JENEI ATTILA ZOLTÁN

TDK DOLGOZAT

„Mivel mindenki a maga módján látja a világot,
a maga módján éli meg nehézségeit és a sikereit.
Tanítani annyi, mint megmutatni a lehetőséget.
Tanulni annyi, mint élni a lehetőséggel.”

Paulo Coelho

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
VILLAMOSMÉRNÖKI ÉS INFORMATIKAI KAR
TÁVKÖZLÉSI ÉS MÉDIAINFORMATIKAI TANSZÉK

TDK DOLGOZAT

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
VILLAMOSMÉRNÖKI ÉS INFORMATIKAI KAR
TÁVKÖZLÉSI ÉS MÉDIAINFORMATIKAI TANSZÉK

JENEI ATTILA ZOLTÁN

TDK DOLGOZAT

Depresszió automatikus beszéd alapú felismerése 2D konvolúciós hálókkal

Témavezető:
Kiss Gábor

Budapest, 2019

NYILATKOZATOK

Elfogadási nyilatkozat

Ezen TDK dolgozat a Budapesti Műszaki és Gazdaságtudományi Egyetem Telekommunikációs és Médiainformatikai Tanszék által a Tanulmányi Diákköri Konferenciára előírt valamennyi tartalmi és formai követelménynek maradéktalanul eleget tesz. E TDK dolgozat a nyilvános bírálatra és nyilvános előadásra alkalmasnak tartom.

A beadás időpontja: 2019.10.25

témavezető

Nyilatkozat az önálló munkáról

Alulírott, *Jenei Attila Zoltán* (NRDG1I), a Budapesti Műszaki és Gazdaságtudományi Egyetem hallgatója, büntetőjogi és fegyelmi felelősségem tudatában kijelentem és saját kezű aláírásommal igazolom, hogy ezt a TDK dolgozatot meg nem engedett segítség nélkül, saját magam készítettem, és a TDK dolgozatban csak a megadott forrásokat használtam fel. Minden olyan részt, melyet szó szerint vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Budapest, 2019.10.24

szigorló hallgató

TARTALOMJEGYZÉK

ÁRBAJEGYZÉK

TÁBLÁZATJEGYZÉK

JELÖLÉSEK JEGYZÉKE

A táblázatban a többször előforduló jelölések magyar nyelvű elnevezése, valamint a fizikai mennyiségek esetén annak mértékegysége található. A ritkán alkalmazott jelölések magyarázata első előfordulási helyüknél találhatók.

Latin betűk

Jelölés	Megnevezés, megjegyzés, érték	Mértékegység
cl	bináris indikátora az osztályoknak (0 vagy 1)	-
f	frekvencia	Hz
p	jósolt valószínűség	-
r	korrelációs együttható	-
x, y	általános változók	-
CNN	Konvolúciós Neurális Háló (Convolutional Nerual Network)	-
DE	depressziós csoport	-
E	rendszerparaméter	-
FN	álnegatív (false negative)	-
FP	álpozitív (false positive)	-
HC	egészséges csoport	-
N	darabszám	-
T	időtartam	Másodperc
TN	valós negatív (true negative)	-
TP	valós pozitív (true positive)	-

Indexek, kitevők

Jelölés	Megnevezés, értelmezés
i	futó index
j	futó index

1. ÖSSZEFOGLALÁS

Kutatásomban depresszió detektálását végeztem 2D konvolúciós háló segítségével. Önmagukban vizsgálva a jellemzőket 83 %-os maximális osztályozási pontosságot sikerült eléri. Ez az eredmény értékében hasonló egy korábbi kutatáshoz, ahol 83-86 %-os felismerési pontosságot értek el a szerzők rendre felolvasott és spontán beszédnél [1].

A feldolgozásban egészséges és depressziós személyek hangfelvételeiből beszédakusztikai jellemzőket nyertem ki. Ezekből korrelációs struktúrákat állítottam elő, amiket mint intenzitásképeket használtam 2D konvolúciós háló bemeneteként. A hálóban teljes kereszt validációt alkalmaztam.

Az eljárásban vizsgáltam több beszédakusztikai jellemzőnél, hogy különböző eltolásokat alkalmazva a korrelációs struktúrában, hogyan változik az osztályozási pontosság. Egyre nagyobb léptékű eltolást alkalmazva eltűnnek korrelációk bizonyos jellemzővektorok között. Továbbá az osztályozási pontosság csökkenése tapasztalható nagyobb eltolások mellett.

Megvizsgáltam az osztályozás pontosságára nézve, hogy milyen hatással van két jellemző kombinációjának alkalmazása. Ennek eredménye rámutat arra, hogy bizonyos jellemzők kombinációjával elérhető jelentősebb javulás az osztályozásban ahhoz képest, mintha önmagukban alkalmaznánk őket.

Több alkalmazott beszédakusztikai jellemző körül az MFCC eredményezte a legnagyobb osztályozási pontosságot.

Jövőbeli célom az osztályozó algoritmus olyan irányú fejlesztése, hogy az a diagnosztikát támogassa. Erre lehetséges irány a depresszió súlyosságának megbecslése, illetve további betegségcsoportok bevonása az osztályozási folyamatba.

2. SUMMARY

In my research, I have executed depression detection with the help of 2D convolutional network. Analyse the features one by one I achieved a maximum classification accuracy of 83 %. This result is corresponding in value to a previous study wherein the authors had acquired a recognition accuracy of 83-86 % for spoken and spontaneous speech [1].

In the processing, I obtained acoustic features from recordings of healthy and depressed persons. From these I created correlation structures which I applied as input to a 2D convolutional network as intensity images. I used full cross validation to evaluate performances.

