

# Deep learning Házi feladat beszámoló

Készítette:

Benedek Zoltán, VZ9AS0

Biró Márton, Z7A244

Vizi Kristóf Levente, GN2VV4

Megajánlott jegyért.

## Bevezető, célkitűzések, motiváció

A projekt célja neurális háló tanítása madarak fajtájának felismerésére a hanguk alapján. Szeretnénk egy olyan rendszert tanítani amely nagy pontossággal tudja felismerni a madárcsiripelés alapján a fajt ezzel segítve az adott régió biodiverzitásának feltérképezését.

A téma választásában nagy szerepet játszott továbbá, hogy mindhármunk számára ismeretlen terület volt a hangfájlokkal való szoftveres munka, így a neurális háló összeállításán kívül további kihívás is rejtett a feladatban.

## Korábbi megoldások, előnyeik, hátrányaik

A projektünk kvázi a BirdCLEF 2024 Kaggle verseny feladványa volt. Mivel ezt a versenyt a Kaggle-ön már többedik éve teszik közzé, így rengeteg megoldás segítségünkre volt.

Ennek persze hatalmas előnye, hogy a korábbi megoldók zsákutcáit az esetek többségében ki tudtuk kerülni, de ez persze nem azt jelenti, hogy teljesen sikerült volna kiküszöbölünk minden pluszmunkát.

A sok korábbi megoldás hátránya, hogy nehéz egy helyes utat választani, hiszen több különféle módszerrel is magas eredmények érhetők el. És persze nem szeretnénk lemásolni egyik megoldást sem.

Az évek során a közzétett adathalmazokban nagyobb változások is történtek, így a hozzájuk tartozó legjobb megoldások is változtak. Ezek miatt mi elsősorban a 2024-es és 2023-as kiíráshoz érkező megoldásokból tudunk szükség esetén ihletet meríteni.

## Rendszerterv

A rendszer az alábbi komponensekből áll:

1. Adatok feltöltése Google Drive-ra
2. Vizualizáció
3. Augmentáció
4. Adatelőkészítés: Az .ogg formátumú hangfelvételek dúsítása, spektrogramokká alakítása, amelyeket a neurális háló bemeneteként használunk.

5. Spektrogramok kiírása Google Drive-ra
6. Modell összerakása és tanítás:
  - a. Modell előtanítása v. Transfer learning (még nem biztos, lehet mindkettő)
  - b. Általunk összerakott CNN alkalmazása az osztályozásra.
7. Kiértékelés

## Adatbázis(ok)

Az adathalmaz madarak hangfelvételeiből áll .ogg fileformátumban és 2 darab .csv file-ok melyek a hangfelvételekhez és a madárfajokhoz tartalmazznak információkat. A hangfelvételek között vannak címkézett és címkézetlen adatok is. A .csv fileok a hangfelvételek helyszínét és a madárfajok megnevezését tartalmazza.

A projektet szinte kizárólagosan a Google Colab és Google Drive alkalmazásol használatával írtuk.

## Architektúra, tanítás, nehézségek és megoldásuk

Architektúra:

- Konvolúciós neurális hálózat.
- Rétegek felépítése: 2D konvolúciós rétegek, pooling rétegek, dense rétegek az osztályozáshoz.

Tanítás: Az adathalmazt minél hatékonyabban kellett feldolgoznunk, mivel a Google Colab memóriája kevésnek bizonyult nagy mennyiségű adatok egyszerre való tárolásához. A talált mintavételezési megoldások közül számunkra leghatékonyabbnak az tűnt, mikor csak a felvételek első 5 másodpercét használjuk fel tanításra. Ez azért előnyösebb a random mintavételezésnél, illetve a teljes hangfelvétel részekre osztásánál, mert így (az adathalmaz tulajdonságai miatt) jelentősen magasabb lesz az adathalmazban a tisztán kivehető madárfütty és helyes címke párosítás.

Ezután a maradék adathalmazt felosztottuk tanuló, validációs és teszt részre. Checkpoint-ot és EarlyStopping-ot is megvalósítottunk.

(A modell teljes architektúrája egyelőre nem biztos, így a draft verzióba nem kerül bele.)

Az első nehézségünk a hangfájlok átalakítása volt feldolgozható formátumra. Itt többválasztási lehetőséget is kiértékelünk: Mel spektrogram, MFCC, Wav2vec. Ezek közül végül a spektrogramos átalakítást választottuk mivel erre jelentősen több referenciát találtunk hasonló feladatok megoldásában, illetve mivel a Kaggle verseny legjobban teljesítő modelljei szinte kivétel nélkül ezt a megoldást használták.

