

Klasifikacija EDM pesama po žanrovima na osnovu spektrograma

Anastasija Đurić

Softversko inženjerstvo i informacione tehnologije, FTN, Novi Sad

PROBLEM KOJI SE REŠAVA

Elektronska muzika sastoji se od veoma velikog broja žanrova i podžanrova, od kojih je svaki definisan posebnom strukturom kompozicije. Prepoznavanje i klasifikovanje EDM pesama može biti veoma izazovan posao za ljudsko uho, imajući u vidu da neki od žanrova pored svojih struktura koje ih jasno definišu, imaju i dosta sličnosti.

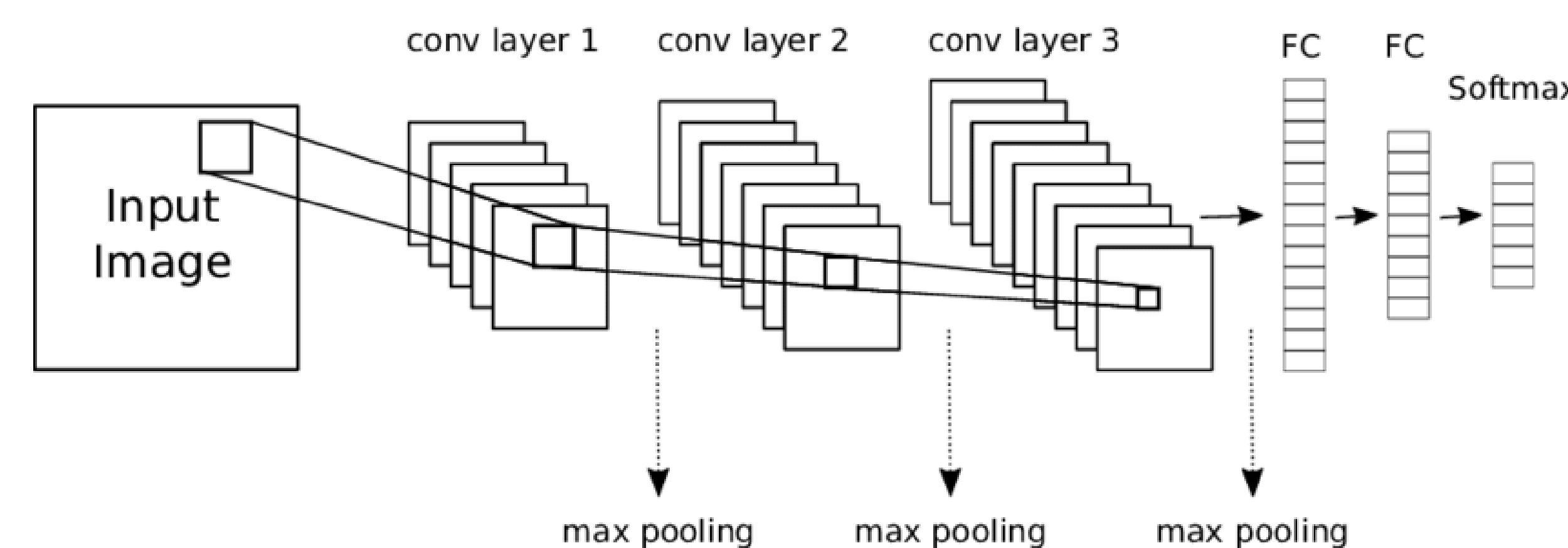
U ovom projektu, za klasifikaciju su odabrana sledeća 4 žanra:

- bigroom
- drum and bass
- house
- techno

Za svaki audio fajl potrebno je generisati njegovu vizuelnu reprezentaciju koja će se koristiti kao ulaz u konvolucionu neuronsku mrežu.