I have investigated that how the classification accuracy may change using different offsets in the correlation structures through several acoustic features. I found that correlations between certain features vectors disappears using increasing offsets. In addition, I experienced decrease in classification accuracy besides larger offsets.

Next to individual features, I examined the impact of applying a combination of two features on the accuracy of classification. The outcome of combining certain features shows that it achieves a significant improvement in classification rather than applying them alone.

The MFCC resulted the highest classification accuracy among the examined features.

My future goal is to improve the classification algorithm in the way of support diagnostics. One of the potential directions is to estimate the severity of depression or the inclusion of other disease groups into the classification process.

3. BEVEZETÉS

Az utóbbi időben a depresszió a világ egyik vezető betegségei közé emelkedett, amitől - a WHO felmérései szerint - több mint 300 millió ember szenved, és évente közel 800 ezren menekülnek öngyilkosságba[2], [3].

A pszichiátria megkülönböztet klinikai-, illetve unipoláris depressziót, továbbá unipoláris és bipoláris zavart. Tüneteik változatosak és sokszor együtt járnak egyéb pszichiátriai rendellenességgel, illetve szellemi betegséggel [4]. A betegség korai állapotában sem egyértelmű a felismerése, ugyanis valamennyi tünete összekeverhető a fáradtsággal. Éppen ezért a betegségben szenvedők jelentős része nem is fordul szakszerű segítséghez. Ennek elmulasztása viszont az állapot romlásakor végzetes is lehet az öngyilkossági hajlam erősödésével.

A betegség korai felismerése kulcsfontosságú, ami műszeres megoldással segíthető. Kutatásomban a depresszió hang alapú automatikus felismerésének lehetőségeit vizsgáltam mély neurális háló segítségével.

A depresszió felismerésének alapját az adja, hogy hatására változások jelennek meg az egyén beszédproduktumában, amik az egészséges állapotától elkülönítik.

A beszédből kinyert akusztikai jellemzők feldolgozásnak alapját Williamson munkássága nyomán alakítottam ki. A cikkben beszédakusztikai jellemzőket határoztak meg, amiknek létrehozták a jellemzőkomponensek közötti korrelációs struktúráját. Majd ennek a mátrixnak sajátértékeire regresszióanalízist végeztek [5].

Ezzel szemben én nem a saját értékeket használtam fel a gépi tanuló eljárás bemeneteként, hanem közvetlenül a korrelációs struktúrát.

Kutatásom célja, hogy automatikusan felismerjük a depressziót és hosszútávon támogassuk a depresszió diagnosztizálását.

Dokumentációmban először ismertetem a felhasznált módszereket és folyamatokat. Majd bemutatom a beszédjelfeldolgozás és osztályozás megvalósítását. Végül bemutatom az eredményeket és a belőlük levont következtetéseket.

4. BESZÉDADATBÁZIS

Egészséges és depressziós alanyok beszéd anyagával dolgoztam, akik az „Északi szél és a Nap” című mesét olvasták fel. A hanganyagok 44,1 [kHz] mintavételezési frekvenciával kerültek felvételre csíptethető mikrofonnal és wav kiterjesztésben lettek tárolva.

Kiegyenlített mennyiséggel dolgoztam:

- 91 egészséges, illetve
- 91 depressziós.

Az egészséges csoportot a továbbiakban ***HC***-vel (healthy control), míg a depressziósat ***DE***-vel (depressed) jelölöm.

5. MÓDSZEREK

Ebben a fejezetben az alkalmazott módszereket ismertetem, illetve bemutatom a kutatási folyamatot. A Beszédakusztikai Laboratórium által készített felvételeken végeztem a jellemzőkinyerést és az osztályozást. Illetve felhasználtam a tanszéken készített programokat a feldolgozás során.

5.1. A beszédet leíró jellemzők

Az emberi kommunikáció egyik jelentős részét képezi a beszéd, ami hanghullámok formájában általában levegő közegben terjed. A beszédet a tüdőből kiáramló levegő hozza létre, ahogy átáramlik a hangszalagokon, illetve arc/orr üregein. A hangszalagok magánhangzók esetében rezgésbe jönnek, míg mássalhangzók képzésekor rezgetés nélkül áramlik át a levegő. A suttogó beszédben a hangszalagok nem rezegnek. Az így kialakított beszédproduktum adott egyénre jellemző, mivel hordozza annak fiziológiás sajátosságait.

A beszéd időben folyamatosan változó jel, aminek feldolgozása bonyolult. Ennek egyik oka, hogy biológiai produktumként függ annak pillanatnyi állapotától. Például az egyén kitartott zöngehangjai is különböző periódusokat tartalmaz időről időre. Továbbá a beszédképzés tartalmaz tranziens, közel állandó és impulzusszerű elemeket. Viszont a beszéd egyes szakaszai egy kis időtartományon belül közel stacionáriusnak tekinthető, így erre az ablakra a jellemzők meghatározhatók.

A beszédjel egyik feldolgozásának módja a spektrumelemzés, amivel meghatározott sávszélességre származtatjuk a teljesítményt vagy az intenzitást. Jelen esetben ezt gördülő ablakkal végezzük, ami adott időszélességben határoz meg teljesítményspektrumot egymás utáni pontokban a beszédjel végéig [6].

Ezek alapján az alábbi jellemzők kerültek meghatározásra.

1. táblázat - Akusztikai jellemzők leírása [6]–[9].

