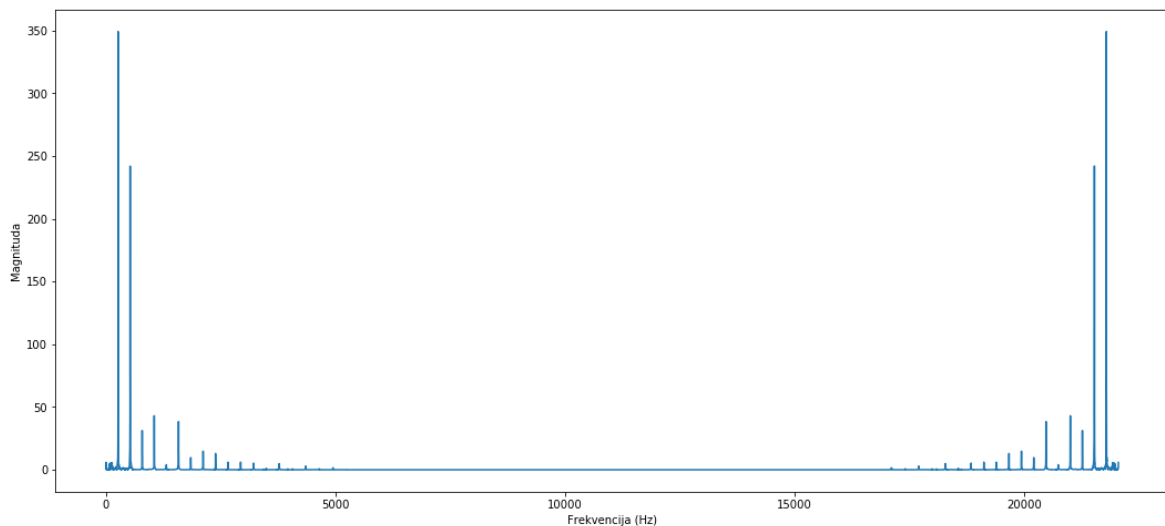
$$d_f = \max_{\varphi \in [0,1>} \left(\int s(t) * \sin(2\pi * (ft - \varphi)) * dt \right) \quad (5)$$

Na slici 8 prikazan je isječak signala i sinusoida s naštimanom amplitudom te faznim pomakom za koje je pronađeno dobro podudaranje. Zbog složenosti signala to podudaranje nikad neće moći biti apsolutno, no to je rezultat istovremenog djelovanja više frekvencija. Magnituda je pomnožena s 10 kako bi prikaz bio ljepši.



Slika 8 Usporedba signala i sinusoide

Taj se postupak ponavlja za sve frekvencije i time se dobiva rezultat Fourierove transformacije, odnosno prikaz svih frekvencija koje djeluju u nekom signalu. Na slici 9 prikazan je rezultat provođenja Fourierove transformacije nad signalom tona c1 na klaviru i vidljivo je da nije aktivna samo jedna frekvencija, već više njih. Ti tiši tonovi, koji se čuju i prate osnovni glasnjiji ton, nazivaju se alikvotni tonovi [8]. Frekvencija njihovog titranja stoji u određenom odnosu prema frekvenciji titranja osnovnog tona. Ako je frekvencija tona n , onda će frekvencije alikvotnih tonova biti $2n$, $3n$ itd. Na slici je vidljivo da su te frekvencije tada 261 Hz, 522 Hz, 783 Hz, itd.

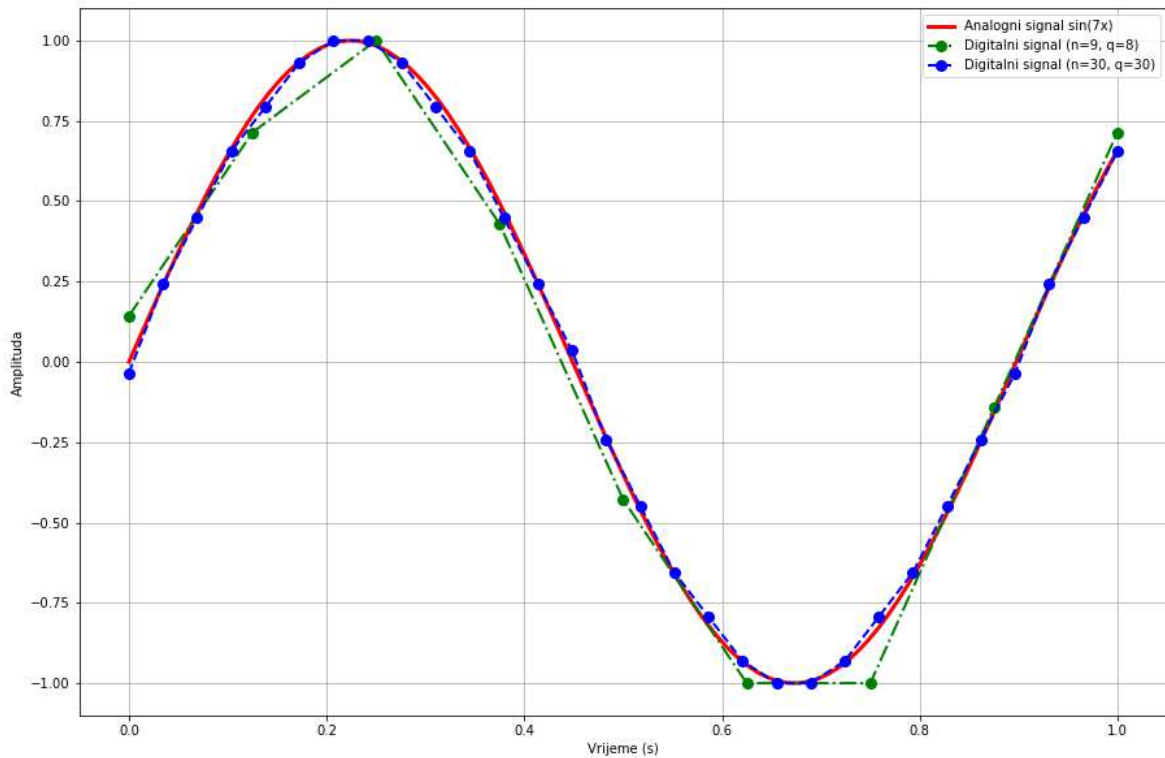


Slika 9 Frekvencije za ton c1 frekvencije 261 Hz na klaviru

U programskom jeziku Python, koristeći paket *SciPy*, provođenje transformacije rezultira poljem kompleksnih brojeva. Kada se ti brojevi zapišu pomoću Eulerove formule, odnosno kao $c = |c| * e^{iy}$, apsolutna vrijednost svakog kompleksnog broja je prikazana magnituda,

a γ opisuje fazni pomak za maksimalnu magnitudu. Na slici je također moguće uočiti centralnu simetriju oko 11025 Hz. Frekvencije iznad te zapravo nisu prisutne i ovdje su dobivene kao rezultat transformacije zbog toga što se prilikom transformacije računa kao da se signal nastavlja u beskonačnost, a da je u transformaciju predan određeni broj perioda. Zbog toga što predani signal nije sinusoida koja održava amplitudu kontinuirano, već zbroj više sinusoida koje slabe s vremenom nakon sviranja tona, uzastopnim povezivanjem tog signala transformacija zaključuje da postoje još neke visoke frekvencije koje omogućuju to spajanje segmenata [19]. Taj fenomen se naziva spektralno curenje [16] i zbog toga se promatraju samo frekvencije do Nyquistove frekvencije, odnosno polovine stope uzorkovanja. Zbog toga što je stopa uzorkovanja ovog isječka 22050Hz, Nyquistova frekvencija iznosi 11025Hz.

Prethodno spomenut izraz (5) za izračun magnitude prikladan je za kontinuirane signale, no kod audiozapisa ti će kontinuirani signali biti diskretizirani. Proces pretvorbe kontinuiranih u diskretne signale naziva se analogno digitalna konverzija i ona se sastoji od uzorkovanja i kvantizacije [17]. Stopa uzorkovanja daje informaciju koliko puta u sekundi se uzima vrijednost iz analognog zapisa i najčešće ima vrijednost 44.1 kHz. Što je ona veća, reprodukcija zapisa bit će bolja, no zapis će biti veći. Kvantizacija opisuje koliko će se bitova koristiti za definiranje spektra amplitude. Primjerice, ako se spektar podijeli na 16 dijelova, tada nam je za definiranje pojedine vrijednosti u tom spektru dovoljno 4 bita. Kao i kod uzorkovanja, veće potpodjele rezultirat će boljom reprodukcijom, no većim zapisom. Na slici 10 crvenom je bojom prikazan signal funkcije $f(x) = \sin(7x)$ i dva primjera provođenja uzorkovanja i kvantizacije. Zeleni točka-crta primjer koristio je stopu uzorkovanja 9 Hz i podjelu amplitudnog spektra na 8 dijelova, a plavi iscrtkani stopu uzorkovanja 30 Hz i podjelu amplitudnog spektra na 30 dijelova. Na slici se jasno vidi da je plavi signal bolje opisao početnu funkciju, no kod digitalnih zapisa uvijek će postojati određena greška, bez obzira na rezoluciju uzorkovanja i kvantizacije, jer se kontinuirano vrijeme nikad neće moći u potpunosti obuhvatiti.



Slika 10 Uzorkovanje i kvantizacija za signal $\sin(7x)$

Fourierova transformacija za digitalne signale je drugačija zbog toga što ovisi o broju uzoraka, a ne o vremenu, i računa se prema izrazu (6).

$$X(e^{2\pi jf}) = \sum_{n=-\infty}^{+\infty} x(n) * e^{-2\pi jfn} \quad (6)$$

Prilikom predstavljanja ove funkcije na računalu pogodno je koristiti njene uzorke, a ne funkciju [18]. Audio zapisi su aperiodski signali konačnog trajanja pa se primjenjuje diskretna Fourierova transformacija opisana izrazom (7) [18].

$$X(k) = \sum_{n=0}^{N-1} x(n) * e^{\frac{-2\pi jnk}{N}} \quad (7)$$

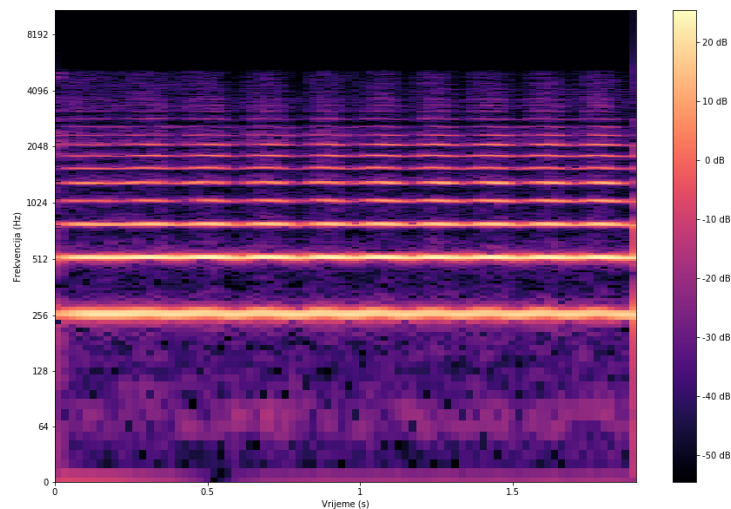
2.1.3. Značajke vremensko-frekvencijske domene

Fourierova transformacija je korisna zbog toga što daje globalni prikaz frekvencija u nekom glazbenom isječku, no bilo bi korisno znati koje su frekvencije aktivne u kojem vremenu kroz taj isječak. Izrazom (8) opisana je kratkotrajna diskretna Fourierova transformacija, gdje je m indeks vremenskog prozora, k indeks dijela frekvencije, N veličina vremenskog prozora, H pomak vremenskog prozora i $w(n)$ funkcija koja daje vrijednost u vremenskom prozoru. Jednostavan primjer funkcije $w(n)$ je pravokutna funkcija koja bi svim

vrijednostima dala vrijednost 1, no često se koriste Hannove ili Hammingove funkcije koje imaju oblik zvona kako bi se početak i kraj signala u prozoru izjednačio.

$$X(m, k) = \sum_{n=0}^{N-1} x(n + mH) * w(n) * e^{\frac{-2\pi jnk}{N}} \quad (8)$$

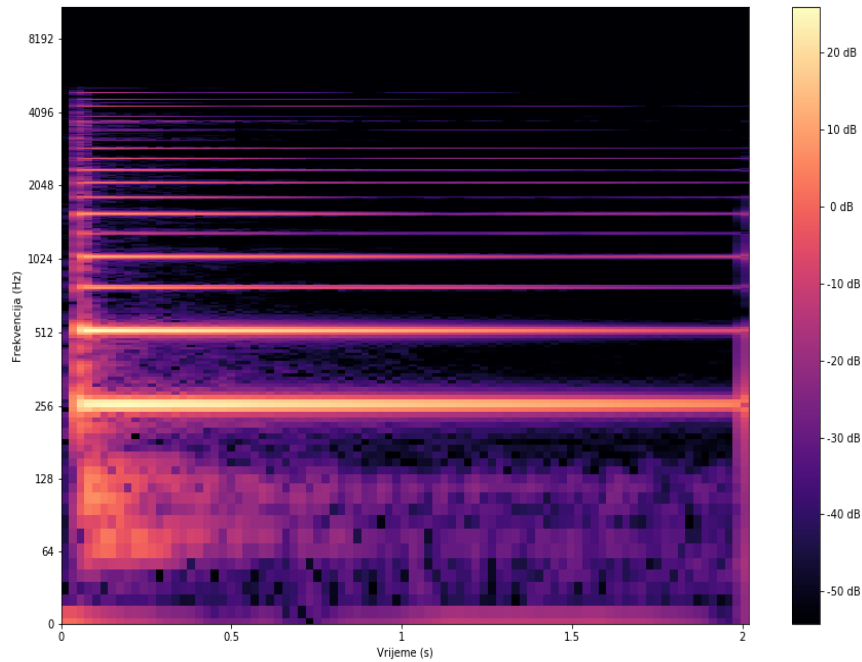
Provedbom algoritma kratkotrajne Fourierove transformacije dobiva se rješenje oblika (k , broj_okvira). Svaki k predstavlja frekvencijski uzorak, koji se kasnije množenjem s frekvencijom uzorkovanja pretvara u frekvenciju, a broj okvira je ukupan broj uzoraka u zapisu podijeljen s udaljenosti između vremenskih okvira. Na slici 11 prikazan je spektrogram koji je rezultat kratkotrajne diskretne Fourierove transformacije tona c1 odsviranog na flauti. Ovdje se, kao i kod diskretne Fourierove transformacije, vidi da postoje alikvotni tonovi, no dodatak je to što se vidi kada su aktivni i kojim intenzitetom. Kod viših frekvencija se mogu uočiti valovi koji nastaju zbog upotrebe vibrata prilikom izvođenja tona. Intenzitet je prikazan u decibelima, mjernoj jedinici glasnoće, a računa se koristeći prethodno spomenutu magnitudu kompleksnog broja koji je rezultat transformacije za pojedini vremenski period i frekvenciju.



