BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM

GÉPÉSZMÉRNÖKI KAR

MECHATRONIKA, OPTIKA ÉS GÉPÉSZETI INFORMATIKA TANSZÉK

NÉMETH ÁRON IMRE

Tudományos Diákköri Konferencia

ANTROPOMETRIAI JELLEMZŐK BECSLÉSE JÁRÁSHANG ALAPJÁN

Konzulens:

Rácz Kristóf

PhD hallgató

Budapest, 2023

Lezárás dátuma: november 05.

Köszönetnyilvánítás

Szeretném megköszönni konzulensemnek, Rácz Kristófnak a rengeteg segítséget, amit a dolgozat elkészítésében és vezetésében nyújtott. Hálás vagyok a folyamatos odaadó és precíz munkájáért, ami nélkül ez a dolgozat nem jöhetett volna létre.

Budapest, 2023

Németh Áron Imre

Tartalomjegyzék

Ábrák jegyzéke

Táblázatok jegyzéke

Kódrészletek jegyzéke

1. Bevezetés

1.1. Célkitűzések, motiváció

Napjainkban egyre nagyobb figyelmet fordítunk arra, hogy privát adatainkat biztonságosan tároljuk, valamint azok ne kerüljenek illetéktelen kezekbe. A klasszikus jelszóval való védekezés egyre inkább elavult, hiszen a jelszó birtokában bárki bárhonnan hozzá férhet a személyes adatainkhoz. Ezért terjedtek el a kétlépcsős azonosítási módszerek, amelyeknél a bejelentkezési folyamat során a jelszó mellett egy második, egyedi kód megadására is szükség van, ami általában egy SMS-ben érkezik a felhasználó telefonjára.

További elterjedt azonosítási módszerek a biometrikus azonosítások, amelyeknél a felhasználó ujjlenyomata, arca, vagy akár a szemének írisze alapján történik azonosítás. A különböző mobilgyártók előszeretettel alkalmazták korábban az ujjlenyomatalapú azonosítást a mobil készülék feloldására. Ezt a módszert az utóbbi években felváltotta az arcalapú azonosítás, amely biztonságosabb [1] és használata kevésbé körülményes a felhasználó számára. A hangalapú felismerés és azonosítás és egyre nagyobb teret nyer napjainkban [2], viszont a hangszintetizáló MI-modellek miatt jogosan merül fel kétség a biztonságosságukkal kapcsolatban.

Általában az emberek az arcuk alapján ismerik fel egymást. Viszont van egy olyan bionikai jellemző, amely alapján még vizuális információ nélkül is felismerhetünk egy adott személyt, ráadásul mesterséges úton utánozni is kifejezetten körülményes. Felmerül a kérdés, hogy vajon melyik jellemző az? Kézen fekvő válasz lenne erre az illető beszédhangja, viszont ez a jellemző nem más, mint az egyén járáshangja és járásképe. Mindenki tapasztalhatta már, hogy egy ismerősét, barátját, vagy akár egy rokonát már anélkül felismerte volna, hogy odanézett volna, csupán az illető a járáshangja alapján.

Felvetődik a kérdés, egyha a járáshang a személyazonosításhoz elég információt hordoz, milyen egyéb tényezőket lehet megállapítani belőle analitikus, vagy akár gépi tanulásra alapozott rendszerekkel is? A járáshang alapú személyazonosítással már több kutatás is foglalkozott [3, 4], még bőven a neurális hálózatok és a gépi tanulás térhódítása előtt. Azonban a gépi tanulás és legfőképpen a deep learning széles körben való elterjedése miatt újabb lehetőségek nyíltak meg a járáshang alapú személyazonosítás és jellemző kinyerés terén [5, 6].

A különböző szolgáltatások igénybevétele előtti ellenőrzés mellett több alkalmazási terület is elképzelhető, mint például a bűnügyi szakértői vizsgálatok, ahol a hangfelvételek és esetleges kamera felvételek segíthetnék a tettes azonosítását. A felvételek alapján antropometriai jellemzők kinyerése

segíthetne a potenciális elkövetők listáját szűkíteni, pontosítani. Az arccal ellentétben a járási stílust nem lehetséges ”eltakarni”, megmásítása pedig rengeteg erőfeszítést igényel. Az elkövető személyek a bűncselekmény során olyan stresszhelyzetben vannak, hogy nagyon nehéz lenne a járásuk megmásítására is odafigyelni.

A fentiek mellett nem nehéz elképzelni, hogy a jövőben a járáshangjából kinyerhető mennyiségekből akár az adott személy egészségügyi állapotára is következtetni lehet majd. Akár preventív jelleggel, akár egy már kialakult betegség diagnosztizálására is alkalmas lehet egy gépi tanulással társított járáshang alapú diagnosztikai módszer a jövőben. Jelenleg megválaszolatlan kérdés, hogy a járáshang alapján lehetséges-e a személy alapvető antropometriai jellemzőire következtetni, azokat megbízhatóan megbecsülni.

Ezen dolgozat célja a járáshangból kinyerhető antropometriai jellemzők vizsgálata, azaz három kiválasztott jellemző (testmagasság, testtömeg, életkor) becslése. A mérési adatok gyűjtése nem képezte a munka részét, azok nyers formában rendelkezésre álltak. A dolgozat során a járáshangokat tartalmazó felvételek megfelelő előfeldolgozása után konvolúciós neurális hálót (CNN) alkalmazva végeztem becsléseket a hangfelvételen szereplő személyek jellemzőire.

1.2. Áttekintés

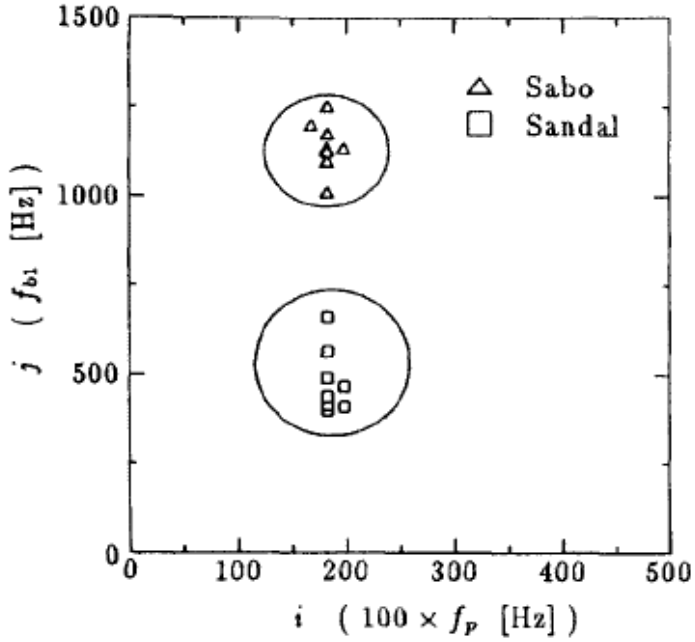
A dolgozat során először azon korábbi tudományos publikációkban megjelent eredményeket foglalom össze, amelyek jelentős elméleti és gyakorlati hátteret biztosítanak a járáshang alapú személyfelismerés és jellemző kinyerés témakörében. Az irodalomkutatás után kifejtésre kerül a rendelkezésre álló hangfelvételek előfeldolgozási eljárása, amely során a felvételek feldarabolásra kerülnek egyetlen lépést tartalmazó szegmentumokra. A szegmentumok Mel-spektogram (a hang vizuális reprezentációja) formájában kerülnek átalakításra, hogy alkalmasak legyenek a konvolúciós neurális hálózat bemeneteként való alkalmazásra. Részletezésre kerül az alkalmazott neurális hálózat felépítése elméleti és gyakorlati hátterével. Ez után a kapott eredmények bemutatása és értékelése következik a különböző eltérések lehetséges okaival együtt, majd a dolgozat összefoglalással és kitekintéssel zárul.

2. Szakirodalmi áttekintés/ előzmények

2.1. Gépi tanulás nélküli hangalapú személyfelismerés

2.1.1. Frekvenciatartomány alapú személyfelismerés

Már jóval a neurális hálózatok és gépi tanulás széles körű elterjedése előtt akadt a járáshangból kinyerhető információkkal kapcsolatos kutatás, amely a lépések által keltett frekvenciák elemzésével foglalkozott [7]. Egy 1999-es japán kutatás két frekvencia jellegű jellemző alapján képes volt két lábbeli alapján létrehozott klaszterbe sorolni a felvételeket. [3]. A járáshangokat két mikrofon segítségével sztereó formátumban vették fel, oly módon, hogy a mikrofonok 15 $[\mathrm{cm}]$ távolságra helyezkedtek el a talaj felett. Minden résztvevő személyről mindkét lábbeliben készültek felvételek, melyek során egy zárt ajtójú szoba egyik végétől a másikig kellett végigsétálniuk a résztvevőknek szandálban és saboban (tradicionális japán lábbeli). A mikrofonok jeleit 20 $[\mathrm{kHz}]$ mintavételezési frekvencián és 12 bit-es A/D konverzió mellett rögzítették, feldolgozását pedig C nyelven írt programmal végezték. A felvételeket két frekvencia alapú értékkel jellemezték, melyek a lépés hangmagassága (f_p) és az első csúcs frekvenciája (f_{bi}) volt. A lépés hangmagassága a felvétel legalacsonyabb frekvenciájú összetevője, az első csúcs frekvencia pedig a hullámforma első csúcsára utal, ami általában a legdominánsabb (legnagyobb) is. A lábbeli klaszterek kialakításához 15 felvételt használtak fel, amelyeket a már ismertetett frekvencia értékek alapján egy koordináta rendszerben ábrázoltak, majd meghatározták az adatpontok alapján a klaszterek középpontját és annak átmérőjét. A középpontokat az adott kategóriába tartozó adatpontok átlagaként határozták meg, a körök átmérőit pedig a pontok Euklidészi távolságainak átlagaiként. Majd a klaszterek létrehozáshoz nem felhasznált felvételekből képzett pontok bekategorizálása következett. A klaszterekbe eső pontokat 83 %-ban sikerült helyesen besorolni csupán kettő frekvencia érték alapján. A koordináta rendszer és a létrehozott klaszterek a 2-1. ábrán láthatók.