Megnevezés	Leírás
Alaphang	Az akusztikai hullám legmélyebb frekvenciaösszetevője.
Formánsok	A hangképző csatorna által létrehozott rezonanciacsúcsok.
Formáns sávszélesség	Az adott rezonanciacsúcsra (és az adott üregre) jellemző érték.
Jitter	Megadja az átlagos abszolút időbeni eltérést két egymást követő periódusban a teljes szakaszra vett átlagos periódussal súlyozva.
Shimmer	Megadja az átlagos abszolút amplitúdó eltérést két egymást követő periódusban a teljes szakaszra vett átlagos amplitúdóval súlyozva.
MelFilter	Szűrősor, ami alacsony frekvenciákon diszkriminatívabb.
MFCC	Diszkrét koszinusz transzformáltja a logaritmikus hangteljesítménynek adott mel-frekvencián.

Az ***1. egyenlet*** mutatja be a jitter számolását a Praat programban.

$$jitter = \frac{\frac{\sum_{j=2}^{N} |T_i - T_{i-1}|}{N-1}}{\frac{\sum_{j=1}^{N} T_i}{N}} \quad (1)$$

A Shimmer számolása megegyezik a Jitter-vel oly módon, hogy idő értékek helyett amplitúdó értékek vannak.

A szubjektív hangmagasság mértékegysége a mel, amelynek skálája az emberi hangmagasságértékelés skálájával azonos. Tehát, amit kétszer olyan hangosnak hallunk, annak a mel skálán is kétszer akkora érték felel meg. 0-16 [kHz]-es tartomány 0-2400 mel értéksorral szokás jellemezni (***2. egyenlet***). Ez alapján a beszédhangot frekvenciasávok szerint szűrjük, amik a mel skála.

$$mel = 2595 \cdot \lg(1 + \frac{f}{700}) \quad (2)$$

5.2. A korrelációs struktúra

A korreláció két halmaz elemei közti kapcsolatot írja le. Lineáris korrelációban a Pearson-féle korrelációs együtthatót (*r*) szokták használni [10].

Az így leírt korrelációs együttható korlátos, -1 és 1 közötti érték. A két szélsőérték felvételénél a két változó között erős lineáris kapcsolat van (-1 esetén a változás ellen-

tétes). 0 korrelációs érték esetén a két változó lineárisan független egymástól. A ***3. egyenlet*** írja ezt le.

$$r = \frac{\sum_{i=1}^{n}(x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x - \bar{x})^2 \cdot (y - \bar{y})^2}} = \frac{\sum_{i=1}^{n}(x - \bar{x}) \cdot (y - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \quad (3)$$

Ahol x és y változók, $\bar{x}$ és $\bar{y}$ a változók átlag értékei, σ_x és σ_y a változók tapasztalati szórása, végül pedig n a változók száma.

A korrelációs együtthatókat felhasználva mátrix formában korrelációs struktúrát hoztam létre a ***2. ábra*** szerint. Első sor, illetve oszlop a 27 darab Melfilter vektort jelölik. Ezek közös celláikban 1-1 korrelációs érték szerepelne.

Viszont én eltolásokat is alkalmaztam, amik egy 10*10-es korrelációs al-mátrixot eredményeznek minden cellában [11]. Ezen korrelációs értékek mindig két vektor közötti korrelációt írják le ebben a vizsgálatban.

Az ***1. ábra*** mutatja az eltolási műveletet egy példán keresztül, ha egyszeres eltolást alkalmazunk, ahol 1.-700.-ig a pozíció látható. Az első eltolásnál a vektor összes eleme egy értékkel eltolódik. A többi sor ennek megfelelően mindig az előző sorhoz képest van eltolva.

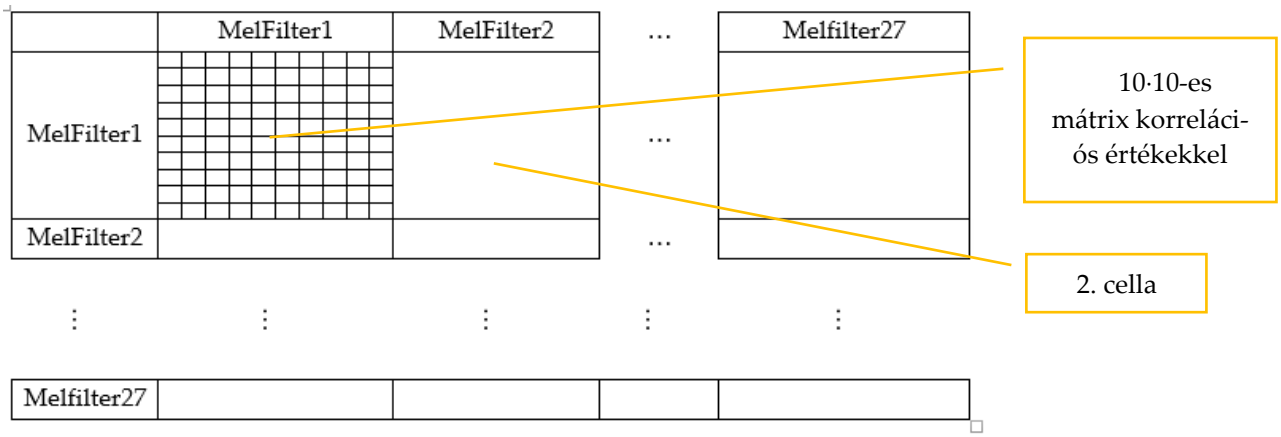
0. eltolás	1.	2.	…	700.
1. eltolás	700.	1.	…	699.
2. eltolás	699.	700.	…	698.
⋮	⋮	⋮		⋮
9. eltolás	692	693.	…	691.

1. ábra – Eltolások alkalmazása a jellemzővektorokban.