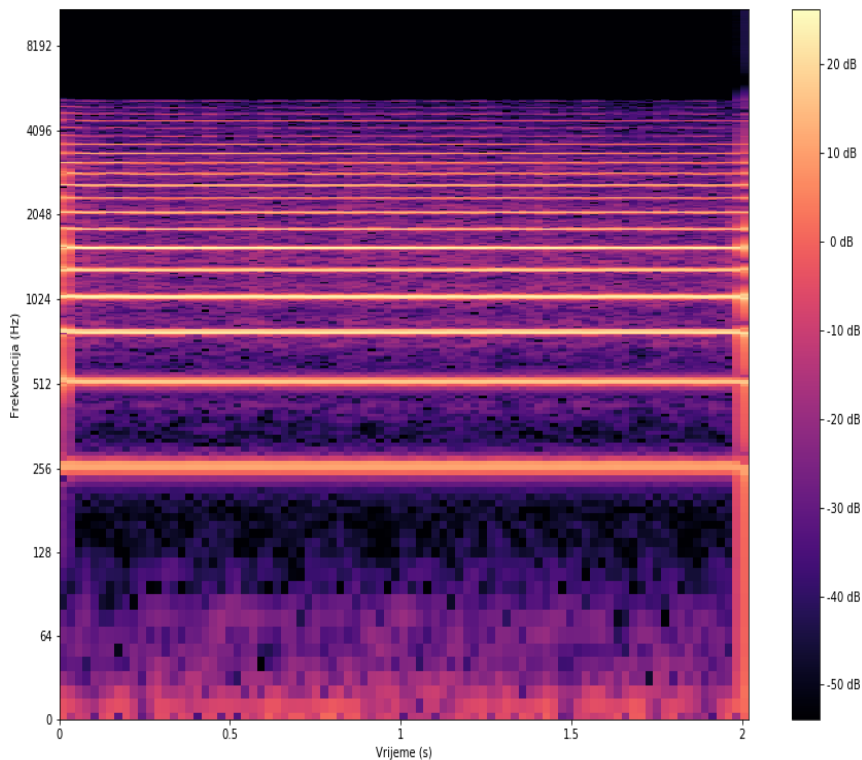
Slika 11 Spektrogram tona c1 na flauti

Na slici 12 prikazan je spektrogram istog tona na klaviru i vidljivo je da su aktivne iste frekvencije kao i kod provođenja diskretne Fourierove transformacije, prikazane na slici 9. Pri višim frekvencijama se vidi da ton slabi s vremenom jer se na njega, nakon što je jednom odsviran, ne može utjecati. Na slici 13 prikazan je spektrogram istog tona na trubi i ovdje se može uočiti kako su puno više izražene više frekvencije, nego kod ostalih instrumenata. Slika 14 prikazuje spektrogram istog tona na violini i ovdje se također mogu primijetiti valovi u višim frekvencijama uzorkovani upotrebom vibrata. Kod spektrograma kombiniranog tona

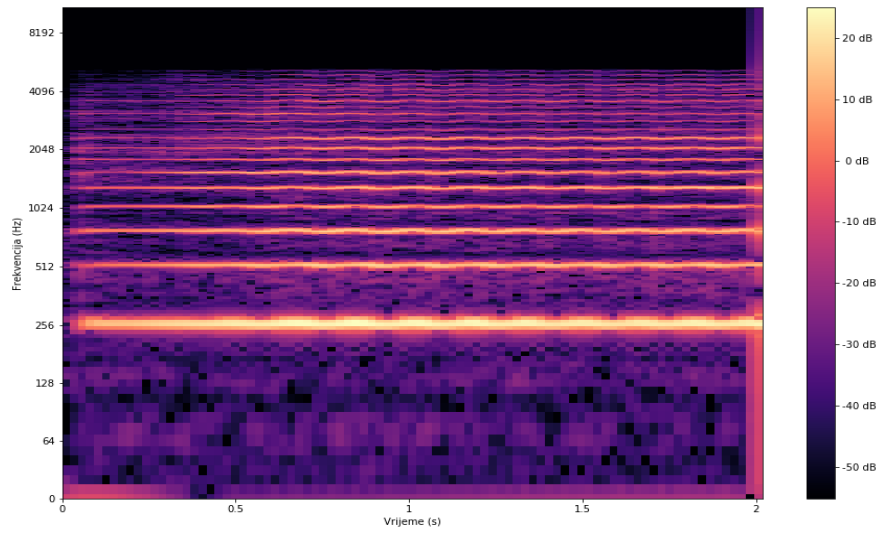
c1, na slici 15, vidljivo je da je frekvencijski spektar najaktivniji, upravo zbog toga što sadrži frekvencije svih svojih komponenti. Na posljednjem primjeru, na slici 16, prikazan je spektrogram isječka iz Beethovenove 5. simfonije. Vidljivo je da je ovdje prisutno više tonova koji stvaraju harmonije i zbog toga se alikvotni tonovi protežu iznad 8000 Hz. Iz viših frekvencija moguće je razlučiti da se u isječku nalaze 3 kratke note i jedna duga.



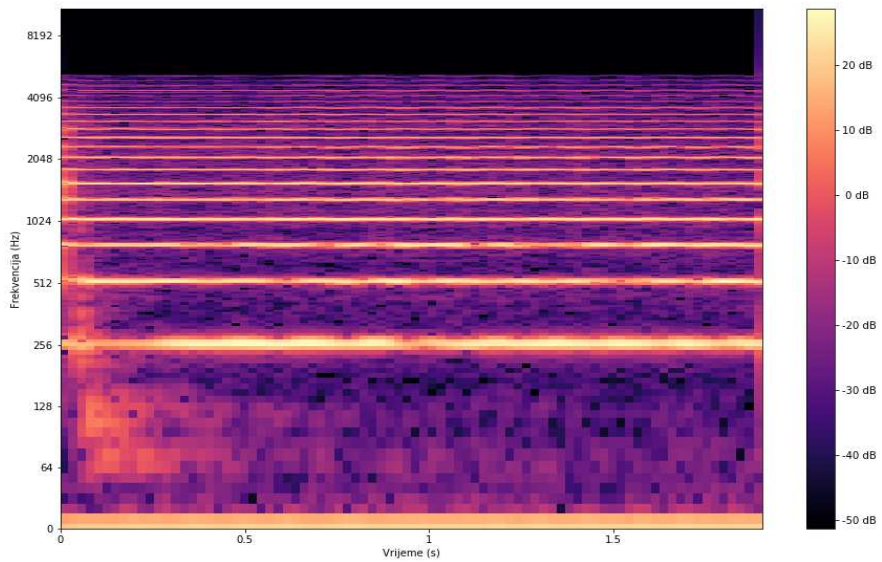
Slika 12 Spektrogram tona c1 na klaviru



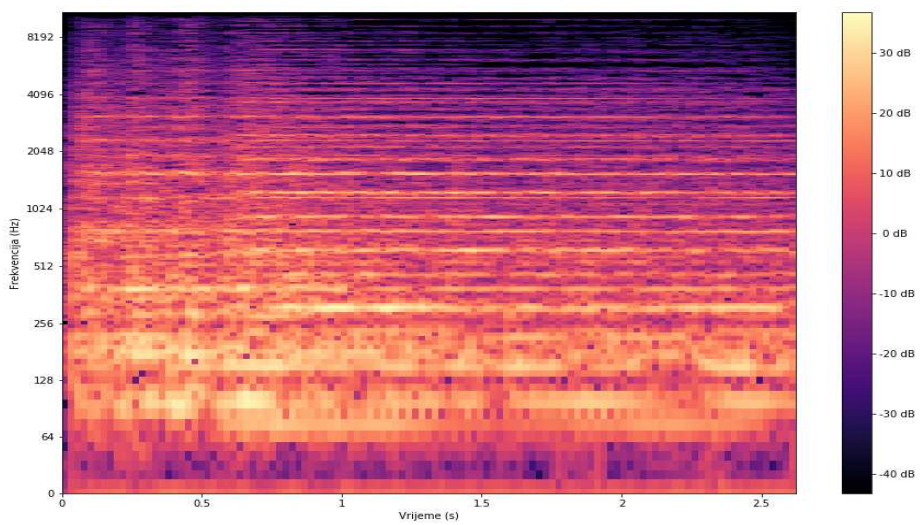
Slika 13 Spektrogram tona c1 na trubi



Slika 14 Spektrogram tona c1 na violini

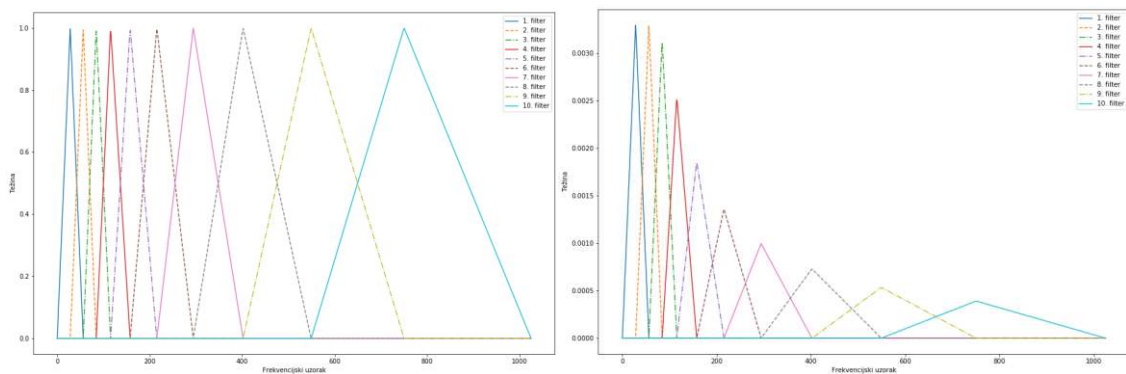


Slika 15 Spektrogram kombiniranog tona c1



Slika 16 Spektrogram isječka iz Beethovenove 5. simfonije

Spektrogram daje dobar pregled aktivnih frekvencija u glazbenom isječku, no postoji još par elemenata koje bi bilo dobro prilagoditi prije upotrebe spektrograma. Prvi problem je da ljudi bolje čuju razliku u nižim, nego u višim frekvencijama i to bi nekako trebalo predstaviti u spektrogramu. Drugi problem je to što ima jako puno frekvencijskih poddjela pa bi bilo korisno da se taj broj umanji, a svejedno očuvaju korisne informacije. Ti problemi mogu se riješiti koristeći Mel-spektrogram kojim se frekvencije dijele u n frekvencijskih skupina. Podjela u frekvencijske skupine se radi tako da se odabere broj skupina i za svaku skupinu se izgenerira filter. Filter radi tako da obuhvaća više frekvencijskih uzoraka, no svakome pridaje odgovarajuću težinu. Svaki filter započinje od sredine prošlog i završava u sredini sljedećeg filtra te tako dolazi do djelomičnog preklapanja kako se ne bi krivo protumačili podaci. Na slici 17 prikazana je podjela frekvencijskih uzoraka na 10 filtara, gdje je prvi filter najljeviji, drugi odmah desno od njega, itd. Prvi filter obuhvaća frekvencijske uzorke od nultog do cca. 55. s maksimumom oko 27. uzorka. Zbog toga što naše uši lošije diferenciraju više tonove, viši filtri su širi te obuhvaćaju veći spektar frekvencija. Na desnom dijelu slike 17 vide se ti filtri kada su normalizirani površinom koju obuhvaćaju te se oni u praksi ponekad ovako primjenjuju na rezultat kratkotrajne Fourierove transformacije.



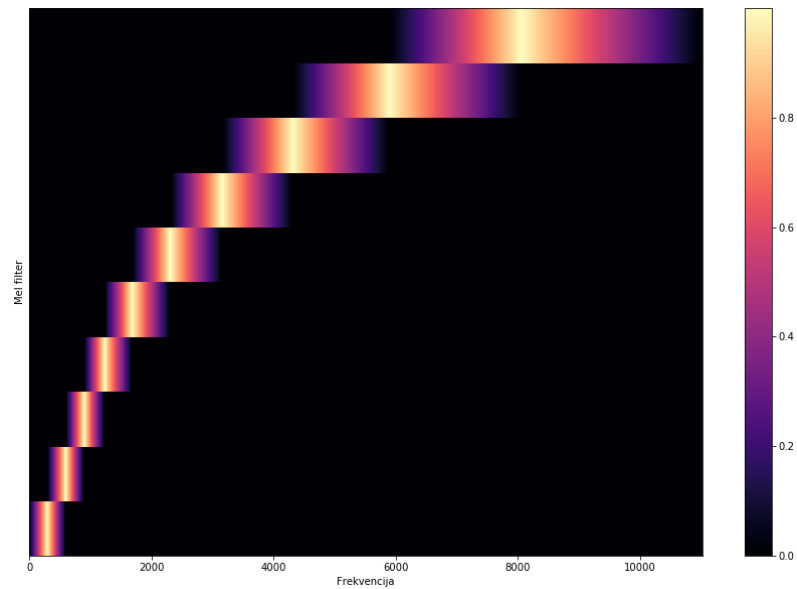
Slika 17 Mel-filtri izvornih vrijednosti (lijevo) i normalizirani površinom (desno)

Osim skaliranja amplituda frekvencija, frekvencije se prebacuju u mel-frekvencije gdje su bolje raspoređene. Povezanost između frekvencija i mel-frekvencija je empirijski utvrđena i zato ne postoji jedinstvena formula. Izrazi (9) i (10) opisuju jednu moguću pretvorbu između dvije frekvencijske domene [20].

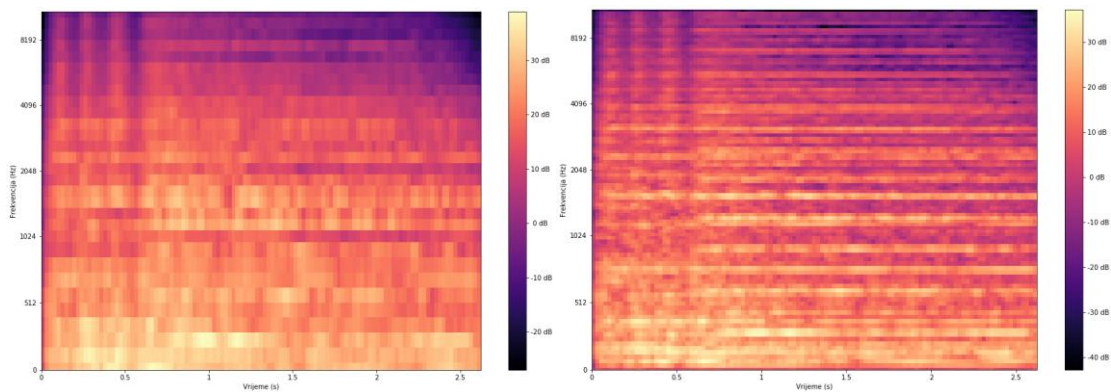
$$f_m = 2595 * \log_{10} \left(1 + \frac{f}{100} \right) \quad (9)$$

$$f = 100 * \left(10^{\frac{f_m}{2595}} - 1 \right) \quad (10)$$

Na slici 18 prikazan je odnos izvornih mel-filtara i frekvencija. Ovdje je radi jednostavnosti korišteno 10 filtara, a u praksi se koristi veći broj koji se zadaje kao hiperparametar kod učenja modela. Na slici 19 prikazani su mel-spektrogrami isječka iz Beethovenove 5. simfonije za 30 i 100 filtara. Veći broj filtara rezultirat će boljim prikazom, no veći broj redundantnih informacija će također biti pohranjen u takvom prikazu.



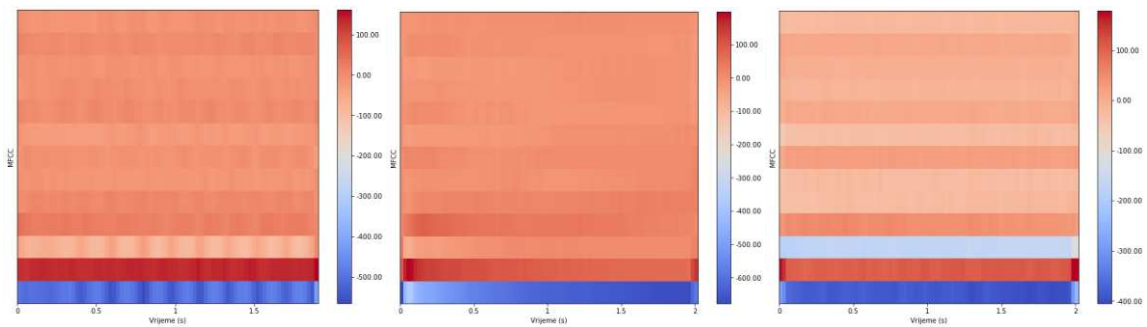
Slika 18 Odnos mel-filtara i frekvencija



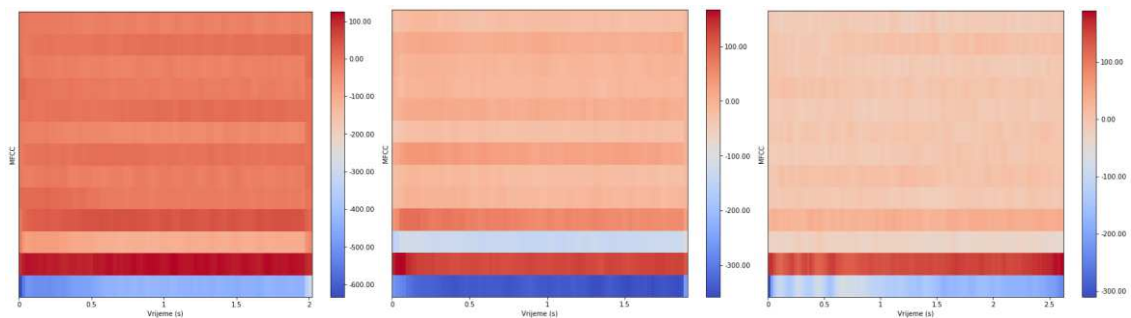
Slika 19 Mel-spektrogram isječka iz Beethovenove simfonije s 30 filtara (lijevo) i 100 filtara (desno)

Sljedeći pristup koji će se koristiti je generiranje kepralnih koeficijenata mel-frekvencije (engl. *Mel-frequency Cepstral Coefficients, MFCC*). Nad vremenskim okvirom se prvo provodi diskretna Fourierova transformacija, potom se radi logaritam amplitude, frekvencije se skaliraju koristeći mel-transformacije i na kraju se primjenjuje diskretna kosinusna transformacija. Umjesto Fourierove transformacije na kraju se koristi diskretna kosinusna transformacija zbog toga što je jednostavnija, rezultat su realni brojevi i umanjuje se dimenzionalnost podataka [21]. Također dolazi do uklanjanja korelacije u mel-skupinama koja nije pogodna za učenje modela. Rezultat provedbe ovih operacija, koristeći paket

Librosa u jeziku *Python*, rezultira s 39 koeficijenata po vremenskom okviru, no nije ih potrebno koristiti sve. Kasniji koeficijenti pružaju sve manje informacija pa se u praksi koristi prvih 12 ili 13 koeficijenata [20]. Dodatni koeficijenti mogu se generirati tako da se u svakom vremenskom okviru dodaju lokalne aproksimacije prve i druge derivacije. Na slici 20 prikazano je 13 MFCC za ton c1 na flauti, klaviru i trubi, a na slici 21 za violinu, kombinirani ton te isječak iz Beethovenove 5. simfonije. Ovdje se može primijetiti vibrato kod flaute i violine i obrisi teme kod Beethovenove simfonije te zaključiti kako različiti instrumenti drugačije aktiviraju ove koeficijente.