2-1. ábra. Szandálban és Sabo-ban történt járások klaszterezése hangfelvételek alapján, ahol f_p a felvételen rögzített lépések hangmagassága, f_{bi} pedig az első csúcs frekvenciája [3]

2.1.2. Rejtett Markov modell alapú személyfelismerés

A személyfelismerés témakörében volt már kutatás a rejtett Markov modell alkalmazásával (HMM - Hidden Markov Modell), igaz ez kutatás leginkább a járáskép alapján való személyazonosításra fókuszált a hang alapú személyazonosítás helyett [8]. Az első csupán hangalapú személyazonosításra fókuszáló publikáció 2014-ben jelent meg, amelyben HMM-et alkalmaztak személyazonosításra [4].

A rejtett Markov-modellek szekvencia modellek. Ez azt jelenti, hogy a HMM egy bemeneti sorozatot, például szavakat adva, meghatározza a bemeneti sorozatokhoz tartozó valószínűségeket. A HMM egy gráf, ahol a csomópontok valószínűségi eloszlások a címkék felett, az élek pedig az egyik csomópontból a másikba való átmenet valószínűségét adják meg. Ezekből együttesen a címkesorozat valószínűségének felhasználásával következtethetünk a bemeneti szekvenciára. [9].

A már említett 2014-es publikációt [4] a Technische Universität München kutatói írták, valamint a járáshangokat tartalmazó adatbázist is ott rögzítették. Összesen 305 személyről készült sztereó hangfelvétel a TUM GAID nevű adatbázisba. A hangfelvételek hossza átlagosan csupán 2 és 3 másodperc között volt, a résztvevők egy $3,5$ méter hosszú folyosón haladtak végig a mérés során. Minden résztvevő összesen háromszor haladt végig a folyosón, egyszer mezítláb, egyszer mezítláb hátizsákkal és egyszer cipőben. A mért adatokat mel-spektogramokon ábrázolták, valamint a lépések modellezésére

ciklikus HMM-eket alkalmaztak az egyes járásciklusokra.

A vizsgálat során azonosították a járáshangok analízisének lehetőségeit egy olyan rendszerben, ahol a lendítő és a támaszfázisokat ekvivalensnek tekintették. Bár tudatában voltak annak, hogy személyspecifikus aszimmetriák is fennállhatnak [10], a kutatást úgy tervezték, hogy egy lépésre összpontosítson, amit egy elkülönített HMM segítségével vizsgáltak. A modell tanításához 150 résztvevő mérési adatait használták fel, a teszteléshez pedig 155 résztvevő adatait. A lépéseket tartalmazó felvételeket MFCC (Mel-frekvencia kepsztrális együttható) formátumba alakították, hogy alkalmasak a legyen a HMM modell bemeneteinek. A felépített modell a beszédfelismerésre használt rejtett Markov modellekre [11] hasonlított, ahol az egyes személyek voltak a szavak, a lépéseik pedig a fonémák. A modellek tanítása során a lépések száma előre rögzített volt, amit egyszerű videófelvételek alapján határoztak meg.

Ezen elveket alkalmazva $41,2$ %-os pontosságot értek el átlagosan a három különböző járásmódot figyelembe véve. Érdekes módon a cipőben végzett mérések során a pontosság csak $9,3$ %-os volt, míg a mezítlábas mérések esetén $69,7$ %-ot értek el. Ezek alapján kijelenthető, hogy a mérési körülményeknek is jelentős hatása lehet a járáshangok felismerésére. Fontos megjegyezni továbbá, hogy egy adott személyről nagyon kevés adat állt rendelkezésre a mérés rövidsége végett, így ez is negatív hatással volt a HMM modell pontosságára.

2.2. Gépi tanulás alapú járáselemzés

2.2.1. Gépi tanulás alapú személyfelismerés

A gépi tanulás térnyerése végett a hangok elemzésével is egyre több kutatás foglalkozik mély tanulás segítségével. A mély tanulás (Deep Learning) olyan gépi tanulási ágazat, amely komplex, adatalapú feladatokat megoldására specializálódott, ahol hagyományos algoritmusok nehézségekbe ütköznének. A mély tanulás olyan neurális hálózatokon alapul, amelyek több rejtett rétegből és nagyszámú neuronokból állnak. Ezek a hálózatok képesek hierarchikus szinteken keresztül reprezentálni és feldolgozni az információkat [12]. Egyik legfontosabb jellemzőjük, hogy a mély tanuló modellek önállóan is képesek megtanulni az adatokban rejlő összefüggéseket, ami lehetővé teszi számukra, hogy magas szintű absztrakt jellemzőket is kinyerjenek a bemeneti adatokból. Ezáltal a mély tanulás különösen hatékony olyan területeken, ahol a komplexitás vagy a sokféleség miatt nehezebb lenne kézzel kifejleszteni a szükséges jellemzőket vagy algoritmusokat.

Konvolúciós neurális hálózatokat (CNN – Convolutional Neural Networks) manapság gyakran al-

kalmazzák audió klasszifikációra [13, 14], pl. zenék műfaji besorolására [15] és beszédfelismerésre. Egy 2017-es kínai kutatás a járáshang alapú személyfelismeréssel foglalkozott, amelyben az adatok gyűjtését és feldolgozását, valamint a modell megépítését is ők végezték [16]. A felvételeket 96 $[\mathrm{kHz}]$ mintavételezéssel és 16 bit-es A/D konverzióval rögzítették egy mikrofon segítségével. A mikrofon 21″ (53,34 $[\mathrm{cm}]$) távolságra volt a talajtól és a folyosó közepén helyezkedett el. A mérés során a résztvevők a folyosó egyik végétől a másikig mentek el, így az első szakaszban közeledve, a második szakaszban távolodva a mikrofontól. A résztvevőknek négy fajta módon kellett végig menniük a folyosón, gyalogolva, kocogva, futva és ugrálva. A mérésben 50 személy vett részt, akikről fejenként 75 darab 10 - 15 másodperc közötti hangfelvétel készült.

A felvételeket demodulálták és aluláteresztő szűrőt alkalmaztak rajta, így a jelek mintavételezési frekvenciája 8 $[\mathrm{kHz}]$-ra redukálódott. Az előfeldolgozás következő lépéseként a szűrt felvételeket sima spektogramokká alakították, így létrehozva a CNN számára alkalmas bemeneti adatokat. Az alkalmazott neurális háló egy konvolúciós neurális háló, amelyet ők fejlesztettek ki. Összesen két darab, egyenként 100 neuront tartalmazó rejtett réteget alkalmaztak, amelyeket a kimeneti réteg követett. A tanítás során 55 felvételt használtak fel minden személytől, tehát nem volt olyan személy, akiről ne került volna felvétel a tanító adathalmazba.

A tesztelés során a modell 97 %-os pontosságot ért el a résztvevő személyek felismerésében, függetlenül a járás mód fajtájától. A magas pontosság oka lehet, hogy a több fajta járásmód miatt a modell jobban tud általánosítani, viszont a tény, hogy minden személyről volt tanító és teszt adat is növeli az adott adatsorra való rátanulás esélyét.

3. Bemeneti és kimeneti adatok létrehozása

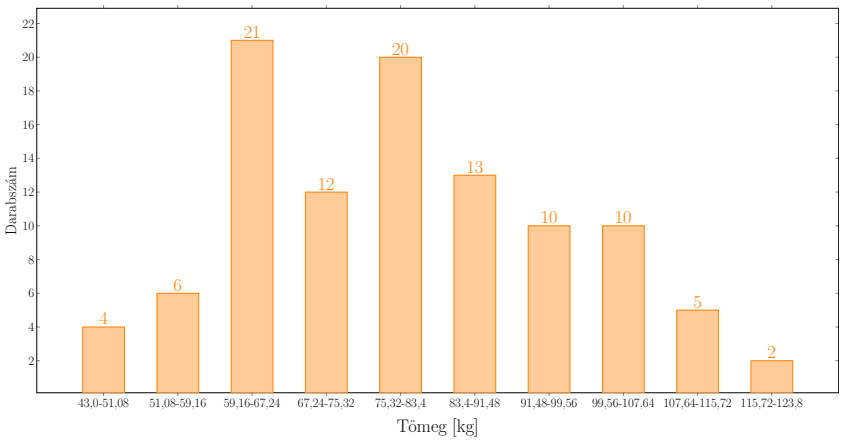
3.1. Az adatok forrása és jellemzői

Jelen dolgozat célkitűzése antropometriai adatok mély neurális hálós becslése futópados járás közben készült hangfelvételek alapján. A felvételek készítése nem képezte a projekt részőt, azok a Szolnoki MÁV Kórház és Rendelőintézet Mozgáslaborjában,2023 nyarán készültek. A mérések során a résztvevőknek egy futópadon kellett sétálniuk mezítláb, a járáshangokat pedig két mikrofon segítségével, sztereóban rögzítették. A felvételek 192 [kHz] mintavételezés és 24 bites felbontással készültek és *.wav* formátumban álltak rendelkezésre összesen 103 különböző személyről.

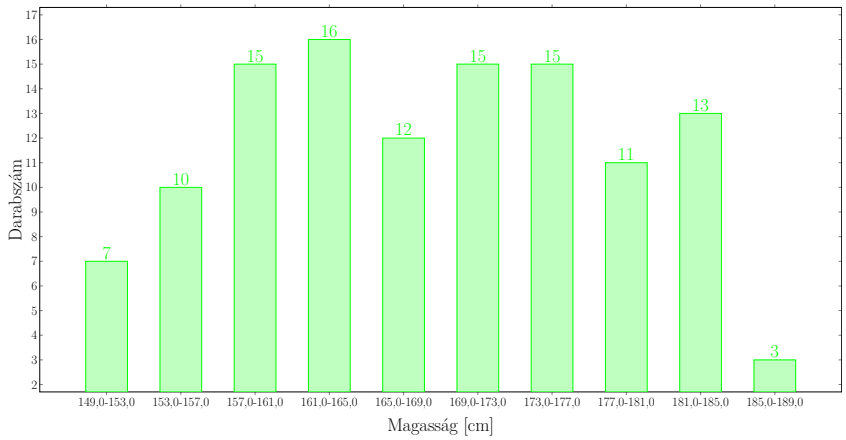
Ez minden személy esetén egy darab, nagyjából 2 perces hangfelvételt jelentett. Emellett elérhetőek voltak minden résztvevő antropometriai jellemzői is táblázatos formában, anonim módon, egyedi azonosítóhoz társítva. A legtöbb jellemző önbevallás alapján lett rögzítve, azonban a testmagasság mérése a helyszínen is megtörtént. Ezekből három antropometriai jellemző került felhasználásra, a testtömeg, a testmagassaság és az életkor. A nemek aránya 63-40 fő a hölgyek javára. A felhasznált jellemzők statisztikai leírása a 3-1. táblázatban láthatók. Az adatok eloszlásának szemléltetése hisztogramok formájában a 3-1., 3-2. és 3-3 ábrákon láthatók.