A ***2. ábra*** illusztrálja, hogy hogyan épül fel egy korrelációs struktúra. Az al-mátrix cellái az ***1. ábra*** alapján meghatározott vektorok közötti korrelációt leíró korrelációs értékek vannak. Ezek alapján felírható, hogy melyik cellában milyen két vektor közötti korrelációs érték szerepel a ***2. ábrán***.

$$r_{i,j} = korreláció(MelFilter1_i, MelFilter1_j) \quad (4)$$

Például a struktúra 2. cellájának az i. sora és j. oszlopa által kijelölt al-cellájában a MelFilter1 i-1-el eltolt vektorának, és a Melfilter2 j-1-el elolt vektorának korrelációs együtthatója szerepel.



2. ábra - Korrelációs struktúra.

5.3. Konvolúciós Neurális Háló

Maga a létrehozott korrelációs struktúra az értékei alapján felfogható egy 1D-s színkódolású képként, ami egy 2D konvolúciós háló bemenete lehet. Így alkalmazható a konvolúciós neurális háló osztályozás megvalósítására, amely az utóbbi években más területeken is alkalmaznak [12]–[14].

A vizsgálatomban az alábbi elemeket használtam a neurális hálóban:

- Konvolúciós réteg: különböző kernelek segítségével információkinyerést végez a bemeneti képeken, amik rejtett rétegekben neuronokhoz kapcsolódnak. A tanítás fázisában e neuronok súlyai kerülnek optimalizálásra.
- Aktivációs függvény: két neuron (vagy két réteg) közötti információáramlást valósít meg a függvény lefutásának megfelelően, amelyhez tartozó súlyok határozzák meg, hogy mennyire érvényesül az értéke a következő neuronnál (vagy rétegben).
    - ReLu: egység meredekségű aktivációs függvény, amely 0 alatti értékeket 0-val feleltet meg (0 alatti értékeket nem „enged át”).
- DropOut: a tanítás során véletlenszerűen meghatározott számú neuront átmenetileg eltávolít a hálóból. A validációs folyamatban az eltávolítás helyett egy valószínűségi értékkel van megszorozva a neuron kimenete. E módszerrel a rendszer komplexitása csökkenthető, ami kisebb futási időt eredményez. Továbbá a háló túltanulásának elkerülésére is alkalmazzák a gyakorlatban [15].

- MaxPooling: csökkenti az előző réteg dimenzióját, amivel hasonlóan csökkenti a rendszer igénybevételét, továbbá szintén egy lehetőség a túltanulás csökkentésére.
- Flatten: az előző réteg több dimenziós kimenetét egy vektorba rendezi.
- Dense: rejtett rétegekből épül fel, ahol egy réteg összes neuronja összeköttetésben áll a következő réteg összes neuronjával.
- SoftMax függvény: más néven normalizált exponenciális függvény, ami a bemeneti K elemű vektorját valószínűségi értékekké alakítja. Általában a neurális háló végén található és az osztályozó kimeneti értékeit alakítja át [16].

A hálót első körben tanítjuk az adathalmazunk egy részével, ami a súlyokat változtatja. Gradiensben mérhető, hogy melyik súly milyen mértékben befolyásolja a hiba változását. A hibák meghatározásához költségfüggvényeket alkalmazhatunk, amiket optimalizációs függvénnyel kombinálva csökkenteni igyekszik a háló.

A neurális hálóval bináris osztályozást végeztem, amihez alkalmazható az alábbi formulával a kereszt entrópia hibafüggvény (***5. egyenlet***) [17].

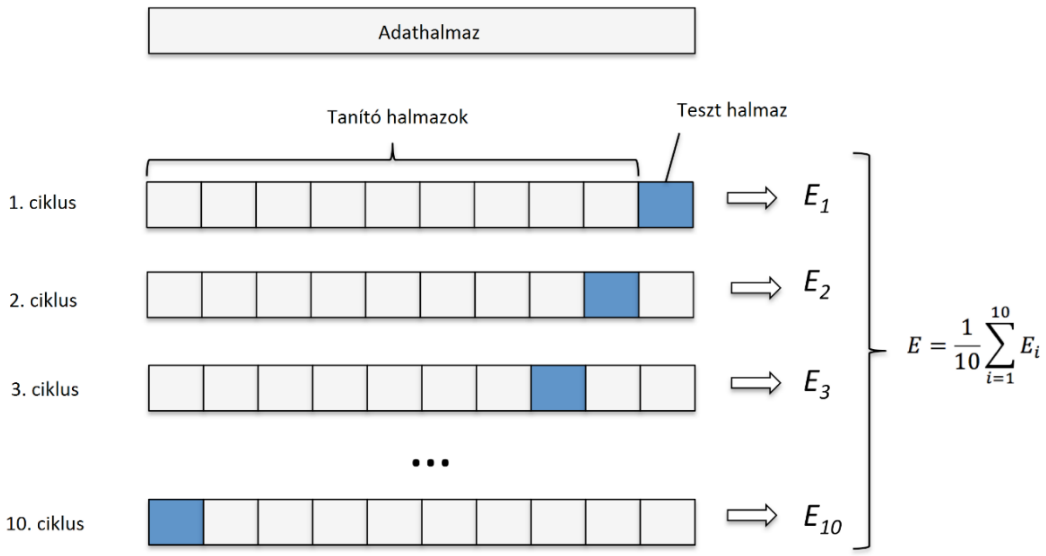
$$Keresztentrópia = -\frac{1}{N} \cdot \sum_{i=1}^{N} (cl \cdot \ln(p) + (1 - cl) \cdot \ln(1 - p)) \qquad (5)$$

Az egyenletben szereplő *cl* jelöli az osztályt bináris értékkel, míg *p* egy valószínűségi értéket.

5.4. Kereszt-validáció

A kereszt-validáció gépi tanulás során alkalmazott adathalmaz felosztási módszer, ami kevés adat esetén is biztosítja az elkülönült tanító-tesztelő halmazok létrehozását [18]. Alapelve, hogy a rendelkezésre álló adathalmazt felosztja több részre, amikből elkülöníti a tesztelő és a tanító halmazt. A munkám során k-Fold kereszt-validációt alkalmaztam, aminek magyarázó ábrája alább látható (***3. ábra)***.