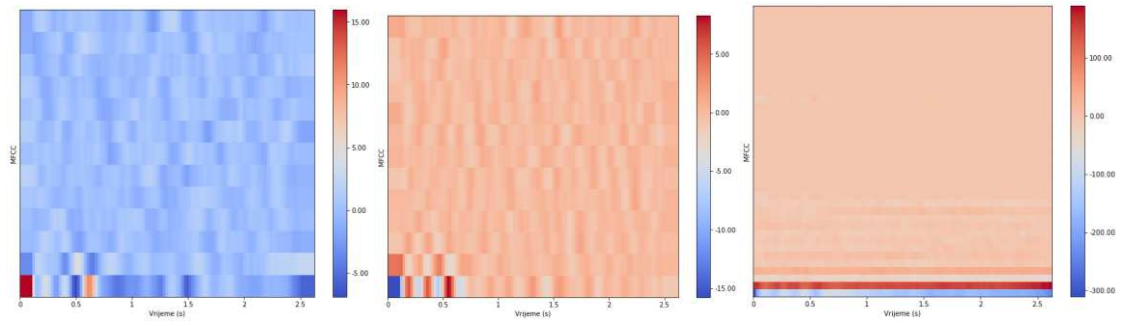


Slika 20 13 MFCC za ton c1 na flauti (lijevo), klaviru (sredina) i trubi (desno)



Slika 21 13 MFCC za ton c1 na violini (lijevo), kombinaciji instrumenata (sredina) i za isječak iz Beethovenove 5. simfonije (desno)

Zbog toga što se u različitim glazbenim razdobljima koriste različiti instrumenti i izvođački sastavi, pretpostavka je da bi ovakve značajke bile korisne. Na slici 22 prikazane su aproksimacija prve i druge derivacije izračunate koristeći paket *Librosa*. Bez obzira na kratki isječak u kojem je iznesena tema, postoje razlike u koeficijentima u vremenskim okvirima. Na desnoj slici su spojene te dvije derivacije s osnovnim koeficijentima u jedan vektor. Viši redovi vektora imaju vrlo male vrijednosti u usporedbi s prvih par redova, no to ne bi trebalo predstavljati problem kod učenja modela zbog različitih težina pojedine vrijednosti vektora.



Slika 22 Aproximacija prve derivacije (lijevo) i druge derivacije (sredina) MFCC te spojeni vektor MFCC (desno)

3. Pribavljanje i priprema podataka

Za uspješnu klasifikaciju ovog visokodimenzionalnog problema bilo je potrebno koristiti veliki broj glazbenih primjera. Zbog toga što nije pronađena pripremljena kolekcija primjera, oni su ručno prikupljeni pretraživanjem po usluzi *YouTube* i preuzimanjem tih zapisa koristeći uslugu *YouTube To Mp3* [22]. Kako bi glazbeno razdoblje bilo što bolje predstavljeno, pronađen je velik broj kompozitora čija djela su preuzeta. Određeni kompozitori skladali su u više razdoblja i zato neka djela imaju obilježja više razdoblja. Prilikom odabira preferirane su kraće kompozicije i težilo se izvedbama na što izvornijim instrumentima te da što veći broj različitih instrumenata bude prisutan u zapisima. Sve ove metode selekcije provedene su uz ograničeno amatersko poznavanje glazbe i zapisi izvedeni na “pogrešnim” instrumentima ili izvedeni u “krivom” stilu mogli bi narušiti postotak ispravne klasifikacije primjera. Također, preuzeti zapisi su zapisi slabe kvalitete i to će također reducirati mogućnosti kvalitetne glazbene analize zbog toga što će značajke biti lošije.

Za pripremu podataka korišten je programski jezik *Python* i paketi *Pydub*, *NumPy* i *Librosa*. Prvi korak pripreme prikupljenih podataka bio je ukloniti tišinu s početka i kraja zapisa. Kao prag za označavanje tišine odabrana je vrijednost -50 dB. Zapisi su pretvoreni u *wav* format kako bi im se pridružila informacija u stopi uzorkovanja te se s njima moglo dalje raditi. Kod nekih zapisa je uočena aktivnost oko frekvencije 0 Hz do koje dolazi zbog pristranosti istosmjerne komponente te je ta pristranost zato uklonjena. Za učenje modela bilo je potrebno da niz značajki bude iste duljine pa su zato sve kompozicije podijeljene u isječke od 10 s. Iz svakog od 5 promatranih razdoblja prikupljeno je 4500 isječaka. U tablici 1 prikazani su brojevi kompozicija i glazbenih isječaka pojedinih kompozitora u razdoblju koji su korišteni za učenje modela.

razdoblje	broj kompozicija i glazbenih isječaka po kompozitoru
srednji vijek	Adam de la Halle (4, 92), Adam de St Victor (3, 68), Anonymous (19, 585), Antonello da Caserta (4, 145), Baude Cordier (4, 63), Colin Muset (4, 92), Comtessa Beatritz de Dia (3, 111), Francesco Landini (5, 100), Gautier de Coincy (3, 88), Gherardello da Firenze (2, 28), Giovanni da Cascia (3, 63), Guido D'Arezzo (1, 6), Guillaume de Machaut (9, 421), Guiraut Riquier (4, 131), Hildegard Von Bingen (3, 92), Jacopo da Bologna (3, 63), Jaufré Rudel (3, 75), Johannes Ciconia (15, 312), John Dunstable (7, 171), Leonin (4, 356), Lorenzo da Firenze (3, 71), Matteo

	da Perugia (3, 100), Moniot d'Arras (1, 24), Neidhart von Reuental (1, 29), Niccolò da Perugia (3, 38), Notker Balbulus (3, 39), Perotin (4, 202), Peter Abelard (2, 83), Petrus de Cruce (2, 25), Philip the Chancellor (1, 38), Philippe de Vitry (7, 153), Richard the Lionhearted (1, 56), Tuotilo (1, 19), Walther von der Vogelweide (15, 376), Étienne de Liège (4, 185)
renesansa	Adrian Willaert (3, 76), Alexander Agricola (3, 159), Alonso Mudarra (3, 45), Andrea Gabrieli (3, 46), Antoine Brumel (3, 69), Antoine Busnois (3, 106), Antonius Divitis (3, 61), Carlo Gesualdo (3, 79), Claude Le Jeune (3, 84), Claudio Monteverdi (5, 135), Clément Janequin (3, 96), Francesco da Milano (3, 47), Gilles Binchois (3, 89), Giovanni Gabrieli (3, 73), Giovanni Pierluigi da Palestrina (10, 1534), Guillaume Dufay (4, 132), Heinrich Isaac (3, 132), Hugh Aston (2, 35), Jacob Obrecht (3, 90), Jacobus Clemens non Papa (3, 97), Jacques Arcadelt (3, 40), Jean Mouton (3, 94), Johann Stamitz (4, 237), Johannes Ockeghem (3, 72), John Dowland (3, 77), John Taverner (3, 116), Josquin des Prez (3, 70), Loyset Compère (3, 115), Luca Marenzio (3, 67), Luis de Narváez (3, 40), Maddalena Casulana (3, 29), Orlando di Lasso (3, 41), Oswald von Wolkenstein (3, 117), Philippe Verdelot (3, 61), Pierre de La Rue (3, 38), Thomas Morley (3, 32), Thomas Tallis (3, 78), Tomás Luis de Victoria (3, 52), William Byrd (3, 75)
barok	Alessandro Scarlatti (3, 78), Alessandro Stradella (3, 88), André Campra (3, 88), Antonio Lotti (3, 112), Antonio Vivaldi (3, 169), Arcangelo Corelli (3, 116), Barbara Strozzi (3, 132), Benedetto Marcello (3, 153), Christoph Graupner (3, 120), Dietrich Buxtehude (3, 130), Domenico Scarlatti (3, 87), Francesco Cavalli (3, 79), Francesco Gasparini (3, 89), Francesco Geminiani (3, 177), François Couperin (3, 99), Georg Philipp Telemann (3, 142), George Frideric Händel (3, 67), Giacomo Carissimi (2, 61), Giovanni Battista Pergolesi (3, 181), Giovanni Legrenzi (3, 66), Giuseppe Tartini (3, 142), Heinrich Ignaz Franz Biber (4, 241), Heinrich Schütz (3, 75), Henry Purcell (3, 71), Jan Dismas Zelenka (3, 142), Jean Marie Leclair (1, 12), Jean-Baptiste Lully (3, 104), Jean-Marie Leclair (2, 128), Jean-Philippe Rameau (3, 72), Johann Hermann Schein (3, 88), Johann Jakob Froberger (3, 74), Johann Joachim Quantz (2, 79), Johann Kuhnau (3, 116), Johann Mattheson (2, 65), Johann Pachelbel (3, 119), Johann Sebastian Bach (3, 105), Luigi Rossi (1, 53), Marc-Antoine Charpentier (3, 120), Pietro Locatelli (3, 148), Samuel Scheidt (3, 97), Tomaso Albinoni (2, 104), Wilhelm Friedemann Bach (1, 50)
klasicizam	Adalbert Gyrowetz (3, 124), Anton Fils (3, 102), Antonio Salieri (3, 142), Carl Ditters von Dittersdorf (3, 99), Carl Friedrich Abel (3, 96), Carl Philipp Stamitz (3, 138), Christian Cannabich (3, 145), Christoph Willibald Gluck (3, 101), Franz Ignaz Beck (3, 77), Franz Joseph Haydn (7, 633), Franz Xaver Richter (2, 49), Georg Anton Benda (2, 110), Giovanni Paisiello (3, 82), Giuseppe Sarti (3, 98), Ignaz Fränzl (1, 48), Johann Baptist Wanhal (3, 85), Johann Christian Bach (4, 192), Josef Mysliveček (3, 125), Joseph Bologne (2, 80), Leopold Koželuch (2, 97), Leopold Mozart (3, 31), Ludwig van Beethoven (4, 241), Luigi Boccherini (4, 314), Muzio Clementi (2, 102), Niccolò Antonio Zingarelli

	(3, 129), Niccolò Piccinni (3, 69), Tommaso Traetta (3, 92), Vicente Martín y Soler (3, 43), Wolfgang Amadeus Mozart (10, 782)
romantizam	Anton Bruckner (4, 78), Antonín Dvořák (5, 798), Bedřich Smetana (3, 117), Camille Saint-Saëns (3, 216), Carl Czerny (1, 106), Carl Maria von Weber (3, 136), Charles Gounod (5, 195), Charles-Marie Widor (2, 75), César Franck (3, 92), Edvard Grieg (3, 132), Felix Mendelssohn (3, 145), Franz Liszt (4, 145), Franz Schubert (3, 77), Franz Xaver Gruber (1, 27), Frédéric Chopin (3, 96), Gabriel Fauré (3, 90), Gaetano Donizetti (3, 84), Gioachino Rossini (3, 104), Giuseppe Verdi (4, 70), Gustav Mahler (3, 163), Hector Berlioz (3, 136), Hugo Wolf (3, 99), Ivan Zajc (3, 57), Jacques Offenbach (4, 104), Johann Strauss I (3, 99), Johannes Brahms (3, 89), Josef Rheinberger (3, 151), Josip Runjanin (1, 11), Nikolai Rimsky-Korsakov (4, 102), Pyotr Ilyich Tchaikovsky (3, 118), Richard Wagner (6, 333), Robert Schumann (3, 105), Vatroslav Lisinski (4, 150)

Tablica 3 Prikaz broja isječaka za svakog kompozitora po razdoblju

Za svaki od isječaka izračunato je 13 koeficijenata $MFCC$ i zbog toga svaki ulazni primjer mreže ima oblik (431, 13). Prva vrijednost opisuje nad koliko su vremenskih okvira izračunati koeficijenti $MFCC$, a računa se kao $n = duljina_zapisa * stopa_uzrokovanja / udaljenost_između_vremenskih_okvira$, gdje je stopa uzorkovanja 22050 Hz, a udaljenost između vremenskih okvira 512 uzoraka.

4. Algoritmi strojnog učenja za klasifikaciju glazbenih isječaka

Strojno učenje prikladno je za rješavanje zadatka za koji nije jednostavno definirati rješenje, već bi ga trebalo iščitati iz pripremljenih podataka. Veći broj podataka dat će bolju reprezentaciju mogućih primjera u klasi i time se povećava izglednost da će model bolje klasificirati neviđene primjere. Cilj strojnog učenja je dobiti model sa što manjom pogreškom, odnosno vrijednosti funkcije gubitka na skupu za testiranje, a to se postiže pravilnim učenjem modela. U svakom koraku učenja, odnosno nakon svake grupe primjera, težine modela se ažuriraju zbrajanjem s negativnom vrijednosti gradijenta gubitka s obzirom na parametre mreže [24]. Veličina grupe može biti cijeli skup primjera, manje grupe primjera ili pojedinačni primjeri. Korištenje cijelog skupa primjera za veličinu grupe rezultirati će najtočnijim gradijentom, no ta tehnika zahtijeva veće memorijske resurse i neće se koristiti zbog velikog broja primjera. Pojedinačni primjeri bi rezultirali velikim oscilacijama iz koraka u korak te je zbog toga odabrano učenje nad manjih grupama podataka. U biblioteci *Keras* uobičajena je vrijednost 32 primjera u grupi, no za učenje modela će grupe biti veličine 128 jer je ispitivanjem zaključeno da su rezultati jednako dobri, a prednost većeg broja primjera u grupi je da će ubrzati učenje modela. Za izračun pogreške odabran je algoritam unakrsne entropije koji za potpuno točnu klasifikaciju svih primjera daje vrijednost 0, a u najgorem slučaju $+\infty$ [24]. Tijekom učenja moguće je koristiti različite algoritme optimizacije, a za ovaj rad odabran je optimizator Adam kod kojeg se prati uprosječeno kretanje gradijenta i kvadrata gradijenta uz eksponencijalno zaboravljanje [23].