3-1. táblázat. A mérésben résztvevő személyek statisztikai jellemzői

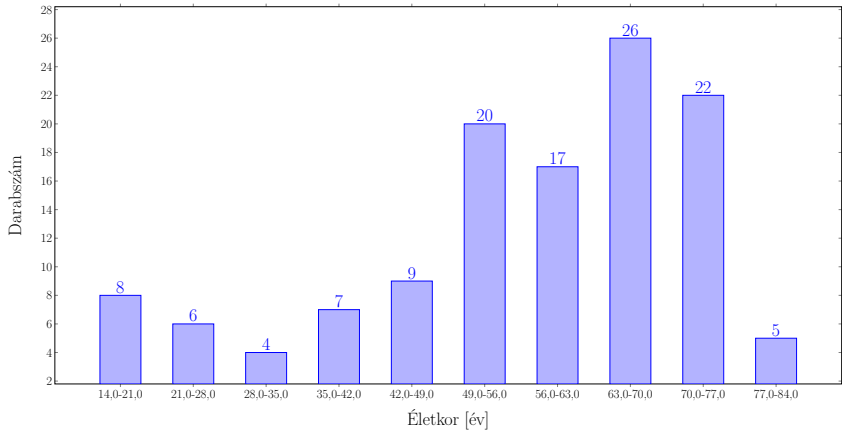
Jellemző	Átlag	Szórás	Medián	Minimum	Maximum
Testtömeg [kg]	79,56	17,55	77,60	43,00	123,80
Testmagasság [cm]	167,69	9,97	168,00	149,00	189,00
Életkor [év]	55,90	17,36	58,00	14,00	84,00



3-1. ábra. A mérésben résztvevő személyek tömegének eloszlása



3-2. ábra. A mérésben résztvevő személyek magasságának eloszlása



3-3. ábra. A mérésben résztvevő személyek életkorának eloszlása

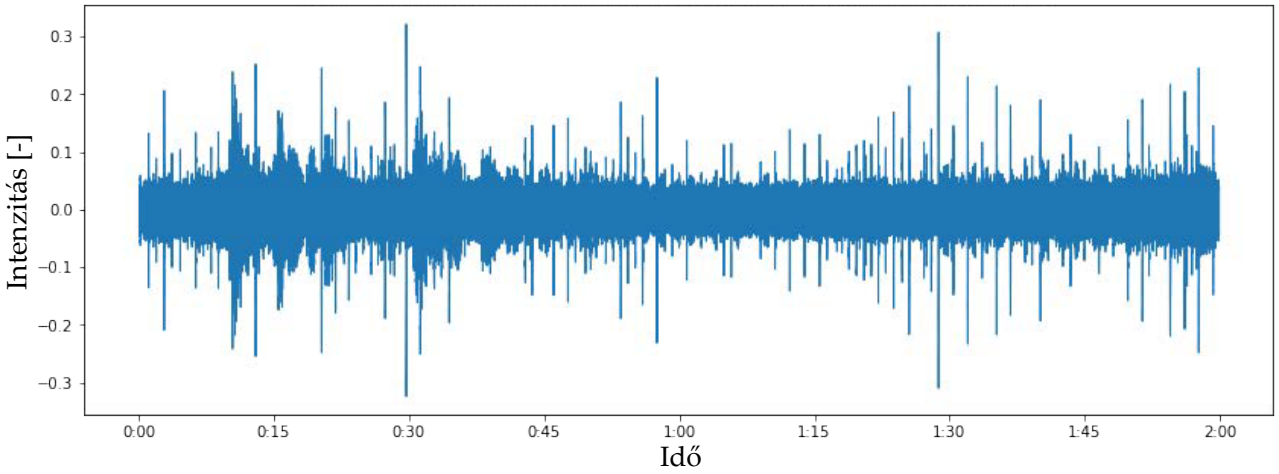
3.2. A hangfelvételek előfeldolgozása

A hangfelvételek előfeldolgozása *Python* programozási nyelvben történt, több nyilvános könyvtár felhasználásával, melyek közül az audió fájlok feldolgozására széles körben erterjedt *Librosa* könyvtárat érdemes kiemelni. Az előfeldolgozás lépései biztosítják az információtartalom megfelelő kiemelése mellett a konvolúciós neurális háló (CNN) bemenete által megkövetelt állandó méretet is.

Az előfeldolgozási lépések részletes leírása a következőkben lesz olvasható. A következő alfejezetekben szereplő egyes függvények saját, a feldolgozó algoritmus egy-egy hosszabb alfeladatát megvalósító programrészek. Az átláthatóság végett a bonyolultabb műveletek hosszabb magyarázattal rendelkeznek, helyenként a lényeges algoritmusokat tartalmazó kód részletek segítségével.

3.2.1. A hangfelvételek betöltése a program számára

A járáshangokat tartalmazó *.wav* fájlok betöltésének teljes folyamatát a saját **open_audio_files** végzi, amely egy előre definiált útvonalról betölti a nyers hangfelvételeket, valamint a felvétel mintavételezési frekvenciáját. Mivel a felvételek sztereó hangfelvételek, ezért egy fájlhoz két hullámforma is tartozik. A több tanító adat elérése érdekében mindkét adatsort elmentettem a betöltés során. Az így kapott adatokat egy tömbben tárolódnak, amelynek a mérete a betöltött fájlok számától függ. Egy betöltött hullámforma ábrázolása az 3-4. ábrán látható.



3-4. ábra. Egy rögzített járáshangfelvétel hullámformája

3.2.2. A felvételek feldarabolása, a lépésszegmentumok kinyerése

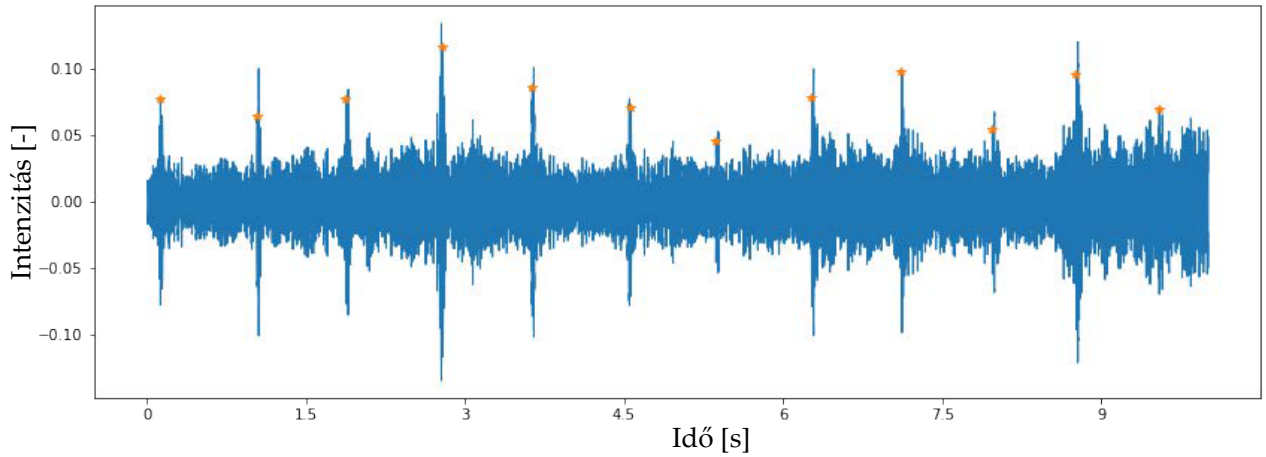
A hangfelvételek betöltése után a következő lépés azok lépésekre való felbontása. A szegmentumok létrehozásának célja, hogy a konvolúciós neurális háló számára megfelelő mennyiségű és méretű bemeneti adat álljon rendelkezésre, hogy az hatékonyan tudjon tanulni. A mérési felvételeket úgy érdemes felosztani, hogy az egyes sarokütések, amik a legtöbb akusztikus információt hordozzák, különkülön szegmentumokba kerüljenek. A 3-4. ábrán látható intenzitás csúcsok a sarokütések helyét jelölik, amelyek segítségével lehetséges az egyes szegmentumok kezdetének és végének meghatározása. Egy szegmentum úgy került meghatározásra, hogy egy sarokütést tartalmazzon, ami a szegmentum teljes hosszának 40%-ánál helyezkedik el (így megfelelő a szeparáció egy adott sarokütés utáni lecsengés és a következő sarokütés között). A korábbi szempontokat figyelembe véve a hangfelvételek feldarabolását és a lépésszegmentumok kinyerését végző algoritmus a **create_segments** függvénybe lettek csoportosítva. A függvény az előfeldolgozási algoritmus egy része, amely a betöltés során kinyert adattömböket használja fel.

A függvény először betölti a már kinyert tömb formátumban tárolt hullámformákat, majd csúcskeresést hajt végre rajtuk az egyes sarokütések lokációjának meghatározásához. Mivel a felvételeket jelentős zaj terhelte, ezért elengedhetetlen volt a csúcskeresési eljárás megfelelő paraméterezése. A csúcskeresést a *scipy.signal* könyvtár **find_peaks** metódusa valósítja meg, amely a megadott paraméterek alapján szelektál az összes lokális maximum közül (3-1 kódrészlet).

```
# Variables for peak finding
peak_height=0.045
peak_distance=sr/2
# Find peaks function
data_peaks, _ = find_peaks(x, height=peak_height, distance=peak_distance)
```

3-1. kódrészlet. A csúcskereső algoritmus megvalósítása és paraméterezése

A csúcskeresés során a **distance** paraméter értékét a mintavételezési frekvencia feleként határoztam meg, hogy yen (a distance paraméter elemszámban kéri a két csúcs közötti minimális távolságot, agy a mintavételezési frekvencia egy másodpercnek felel meg), ezzel kiküszöbölve a túl rövid, hibás szegmentumok létrehozását. A paraméterezés alapján az algoritmus az 3-5 ábrán látható módon jelölte ki a lokális maximumokat.