Az ábra alapján megtörténik az adathalmaz felosztása, jelen esetben 10 részre. Az első lefutásban 9 halmazt használ tanító mintának, míg a 10.-el a tesztelést tudjuk elvégezni. A többi lefutásban is ugyanilyen mennyiségekkel dolgozik azzal a különbséggel, hogy teszthalmazként egy másik tizedét használja az eredeti adathalmaznak. Ezáltal minden ciklusban úgy valósul meg a tanítás és tesztelés, hogy a halmazok nem fednek át egymással. Minden ciklusban meghatározható egy rendszerparaméter (E_i - például a pontosság), ami használható modellszelekcióra.



3. ábra - k-Fold validáció folyamata [19].

Teljes kereszt-validációt alkalmaztam, ahol K érték a minta mennyiségével egyezik meg. Ebből kifolyólag 1 elemű a teszthalmaz, míg n-1 a tanító halmaz.

5.5. *Az osztályozást leíró változók*

A létrehozott 2D konvolúciós háló pontosságának leírására tévesztési mátrixot, illetve belőle származtatható értékeket alkalmaztam. A táblázat számszerűen tartalmazza, hogy az osztályozó melyik csoportra döntötte a tesztmintákat. A táblázat jelölési az alábbiak:

- TP (valós pozitív): Eredetileg depressziós mintát depressziósnak döntött.
- TN (valós negatív): Eredetileg egészséges mintát egészségesnek döntött.
- FP (álpozitív): Eredetileg egészséges mintát depressziósnak döntött.
- FN (álnegatív): Eredetileg depressziós mintát egészségesnek döntött.

2. táblázat - Tévesztési mátrix felépítése.

		Eredeti	
		Depressziós	Egészséges
Becsült	Depressziós	TP	FP
	Egészséges	FN	TN

A táblázat értékei alapján a klinikai gyakorlatban is levonhatók következtetések és megfontolások. Az álnegatív értékből következik, hogy a depressziós páciens nem kap kezelést. Ez a legrosszabb esetben a páciens életébe kerülhet. Ezzel szemben az álpozitív esetben az egészséges alany átesik további vizsgálatokon, amire külön nincs szükség. Ez utóbbinak költségvonzata, illetve felesleges kapacitáslekötése van az egészségügyben.

A táblázat értékeiből további leíró jellemzők származtathatók, mint a pontosság, specificitás, szenzitivitás és az F1 érték. Ezek számolási képletét alább ismertetem (**6 – *9. egyenletek***).

$$pontosság = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{\text{összes teszt elem}} \tag{6}$$

A pontosság megadja, hogy az osztályozó milyen százalékban döntött helyesen.

$$specificitás = \frac{TN}{TN + FP} = \frac{TN}{\text{összes eredeti egészséges}} \tag{7}$$

A specificitás a valós negatív megfelelő elkülönítését mutatja meg.

$$szenzitivitás = \frac{TP}{TP + FN} = \frac{TP}{\text{összes eredeti depressziós}} \tag{8}$$

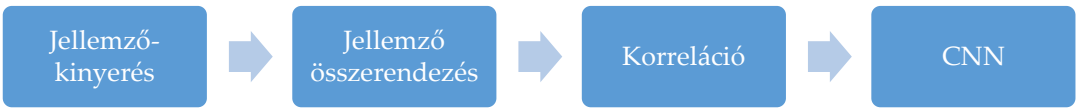
Hasonló a specificitáshoz azzal a különbséggel, hogy ez esetben a depresszió felismerésére ad értéket.

$$F1\ érték = \frac{2TP}{2TP + FP + FN} \tag{9}$$

Az F1 érték egy komplexebb mérőszám, ami közvetlen nem veszi figyelembe a valós negatív értéket.

5.6. *Megvalósított folyamat*

A hanganyagokat feldolgoztam, az *5.1* fejezetben leírt jellemzőket meghatároztam, majd összerendeztem a származtatott jellemzők szerint (***4. ábra***). Kialakítottam a korrelációs mátrixokat, majd osztályozást végeztem rajtuk.



4. ábra – Az előfeldolgozás folyamata.

Először feldolgoztam a beszédhangokat, amikből jellemzőket nyertem ki egy tanszéki program segítségével. Ez a program a Praat szabad felhasználású beszédelemző software segítségével jellemzőket számolt. Ezt megelőzően a felvételeket amplitúdó szerinti csúcsértékre normalizálta. Az akusztikai jellemzők kiszámolásánál egységesnek 50 [ms]-os időablakot állítottam be, amivel a frekvenciafelbontás jobban vizsgálható. A programban további beállításokat alkalmaztam, melyeket a táblázatban foglalok össze (***3. táblázat***).

3. táblázat - Jellemparaméterek.

Jellemző	**Paraméter(ek)**
Alaphang	Maximális érték: 400 [Hz].
Formánsok	Maximális frekvencia: 5500 [Hz]. Formánsok száma: 3
Sávszélesség	Sávszélességek száma: 3
Jitter és Shimmer	Lehetséges intervallumok: 1-20 [ms]. (T_{min} - T_{max}) Maximális periódus faktor: 1.3 (egymást követő periódusidők aránya)
MelFilter	Értékek száma: 27 Minimum frekvencia: 60 [Hz]. Lépésköz: 100 mel.
MFCC	Értékek száma: 14. Minimum frekvencia: 60 [Hz] Lépésköz: 100 mel.