Prilikom učenja modela treba paziti na niz hiperparametara, odnosno vrijednosti koje se testiraju prije odabira najboljeg modela, koji će odrediti hoće li model predviđati bolje ili lošije. Prekompleksan model, odnosno model s previše neurona, moguće je prenaučiti na primjerima za učenje te će on loše predviđati, a prejednostavan model moguće je podnaučiti te on neće moći uhvatiti srž problema i zbog toga će i on loše predviđati [24]. Optimizator prema zadanoj stopi učenja određuje koliko će se parametri ažurirati u svakom koraku. U praksi su to vrijednosti manje od 0.1 jer manji pomaci daju bolje rezultate. Velike težine nisu poželjne u mreži jer mala razlika u ulaznim podacima daje velike promjene na izlazu te se zbog toga te težine dodaju funkciji gubitka modela [24]. Taj postupak naziva se regularizacija i za učenje modela u ovom radu koristi se L2 regularizacija, odnosno gubitku

modela se dodaju kvadratne vrijednosti težina [24]. Time se određeni broj težina smanjuje do 0 i dobivaju se modeli koji su manje skloni prenaučnosti, računalno manje zahtjevni i interpretabilni jer je vidljivo koje težine utječu na klasifikaciju [24]. Ta vrijednost regularizacije mogla bi biti poprilično velika i zato je ona pomnožena regularizacijskim faktorom koji ima malu vrijednost. Normalizacija grupe (engl. *batch normalization*) je metoda za ubrzavanje učenja tako što se standardizacija podataka integrira u arhitekturu mreže [25]. Ima razne prednosti poput djelovanja kao regularizacijski element i korištenja veće stope učenja bez problema s divergencijom [25]. Regularizacija grupe bit će postavljena nakon svakog konvolucijskog i potpuno povezanog sloja testiranih mreža. Za sprječavanje prenaučnosti često se koristi ispuštanje neurona (engl. *dropout*), odnosno u nekim slojevima se zadani postotak neurona ne koristi kako mreža ne bi bila previše ovisna o rezultatu tih neurona [23]. Prilikom klasifikacije podataka za ispitivanje ova metoda se ne koristi te to ponekad rezultira većom točnošću na skupu za ispitivanje od skupa za učenje. Normalizacija grupe obavlja sličnu funkciju kao i sloj ispuštanja te za njim često nema potrebe, no on će svejedno biti korišten s vjerojatnošću ispuštanja 10% [26]. Aktivacijska funkcija koristi se za modeliranje izlaza sloja. Primjerice funkcija zglobnice propušta samo pozitivne vrijednosti, a negativne pretvara u 0, dok sigmoidalna funkcija vrijednosti prebacuje u interval [0,1] koristeći eksponencijalnu funkciju [27]. Na izlaznom sloju mreže koristit će se funkcija *softmax*, koja je generalizacija sigmoidalne funkcije i koristi se za višeklasnu klasifikaciju [27]. Taj izlaz se onda može gledati kao vjerojatnost da primjer pripada određenoj klasi. U ostalim slojevima mreže koristit će se zglobnica jer omogućava brže i bolje učenje modela zbog toga što uvodi nelinearnost [27]. Umjesto zglobnice moguće je koristiti razne druge funkcije poput eksponencijalne zglobnice, no taj hiperparametar nije optimiran. Jedinstveni način odabira hiperparametara ne postoji i zbog toga je kod odabira najboljeg modela provedena optimizacija hiperparametara, odnosno testirani su razni modeli s varijabilnim parametrima.

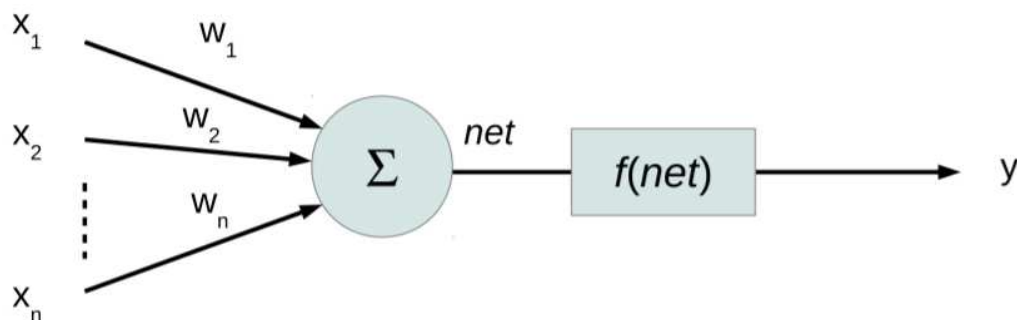
Prilikom učenja modela treba na neki način provjeriti koliko je taj model dobar u klasifikaciji primjera koje je već vidio, a pogotovo onih koje još nije vidio jer će mu to biti glavna funkcionalnost. Zbog toga je podatke potrebno podijeliti na skup koji se koristi tijekom učenja, odnosno skup za učenje, i skup nad kojim se testira, odnosno skup za testiranje. Problemi koji se rješavaju dubokim neuronskim mrežama često su složeni te je zbog toga učenje potrebno provesti u puno iteracija. Svaka iteracija naziva se epoha i u njoj se model uči na primjerima za učenje i validira na malom skupu iz primjera za učenje, odnosno

validacijskom skupu. Broj epoha ovisi o složenosti problema i učenje će biti provedeno kroz 200 epoha kako bi se što bolje proučilo ponašanje modela. Pripremljeni skup podataka podijeljen je tako da je kod učenja korišteno 80% primjera, a kod testiranja 20%. Iz skupa za učenje izlučen je skup za validaciju veličine 25% skupa za učenje, odnosno 20% ukupnih primjera. Skup za učenje sadrži 60% ukupnih primjera. U sva tri skupa ima jednako primjera iz svih klasa kako bi učenje i ispitivanje bilo uravnoteženo.

Problem klasifikacije glazbenih isječaka je prilično složen te se zbog toga u ovom radu koriste sljedeći duboki modeli: višeslojni perceptron, konvolucijska neuronska mreža i povratna mreža.

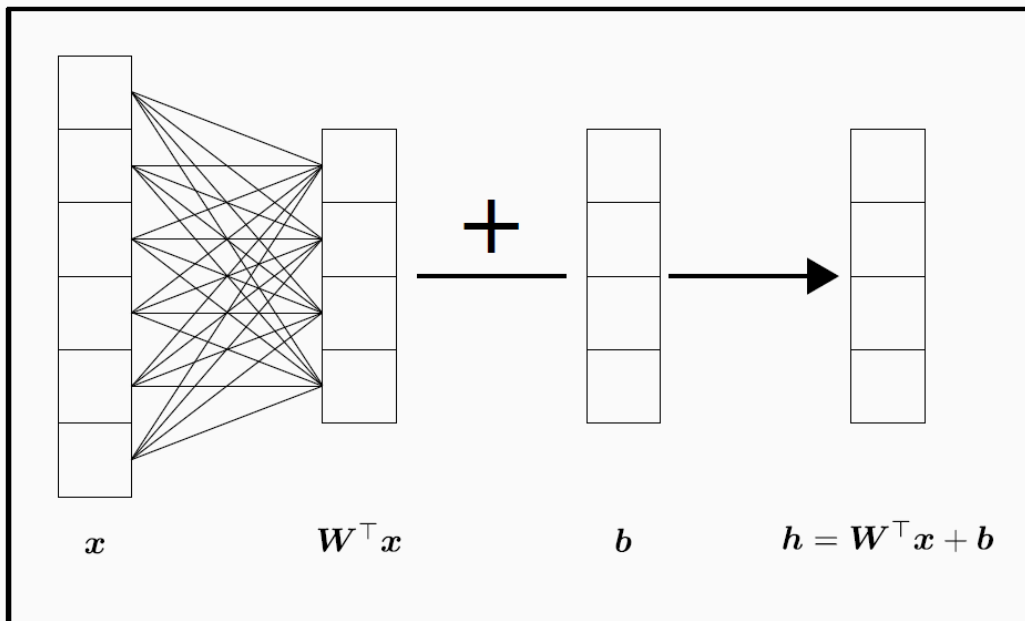
4.1. Potpuno povezana neuronska mreža

Potpuno povezana neuronska mreža je prvi tip duboke mreže nad kojim će se provoditi učenje zbog toga što je on osnovni duboki model, iako se u praksi rijetko koristi zbog svoje složenosti i velikog broja parametara [27]. Na slici 23 prikazan je umjetni neuron koji se sastoji od n ulaza i 1 izlaza, gdje su s x označeni ulazi u neuron, a s w težine za te ulaze. Neuron radi tako da svaki od ulaza množi s odgovarajućom težinom, sve te umnoške zbraja i taj zbroj prosljeđuje u aktivacijsku funkciju označenu s $f(net)$ [28].



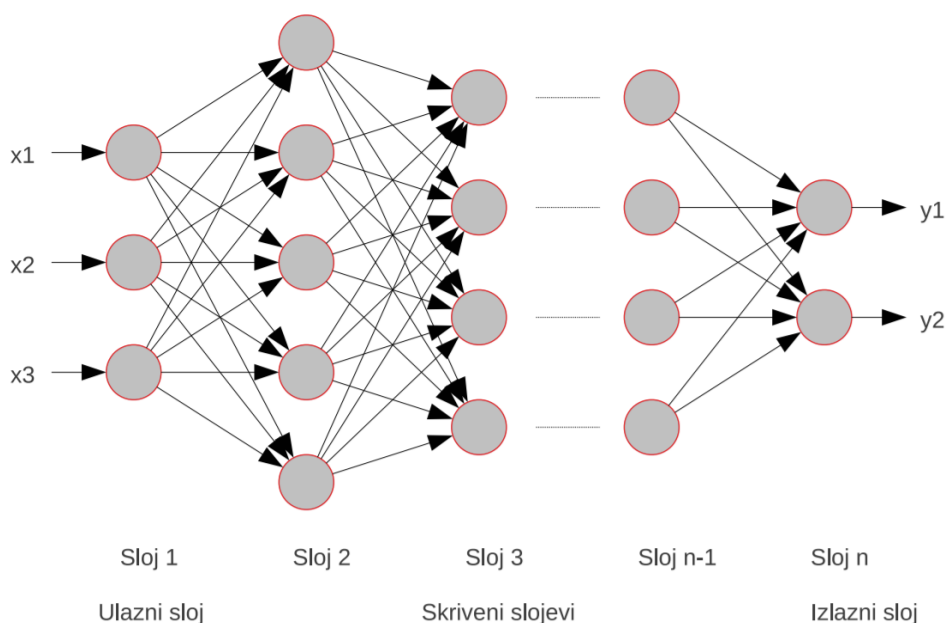
Slika 23 Umjetni neuron s aktivacijskom funkcijom [28]

Na slici 24 prikazan je niz neurona složen u neuronski sloj. Vektor x predstavlja ulaze u neuronski sloj i svaki član vektora povezan je sa svakim neuronom u drugom sloju. Zbog toga se ovakav sloj naziva potpuno povezani sloj. Na poveznicama ulaza i neurona nalaze se težine W i umnožak tih težina i ulaza rezultira vektorom $W^T x$. Taj se izraz zbraja s vektorom b , koji predstavlja pomak od ishodišta i nad njime se ne provodi regularizacija, i to je izlaz iz neuronskog sloja [24]. Izlaz sloja je vektor duljine broja neurona u sloju, koji se može provesti kroz aktivacijsku funkciju, i u dubokoj mreži služi kao ulaz u sljedeći sloj.



Slika 24 Potpuno povezani sloj [29]

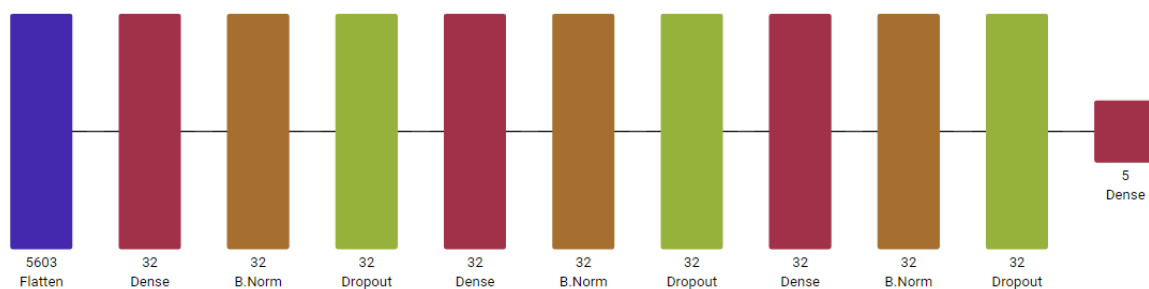
Na slici 25 prikazana je neuronska mreža koja prima vrijednosti x_1 , x_2 i x_3 i kao rezultat vraća rezultate y_1 i y_2 . Što je veći broj slojeva i neurona u sloju, veći je broj težina u mreži i mreža će bolje moći naučiti iz primjera za učenje [27].



Slika 25 Neuronska mreža [28]

Na slici 26 prikazana je jedna od testiranih potpuno povezanih mreža izrađena pomoću alata *Net2Vis* [30]. Na ulazu je vektor dimenzija $N \times 431 \times 13$, gdje je N veličina grupe, a 431 broj vremenskih isječaka za koje je izlučeno 13 koeficijenata *MFCC*, i on se splošćuje u vektor dimenzija $N \times 5603$ u prvom sloju, na slici označenom imenom "Flatten". Nakon njega slijede tri jednake grupe koje se sastoje od potpuno povezanog sloja, normalizacije grupe te

ispuštanja neurona. Na svakom od ova tri potpuno povezana sloja primijenjena je regularizacija s faktorom 0.01. Broj neurona u potpuno povezanom sloju je 32, a normalizacija i ispuštanje neurona ne utječu na izmjenu dimenzionalnosti. Za ispuštanje neurona je odabrana vrijednost 10%. U zadnji sloj ulazi vektor dimenzija $N \times 32$, a na izlazu modela primijenjena je funkcija *softmax* s ciljem klasifikacije primjera u pet glazbenih razdoblja.



Slika 26 Testirana potpuno povezana mreža

U tablici 2 prikazani su hiperparametri koji su optimirani tijekom učenja.

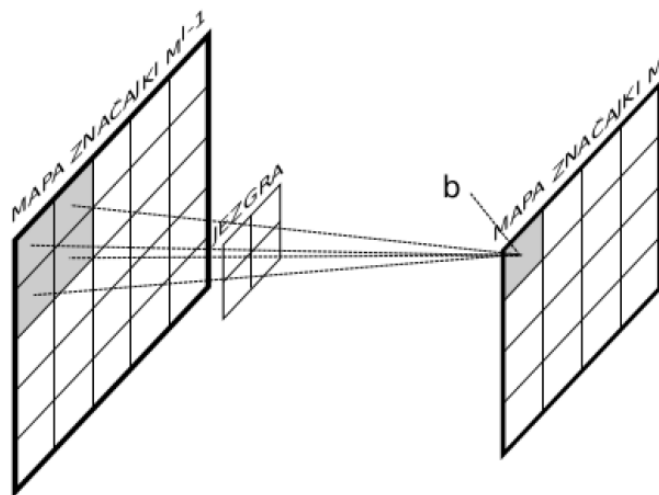
hiperparametar	vrijednosti
regularizacijski faktor	0.0001, 0.001, 0.01
broj potpuno povezanih slojeva	1, 2, 3, 4, 5
broj neurona u potpuno povezanom sloju	8, 32, 128, 512

Tablica 1 Hiperparametri korišteni tijekom optimizacije potpuno povezane mreže

4.2. Konvolucijska neuronska mreža

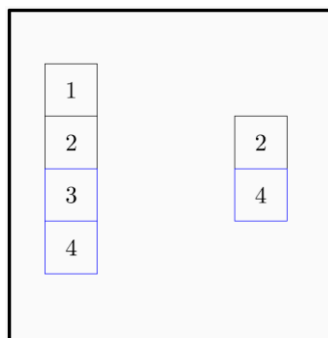
Konvolucijska neuronska mreža razlikuje se od potpuno povezane neuronske mreže u tome što sadrži barem jedan konvolucijski sloj umjesto potpuno povezanog sloja [29]. Konvolucijski sloj je jednostavniji od potpuno povezanog zbog toga što ne postoje veze sa svim neuronima u sljedećem sloju, već se vrijednosti sljedećeg sloja dobivaju konvolucijom jezgre i vrijednosti neurona [29]. Na slici 27 prikazan je konvolucijski sloj gdje se vidi da je na vrijednost sivog polja na desnoj mapi značajki ovisi o umnošku sivih polja na lijevoj mapi značajki i jezgre te pomaku b . Nakon tog izračuna jezgra se pomiče i tako se računaju sve vrijednosti sljedeće mape značajki. Vrijednost za koju se jezgra pomiče po ulaznoj mapi značajki naziva se pomak (engl. *stride*) i na prikazanoj slici on iznosi $(1, 1)$. U svakom koraku jezgra se pomiče jedno polje u desno do kraja retka i tada se vraća na početak sljedećeg retka. To je i uobičajena vrijednost pomaka koja se koristi u biblioteci *Keras* pa će se ona koristiti

u izgradnji modela. Kod konvolucije se osim dimenzionalnosti podataka zadaje i broj kanala za te podatke. Primjerice, slika u crvenoj, zelenoj i plavoj boji ima tri kanala, a crno-bijela slika ili standardni tipovi podataka za učenje neuronske mreže imaju jedan kanal.