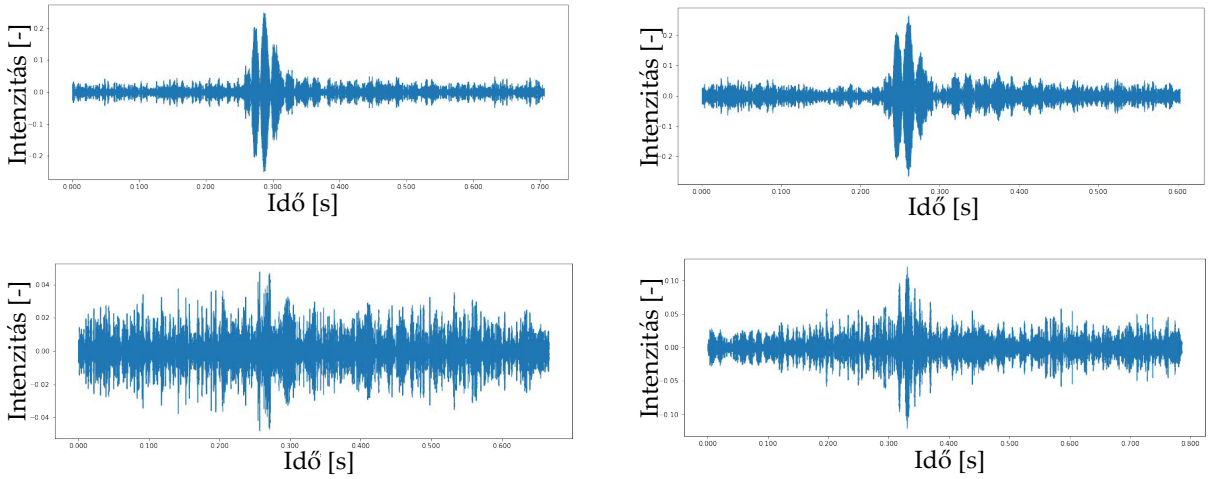
3-5. ábra. Csúcskereséssel azonosított sarokütések egy 10 másodperces felvételszakaszon

A csúcskeresés során a függvény a csúcsok indexeit egy tömbben adja vissza, amelyek alapján szegmentálhatók a felvételek. az egyes felvételek. A szegmentumok elválasztásának pontja a két sarokütések közötti részt 40-60 %-os felosztásával lett meghatározva (3-2. kódrészlet).

```
# Set start and end of the segment
segment_rate = 0.6 # 40%-60% of the cycle
start = int((peak_locs[i]-peak_locs[i-1])*segment_rate)+peak_locs[i-1]
end = int((peak_locs[i+1]-peak_locs[i])*segment_rate)+peak_locs[i]
# Extract the segment
segment = amplitude_array[start:end]
```

3-2. kódrészlet. Hangfelvételek szegmentálása sarokütések helye alapján

A hibásan szegmentált ciklusok kiszűréséhez kiszámításra kerül az azonos felvételből létrehozott szegmentumok átlagos hossza és szórása. Azon elemek, amelyek hossza az átlag ± 3* szórás tartományon kívül esik kiugró értékeknek tekinthetők, ezért eltávolításra kerültek. A 3-6. ábrán látható pár, szegmentált sarokütés.



3-6. ábra. Szegmentált sarokütések hullámformája négy különböző felvételből

3.2.3. Mel spektogramok létrehozása a szegmentumokból

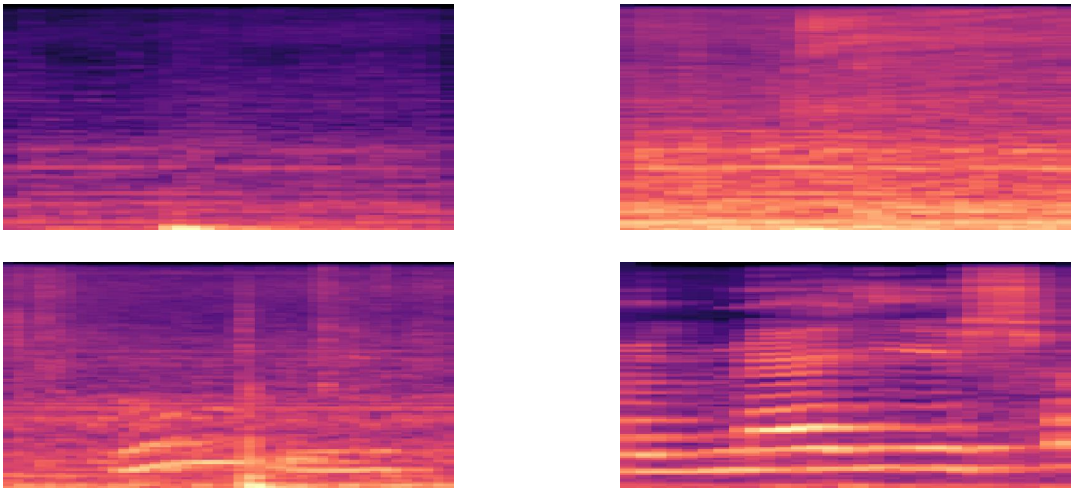
A korábban létrehozott lépésszegmentumok jelenlegi formátumban még nem alkalmasak a konvolúciós neurális háló tanítására, ezért olyan eljárást kell végrehajtani, ami a hangfelvétel részleteket információ vesztés nélkül átalakítja egy olyan formátumba, amely már felhasználható bemenetként. A hangnak több vizuális reprezentációja is létezik, ezek közül az egyik legelterjedtebb a spektrogram, amely az adott felvétel frekvencia tartománybeli, időben változó jellemzőit ábrázolja. A szakirodalomban a hangfeldolgozás témakörében manapság egyre elterjedtebb a Mel spektrogramok használata, még köhögések elemzésében is alkalmazták őket [17]. Ezek a normál spektrogramokhoz hasonlóan a

frekvenciatartománybéli jellemzőket ábrázolják, viszont a frekvencia tengely a szokásos logaritmikus Decibel skála helyett a Mel-skálát követi, amely az emberi hallás érzékenységét modellezi. A neurális hálók terén pedig hangalapú személyfelismerés téma körében átlagosan jobb eredményt érnek el, mint a hagyományos spektrogramok [18]. A Mel spektrogramok létrehozását és azok mentését a saját fejlesztésű **create_mel_spectrogram** nevű függvény végzi, a megfelelő paraméterek beállítása után. A függvény a Mel spektogramok létrehozásához a *Librosa* könyvtár egy beépített függvényét használja, majd az egyes frekvenciakomponensek intenzitását értékeit teljesítményről Decibelbe konvertálja, hogy az alacsonyabb intenzitású komponensek jobban megfigyelhetők legyenek. Ennek menete a 3-3. kódrészletben látható, ahol i és j a ciklusváltozók:

```
    # Get current segment
    current_segment = np.array(segments_tensor[i][j]).flatten()
    mel_spectogram = librosa.feature.melspectrogram(y=current_segment, sr=sr[i])
    # Convert to log scale (dB)
    mel_spectogram_dB = librosa.power_to_db(mel_spectogram, ref=np.max)
```

3-3. kódrészlet. A Mel spektrogramok létrehozása

Az így létrehozott Mel-spektogramok a felvételszegmentumok időbeli hosszától függetlenül fix pixelszámmal kerülnek mentésre. Nyilvánvalóan egy hosszabb részlet esetén több információt hordoz az adott ábra, viszont a bemenetek állandó mérete szükséges a neurális hálók használatakor. A létrehozott képek neve tartalmazza a hangfájl egyedi azonosítóját, valamint azon szegmentum sorszámát, amelyet a Mel spektogram reprezentál. Pár a függvény által létrehozott Mel spektrogram látható az 3-7 ábrán. A felvételekből összesen 6649, tanításra felhasználható szegmentum jött létre.



3-7. ábra. Különböző szegmentumokhoz tartozó Mel spektrogramok

3.2.4. Tanító, validációs és teszt bemeneti adathalmazok létrehozása

Következő lépésként a létrehozott Mel-spektrogramokat fel kell osztani tanító, validációs és teszt adathalmazokra, amelyet **train_val_test_splitter** nevű függvény végez. A véletlenszerű felosztás egy *seed* segítségével lett reprodukálhatóvá téve. A felosztás az összes bemenetet tartalmazó adathalmaz szintjén történt a felvételhez tartozó résztvevőre való tekintet nélkül. A létrehozott három adathalmaz méretei a teljes adathalmazhoz képest az alábbi módon oszlik meg:

- Tanító adathalmaz: 70% (4654 adatpont)
- Validációs adathalmaz: 15% (997 adatpont)
- Teszt adathalmaz: 15% (998 adatpont)

Az így létrehozott adathalmazok külön könyvtárakba kerültek mentésre, ahonnan a **data_generator** nevű saját fejlesztésű eljárás a neurális háló bemenetének megfelelő formátumban tudja majd betölteni azokat. A betöltött spektogramokat tartalmazó képeket a *Keras* könyvtárban szereplő **img_to_array** függvény transzformálja nyers RGB pixel értékeket tartalmazó tömbökké, nyers adattömbbé konvertálja, majd azok értékét és 0 és 1 közé skálázza. Így minden spektogramból egy háromdimenziós tömb jön létre, ahol az egyik dimenzió a három színcsatornát, míg a másik kettő az adott pixel képen belüli pozícióját reprezentálja. A fenti folyamat a 3-4 kódrészletben látható.

```
def data_generator(dir, target_size=(299, 299)):
    # Get a list of all files in the directory
    paths = [os.path.join(dir, mel) for mel in os.listdir(dir) if mel.endswith('.jpg')]
    # Load and preprocess all images in the directory
    X_data = []
    for mel_path in paths:
        # Load image and resize
        img = load_img(file_path, target_size=target_size)
        # Convert image to array and normalize pixel values
        img_array = img_to_array(img) / 255.0
        X_data.append(img_array)
    return np.array(X_data)
```

3-4. kódrészlet. Mel-spektrogramok betöltése és transzformálása a neurális háló bemenetének megfelelő alakra

3.3. Az antropometriai adatok előfeldolgozása

Míg a felvételekből a bemeneti adatokat, addig az antropometriai jellemzőket tartalmazó táblázatból az elvárt kimeneti értékeket kell létrehozni és azokat a bemeneti adatokkal párosítani. Egy adott felvételhez több szegmentum tartozik, ezért az adott személyhez tartozó rekordokat az összes felvételrészlethez hozzá kell rendelni. A teljes adathalmazt tartalmazó *Excel* táblázatból csak a kívánt antropometriai jellemzőket tartalmazó *.csv* fájlt exportálva, a *Pandas* könyvtár segítségével egy *DataFrame* objektumként kerülnek betöltésre a rekordok. Az adatok összepárosítása a 3-5. kódrészletben látható módon történik: a soron következő spektrogram elérési útvonalából kinyerésre kerül a megfelelő azonosító, és ez alapján az antropometriai adatok hozzáadásra kerülnek egy másik *DataFrame* objektumhoz új sorként. Ez a DataFrame így a szegmentumoknak megfelelő sorrendben fogja a kimeneti értékeket tartalmazni. Az így kapott értékeket végül a 3-6 kódrészletben látható módon a tanításhoz már felhasználható egyszerű tömbbé konvertáljuk.

```
for i in range(mel_number):
    # Getting the name of the file
    mel_name = os.path.basename(mel_spectograms[i]).strip('.jpg')
    object_name = mel_name[:9]  # First 9 characters of the name (ID)
    # Getting the row of the dataframe
    df_index = df['ID'] == object_name
    df_row = df[df_index]
    df_set = pd.concat([df_set, df_row], ignore_index=True)
```

