

# *Szakmai gyakorlat*

Név: Nagy Csaba

Témavezető: Dr. Iclănzan David Andrei

Téma: Automatikus hangszerfelismerés zenefelvételek alapján

Számítástechnika III  
2024.08.29

## ***Tartalom***

1. Bevezető	3
2. A szakmai gyakorlat menete	3
2.1 Alapvető audio-feldolgozási műveletek elsajátítása	3
2.2 Alacsony szintű audio jellemzők kinyerése és vizualizálása	3
2.3 Fourier-transzformáció	3
2.4 Spektrogramok létrehozása és vizualizálása	4
2.5 Deep learning alapjainak átvizsgálása	4
2.6 Monofón hangszerosztályozás alapjai	4
3. Elméleti összefoglaló és tapasztalatok	4

# 1. Bevezető

A szakmai gyakorlatom célja az Államvizsga-dolgozat előkészítése során elsajátított elméleti alapokra való építkezés és folytatás. A szakdolgozatom témája a számítógépes audiófeldolgozás és a hangminták elemzése, ezért gyakorlatom középpontjában olyan kísérletek és implementációk álltak, amelyek közvetlenül építhetők a későbbi munkámba.

## 2. A szakmai gyakorlat menete

A gyakorlat során az elméleti tanulmányokat folyamatosan kísérték kódimplementációk és kisebb gyakorlati projektek, amelyek egymásra épülve vezettek el az összetettebb problémák megoldásához. A főbb lépések az alábbiak voltak:

### *2.1 Alapvető audio-feldolgozási műveletek elsajátítása*

Ebben a részben a hangfájlok kezelésének alapjait ismertem meg, valamint az audiojelek legfontosabb alapjellemzőit vizsgáltam

- Hangfájlok betöltése és alapvető jellemzőik vizsgálata (mintavételi frekvencia, amplitúdó, jelalak)
- Az időtartománybeli és frekvenciatartománybeli reprezentációk megismerése, összekapcsolása

### *2.2 Alacsony szintű audio jellemzők kinyerése és vizualizálása*

Itt az egyszerűbb matematikai és statisztikai jellemzők kiszámítására és azok grafikus megjelenítésére helyeztem a hangsúlyt.

- Időtartománybeli mutatók (amplitude envelope, root mean square, zero crossing rate)
- A jellemzők grafikus vizualizálása, értelmezése

### *2.3 Fourier-transzformáció*

Ennél a lépésnél a jelek frekvencia-összetevőit elemeztem, ami lehetővé teszi a hang szerkezetének pontosabb feltárását.

- Diszkrét Fourier-transzformáció (DFT) és gyors Fourier-algoritmus (FFT) alkalmazása
- Harmonikus összetevők azonosítása, alaphang és felhangok vizsgálata.
- Fourier-transzformáció megvalósítása és vizualizálása beépített függvény használata nélkül

## ***2.4 Spektrogramok létrehozása és vizualizálása***

A különböző zeneminták frekvenciaösszetevőinek időbeli változását spektrogramok segítségével ábrázoltam.

- Több zeneminta spektrogramját készítettem el a rövid időablakú Fourier-transzformáció (STFT) felhasználásával
- A teljesítményspektrumokat decibel-skálára alakítottam a jobb vizualizáció érdekében
- Készítettem spektrogramokat lineáris és logaritmikus frekvenciatengelyekkel, hogy különböző nézőpontból elemezhessem a hangminták szerkezetét

## ***2.5 Deep learning alapjainak átismétlése***

Felépíteni Python kódokat egyszerű neurontól, többrétegű hálókig.

- Egyszerű neuron aktivációs függvény implementálása
- Többrétegű perceptron (MLP) felépítése és előre felé terjedő számítás
- Egyszerű neurális hálózat implementálása TensorFlow-val

## ***2.6 Monofón hangszerosztályozás alapjai***

Létrehoztam egy kezdetleges neurális hálózatot, amely képes megkülönböztetni a tiszta monofón zongora és a hegedű hangmintáit. Az adatokat MFCC jellemzőkké alakítottuk, majd tréning, validáció és teszt halmazokra osztottuk. A modell konvolúciós és LSTM rétegeket tartalmaz, amelyek hatékonyan kezelik az időbeli és spektrális jellemzőket. EarlyStopping segítségével a tanítás automatizáltan leáll, amikor a teljesítmény nem javul. A teszt pontosság közel 93%, ami jó alapot ad további fejlesztésekhez és alkalmazásokhoz.

A modell a teszt halmaz eredményei alapján elkészíti a konfúziós mátrixot, és a felhasználó számára egyszerre jeleníti meg az osztályozás pontosságát vizuálisan, valamint lejátsza minden hangfájl első két másodpercét, így ellenőrzés közben egyszerre hallhatja és láthatja, hogy mit melyik hangszernek ítélt a rendszer.

*(Ehhez a ponthoz, a kód megvalósításánál kivételesen AI segítségét használtam a szűkös időkeret, a nagy és összetett anyagrész, valamint a szakmai gyakorlat végső céljának elérése érdekében.)*

## **3. Elméleti összegzés és tapasztalatok**

A szakmai gyakorlat során sikerült egy lépéssel mélyebbre kerülnöm a számítógépes audiofeldolgozás és gépi tanulás alapjaiban. A gyakorlat lehetőséget adott arra, hogy az elméleti ismereteket közvetlenül kipróbáljam és alkalmazzam különböző kódimplementációkon keresztül. A folyamat során több fontos készséget fejlesztettem, mint például a hangfájlok kezelése, frekvenciaelemzés, spektrogramok

létrehozása, valamint az egyszerűbb és haladóbb gépi tanulási modellek implementálása Python-ban és TensorFlow-ban.

A legnagyobb kihívást a mélytanulási technikák megértése és alkalmazása jelentette, ugyanakkor itt láttam meg leginkább a jövőbeli szakdolgozatom fejlődési irányait és lehetőségeit. A monofón hangszerfelismerő modell kidolgozása alapot teremtett további kutatási lépésekhez, amelyek révén később komplexebb, polifon feladatokat is meg tudok majd közelíteni.

Összességében a szakmai gyakorlat során szerzett tapasztalatok jelentős mértékben hozzájárultak szakmai tudásom bővítéséhez, és segítettek abban, hogy tudatosabban és felkészültebben álljak neki az államvizsga-dolgozat megírásának.

***Felhasznált forrás:***

***Valerio Velardo - Deep Learning (for Audio) with Python - The Sound of AI***

<https://www.youtube.com/playlist?list=PL-wATfeyAMNrtbkCNSLcpoAyBBRJZVlnf>

***Github:***

<https://github.com/nCsab/Instrument-Recognition>