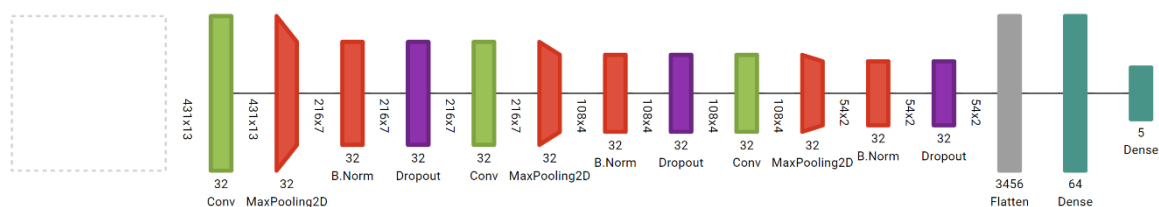
Slika 27 Konvolucija jezgre s mapom značajki [29]

Na ulaznu mapu značajki moguće je dodati nadopunu (engl. *padding*) tako da se ona uokviri nulama i time se može utjecati na veličinu izlaznog sloja. Prednost nadopune je to što će krajevi ulazne mape značajki utjecati na više značajki na izlaznom sloju i to što veličina slojeva ostaje konstantna [29]. U biblioteci *Keras* uobičajeno se ne koristi nadopuna, no zbog prethodno spomenutih razloga će se tijekom učenja ona koristiti. Funkcija sažimanja pretvara bliske značajke iz ulaznog sloja u jednu značajku na izlaznom sloju [29]. Korisna je jer povećava invarijantnost na pomak, odnosno kad je bitnije detektirati prisutnost određene značajke, nego njezinu točnu lokaciju [29]. Sažimanje se može provoditi maksimalnom vrijednosti, srednjom vrijednosti, L2 normom ili težinskim usrednjavanjem [29]. Tijekom učenja će se koristiti sažimanje maksimalnom vrijednosti (engl. *max pooling*), a za veličinu jezgre koristit će se ista vrijednost kao i za konvolucijski sloj i primjenjivat će se nadopuna. Nadopuna neće rezultirati vektorom istih dimenzija kao ulazni vektor, nego će dodati dovoljno nula kako bi se jezgra sažimanja mogla preklopiti sa svim vrijednostima ulaznog sloja. Kod sažimanja se također postavlja vrijednost pomaka i koristit će se uobičajeno pravilo u biblioteci *Keras* prema kojem je veličina pomaka jednaka veličini jezgre. Na slici 28 prikazano je sažimanje maksimalnom vrijednosti s veličinom jezgre 2 i pomakom 2.



Slika 28 Sažimanje maksimalnom vrijednosti [29]

Na slici 29 prikazana je jedna od testiranih konvolucijskih mreža izrađena pomoću alata *Net2Vis* [30]. Na ulazu u mrežu je skup od N primjera dimenzija $N \times 431 \times 13 \times 1$, gdje zadnja vrijednost opisuje broj paralelnih kanala. Nakon ulaznog sloja slijedi grupa koja se ponavlja 3 puta i sastoji se od konvolucijskog sloja, sažimanja maksimumom, normalizacije grupe i ispuštanja neurona. Konvolucijski sloj ima 32 jezgre veličine $(2, 2)$ s nadopunom i njegov izlaz imat će dimenzije $N \times 431 \times 13 \times 32$. Sažimanje maksimumom s jezgrom $(2, 2)$ prepolovit će dimenzije i izlaz tog sloja imat će dimenzije $N \times 216 \times 7 \times 32$. Normalizacija grupe ne utječe na dimenzionalnost pa će dimenzije na sljedećem ulazu biti jednake. Za ispuštanje neurona odabrana je vrijednost 10% i ovaj sloj također ne utječe na izmjenu dimenzionalnosti. Nakon ove grupe ista stvar se ponavlja još dva puta nakon čega je izlazni vektor dimenzija $N \times 54 \times 2 \times 32$. Sljedeći sloj, na slici označen s “Flatten”, služi za spljoštavanje dimenzija s ciljem pripreme za potpuno povezani sloj i njegov izlaz imat će dimenzije $N \times 3456$. Sljedeći sloj je potpuno povezani sloj, na slici označen s “Dense”, i na njemu je primijenjena regularizacija s faktorom 0.01. Izlaz tog sloja ima dimenzije $N \times 64$ i to je ulaz u posljednji potpuno povezani sloj, nad kojim je primijenjena funkcija *softmax*, i izlaz modela su vjerojatnosti za svako od pet glazbenih razdoblja.



Slika 29 Testirana konvolucijska mreža

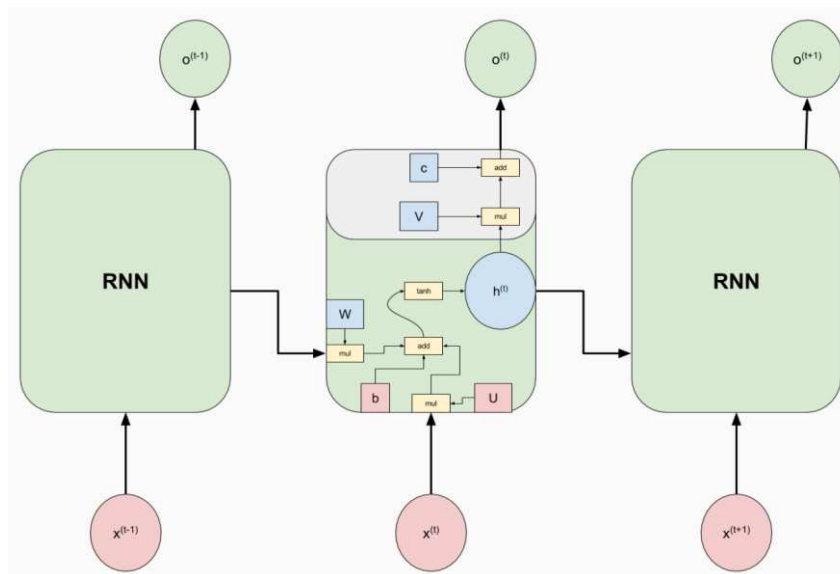
U tablici 2 prikazani su svi hiperparametri koji su korišteni tijekom optimizacije i njihove vrijednosti.

hiperparametar	vrijednosti
regularizacijski faktor	0.0001, 0.001, 0.01
broj konvolucijskih slojeva	1, 2, 3, 4
broj jezgri u konvolucijskom sloju	16, 32, 64
veličina jezgre u konvolucijskom sloju i sloju sažimanja	(2,2), (3,3)

Tablica 2 Hiperparametri korišteni tijekom optimizacije konvolucijske neuronske mreže

4.3. Povratna neuronska mreža

Prethodno opisani tipovi neuronskih mreža dobro rade za veliki broj problema, no za neke probleme te mreže nisu prikladne. Primjerice kod prevođenja rečenica na drugi jezik ne možemo učiti model ako unaprijed nemamo fiksne duljine rečenica. Kod prevođenja je također bitan poredak riječi u rečenici i da se značenje očuva od početka do kraja rečenice. Slično se može reći i kod klasifikacije audio zapisa. U sklopu ovog rada audio zapisi su podijeljeni na jednake isječke pa je time riješen problem različite duljine, no još uvijek nije riješen problem pohranjivanja bitnih informacija prethodnih podataka. Kod povratnih neuronskih mreža obrađuje se jedan po jedan ulaz tako da ga se prosljeđuje u povratnu ćeliju. Niz povratnih ćelija prikazan je na slici 30 i na njoj je vidljivo da se, za razliku od potpuno povezanih i konvolucijskih neuronskih mreža, vrijednosti prenose iz jedne ćelije u drugu.

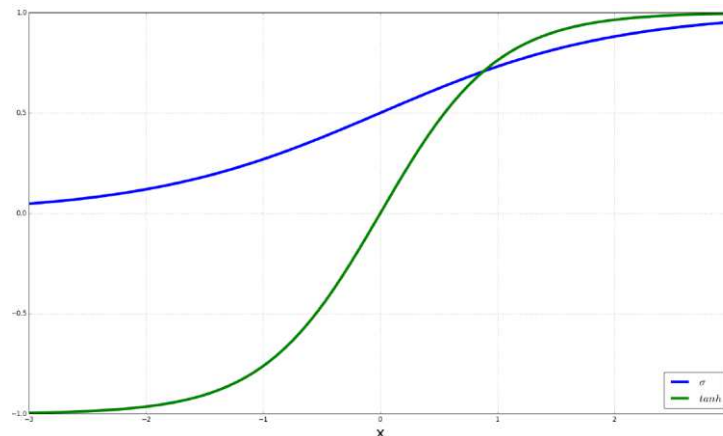


Slika 30 Niz povratnih ćelija [31]

Skriveno stanje označava se s h i ono je za prvu ćeliju inicijalizirano na 0. Skriveno stanje u trenutku t računa se kao $h_t = \tanh(Wx_{t-1} + Ux_t + b)$, gdje je \tanh funkcija tangens

hiperbolni, W težine za skriveni sloj, U težine za ulazni sloj i b pomak. Skriveno stanje prosljeđuje se u sljedeću ćeliju, a može ga se koristiti i za izračun izlaza te ćelije. Ovisno o problemu koji povratna mreža rješava, moguće je da bude jedan ili više izlaza. Više izlaza bit će primjerice kod prevođenja na drugi jezik, a prilikom rješavanja problema klasifikacije glazbenih isječaka bit će bitan samo izlaz posljednje ćelije. Izlaz ćelije u trenutku t se računa kao $o^t = Vh^t + c$, gdje su V težine za skriveni sloj, a c pomak.

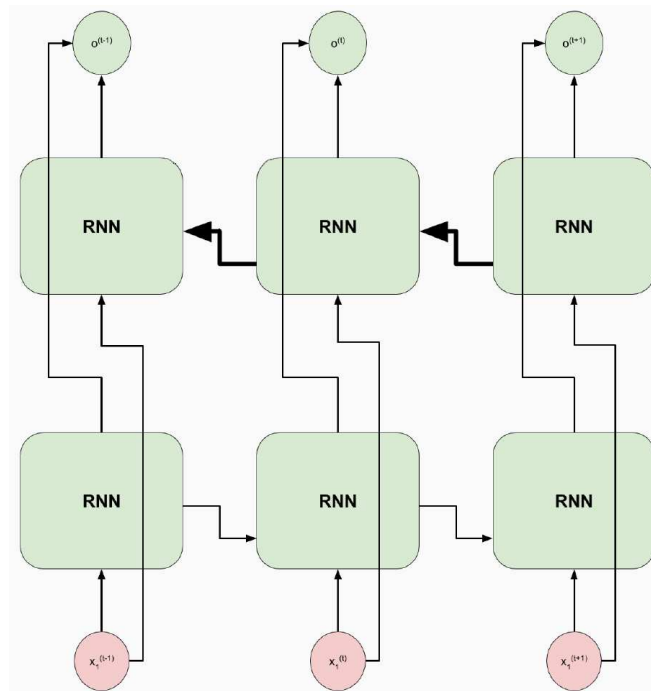
Tangens hiperbolni i sigmoida prikazani su na slici 31 i na njoj se vidi da se razlikuju u tome što tangens hiperbolni ograničava ulaze u interval $[-1, 1]$. Ovo svojstvo je potrebno zbog toga što mreža može biti jako duboka i propagacijom gradijenata kroz ćelije taj gradijent bi mogao postati jako velik. Ovaj problem naziva se eksplodirajući gradijent i zbog njega mreža može postati nestabilna i preskočiti lokalne minimume, a može doći i do preljeva i to će rezultirati greškom [32].



Slika 31 Usporedba sigmoide i tangensa hiperbolnog [31]

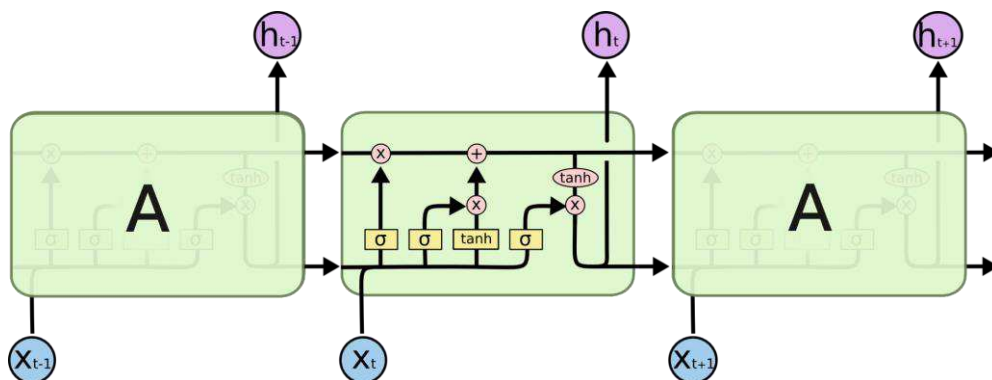
Zbog toga što svaki ulaz ima svoju ćeliju, ovaj tip neuronske mreže još je dublji od prethodno spomenutih i zbog toga bi imao još više parametara. Broj parametara se smanjuje tako da se parametri dijele između ćelija tako da se tri prethodno spomenute matrice težina, U , W i V , koriste se u svakoj ćeliji. U povratnoj mreži bilo bi dobro da i raniji ulazi imaju informaciju o vrijednostima i izračunima kasnijih ulaza te zbog toga postoje dvosmjerne povratne neuronske mreže. Takva mreža prikazana je na slici 32 i na njoj se vidi da je to zapravo dvostruka jednosmjerna povratna mreža čiji izlaz je konkatencija izlaza pojedinih slojeva ili uprosječivanje ako postoje memorijska ograničenja [31]. Dvosmjernost bi mogla pomoći da se model bolje nauči, no mana je dodatna vremenska složenost. Kod jednosmjerne povratne mreže vremenska složenost iznosi $T + D - 1$, gdje je T broj vremenskih koraka, a D broj povratnih slojeva [33]. Nakon što je iz prve ćelije generiran izlaz, moguće ga je prosljediti u prvu ćeliju sljedećeg sloja. Kod dvosmjerne povratne mreže vremenska

složenost je $T * D$ zbog toga što je potrebno proći kroz sve vremenske korake kako bi se izlaz prosljedio u sljedeći povratni sloj [33].



Slika 32 Dvosmjerna povratna neuronska mreža [33]

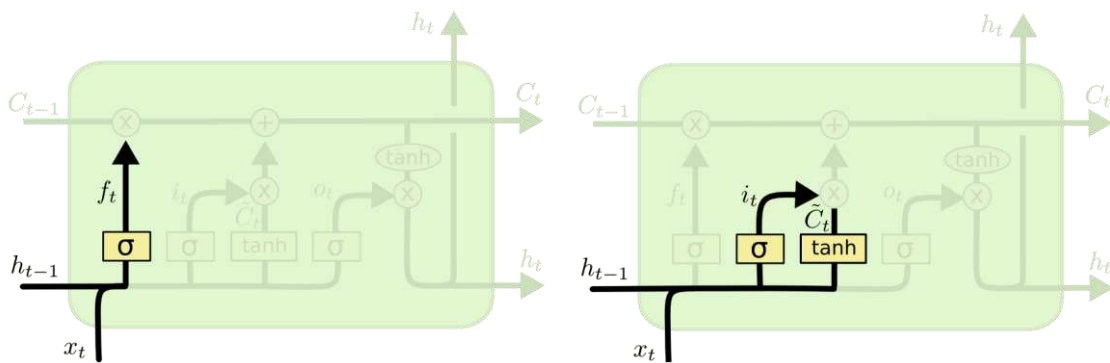
Kroz niz vremenskih koraka neka informacija može izgubiti svoju važnost i zato se razvija kompleksnija povratna ćelija s dugoročnom memorijom (*LSTM*). Ova ćelija prikazana je na slici 33 i pomoću nje je moguće zaboraviti određeni dio informacija iz prošlog stanja te propustiti samo određeni dio informacija iz ulaza. Spajanje dvije strelice označava konkatenciju matrica, razdvajanje strelice označava kopiranje matrice, rozi elementi su funkcije koje se provode nad istim elementima ulaza, a žuti elementi su aktivacijski slojevi neuronske mreže, sigmoida i tangens hiperbolni.