A korrelációhoz szükséges struktúra kialakításához egy már meglévő C# programot használtam és fejlesztettem, ami egy felvételhez egy jellemzőmátrixot társít (Például 14 MFCC szöveges fájlból készít 1-et).

Ebben a mátrixban eltárolom az alany/páciens azonosítóját és csoportját (DE/HC) és a jellemzőkomponensek hosszát. Továbbá amit a jellemzőkinyerő program nem tudott meghatározni, oda "—undefined—„ került. Ezeket az összerendező prog-

rammal eltávolítottam az egységes fájl létrehozásánál, oly módon, hogy ugyanazon indexen a többi komponensél is törtöltem az érték akkor is, ha az számérték volt.

Harmadik lépésben előállítottam a korrelációs mátrixokat. Négy eltolási értéket alkalmaztam, amelyeknél rendszerint eggyel, kettővel, néggyel és nyolccal történt meg a pozícióeltolás.

Az osztályozó algoritmus és az ahhoz szükséges korrelációs mátrixok feldolgozása *python* kódban lett kifejlesztve, ami alapjaiban a rendelkezésemre állt. A továbbiakban ezt a kódot fejlesztettem. Tensorflow környezetet használtam, amit kiegészítettem több programkönyvtárral, mint numpy, keras, pandas [20].

A program első része a korrelációs adatmátrixok előkészítése a konvolúciós neurális hálóhoz. Itt történik meg a fájlok beolvasása, amikből egyénenként 1-1 vektort állítottam elő. Ezeket véletlenszerűen megkevertem.

Két halmazra választottam szét a fent létrehozott objektumot: korrelációs értékeket tartalmazó vektorokra, illetve címkékre. A vektorokat 0 és 1 érték közé normalizáltam, majd visszarendeztem mátrix alakba.

A CNN-t egy szekvenciális modellel hoztam létre, amihez a ***5.3 fejezetben*** ismertetett elemeket adtam hozzá a ***4. táblázatban*** leírt módon.

4. táblázat - CNN rétegei.

Réteg	**Paraméter(ek)**
Konvolúciós	filterek száma: 32 db, konvolúciós ablak mérete: $10 \cdot 10$, lépésköz: 10
ReLU	-
DropOut	25 %
Konvolúciós	filterek száma: 32 db, konvolúciós ablak mérete: $10 \cdot 10$, lépésköz: 10
ReLU	-
DropOut	25 %
MaxPooling	Pool méret: $2 \cdot 2$
DropOut	25 %
Flatten	-
Dense	kimeneti tér dimenziója: 2
SoftMax	-

Az első konvolúciós réteg esetén a bemeneti mátrix méretét is meg kellett adnom, amit mindig az éppen vizsgált jellemző (vagy jellemzőkön) határozott meg. A konvo-

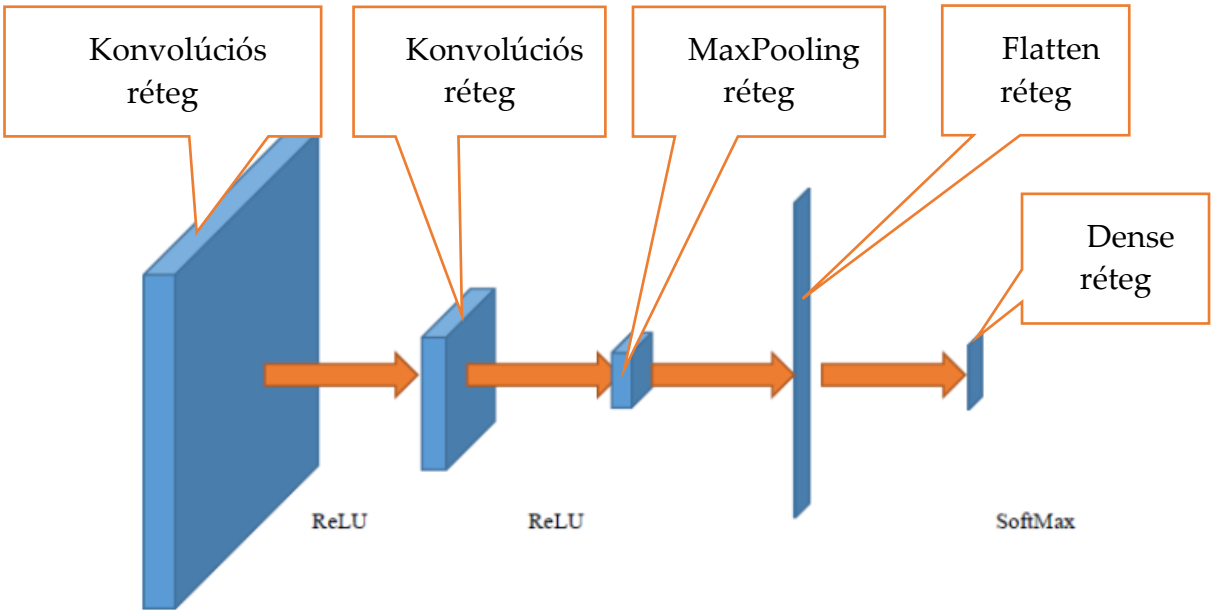
lúciós ablak mérete az egy-egy jellemzőkomponens korrelációs almátrixa miatt lett 10·10-es érték, míg a 10 lépésköz is az al-mátrixról al-mátrixra való áttérést hivatott megoldani.

A ReLU aktivációs függvény használata mögötti gondolat azt volt, hogy a 0 feletti értékeket transzformáció nélkül átviszi a következő rétegbe, míg az az alatti értékeket nem.

A DropOut komponenssel véletlenszerűen a tanítás alatt 25 %-át a neuronoknak figyelmen kívül hagytam, így csökkenthető a túltanulás esélye a neurális hálónál.