3-5. kódrészlet. A neurális háló kimeneteként elvárt antropometriai adatok párosítása a bemeneti spektrogramokkal

```
# Column names
column_names = ['BODY MASS [Kg]', 'HEIGHT [cm]', 'AGE']
# Get the columns
Y_data_array=[]
for column in column_names:
    Y_data_array.append(np.array(df_set[column]))
```

3-6. kódrészlet. A kimeneti adatok konvertálása a tanításhoz használható tömb formátumba

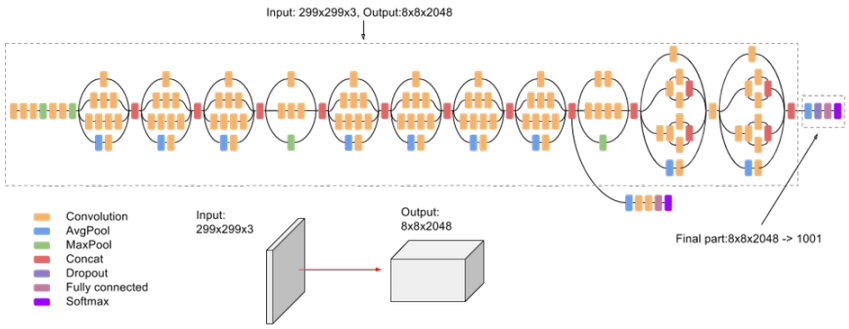
A bemeneti adatok mellett a kimeneti adatokat is standardizálni érdemes, a konvolúciós neurális háló hatékonyabb tanítása érdekében. A standardizálás során a kimeneti adattömb minden változóját

egy nulla középértékű és egységnyi szórású változóvá alakítjuk. Ezen műveletek elvégzése után rendelkezésre állnak a bemeneti és kimeneti adatok, amelyeket a modell tanítására, valamint tesztelésére lesznek felhasználva. A létrehozott adattömbök sorrendisége megegyezik, az azonos indexű elemek ugyanahhoz a mérési felvételszelethez tartoznak.

4. A Deep Learning modell implementációja

4.1. A neurális hálózat architektúrája

A hangfelvételek Mel-spektogramokká alakítása után, egy olyan összetett neurális hálózatra van szükség, ami alkalmas a képekkel reprezentált hangfelvételek információtartalmának kinyerésére, és azok alapján a vonatkozó kimenetek értékeinek becslésére. A modell két fő részből áll, egy a képek feldolgozásáért felelős konvolúciós hálóból, illetve egy jelentősen kisebb, a konvolúciós háló által kinyert jellemzők alapján a kimeneteket becslő szekvenciális hálózatból. A konvolúciós szakasz a Google által fejlesztett InceptionV3 [19] architektúrát alkalmazza, amely eredetileg az ImageNet nevű képfelismerési versenyre készült. A tervezés során fontos szempont volt a paraméterek számának kordában tartása. Az InceptionV3 25 millió paraméterrel rendelkezik szemben a hasonló eredményeket elért konkurens AlexNet 60 millió paraméterével szemben [20]. A háló eredeti elvi felépítése a 4-1. ábrán látható.



4-1. ábra. Az InceptionV3 architektúra felépítése [21]

A teljes háló második, a kimenetek becsléséért felelős szekvenciális rész, aminek bemenete az InceptionV3 modell konvolúciós szakaszának kimenete. A bemeneteket egy *Flatten* réteg alakítja egydimenzióssá, így téve alkalmassá az azt követő két rejtett réteg bemenetének. A kimenet három párhuzamos, egy neuront tartalmazó *Dense* rétegből áll, amelyek egyenként a három becsülni kívánt antropometriai jellemzőkhöz tartoznak. A teljes modell felépítése a 4-1. táblázatban látható.

4-1. táblázat. Az alkalmazott neurális háló felépítése

Réteg neve	Kimeneti méret	Aktivációs függvény	Paraméterek száma	Előző réteg
InceptionV3	(8,8,2048)	-	21802784	-
Flatten	(131072)	-	0	InceptionV3
Dense_128	(128)	LeakyReLu	16777344	Flatten
Dense_32	(32)	LeakyReLu	4128	Dense_128
Dense_mass	(1)	linear	33	Dense_32
Dense_height	(1)	linear	33	Dense_32
Dense_mass	(1)	linear	33	Dense_32

4.2. A tanítási eljárás megtervezése

A háló tanításához szüksége van egy tanító algoritmusra, az optimalizálandó célfüggvény meghatározására, illetve az tanítás során követni kívánt egyéb mérőszámok (metrikák) megadására. Utóbbiak a modell aktuális teljesítményének mérésére szolgálnak, ezzel betekintést engedve a probléma megoldásának hatékonyságába. A cél-/költség-/hibafüggvény a becsült és a valós kimeneti értékek közötti eltérést méri, regressziós modellek esetén általában a négyzetes (MSE) vagy az átlagos abszolút (MAE) hibát használják. A gépi tanulás tanító alogritmusai olyan, általában a gradiens módszerre épülő algoritmusok, amelyek a modell belső paramétereit iteratív módon állítják be úgy, hogy azzal a hibafüggvényt minimalizálják, ezáltal javítva a modell prediktív pontosságát. A hálózat tanításához optimalizáló algoritmusként az *Adam*, költségfüggvényként a *Mean Squared Error* (MSE) és metrikaként a *Mean Absolute Error* (MAE) kerültek megadásra (4-7).

```
# Setting the losses
losses = { "mass": "mean_squared_error",
        "height": "mean_squared_error",
        "age": "mean_squared_error",}
# Compile the model
model.compile(optimizer = Adam(learning_rate=1e-4), loss=losses, metrics=['mae'])
```

4-7. kódrészlet. A neurális haló tanító paramétereinek megadása

Lehetőség van még úgynevezett Callback függvények megadására, amelyek a tanításhoz nem feltétlenül szükséges, de annál hasznosabb funkciókat látnak el, mint a tanítás korai leállítása, túltanulás

jelei esetén, vagy a modell legjobban teljesítő állapotához tartozó súlyok mentése. Ezeket a funkciókat úgy tervezték, hogy figyelemmel kísérjék és kölcsönhatásba lépjenek a betanítási folyamattal anélkül, hogy ténylegesen befolyásolnák a modell paramétereit. Az alkalmazott Callback-ek paraméterei a 4-8. kódrészletben láthatók.

```
# Early stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=50,
                               restore_best_weights=True, verbose=1)
# Checkpointer
checkpointer = ModelCheckpoint(filepath=filepath, monitor='val_loss', verbose=1,
                               save_best_only=True, mode='min')
```

4-8. kódrészlet. A neurális háló tanításához használt Callback függvények beállítása

A megfelelő előkészületek és paraméterek beállítása után a következő lépés a tanítás futtatása. Az első tanítás előtt a súlyok az előtanított, az *imagenet* adatbázison optimalizált modell súlyaiként inicializálódnak, majd a tanítás során folyamatosan mentésre kerültek a legjobb validációs eredményeket elérő értékek. A tanítás Google Colab platformon történt a folyamat erőforrás igényének kielégítése és felgyorsítása érdekében. Az első előtanítás 200 iteráción keresztül futott (nem volt korai leállítás). A többi tanítás mindig az előzőleg legjobb eredményt elérő súlyokkal kezdődött, és összesen ötször lett futtatva. Az utolsó tanítás összesen 54 epochig tartott, a korai leállítás miatt.

4.3. A modell tesztelése, eredmények

4.3.1. A modell statisztikai eredményei

A tanító folyamatok elvégzése közben is kapunk visszajelzést a modell teljesítményéről, azonban a valódi teljesítmény a tesztelésre elkülönített adatok alapján jellemezhető a legjobban. A modell eredményességét reprezentáló statisztikák a 4-2. táblázatban láthatók, ahol a teljesítményt már említett átlagos abszolút hiba (MAE), az átlagos négyzetes hiba (MSE), valamint a becsült és a valódi adatsor közötti korreláció értéke jellemezi. A mérőszámok meghatározása az *sklearn* könyvtár tagfüggvényeivel történt.

4-2. táblázat. A modell teljesítménye a prediktált és valós antropometriai adatok közötti eltérések jellemzésével (MAE – valós és prediktált jellemzők közötti átlagos abszolút hiba, MSE - valós és prediktált jellemzők közötti átlagos négyzetes hiba, R^2 valós és prediktált jellemzők közötti korrelációs együttható)

Antropometriai jellemző	MAE	MSE	R^2
Testtömeg $[\mathrm{kg}]$	0,672	10,226	0,959
Testmagasság $[\mathrm{cm}]$	0,458	2,829	0,966
Életkor $[\mathrm{év}]$	0,358	2,587	0,990

A fenti eredményeket figyelembe véve kijelenthető, hogy a modell jól teljesített az adatok előrejelzésében. Érdekes módon életkor esetén van a legnagyobb korreláció a becsült és a valós adatok között, valamint a legkisebb névleges hibák is az életkor becslésekor adódtak. Egy átlagos becslés esetén az előre jelzett és a valódi életkor közötti különbség csupán 0,358 év, ami kevesebb mind az adatok rögzítésének felbontása. A becslések pontosságát akár az is okozhatja, hogy az életkor egy olyan jellemző, amelyet nem kell közvetlenül mérni, ezért azt mérési hibák nem terhelik.

A testtömeg prediktálása volt a legpontatlanabb, viszont itt is csak 0,672 $[\mathrm{kg}]$-os az átlagos abszolút hiba, ami kevesebb, mint egy ember szokásos egy napon belüli tömegingadozása (1-2 kg). A becslések és a valós adatok közötti korreláció itt is magas. Az előrejelzés pontatlanságában közre játszhat, hogy a mérésben résztvevő személyek testtömege a helyszínen nem lett lemérve, hanem önbevallás alapján kerültek bejegyzésre, amely adatok elavultak és pontatlanok lehetnek, , illetve gyakori az valósnál alacsonyabb tömeg bevallása. A fenti jellemzők közül a tömeg mérése a legpontatlanabb a különböző kisebb mérési hibák miatt.