Slika 33 Ćelija s dugoročnom memorijom [33]

Na slici 34 prikazana su vrata zaboravljanja i ulazna vrata. Izlaz vrata zaboravljanja u ćeliji t računa se formulom $f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$, gdje je σ sigmoidalni sloj, W_f matrica

težina vrata zaboravljanja, $[h_{t-1}, x_t]$ konkatenerana matrica skrivenog stanja iz prošle ćelije i trenutnog ulaza i b_f pomak vrata zaboravljanja. Izlaz ovih vrata opisuje koliko je bitno zadržati informacije i ta značajnost je prikazana kao postotni izlaz iz sigmoide [34]. Ulazna vrata su kompleksnija te imaju dva međurezultata, izlaz sigmoide i izlaz tangensa hiperbolnog. Izlaz sigmoide u ćeliji t računa se formulom $i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$, gdje je W_i matrica težina, a b_i pomak ulaznih vrata. Izlaz tangensa hiperbolnog u ćeliji t računa se formulom $\hat{C}_t = \tanh(W_C * [h_{t-1}, x_t] + b_C)$, gdje je W_C matrica težina, a b_C pomak operatora tangensa hiperbolnog. Izlaz ovih vrata opisuje nove vrijednosti koje će biti dodane u stanje ćelije [34].

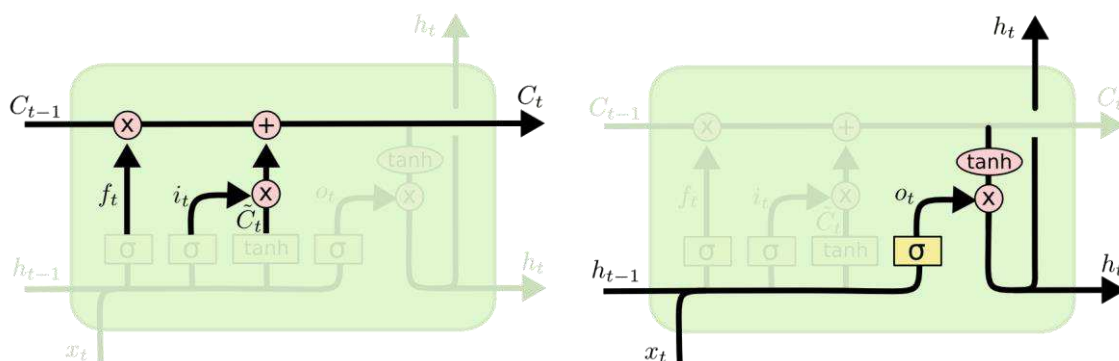


Slika 34 Vrata zaboravljanja (lijevo) i ulazna vrata (desno) [33]

Stanje ćelije prikazano je lijevo na slici 35 i u trenutku t računa se formulom $C_t = f_t * C_{t-1} + i_t * \hat{C}_t$. Stanje ćelije prosljeđuje se u sljedeći sloj i tako je ostvareno dugotrajno pamćenje modela. Izlazna vrata ćelije prikazana su desno na slici 35 i prvo je potrebno izračunati izlaz iz označene sigmoide prema formuli $o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$. Taj se izraz množi s tangensom hiperbolnim stanja ćelije, kako bi vrijednosti bile u intervalu $[-1,1]$, prema formuli $h_t = o_t * \tanh(C_t)$. Izlaz ćelije prosljeđuje se u sljedeću ćeliju, a moguće je proslijediti ga i u sljedeći sloj. Postoje razne varijacije ćelija s dugoročnom memorijom među kojima je jedan od poznatijih Gated Recurrent Unit, koji je jednostavniji, a ima slične rezultate kao standardni LSTM [35].

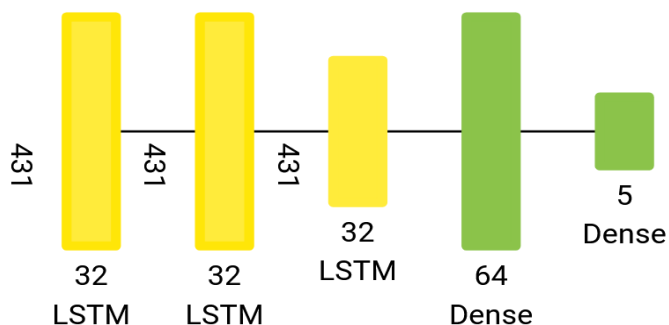
Tijekom učenja modela neće biti korištena normalizacija grupe zbog toga što ona kod povratnih neuronskih mreža može biti štetna i rezultirati eksplodirajućim ili nestajućim gradijentom [36].

Na slici 36 prikazana je jedna od testiranih povratnih mreža izrađena pomoću alata *Net2Vis* [30].



Slika 35 Stanje ćelije (lijevo) i izlazna vrata (desno) [33]

Ova mreža izgleda jednostavnije od prethodnih, no detaljnijom analizom jednostavno se zaključuje da to nije tako. Na ulazu u mrežu je vektor dimenzija $N \times 431 \times 13$ i u prvom sloju se za svaki od 431 vremenskih koraka stvaraja povratna ćelija s 32 neurona koja prosljeđuje informacije prema kasnijim vremenskim koracima. Izlaz tog sloja je vektor dimenzija $N \times 431 \times 32$. Na svaki sloj primjenjeno je ispuštanje 10% neurona. Drugi sloj je jednak prvom po svemu, osim po dimenzijama ulaznog vektora. Treći sloj razlikuje se od prva dva po tome što je na njegovom izlazu vektor dimenzija $N \times 64$ koji se može spojiti na sljedeći potpuno povezani sloj. Taj predzadnji sloj ima 64 neurona i nad njime je provedena regularizacija s faktorom 0.01. Zadnji sloj je ponovo potpuno povezani s 5 neurona i funkcijom *softmax*.



Slika 36 Testirana povratna mreža

U tablici 4 prikazani su hiperparametri koji su optimirani tijekom učenja modela i njihove vrijednosti.

hiperparametar	vrijednosti
regularizacijski faktor	0.0001, 0.001, 0.01
broj povratnih slojeva	1, 2, 3, 4
broj neurona u sloju	16, 32, 64
dvosmjernost	ne, da

Tablica 3 Hiperparametri korišteni tijekom optimizacije povratne mreže LSTM

5. Rezultati klasifikacije

Programski kod za učenje i testiranje modela napisan je u programskom jeziku *Python* koristeći biblioteku *Keras*. Učenje puno modela zahtijeva velike računalne resurse i zbog toga je korišten *Google Colab* za provedbu testiranja modela [37]. Prije početka obrade rezultata učenja treba napomenuti da je inicijalna podjela u skup za učenje, skup za validaciju i skup za testiranje sigurno utjecala na rezultate pojedinog modela. Drugačija podjela mogla bi rezultirati boljim modelima, no testiranje s različitim podjelama skupova nije obuhvaćeno ovim radom. Za svaku konfiguracija hiperparametara provedeno je 200 epoha i spremljen je model koji je postigao najbolju točnost na skupu za validaciju. Na tom modelu je izračunata točnost na skupu za testiranje. Ova strategija je odabrana zbog toga što nakon određenog broja epoha model postaje prenaučeni i točnost na skupu za validaciju počinje padati. Tako se model testira u trenutku kada je najbolji.

U sljedećim tablicama prikazani su rezultati učenja modela na skupu za testiranje za različite hiperparametre. Točnost klasifikacije prikazana je u intervalu $[0,1]$, odnosno moguće je tumačiti ju kao postotnu točnost modela kada ju se pomnoži sa 100. Primjerice, vrijednost 0.2333 označava da je model 23.33% primjera smjestio u ispravno razdoblje.

Za najbolje i najlošije modele svakog tipa mreže prikazani su grafovi točnosti i gubitaka na skupu za učenje i skupu za validaciju. Horizontalna os označava indeks epohe, a vertikalna točnost, odnosno gubitak. Gubitak je prikazan u logaritamskom mjerilu zbog toga što se u njemu jasnije vidi kretanje, nego u linearnom mjerilu. U slikama se mogu uočiti oscilacije, umjesto konzistentnog kretanja prema višoj točnosti i manjem gubitku. Te oscilacije moguće je ublažiti optimizacijom stope učenja i veličine grupe, no ta optimizacija u sklopu rada nije provedena. Za ta dva hiperparametra odabrane su vrijednosti za koje je veliki postotak modela na skupu za učenje postigao prenaučeniost, a da se izračun završi u što kraćem vremenu. Kada ne bi postojala vremenska i računalna ograničenja, velika je vjerojatnost da bi modeli imali bolje rezultate.

5.1. Potpuno povezana neuronska mreža

U tablicama 4, 5 i 6 prikazani su rezultati klasifikacije na testnom skupu za različite regularizacijske faktore, brojeve slojeva i brojeve neurona u sloju. Za stopu učenja korištena je vrijednost 0.00005, a veličinu grupe 256.

slojevi \ neuroni	8	32	128	512
1	0.4633	0.5498	0.5531	0.5720
2	0.4771	0.5618	0.5947	0.5927
3	0.4858	0.5660	0.6058	0.6420
4	0.4584	0.5829	0.6278	0.6740
5	0.4760	0.5740	0.6313	0.6669

Tablica 4 Rezultati klasifikacije potpuno povezane mreže na testnom skupu za regularizacijski faktor 0.0001

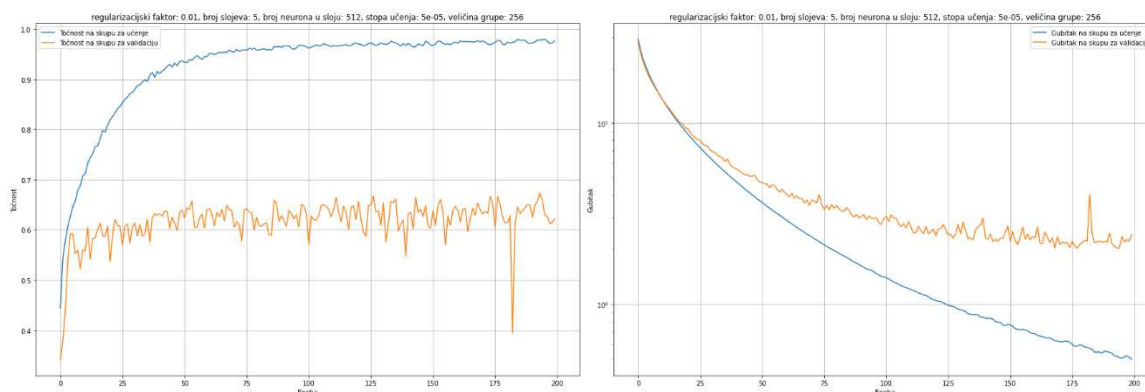
slojevi \ neuroni	8	32	128	512
1	0.3993	0.5418	0.5627	0.5711
2	0.4629	0.5633	0.5891	0.5993
3	0.4902	0.5689	0.6140	0.6187
4	0.5093	0.5649	0.6360	0.6756
5	0.5053	0.5778	0.6322	0.6722

Tablica 5 Rezultati klasifikacije potpuno povezane mreže na testnom skupu za regularizacijski faktor 0.001

slojevi \ neuroni	8	32	128	512
1	0.4738	0.5191	0.5449	0.5671
2	0.4449	0.5520	0.5733	0.5727
3	0.4893	0.5680	0.5973	0.6244
4	0.4507	0.5740	0.6120	0.6658
5	0.4729	0.5616	0.6147	0.6796

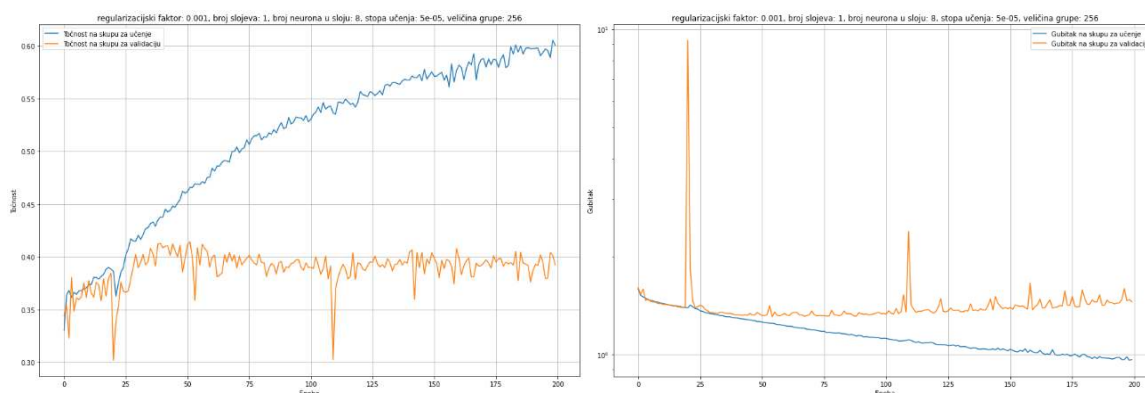
Tablica 6 Rezultati klasifikacije potpuno povezane mreže na testnom skupu za regularizacijski faktor 0.01

Iz tablica je vidljivo da je najbolji rezultat 67.96% postignut za model s pet slojeva od 512 neurona i regularizacijskim faktorom 0.01. Na slici 37 prikazani su točnost i gubitak tog modela kroz epohe za skup za učenje i validacijski skup. Već oko 55. epohe postignuta je gotovo najbolja točnost, a tek u 193. epohi je postignuta najbolja točnost.



Slika 37 Točnost i gubitak na skupu za učenje i validacijskom skupu za najbolju potpuno povezanu neuronsku mrežu

Na slici 38 prikazani su točnost i gubitak za najbolji model s jednim slojem od 8 neurona s regularizacijskim faktorom 0.001. Taj model ostvario je točnost na skupu za testiranje od 39.93%, a tijekom učenja najveća točnost na validacijskom skupu bila je oko 42%.



Slika 38 Točnost i gubitak na skupu za učenje i validacijskom skupu za najboljšu potpuno povezanu neuronsku mrežu

5.2. Konvolucijska neuronska mreža

U tablicama 7, 8 i 9 prikazani su rezultati klasifikacije na testnom skupu za različite regularizacijske faktore, brojeve slojeva, brojeve jezgri i veličine jezgri. Za stopu učenja korištena je vrijednost 0.001, a veličinu grupe 256.

slojevi \ jezgre	veličina jezgre	16		32		64	
		(2,2)	(3,3)	(2,2)	(3,3)	(2,2)	(3,3)
1		0.5316	0.5582	0.4644	0.548	0.4347	0.3944
2		0.5804	0.6258	0.574	0.6602	0.5747	0.6811
3		0.634	0.6822	0.6533	0.7318	0.6778	0.7184
4		0.6436	0.7316	0.6833	0.7738	0.7271	0.7747

Tablica 7 Rezultati klasifikacije konvolucijske mreže na testnom skupu za regularizacijski faktor 0.0001

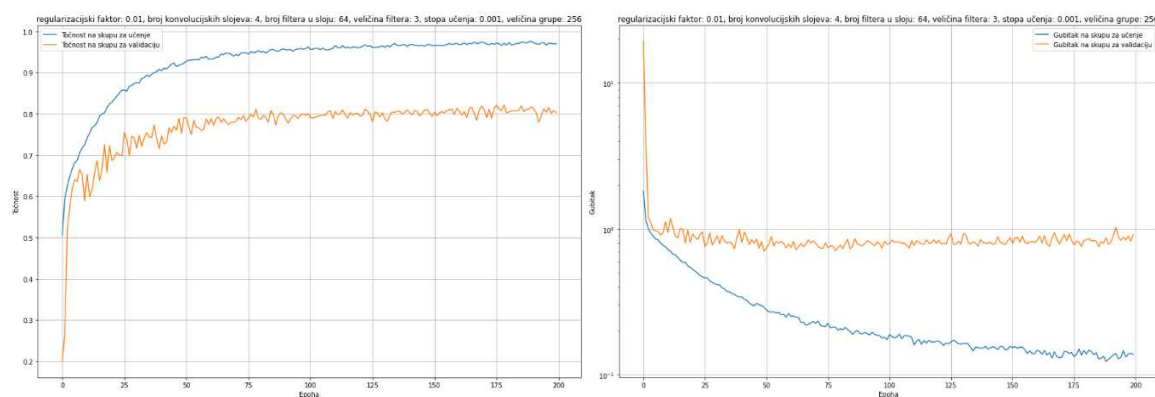
slojevi \ jezgre	veličina jezgre	16		32		64	
		(2,2)	(3,3)	(2,2)	(3,3)	(2,2)	(3,3)
1		0.5178	0.5547	0.4644	0.5262	0.3698	0.5131
2		0.5847	0.63	0.5973	0.66	0.5831	0.6773
3		0.638	0.6971	0.682	0.7242	0.6718	0.7251
4		0.63	0.7198	0.6964	0.7704	0.7184	0.7891