Viszonylag kevésbé bevett megoldás a Wav2vec, de a Colab-os számítási kapacitás miatt ezt sajnos nem engedhetjük meg magunknak.

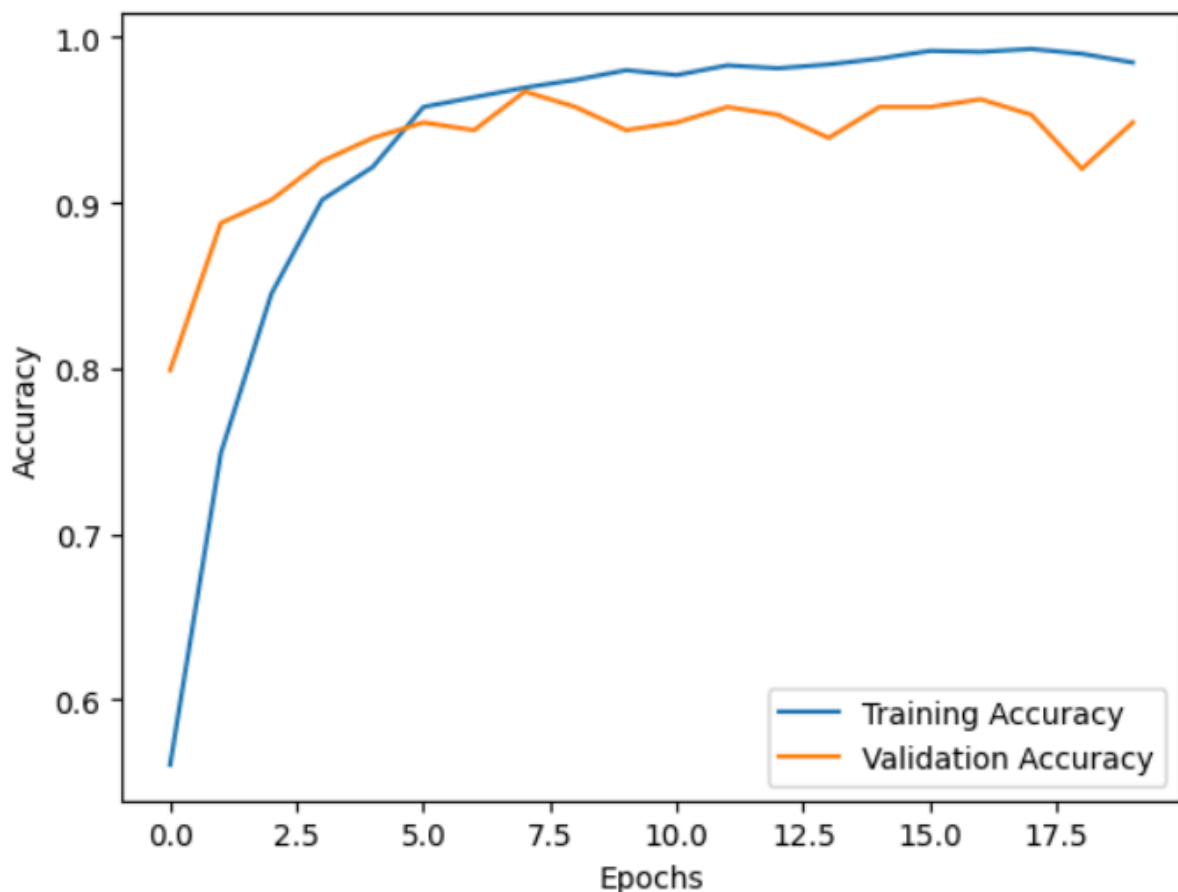
# Eredmények és ezek kiértékelése

A háló pontosságának meghatározására a kereszt-entrópia módszerét használjuk. Ezen kívül eddig a Starfield K-fold Cross Validation tűnik a leghatékonyabb hibamértéknek, de ez további tesztelést igényel.

Egyes fajok nehezebben oszthatóak a hasonló hangzásuk miatt, erre megpróbálunk külön figyelmet fordítani.

## DEMO

A képen a legutolsó sikeres tanításunk futtatásának eredménye szerepel - osztályonként sok adat, de csak 5 osztály. Innen persze a teljes feladat 182 osztályától már csak egy kis munka, és annál jelentősen több számítási kapacitás és futási idő választ el.



## Összefoglaló

- A projekt bemutatja a deep learning hatékonyságát a bioakusztikai elemzésben.
- Egy robusztus osztályozási modell és kifejlesztése.
- Jövőbeli tervek: Az adathalmaz diverzitásának bővítése, valós idejű osztályozási képességek beépítése, és akár további alkalmazások felfedezése az ornitológián túl.