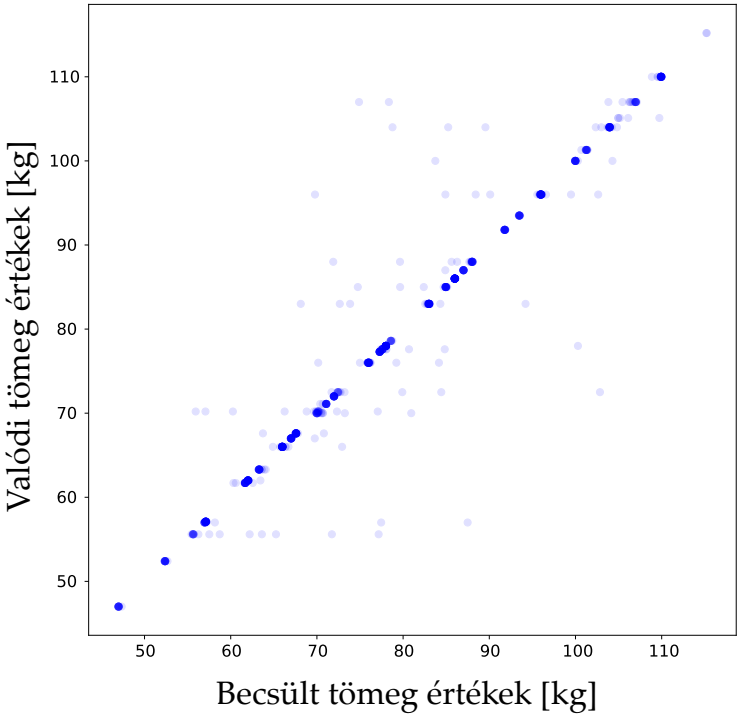
A testmagasság becslése hasonlóan jó eredményeket ért el, mint az életkoré, igaz ezen jellemző szórása volt a legkisebb a vizsgált populációban (lásd 3-2. ábra). A korreláció értéke itt is közel áll az egyhez és a négyzetes hibák sem magasak. A testtömeg mérésével ellentétben a résztvevők magasságát a helyszínen lemérték, így a lejegyzett adatok pontosabbak és kevesebb hibával terheltek.

4.3.2. Az eredmények vizualizációja

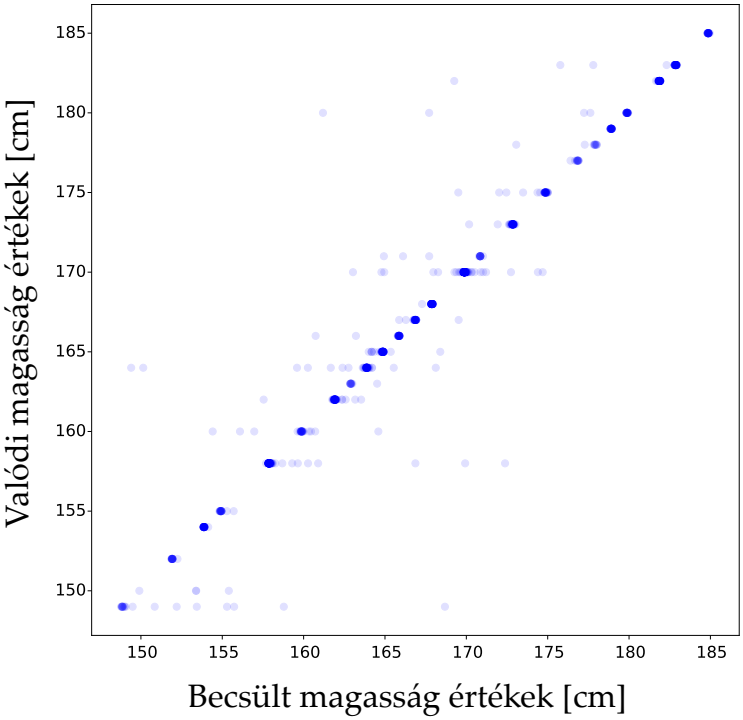
A becslések és a valós adatok összehasonlítása

Az eredmények kiértékelésének egy fontos lépése az eredmények szemléletes ábrázolása, amely segítségével jobban megérthetőek az eltéréseket és a becslések pontosságát befolyásoló tényezőket. A

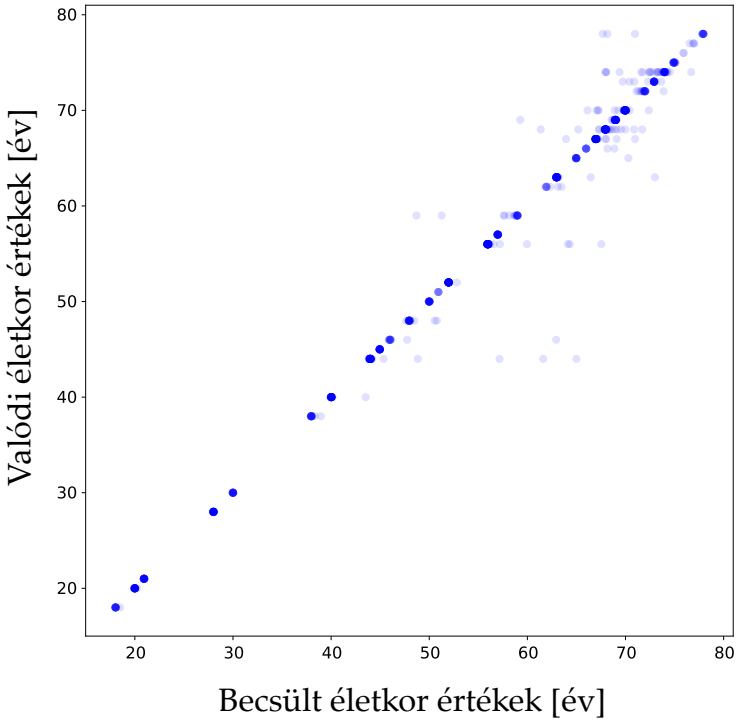
korreláció vizsgálat miatt már tudjuk, hogy a becslések és a valós adatok között minden jellemző esetén lineáris kapcsolat van, de ha ábrázoljuk a valódi értékeket a becsült értékek függvényében, akkor jobb betekintést kapunk a két adatsor közötti kapcsolatra. A kapcsolatokat vizualizáló ábrák a 4-2., 4-3. és a 4-4. ábrákon láthatóak, ahol sötétebb színű pontok láthatóak az ábrán ott több adatpont fedi egymást.



4-2. ábra. A becsült és a valós testtömeg közötti kapcsolat



4-3. ábra. A becsült és a valós testmagasság közötti kapcsolat



4-4. ábra. A becsült és a valós életkor közötti kapcsolat

Kijelenthető, hogy a becsült és valós adatok között mindhárom jellemző esetén lineáris kapcsolat van, azonban vannak adatpontok, amelyek kívül esnek az egyenesen. A túl és az alábecslési hibák az ábrák alapján kiegyenlítettnek nevezhetőek, nem látható, hogy a modell az egyik irányt preferálná.

A tömegbecslés esetén a legtöbb hiba a 60 és 90 $[\mathrm{kg}]$ közötti tömegtartományban figyelhető meg, ahol több az lineáris egyenestől eltérő adatpont. A tömegtartomány szélein viszont az előre jelzett adatok jól követik a valósokat, kevés eltéréssel. Ezt okozhatja az, hogy a mérésben résztvevők többsége a fenti (60 és 90 $[\mathrm{kg}]$ közötti) tömegtartományba esik, így a tartományon kívül eső adatpontok becslése könnyebben kivitelezhető. Egy másik elképzelhető ok, hogy a nagy tömegű személyek járásának jóval karakteresebb, erőteljesebb hangja van, mint az átlagos személyeknek, ezért a modell nagyobb bizonyossággal tudja azonosítani őket. Az alacsonyabb tömegű személyeknél is hasonló a helyzet áll fent, csak itt a hangjuk gyengébb, mint az átlag emiatt könnyebben azonosítani tudja a modell.

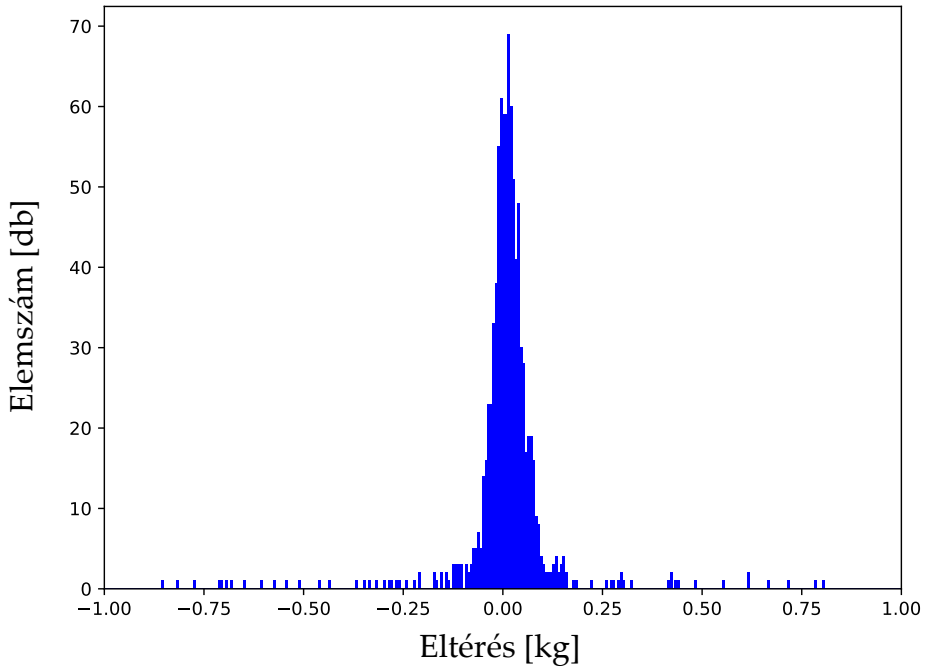
A testmagasság becslés során kevesebb kiugró érték figyelhető meg, mint a tömeg esetén, valamint az egész tartományon belül közel azonos az eltérések mértéke, sőt a tartomány alsó határán a legkisebb magasság esetén volt a legtöbb hibás előrejelzés. A nagyobb pontosságokat okozhatja a már említett helyszíni mérés, vagy akár az is, hogy a mérésben résztvevők magasságának terjedelme és szórása kisebb, mint a tömegé.