Tablica 8 Rezultati klasifikacije konvolucijske mreže na testnom skupu za regularizacijski faktor 0.001

slojevi \ jezgre	veličina jezgre	16		32		64	
		(2,2)	(3,3)	(2,2)	(3,3)	(2,2)	(3,3)
1		0.5598	0.5953	0.5364	0.5709	0.3878	0.5291
2		0.5862	0.6478	0.5687	0.672	0.5678	0.6869
3		0.6369	0.678	0.7027	0.7873	0.7273	0.7924
4		0.6618	0.736	0.7558	0.7976	0.7776	0.8133

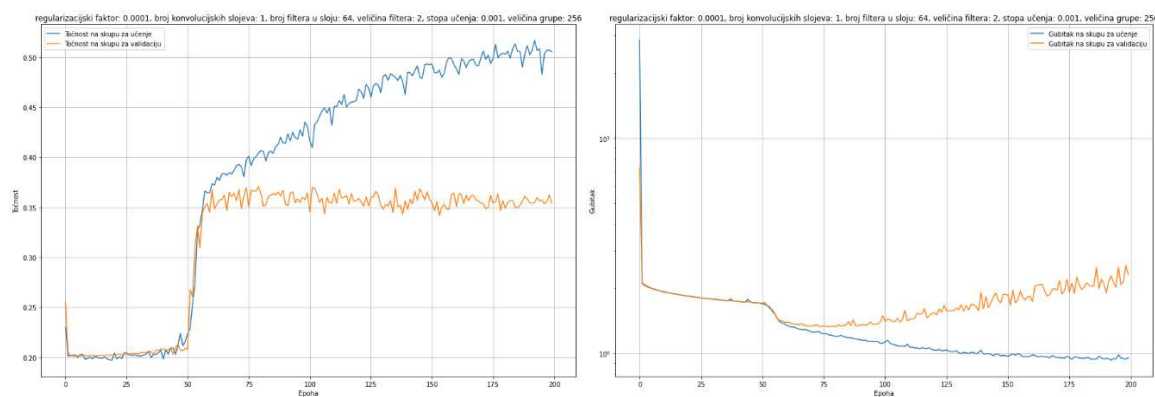
Tablica 9 Rezultati klasifikacije konvolucijske mreže na testnom skupu za regularizacijski faktor 0.01

Najbolja ostvarena točnost na skupu za testiranje iznosi 81.33% i postignuta je za konvolucijsku mrežu s regularizacijskim faktorom 0.01, 4 sloja sa 64 jezgre veličine (3,3). Kao i kod potpuno povezane mreže, i ovdje je najbolji rezultat ostvaren za najkompleksniju konfiguraciju te bi veći broj slojeva sa više većih jezgri potencijalno postigao bolje rezultate. Točnost i gubitak najboljeg modela na skupu za učenje i validacijskom skupu prikazani su na slici 39. Najbolja točnost na validacijskom skupu ostvarena je u 178. epohi te zato postoji mogućnost da bi dodatne epohe učenja dodatno poboljšale najbolji rezultat.



Slika 39 Točnost i gubitak na skupu za učenje i validacijskom skupu za najbolju konvolucijsku neuronsku mrežu

Najlošiji model, s regularizacijskim faktorom 0.001, jednim slojem sa 64 jezgre veličine (2,2), ostvario je točnost na skupu za testiranje od 36.98. Točnost i gubitak na skupu za učenje i validacijskom skupu prikazani su na slici 40. Najveća točnost postignuta je u 79. epohi i nakon toga ona pada, bez obzira na slabu točnost skupa za učenje.



Slika 40 Točnost i gubitak na skupu za učenje i validacijskom skupu za najlošiju konvolucijsku neuronsku mrežu

5.3. Povratna neuronska mreža

U tablicama 10, 11 i 12 prikazani su rezultati klasifikacije na testnom skupu za različite regularizacijske faktore, brojeve slojeva, brojeve neurona u sloju i dvosmjernost mreže za mrežu LSTM. Za stopu učenja korištena je vrijednost 0.001, a veličinu grupe 1024. Ovdje je korištena veća grupa zbog toga što bi učenje s manjim grupama trajalo jako dugo, pogotovo kod dvosmjernih mreža.

neuron slojevi	dvosmjernost	16		32		64	
		ne	da	ne	da	ne	da
1		0.3422	0.516	0.4984	0.4378	0.4076	0.5464
2		0.5713	0.6387	0.59	0.6247	0.694	0.6958
3		0.4618	0.5996	0.5678	0.6776	0.6633	0.7093
4		0.53	0.6131	0.5744	0.6573	0.6678	0.7276

Tablica 10 Rezultati klasifikacije povratne mreže na testnom skupu za regularizacijski faktor 0.0001

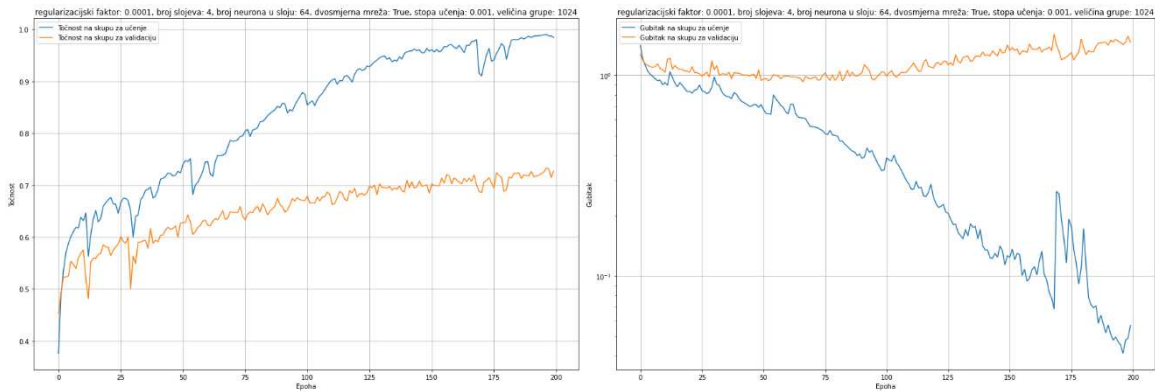
neuron slojevi	dvosmjernost	16		32		64	
		ne	da	ne	da	ne	da
1		0.512	0.5269	0.4553	0.4756	0.5236	0.462
2		0.5724	0.6082	0.5164	0.6647	0.5482	0.7084
3		0.4642	0.5376	0.5756	0.6362	0.6071	0.7267
4		0.4862	0.6169	0.5944	0.6622	0.6211	0.706

Tablica 11 Rezultati klasifikacije povratne mreže na testnom skupu za regularizacijski faktor 0.001

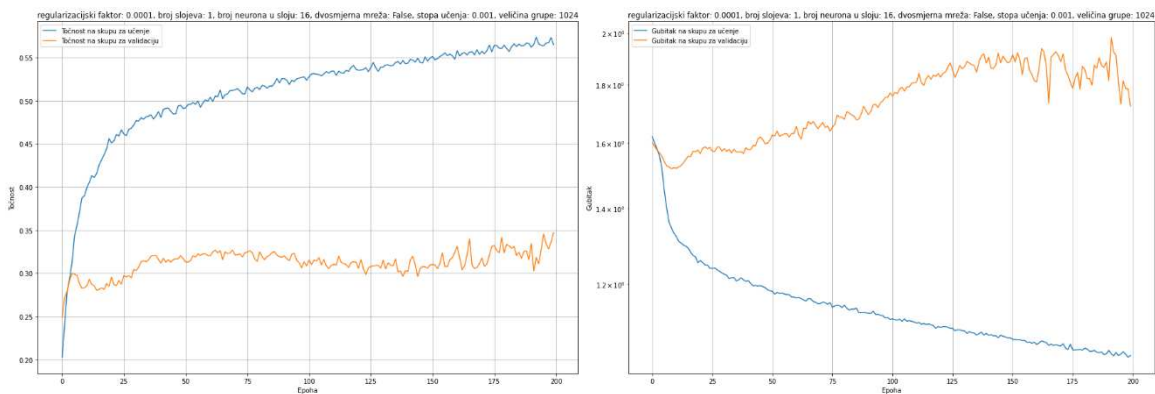
neuron slojevi	dvosmjernost	16		32		64	
		ne	da	ne	da	ne	da
1		0.3913	0.4611	0.4938	0.5456	0.5182	0.5104
2		0.5589	0.5847	0.5498	0.6071	0.5642	0.7027
3		0.4367	0.5769	0.4742	0.6827	0.6091	0.6987
4		0.3787	0.6142	0.5129	0.6533	0.5798	0.6429

Tablica 12 Rezultati klasifikacije povratne mreže na testnom skupu za regularizacijski faktor 0.01

Najbolji ostvareni rezultat klasifikacije je 72.76% za regularizacijski faktor 0.0001 i 4 dvosmjerna sloja sa 64 neurona. Kod ove mreže najbolja točnost također je ostvarena za najkompleksniji model, a iz rezultata je moguće zaključiti da bi dodatni neuroni u sloju mogli još povećati točnost. Na slici 41 prikazani su točnost i gubitak na skupu za učenje i validacijskom skupu i ovdje je najbolja točnost na validacijskom skupu ostvarena u 196. epohi i moguće je da bi daljnjim učenjem ta točnost još porasla. Najniža točnost na validacijskom skupu iznosi 34.22% i ostvarena je za model s regularizacijskim faktorom 0.0001 i jedan jednosmjerni sloj sa 16 neurona. Točnost i gubitak za tog modela na skupu za učenje i validacijskom skupu prikazani su na slici 42.



Slika 41 Točnost i gubitak na skupu za učenje i validacijskom skupu za najbolju povratnu neuronsku mrežu

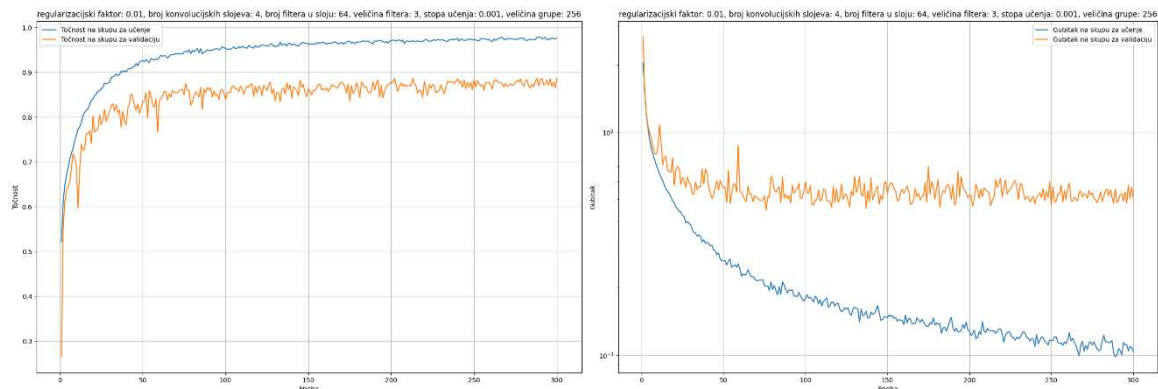


Slika 42 Točnost i gubitak na skupu za učenje i validacijskom skupu za najlošiju povratnu neuronsku mrežu

5.4. Odabrani model

Konvolucijska mreža s regularizacijskim faktorom 0.01 i 4 sloja sa 64 jezgre veličine (3,3) ostvarila je najbolje rezultate i zbog toga će se ona koristiti u aplikaciji. Ta mreža je naučena u 300 epoha nad skupom za učenje i skupom za validaciju, odnosno nad 80% svih isječaka. Posljednjih 20% je ostavljeno za testni skup kako bi se pratila točnost i pogreška na tom skupu. Na slici 43 prikazani su točnost i gubitak na skupu za učenje (sada svih 80% skupa) i testnom skupu i najbolja točnost na testnom skupu iznosi 88.73% i ostvarena je u 213. epohi. Sa slike je vidljivo da ne dolazi do prenaučivosti, odnosno pada točnosti na testnom skupu, a točnost na skupu za učenje je gotovo 100%. Pretpostavka je da je uzrok tome način na koji su isječci podijeljeni u skupove. Zbog toga što su sve kompozicije prvo podijeljene i nakon toga isječci nasumično razdijeljeni po skupovima u konzistentnim omjerima, u testnom skupu se sigurno nalaze dijelovi kompozicija nad kojima je mreža naučena. To nisu isti isječci, no slične primjere je lakše klasificirati od potpuno novih. Moguće je da bi rezultati bili lošiji, i time realističniji, kada bi se podjela radila tako da se isječci iste

kompozicije ne nalaze u različitim skupovima. Takvu podjelu bilo bi zanimljivo ispitati, no ona nije provedena u sklopu ovog rada.



Slika 43 Točnost i gubitak na skupu za učenje i validacijskom skupu za odabrani model povratne neuronske mreže

Konfiguracija i parametri najboljeg modela su spremljeni i izvršena je klasifikacija svih generiranih isječaka. Ovakva analiza daje uvid u to za koje kompozitore mreža zaključuje da bolje reprezentiraju pojedino razdoblje. Točnost klasifikacije po kompozitorima prikazana je u tablici 13.

razdoblje	točnost klasifikacije isječaka kompozitora
srednji vijek	Adam de la Halle (1.0000), Adam de St Victor (0.8750), Anonymous (0.9565), Antonello da Caserta (1.0000), Baude Cordier (0.9000), Colin Muset (0.9000), Comtessa Beatrix de Dia (0.9615), Francesco Landini (0.8947), Gautier de Coincy (0.9500), Gherardello da Firenze (1.0000), Giovanni da Cascia (1.0000), Guido D'Arezzo (1.0000), Guillaume de Machaut (0.9759), Guiraut Riquier (1.0000), Hildegard Von Bingen (0.7333), Jacopo da Bologna (0.8125), Jaufré Rudel (0.9500), Johannes Ciconia (0.9306), John Dunstable (0.8966), Leonin (0.9620), Lorenzo da Firenze (1.0000), Matteo da Perugia (1.0000), Moniot d'Arras (1.0000), Neidhart von Reuental (1.0000), Niccolò da Perugia (1.0000), Notker Balbulus (1.0000), Perotin (0.9459), Peter Abelard (1.0000), Petrus de Cruce (1.0000), Philip the Chancellor (1.0000), Philippe de Vitry (0.9444), Richard the Lionhearted (1.0000), Tuotilo (1.0000), Walther von der Vogelweide (1.0000), Étienne de Liège (1.0000)
renesansa	Adrian Willaert (0.6429), Alexander Agricola (0.8235), Alonso Mudarra (1.0000), Andrea Gabrieli (0.8333), Antoine Brumel (0.8947), Antoine Busnois (0.8333), Antonius Divitis (0.7273), Carlo Gesualdo (0.7000), Claude Le Jeune (0.5556), Claudio Monteverdi (0.3636), Clément Janequin (0.8000), Francesco da Milano (1.0000), Gilles Binchois (0.8235), Giovanni Gabrieli (0.7143), Giovanni Pierluigi da Palestrina (0.9652), Guillaume Dufay (0.8750), Heinrich Isaac (0.6818), Hugh Aston (0.8000), Jacob Obrecht (0.7727), Jacobus Clemens non Papa (0.7917), Jacques Arcadelt (0.8571), Jean Mouton (0.6875), Johann Stamitz (0.0000), Johannes Ockeghem (0.6190), John Dowland (1.0000), John Taverner (0.7500), Josquin des Prez (0.8750), Loyset Compère (0.8462), Luca Marenzio (1.0000), Luis de Narváez (1.0000), Maddalena Casulana (1.0000), Orlando di Lasso (0.9091), Oswald von Wolkenstein