ARHITEKTURA KONVOLUCIONE NEURONSKE MREŽE

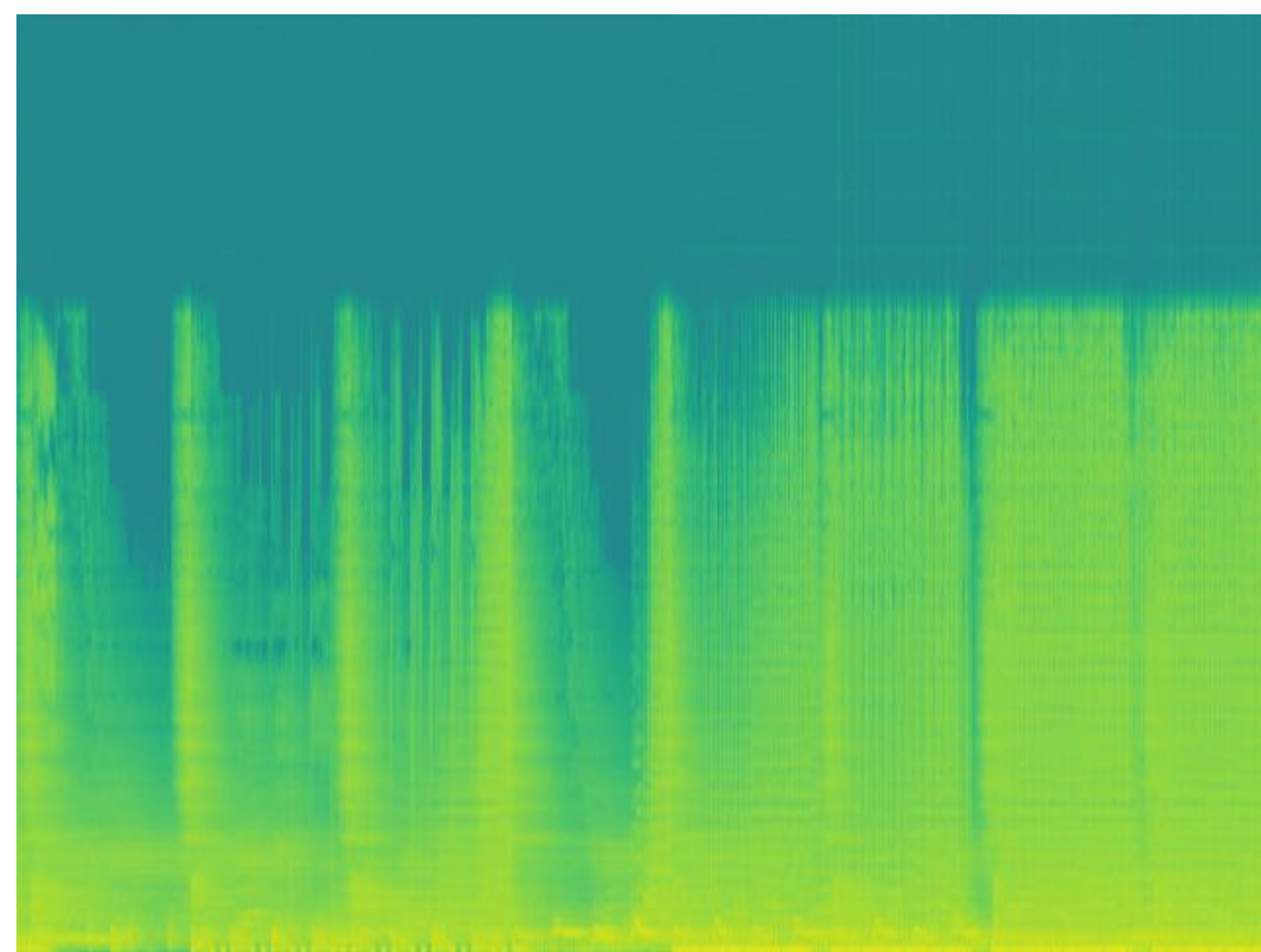
Upotrebom **keras** biblioteke kreiran je sekvencijalni model. On se sastoj od tri konvoluciona sloja. Prvi sloj sadrži 32 filtera, dimenzija 3x3. Aktivaciona funkcija koja se primenjuje na ovom sloju je relu. Zatim je dodat MaxPooling sloj sa pool dimenzijama 3x3 i strides 2x2. Finalno, dodat je i Batch Normalization sloj zadužen za normalizaciju težine neurona čime se ubrzava proces treniranja, i sam model postaje pouzdaniji. Drugi konvolucionni sloj u potpunosti je isti kao i prvi, dok su kod trećeg dimenzije jezgra i pool size 2x2. Izlaz iz konvolucionih slojeva je dvodimenzionalni niz koji se zatim pretvara u jednodimenzionalni prolaskom kroz Flatten sloj. Sledeći slojevi su Dense i Dropout. Izlazni sloj je Dense sloj sa 4 neurona od kojih svaki predstavlja po jedan žanr, i softmax aktivacionom funkcijom. Dimenzije ulaza su 128x128x1.