Az életkor becsléséről már a statisztikai eredményeknél is láthattuk, hogy a legpontosabb, azonban az adatpontok eloszlásáról nem adott pontos képet. Érdekes módon a legtöbb egyenestől eltérő pont az idősebb korosztály esetén megfigyelhető, viszont nagyon kiugró eltérések nem figyelhetőek meg. Az 3-3. ábrán látható, hogy az idősebb korosztály nagyobb számban képviseltette magát a mérésben, a legtöbb résztvevő 65 és 74 év közötti volt, ami szintén okozhatja a sok kis eltérést a prediktálás során. Az idősebbek személyek egészségügyi állapotában általában nagyobb a szórás, mint a fiatalabbaknál. Aktivitástól és életmódtól függően, akár egy 80 éves személy járása is lehet olyan, mint egy kevésbé egészséges 60 évesé, ami szintén okozhat eltérést a becsült és valós életkor között. Ezek alapján feltételezhető, hogy a modell inkább a biológiai életkor (hány éves átlagos embernek felel meg az egészségügyi állapot), mintsem a tényleges kronológiai életkort (születés óta eltelt évek) becsli.

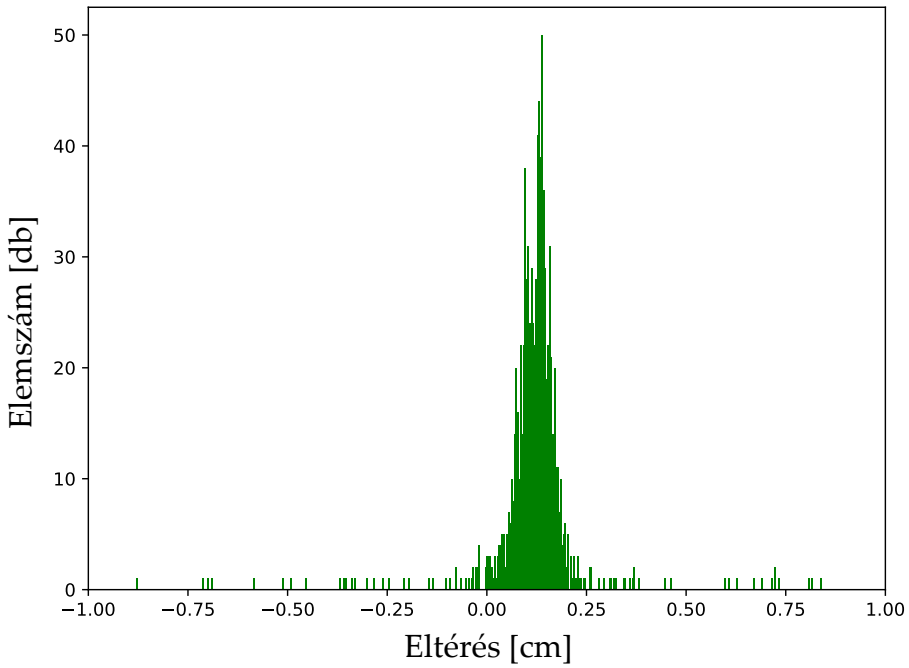
A becslések hibáinak vizualizációja hisztogramokon

A hibák, vagyis a tényleges és becsült paraméterek közötti eltérés értékelésére célszerű a különbségeket külön ábrázolni. A hibák értékét a teszt és a becsült adatok különbsége adja meg, amely eloszlását a 4-5., 4-6. és 4-7 ábrák mutatják, ahol a hibák eloszlása a becslések pontosságát jellemzi. A hisztogramokon ± 1 tartomány került részletes ábrázolásra, mivel a hibák nagy része ebben a tartományban

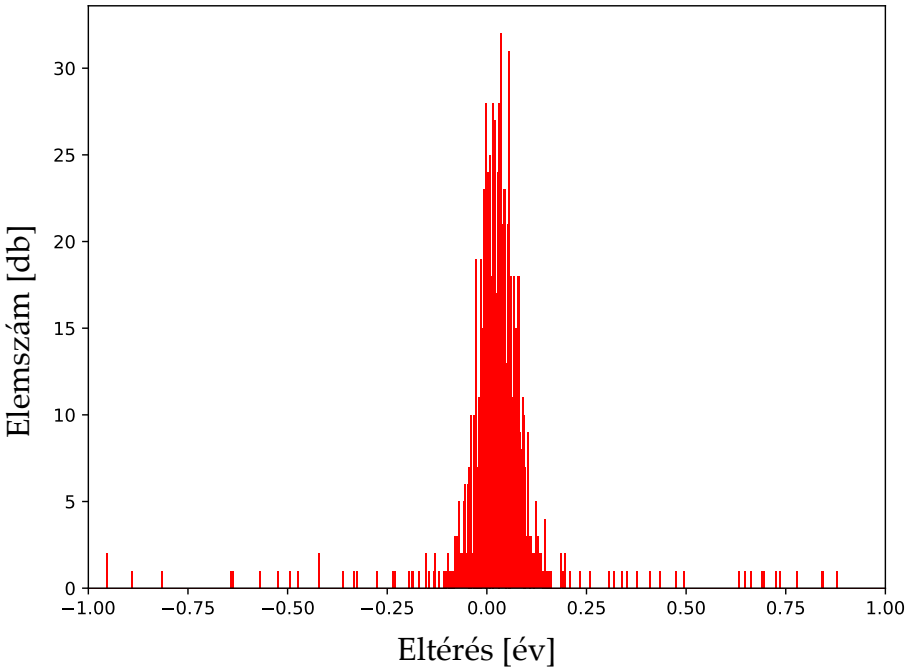
helyezkedik el (A kiugró hiba értékek a 4-8. ábrán láthatók).



4-5. ábra. A tömeg becslés hiba hisztogramja



4-6. ábra. A magasság becslés hiba hisztogramja

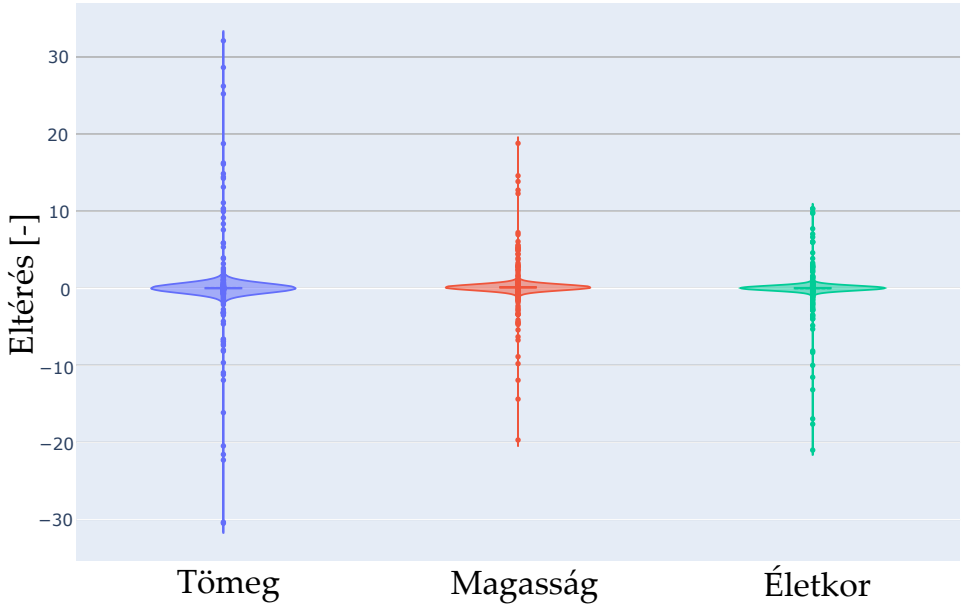


4-7. ábra. Az életkor becslés hiba hisztogramja

A hisztogramok alapján látszik, hogy a legtöbb hiba közel zérus mindhárom esetben. A testtömeg és az életkor esetén az eloszlás középértéke közel nulla, míg a testmagasság esetén a hisztogram jobbra tolódik. Az eloszlások közel szimmetrikusak, ami azt sugallja, hogy a hibák jelentős része egyensúlyban van a pozitív és negatív tartományokban. Fontos azonban megemlíteni, hogy a tömeg becslés hiba esetén több negatív előjelű hiba figyelhető meg, mint pozitív. Ez utalhat arra, hogy a modell hajlamos lehet az alul becslésre a tömeg esetén. A becslési hibák száma és azok mértéke elenyészőek, a hibák általában a $\pm 0,25$ tartományon helyezkednek el, kevés kiugró értékkel.

A becslések hibáinak vizualizációja hegedű diagramokon

A kevésszer előforduló, de átlagtól jelentősen eltérő becslések szemléltetésére a hisztogram nem alkalmas, ezek a boxplothoz hasonló hegedű diagramon (violin plot) jobban megfigyelhetők (4-8. ábra).



4-8. ábra. A becsült és a valós adatok közötti hiba violin plotjai

A hisztogramokhoz hasonlóan a violin plotok alapján is kijelenthető, hogy a hibák eloszlása közel szimmetrikus, a legtöbb hiba a zérus körül helyezkedik el. A legnagyobb hibák a tömeg meghatározása során léptek fel mindkét irányban. Ez összhangban van a 4-2. táblázattal, ahol a legnagyobb MAE és MSE is a tömegbecslése esetén volt. A magasság és az életkor esetén a legnagyobb tévedések hasonló nagyságúak, és megfigyelhető, hogy az életkor becslés esetén a kiugró hibák javarészt túlbecsülték az illető korát, a negatív előjelű hibák száma alapján. A magasság esetén a kiugró hibák eloszlása a tömegéhez hasonlóan szimmetrikus, nem figyelhető meg túlbecslésre vagy alul becslésre való hajlam. A hibák statisztikai jellemzése a 4-3 táblázatban látható.

4-3. táblázat. A hibákat jellemző statisztikai adatok

Antropometriai jellemző	Átlagos hiba	Medián hiba	Maximum hiba	Minimum hiba
Testtömeg [kg]	0,061	0,012	32,114	-30,487
Testmagasság [cm]	0,155	0,126	18,811	-19,683
Életkor [év]	-0,019	0,027	10,331	-20,992

5. Összefoglalás/ eredmények kiértékelése

5.1. Összefoglalás

Ezen dolgozat személyek három különböző antropometriai jellemzőjének járáshang alapján történő becslését vizsgálta. A cél egy olyan deep learning alapú modell elkészítése volt, amely képes egy adott személy tömegét, magasságát és életkorát megbízhatóan megbecsülni.