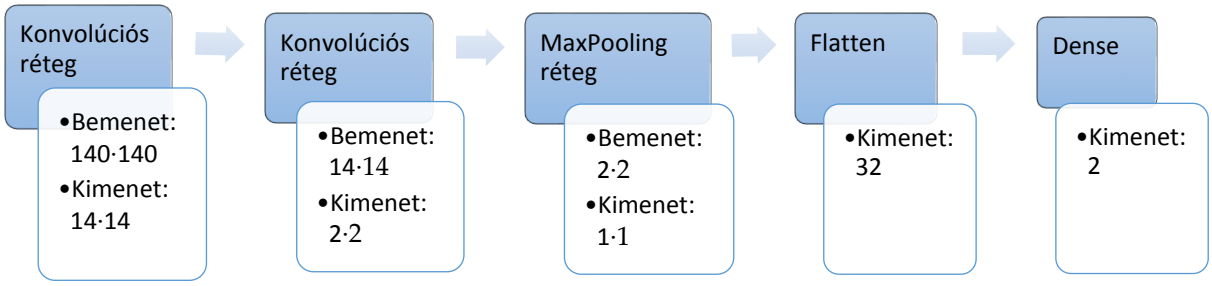
A Dense rétegben a kimeneti két csoportnak megfelelően adtam meg az értéket, és a SoftMax függvénnyel megkaptam az osztályozó döntését valószínűségi skálán mindkét csoportra. Kerekítésnek megfelelően a valószínűségi értékek bináris döntéssé alakíthatók.

A CNN-t vizuálisan a ***5. ábra*** szemlélteti megnevezve a különböző rétegeket.



5. ábra - CNN felépítése.

A bemeneti, illetve kimeneti mátrixok méretének alakulását az MFCC jellemzőn keresztül szeretném érzékeltetni az ***6. ábrán***. A CNN bemeneti mátrixa páciensenként 140·140 (14 jellemző; 10·10-es korreláció mátrix), ami a Dense réteghez érve 32 értékké alakul (filterek száma).



6. ábra – Dimenzióváltozás a feldolgozás során.

A létrehozott modellben a súlyok állításához ADAM féle optimalizációt alkalmaztam, mely a költségfüggvényt igyekszik minimalizálni. A tanítás alatt egy validációs halmazon a modell kiszámolja minden iterációban a bináris keresztentrópiát, mint költségfüggvény.

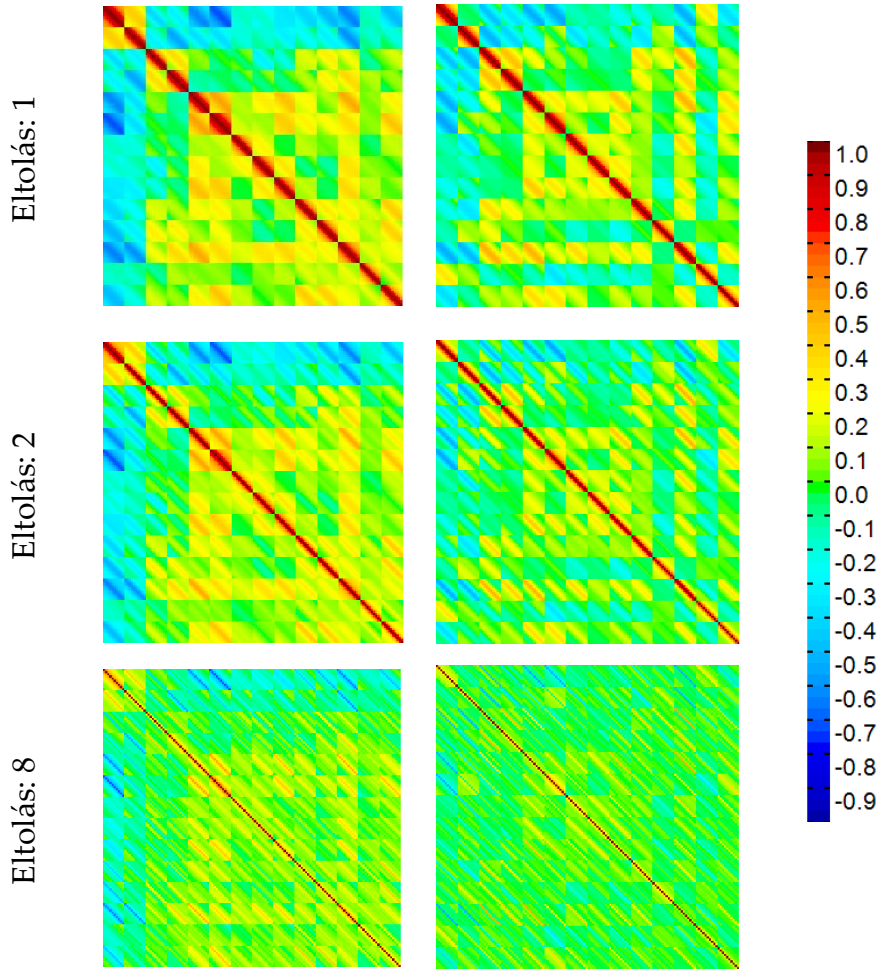
A formánsok, a sávszélességek és kettőjük kombinációja esetén a második konvolúciós réteget, a ReLU függvénynél és a DropOut réteggel együtt kivettem a CNN-ből. Erre azért volt szükség, mert a mátrixok dimenziója a feldolgozás során 0-ra eliminálódott (kisebb, mint 100·100-as korrelációs struktúrájuk van).