Slika 3. Arhitektura

SKUPOVI PODATAKA

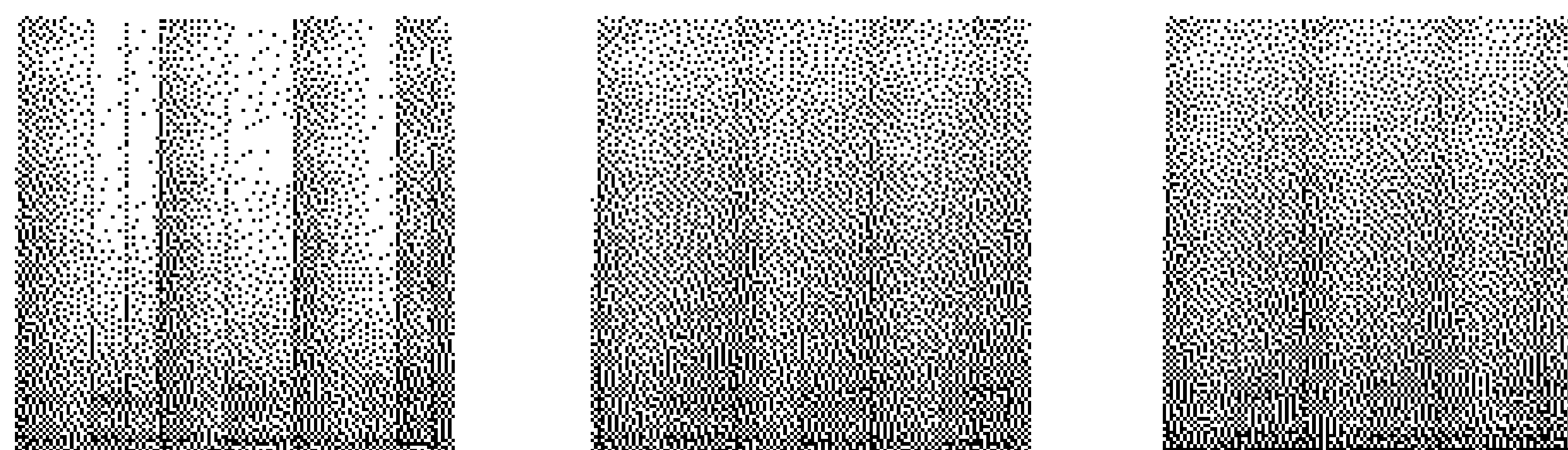
Inicijalni skup podataka kreiran je od 912 audio fajlova mp3 formata, svaki dužine 1 minut. Za učitavanje mp3 fajlova upotrebljena je biblioteka **pydub**. Nakon učitavanja, audio signal prevodi se u vizuelnu reprezentaciju – **spektrogram** koji predstavlja promenu frekvencije audio signala u vremenu. Kreirani spektrogram čuva se u direktorijumu sa nazivom žanra kome pripada u png formatu.



Slika 1. Primer spektrograma za audio signal koji pripada trance žanru

Zbog nedovoljnog broja audio semplova, inicijalni skup podataka nije davao naročite rezultate (accuracy: 40,09% za test skup).

U pokušaju da se treniranje mreže ubrza, a rezultati poboljšaju, spektrogrami pojedinačnih pesama podeljeni su na više manjih delova i pretvoreni su u monohromatske slike. Prezet je postojeći dataset.