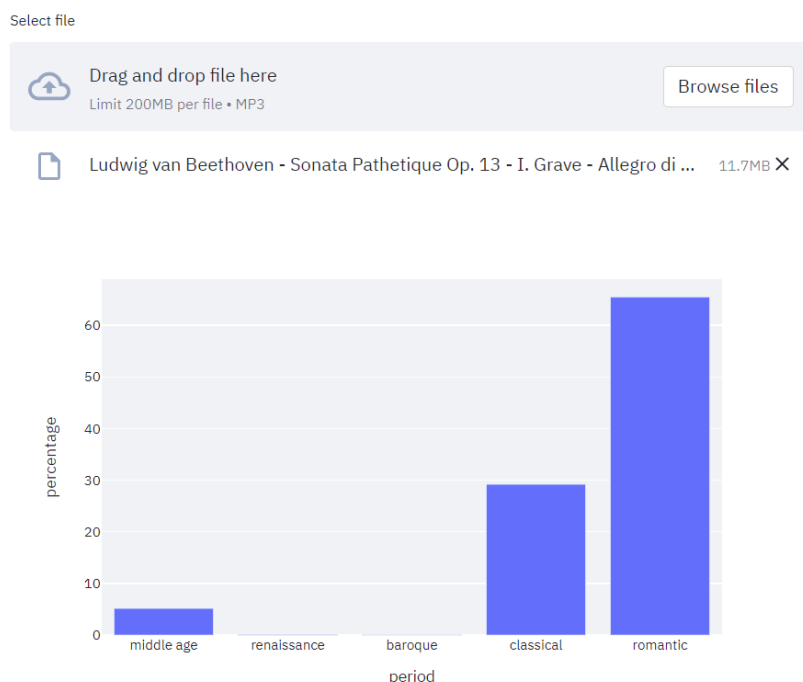
	(0.9474), Philippe Verdelot (0.8667), Pierre de La Rue (0.8571), Thomas Morley (1.0000), Thomas Tallis (1.0000), Tomás Luis de Victoria (1.0000), William Byrd (1.0000)
barok	Alessandro Scarlatti (1.0000), Alessandro Stradella (1.0000), André Campra (0.8889), Antonio Lotti (0.7917), Antonio Vivaldi (0.8611), Arcangelo Corelli (0.8333), Barbara Strozzi (0.9630), Benedetto Marcello (1.0000), Christoph Graupner (0.8966), Dietrich Buxtehude (0.8571), Domenico Scarlatti (1.0000), Francesco Cavalli (0.4615), Francesco Gasparini (0.8235), Francesco Geminiani (1.0000), François Couperin (0.8500), Georg Philipp Telemann (0.9643), George Frideric Händel (1.0000), Giacomo Carissimi (0.5714), Giovanni Battista Pergolesi (0.8788), Giovanni Legrenzi (0.7273), Giuseppe Tartini (0.8000), Heinrich Ignaz Franz Biber (0.9535), Heinrich Schütz (0.8182), Henry Purcell (0.8947), Jan Dismas Zelenka (0.7692), Jean Marie Leclair (1.0000), Jean-Baptiste Lully (0.9000), Jean-Marie Leclair (0.9286), Jean-Philippe Rameau (1.0000), Johann Hermann Schein (0.8333), Johann Jakob Froberger (0.9375), Johann Joachim Quantz (0.8333), Johann Kuhnau (0.6552), Johann Mattheson (0.8889), Johann Pachelbel (0.8571), Johann Sebastian Bach (0.8889), Luigi Rossi (0.5556), Marc-Antoine Charpentier (0.7727), Pietro Locatelli (0.9667), Samuel Scheidt (0.6667), Tomaso Albinoni (1.0000), Wilhelm Friedemann Bach (1.0000)
klasicizam	Adalbert Gyrowetz (0.9200), Anton Fils (0.8421), Antonio Salieri (0.7857), Carl Ditters von Dittersdorf (0.6000), Carl Friedrich Abel (0.9474), Carl Philipp Stamitz (0.8000), Christian Cannabich (0.7200), Christoph Willibald Gluck (0.7727), Franz Ignaz Beck (1.0000), Franz Joseph Haydn (0.8968), Franz Xaver Richter (0.7778), Georg Anton Benda (1.0000), Giovanni Paisiello (0.7500), Giuseppe Sarti (0.6111), Ignaz Fränzl (0.8889), Johann Baptist Wanhal (0.8750), Johann Christian Bach (0.9688), Josef Mysliveček (0.8621), Joseph Bologna (0.8235), Leopold Koželuch (0.8667), Leopold Mozart (0.8750), Ludwig van Beethoven (0.8333), Luigi Boccherini (0.7903), Muzio Clementi (0.9062), Niccolò Antonio Zingarelli (0.8000), Niccolò Piccinni (0.8333), Tommaso Traetta (0.7222), Vicente Martín y Soler (0.4000), Wolfgang Amadeus Mozart (0.8601)
romantizam	Anton Bruckner (0.7778), Antonín Dvořák (0.9398), Bedřich Smetana (1.0000), Camille Saint-Saëns (0.9737), Carl Czerny (1.0000), Carl Maria von Weber (0.8462), Charles Gounod (0.8000), Charles-Marie Widor (0.8571), César Franck (1.0000), Edvard Grieg (0.8333), Felix Mendelssohn (0.9167), Franz Liszt (1.0000), Franz Schubert (0.9375), Franz Xaver Gruber (0.6667), Frédéric Chopin (1.0000), Gabriel Fauré (0.7857), Gaetano Donizetti (0.5455), Gioachino Rossini (1.0000), Giuseppe Verdi (0.7500), Gustav Mahler (0.9118), Hector Berlioz (0.8750), Hugo Wolf (0.8667), Ivan Zajc (0.9286), Jacques Offenbach (1.0000), Johann Strauss I (0.8696), Johannes Brahms (0.9444), Josef Rheinberger (0.7333), Josip Runjanin (1.0000), Nikolai Rimsky-Korsakov (0.9444), Pyotr Ilyich Tchaikovsky (0.8889), Richard Wagner (0.9697), Robert Schumann (0.9500), Vatroslav Lisinski (0.7750)

Tablica 13 Rezultati klasifikacije isječka kompozitora

6. Web-aplikacija

Web aplikacija izrađena je u programskom jeziku *Python* pomoću radnog okvira *Streamlit*. *Streamlit* omogućuje brzu i jednostavnu izradu web aplikacija u kojima je lako raditi s podacima [38]. U aplikaciju je moguće unijeti datoteke u formatu *mp3* ispuštanjem datoteka u aplikaciju ili pretraživanjem po direktorijima. Za ulaznu datoteku se provjerava da ima duljinu veću od duljine isječka korištenog za učenje mreže, odnosno 10 sekundi, i pojavljuje se greška ako taj uvjet nije zadovoljen. Nakon toga se s početka i kraja datoteke obrezuje tišina i potom se ponovo provjerava duljina datoteke. Datoteka se potom sprema u *wav* formatu te joj se uklanja istosmjerna pristranost. Potom slijedi podjela na isječke duljine 10 sekundi te računanje koeficijenata *MFCC* za svaki od segmenata. Koeficijenti se spremaju u datoteku u formatu *json* i učitavaju se prilikom klasifikacije. Parametri i konfiguracija odabranog modela spremljeni su u aplikaciju te se model učitava prije klasifikacije. Klasifikacija se provodi nad svim isječcima zapisa istovremeno i dobiveni rezultati se uprosječuju i množe sa 100 kako bi bili u formatu postotka. Za prikaz klasifikacije koristi se biblioteka *plotly*. Za pripremu podataka korištene su biblioteke *json*, *os*, *Librosa*, *NumPy*, *SoundFile* i *Pydub*. Tijekom klasifikacije korištene su biblioteke *json*, *NumPy* i *TensorFlow*. Sučelje aplikacije prikazano je na slici 37.

Audio Clip Classifier



Slika 44 Sučelje aplikacije

Zaključak

Tema ovog rada je razvoj aplikacije za klasifikaciju glazbenih isječaka u glazbena razdoblja. Strojno učenje prikladno je za rješavanje ovog problema zbog toga što ne postoji jasno definiran odnos između značajki kompozicije i glazbenog razdoblja pa je iste potrebno naučiti iz prikupljenih glazbenih isječaka. Za potrebe klasifikacije ispitane su potpuno povezane, konvolucijske i povratne neuronske mreže i najbolji rezultati postignuti su kod konvolucijskih mreža. Točnost najboljeg modela, koji se koristi u aplikaciji, na testnom skupu iznosi 88.73%. Za pet razreda klasifikacije nasumični model bi ostvario točnost 20% te su zato prethodno navedeni rezultati zadovoljavajući. Bolja točnost mogla je biti ostvarena većom optimizacijom hiperparametara, poput stope učenja, veličine grupe ili ispuštanja neurona, no u tu optimizaciju se u sklopu ovog rada nije ulazilo. Glazbeni isječci skupljeni su ručno te zbog toga nisu vrhunske kvalitete i to je sigurno utjecalo na reprezentabilnost razdoblja. Za ostvarenje bolje točnosti trebalo bi ispitati i ostale značajke spomenute u ovom radu, primjerice omotnicu amplitude, korijen srednje kvadratne energije i učestalost prelaska nule. Aproksimacije prve i druge derivacije koeficijenata *MFCC* bi se također mogli dodati u ulazne vektore i tako istražiti hoće li ta vremenska povezanost rezultirati kvalitetnijim modelom. Na točnost je vjerojatno utjecao i način na koji su isječci podijeljeni u skup za učenje, skup za validaciju i skup za testiranje. Zbog toga što su isječci iste kompozicije mogli završiti u više skupova, vjerojatno su rezultati na testnom i validacijskom skupu bolji nego što bi bili na potpuno novim primjerima. Bolji rezultati na neviđenim kompozicijama mogli bi biti postignuti kada bi se podjela radila tako da se isječci iste kompozicije ne nalaze u različitim skupovima.

Literatura

- [1] Baltzell, W. J. *A Complete History of Music*. 2017.
- [2] *The History of Vocal Music*. Poveznica: <https://ourpastimes.com/the-history-of-vocal-music-12200138.html>; pristupljeno 17. svibnja 2021.
- [3] Spalding, W. R. *Music: an ART and a LANGUAGE*. 2009.
- [4] *Concert Pitch*. Poveznica: https://en.wikipedia.org/wiki/Concert_pitch; pristupljeno 17. svibnja 2021.
- [5] Šikić, Z., Šćekić, Z. *Matematika i muzika*. 1. izdanje. Zagreb: Profil
- [6] *Consonance and Dissonance*. Poveznica: <https://www.earmaster.com/music-theory-online/ch05/chapter-5-3.html#:~:text=In%20simplest%20terms%2C%20the%20sound,wave%20of%20the%20other%20note>; pristupljeno 18. svibnja 2021.
- [7] *Dynamics and Dynamic Changes*. Poveznica: https://courses.lumenlearning.com/musicappreciation_with_theory/chapter/dynamics-and-dynamics-changes/; pristupljeno 18. svibnja 2021.
- [8] *Muzička enciklopedija*. 2. izdanje. Zagreb: Jugoslavenski Leksikografski Zavod
- [9] Apel, W. *Harvard Dictionary of Music*. Cambridge: Harvard University Press, 1950.
- [10] *History of Classical Music*, Naxos, Poveznica: https://www.naxos.com/education/brief_history.asp; pristupljeno 12. svibnja 2021.
- [11] *Instruments used in Renaissance Music*. Poveznica: https://courses.lumenlearning.com/musicappreciation_with_theory/chapter/instruments-used-in-renaissance-music/; pristupljeno 18. svibnja 2021.
- [12] Cormack, J. *30-Second Classical Music*. London: 2019.
- [13] *Musical Instruments*, Poveznica: <https://www.ee.columbia.edu/~dpwe/sounds/instruments/>; pristupljeno 20. svibnja 2021.
- [14] *Amplitude Envelope*, Poveznica: <https://maplelab.net/overview/amplitude-envelope/#:~:text=Amplitude%20envelope%20refers%20to%20the,distinguish%20them%20from%20other%20sounds>; pristupljeno 22. svibnja 2021.
- [15] *Zero Crossing Rate*, Poveznica: <https://www.sciencedirect.com/topics/engineering/zero-crossing-rate>; pristupljeno: 20. svibnja 2021.
- [16] *Spectral Leakage and Zero-Padding of the Discrete Fourier Transform*, DSP Illustrations. Poveznica: <https://dspillustrations.com/pages/posts/misc/spectral-leakage-zero-padding-and-frequency-resolution.html#:~:text=Spectral%20leakage%20occurs%20when%20a,actual%20frequency%20of%20the%20signal>; pristupljeno 5. lipnja 2021.

- [17] *Analog to Digital Conversion*. Poveznica: [https://wiki.analog.com/university/courses/electronics/text/chapter-20#:~:text=Analog%2Dto%2DDigital%20converters%20\(,%2C%20computed%2C%20transmitted%20or%20stored](https://wiki.analog.com/university/courses/electronics/text/chapter-20#:~:text=Analog%2Dto%2DDigital%20converters%20(,%2C%20computed%2C%20transmitted%20or%20stored); pristupljeno 7. lipnja 2021.
- [18] Signali i sustavi, 2. vježba – *Diskretna Fourierova Transformacija (DFT)*
- [19] *Window Functions in Spectrum Analyzers*, Tektronix. Poveznica: <https://www.tek.com/blog/window-functions-spectrum-analyzers>; pristupljeno 6. lipnja 2021.
- [20] Ghosal, A., Chakraborty R. et al. *Music Classification based on MFCC Variants and Amplitude Variation Pattern: A Hierarchical Approach*, International Journal of Signal Processing, Image Processing and Pattern Recognition, 5,1 (2012.), str. 131-150.
- [21] Asmara, R. A. et al. *Comparison of Discrete Cosine Transforms (DCT), Discrete Fourier Transforms (DFT), and Discrete Wavelet Transforms (DWT) in Digital Image Watermarking*, International Journal of Advanced Computer Science and Applications, Vol. 8, No. 2, 2017
- [22] *YouTube To Mp3*, Poveznica: <https://ytmp3.cc/youtube-to-mp3/>; pristupljeno 10. travnja 2021.
- [23] Duboko učenje, *Optimizacija parametara modela*
- [24] Šnajder, J., Dalbelo Bašić, B. *Strojno učenje*. Zagreb, 2012.
- [25] Schilling, F. *The Effect of Batch Normalization on Deep Convolutional Neural Networks*, 2016.
- [26] Ioffe, S., Szegedy, C. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015.
- [27] Duboko učenje, *Duboke unaprijedne mreže*
- [28] Čupić, M., Dalbelo Bašić, B., Golub, M. *Neizrazito, evolucijsko i neuroračunarstvo*. 2013.
- [29] Duboko učenje, *Konvolucijski modeli*
- [30] *Net2Vis*. Poveznica: <https://viscom.net2vis.uni-ulm.de/>; pristupljeno 13. lipnja 2021.
- [31] Duboko učenje, *Povratne neuronske mreže*

- [32] Helfrich, K. E., *Orthogonal Recurrent Neural Networks and Batch Normalization in Deep Neural Networks*. Theses and Dissertations--Mathematics. 70. https://uknowledge.uky.edu/math_etds/70. 2020.
- [33] Duboko učenje, *Povratne neuronske mreže 2*
- [34] Olah, C., *Understanding LSTM Networks*. Poveznica: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>; pristupljeno 14. lipnja 2021.
- [35] Chung, J., Gulcehre, C., Cho, KH, Bengio, Y. *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*, 2014.
- [36] Laurent, C. et al. *Batch Normalized Recurrent Neural Networks*, 2015.
- [37] *Google Colab*. Poveznica: <https://colab.research.google.com/>; pristupljeno 7. lipnja 2021.
- [38] *Streamlit*. Poveznica: <https://streamlit.io/>; pristupljeno 17. lipnja 2021.

Aplikacija za klasifikaciju glazbenih isječaka u glazbena razdoblja

Sažetak

Cilj ovog diplomskog rada je razvijanje aplikacije za klasifikaciju glazbenih isječaka u jedno od pet glazbenih razdoblja: srednji vijek, renesansa, barok, klasicizam i romantizam. U radu su istražene razne vremenske i frekvencijske značajke koje je moguće izlučiti iz isječaka, primjerice spektrogram i kepralni koeficijenti mel-frekvencija (MFCC). Audio isječci su sakupljeni ručno s platforme YouTube, podijeljeni u isječke od 10 sekundi i iz njih su izlučeni koeficijenti *MFCC*. Za klasifikaciju su ispitane potpuno povezane, konvolucijske i povratne neuronske mreže i najbolji rezultati ostvareni su s konvolucijskim neuronskim mrežama. Parametri i konfiguracija optimalnog modela su spremljeni i korišteni su u izrađenoj web aplikaciji ostvarenoj pomoću radnog okvira *Streamlit* koja omogućava predviđanje na novim primjerima.

Ključne riječi: klasifikacija, spektrogram, neuronska mreža, aplikacija

Web Application for Audio Clip Classification into Musical Periods

Summary

The aim of this thesis is development of a web application for audio clip classification into five musical periods: middle age, renaissance, baroque, classical and romanticism. Multiple time and frequency features, which can be extracted from an audio segment, were explored such as spectrogram and Mel-Frequency Cepstral Coefficients (*MFCC*). Audio clips were gathered by hand from YouTube platform, split into 10 second segments and *MFCCs* were extracted from them. Fully connected, convolutional and recurrent neural networks have been tested for classification purposes and convolutional networks achieved the best results. Parameters and configuration of the optimal model have been saved and used in the created *Streamlit* app to predict on unseen clips.

Keywords: classification, spectrogram, neural network, application