Először szakirodalmi áttekintés történt a hang alapú járásvizsgálat és az abból kinyerhető jellemzőkkel kapcsolatban, majd a különböző hang alapú osztályozási és regressziós módszerekről. Ismertetésre kerültek a deep learning elterjedése előtti, majd az azt alkalmazó megoldások is.

Ezek után ismertetésre kerültek az felhasznált tanító adatok, azok forrásának bemutatásával, valamint statisztikai elemzésével. Ezután azon előfeldolgozási eljárás lépései kerültek ismertetésre, amelyekkel a hangfelvételek a konvolúciós neurális hálózat bemeneteként használható formába kerültek. A felvételek lépésekre szegmentálása, majd azokból Mel spektogramok készítése részletesen bemutatásra került a feldolgozási elvek bemutatása mellett magyarázó ábrákkal és kódrészletekkel. A bemenetként szolgáló spektogramokon túl a becsülendő, táblázatos formában rendelkezésre álló antropometriai paraméterek előkészítésének lépései is kifejtésre kerültek.

A be és kimeneti adatok létrehozása után, a neurális háló felépítésének és architektúrájának bemutatása mellett, a tanítás és tesztelés lépéseinek leírása következett. A modell tesztelés utáni eredményeinek jellemzése, mind statisztikai, mind pedig vizuális formában megtörtént mindhárom antropometriai jellemző esetében. Utolsó lépés a hibák potenciális okainak leírása, valamint azok jellemzése volt.

5.2. Eredmények értékelése

A dolgozat arra a kérdésre kereste a választ, hogy egy mély tanulás alapú modell képes-e nagy pontossággal megbecsülni a járáshangból kinyerhető jellemzők alapján a személyek tömegét, magasságát és életkorát. A modell eredményeinek elemzése után kijelenthető, hogy az implementált modell mindhárom antropometriai jellemző mérésére alkalmas, amiket nagy pontossággal képes megbecsülni, a hibák száma és nagysága elenyésző, kevés esetben fordul elő kiugró érték. Ez az eredmény egyértelműen mutatja, hogy a mély tanulás alapú megközelítés hatékony eszköz lehet az antropometriai adatok megbízható és pontos becslésére a járáshang elemzése révén. A modellek alkalmazása ezen a területen ígéretes lehetőségeket kínál a kutatásokban és a gyakorlati alkalmazásokban egyaránt.

5.3. Továbbfejlesztési lehetőségek

modell elkészítése közben voltak olyan elhanyagolások és tervezési döntések, amelyek az eredmények megbízhatóságát csökkentik. Egy hitelességet potenciálisan befolyásoló döntés a tanító, validációs és teszt adatok létrehozásának módja, hiszen a létrehozott teszt adatok nem teljesen függetlenek a tanító adatoktól, a tesztelés és validáció során feltehetően nem volt olyan adatpont, ami a tanítóadatkban egyáltalán nem megtalálható személytől származott volna. Ez felveti a kérdést, hogy a modell valóban képes-e az adott paramétereket becsülni, vagy egy extrém mértékű túltanulásnak vagyunk tanúi. Ezen felül a felvételek szegmentálása során keletkezhettek fals (több lépést tartalmazó) szegmentumok, amelyeket a statisztikai eljárás sem tud kiszűrni, így terhelve hibás adatokkal a modell teljesítményét.

A hibás becslések potenciális indokai lehetnek, az esetlegesen hibásan szegmentált felvételek, amelyek zajosságuk révén nehezítik a predikciós eljárást, továbbá a már említettek szerint a mérési hibák is okozhatnak pontatlan becsléseket. A neurális háló regresszióért felelős részének bővítése, változtatása javíthatja a kapott eredményeket. Célszerű a modellen hiperparaméter optimalizálást végezni a jövőben, amely során a modell finomítása történik meg, ezáltal elérhető lehet a nagyobb pontosság elérése vagy a tanítási ciklusok számának csökkentése. Az adatfeldolgozás átdolgozása is javíthat az eredményeken. Érdemes lehet megvizsgálni az egy bemenethez használt lépések számát, a felvételek mintavételi frekvenciájának csökkentését, valamint más spektogram típusok alkalmazását. Érdemes lehet még megvizsgálni egy LSTM modell felhasználását, amely esetben közvetlenül a hangfelvételek alkothatnák a bemeneteket. A modell megváltoztatása természetesen a paraméterek és az eljárások áttervezését igénylik. Autoenkóderek alkalmazásával is érdemes lehet kísérletezni, amelyek a bemeneteket legjobban leíró absztrakt jellemzők kinyerésére használhatók, amelyeket később a regressziós réteg bemeneteként felhasználva, a tanítás történhet.

A jelenlegi szoftver továbbfejlesztési lehetőségei közé tartozik egy Real-Time funkció hozzáadása, amely során a modell képes lenne a már korábban feldolgozott hangfelvételek alapján, valós időben becsülni a személyek antropometriai jellemzőit, akár egy telefon mikrofonjának használatával. Ezen felül egy felhasználóbarát kezelőfelületű program lefejlesztése is előnyös lehet, amely a korábban fejlesztett *Python* komponenseket használja fel.

5.4. Kitekintés

A mély tanulás alapú járáshang elemzésnek hatalmas potenciálja van a jövőben. Nem csupán az emberek azonosításában vagy egészségi állapotuk felmérésében lehet kiváló segítség, hanem a járásmintázatok elemzésében is számos új lehetőség körvonalazódik. Ezen túlmenően, a módszer alkalmazható lehet az állatok mozgás hangjainak vizsgálatára is, megnyitva ezzel az utat új kutatási területek előtt. A gépi és mély tanulás területén tapasztalható folyamatos fejlődés további lehetőségeket teremthet a modellek finomhangolására és a megbízhatóságuk növelésére. Ezáltal a technológia egyre hatékonyabb és pontosabb eszközöket kínál majd az antropometriai adatok megbízható elemzésére és értelmezésére, ami számos területen hozhat értékes eredményeket a tudományban és a gyakorlatban egyaránt.

Irodalomjegyzék

[1] Adjabi Insaf, Ouahabi Abdeldjalil, Benzaoui Amir, and Taleb-Ahmed Abdelmalik. Past, present, and future of face recognition: A review. *Electronics*, 9(8), 2020. ISSN 2079-9292. doi: 10.3390/electronics9081188. URL `https://www.mdpi.com/2079-9292/9/8/1188`.

[2] Abdel-Hamid Ossama, Mohamed Abdel-rahman, Jiang Hui, Deng Li, Penn Gerald, and Yu Dong. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10):1533–1545, 2014. doi: 10.1109/TASLP.2014.2339736.

[3] M. Tanaka and H. Inoue. A study on walk-recognition by frequency analysis of footsteps. *IEEJ Trans. Electron. Inf. Syst.*, 119:762–763, 1999.

[4] G. Rigoll J. Geiger, B. Schuller and M. Kneißl. Acoustic gait-based person identification using hidden markov models. *Workshop on Mapping Personality Traits Challenge and Workshop*, 2014.

[5] Bodhibrata Mukhopadhyay, Sahil Anchal, and Subrat Kar. Person identification using seismic signals generated from footfalls. 09 2018.

[6] Jazi Istiyanto, Jan Riwurohi, and Agfianto Putra. People recognition through footstep sound using mfcc extraction method of artificial neural network back propagation. 04 2018.

[7] M. Tanaka and H. Inoue. A study of feature extraction of footstep by frequency analysis (in japanese). *Transcript of IEE Japan*, 117(4):483–484, 1997.

[8] M. Cheng, M. Ho, and C. Huang. Gait analysis for human identification through manifold learning and hmm. *Pattern Recognition*, 41(8):2541–2553, 2008.

[9] Sicheng Zhou Jake Vasilakes and Rui Zhang. Natural language processing. *Machine Learning in Cardiovascular Medicine*, pages 123–148, 2021.

[10] M. Nixon, T. Tan, and R. Chellappa. Human identification based on gait. *IET Computer Vision*, 1(2):70–83, 2006.

[11] J. Picone. Continuous speech recognition using hidden markov models. *IEEE ASSP Magazine*, 7(3):26–41, 1990. doi: 10.1109/53.54527.

[12] Jinglan Zhang Laith Alzubaidi and Ye Duan. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, 8(53), 2021.

[13] Sander Dieleman and Benjamin Schrauwen. End-toend learning for music audio. *Acoustics, Speech and Signal Processing (ICASSP),*, pages 6964–6968, 2014.

[14] George Fazekas Keunwoo Choi and Mark Sandler. Automatic tagging using deep convolutional neural networks. In ISMIR, editor, *International Society of Music Information Retrieval Conference*, 2016.

[15] Mark Sandler Keunwoo Choi, George Fazekas and Kyunghyun Cho. Convolutional recurrent neural networks for music classification. In IEEE, editor, *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[16] Yingxue Wang, Yanan Chen, Md Zakirul Alam Bhuiyan, Yu Han, Shenghui Zhao, and Jianxin Li. Gait-based human identification using acoustic sensor and deep neural network. *Future Generation Computer Systems*, 86:1228–1237, 2018. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2017.07.012. URL `https://www.sciencedirect.com/science/article/pii/S0167739X17314760`.

[17] Zhou Quan, Shan Jianhua, Ding, Wenlong, Wang, Chengyin, Yuan Shi, Fuchun Sun, Li Haiyuan, and Fang Bin. Cough recognition based on mel-spectrogram and convolutional neural network. *Frontiers in Robotics and AI*, 8, 2021. doi: 10.3389/frobt.2021.580080. URL `https://www.frontiersin.org/articles/10.3389/frobt.2021.580080`.

[18] Keunwoo Choi, György Fazekas, Kyunghyun Cho, and Mark Sandler. A comparison of audio signal preprocessing methods for deep neural networks on music tagging, 2018.

[19] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision, 2015.

[20] Benedicta Nana Esi Nyarko, Wu Bin, Jinzhi Zhou, George K. Agordzo, Justice Odoom, and Ebenezer Koukoyi. Comparative analysis of alexnet, resnet-50, and inception-v3 models on masked face recognition. In *2022 IEEE World AI IoT Congress (AIIoT)*, pages 337–343, 2022. doi: 10.1109/AIIoT54504.2022.9817327.