Slika 2. Tri segmenta jedne trance pesme

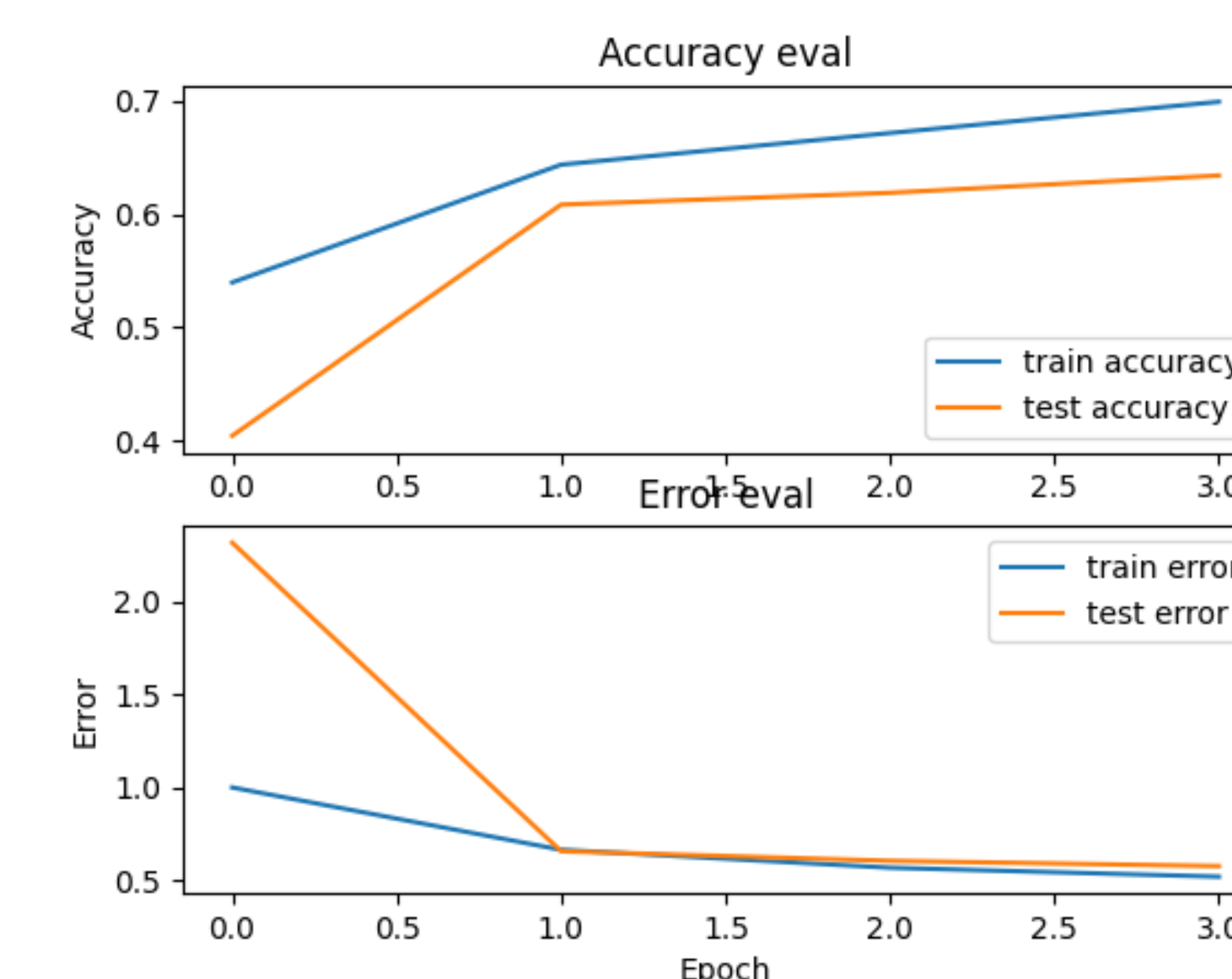
ANALIZA REZULTATA I ZAKLJUČAK

Zbog nedovoljnog broja audio semplova, inicijalni skup podataka nije davao naročite rezultate (accuracy: 40,09% za test skup).

U pokušaju da se treniranje modela ubrza, a rezultati poboljšaju, spektrogrami pojedinačnih pesama podeljeni su na više manjih delova i pretvoreni su u monohromatske slike. Za tako kreiran skup od 6000 slika (150 pesama), podeljen na trening, test (10%) i validacioni (20%) deo dobijeni su sledeći rezultati (nakon 4 epohe):

- Trening: 70,02%
- Validacija: 63,63%
- Test: 61,09%

Rezultati bi se mogli poboljšati dodatnim proširenjem i filtriranjem dataset-a tako da se izdvoje karakteristični delovi audio signala koji definišu žanr (veći broj pesama, uz po nekoliko značajnijih segmenata za svaku). Povećavanje broja epoha prilikom treniranja modela takođe bi poboljšalo rezultate, ali ti rezultati trenutno nisu dostupni zbog hardverskih ograničenja.



Slika 4. Rezultati

REFERENCE

- <https://towardsdatascience.com/musical-genre-classification-with-convolutional-neural-networks-ff04f9601a74>
- <https://www.analyticssteps.com/blogs/music-genre-classification-using-machine-learning>