6. EREDMÉNYEK ISMERTETÉSE

Vizsgálati eredményként bemutatom a létrehozott korrelációs mátrixok intenzitásképeit, a teszthalmazon végzett osztályozás eredményét, illetve a származtatott metrikákat. A jellemzők, amire az osztályozást elvégeztem az MFCC, Melfilter értékek, a Formáns frekvenciák (1., 2., 3. formáns), a hozzájuk tartozó sávszélesség, és az utóbbi kettő együttese.

6.1. Korrelációs mátrixok

A korrelációs struktúra intenzitásértékek szerinti megfeleltetésben egy-egy alanyra a ***7. ábra***-n láthatók az MFCC jellemző esetén. Az ábra jobb oldalán látható a skálázás, ami alapján vörös szín jelöli az 1-es, illetve sötétkék a -1 korrelációértéket.



7. ábra - Korrelációs struktúrák MFCC jellemzőre depressziós (bal) és egészséges (jobb) alany esetén.

A bal oldali három kép a depressziós, a jobb oldaliak pedig az egészséges személy korrelációs értékeit tartalmazza.

A Melfilter és a Formáns + sávszélesség intenzitásképeket az MFCC mintájára a függelékekben csatolom.

6.2. Tévesztési mátrixok

Az ***5. táblázat*** mutatja a tévesztési mátrixokat jellemzők szerint. Az MFCC értékek esetén az egymáshoz képesti 8-al való eltolás eredményezte a legnagyobb pontosságot. A többi jellemzőnél a legnagyobb pontosságot az 1-es való eltolásnál kaptam.

A formáns-sávszélesség kombinálása több depressziós felvétel helyes felismerését eredményezte a tévesztési mátrixok alapján, mintha elkülönítve alkalmaznánk őket.

5. táblázat - Téveszéti mátrixok.

MFCC			**MelFilter**		
8-as eltolás			1-es eltolás		
	Eredeti			Eredeti	
	HC	DE		HC	DE
Becsült HC	80	18	Becsült HC	77	16
Becsült DE	11	73	Becsült DE	14	75

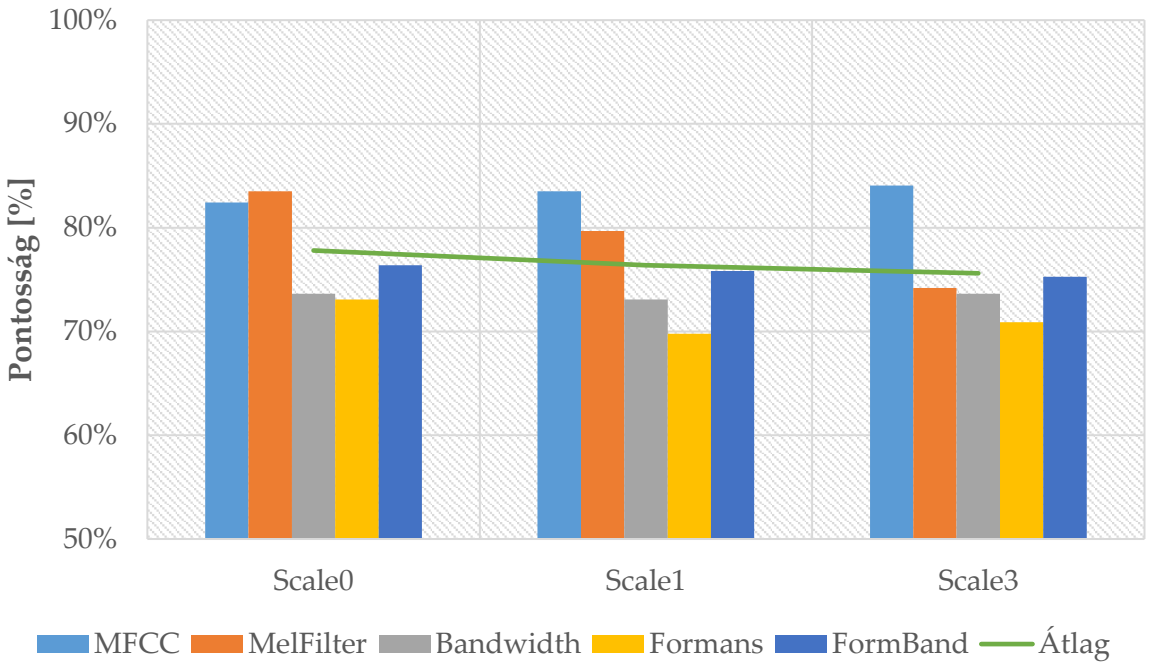
Sávszélesség			**Formáns**			**Formáns - Sávsz.**		
1-es eltolás			1-es eltolás			1-es eltolás		
	Eredeti			Eredeti			Eredeti	
	HC	DE		HC	DE		HC	DE
Becsült HC	71	28	Becsült HC	70	28	Becsült HC	70	22
Becsült DE	20	63	Becsült DE	21	63	Becsült DE	21	69

6.3. *Metrikák*

Az alábbi metrikákat a tévesztési mátrixokból származtattam és ábrázoltam. Segítségükkel az osztályozást egy-egy mérőszámmal jellemezhetők.

6.3.1. Pontosság

A pontosságok alakulását láthatjuk a ***8. ábra**-n*, ahol a három diagramcsoport rendre az 1-el, 2-vel és 8-al való eltolást mutatja. Az MFCC érték mindhárom eltolásnál 80 % fölött marad, míg a MelFilter látványosan csökken. A sávszélesség hasonlóan stabil pontosságot mutat, míg a formáns kis léptékben csökken. A zöld egyenes jelöli az átlagos pontosságot az adott eltolásnál, ami csökkő tendenciát mutat. Nagyobb eltolások alkalmazásakor (16, 32, illetve 64) az értékek látványosabb csökkenése volt tapasztalható.

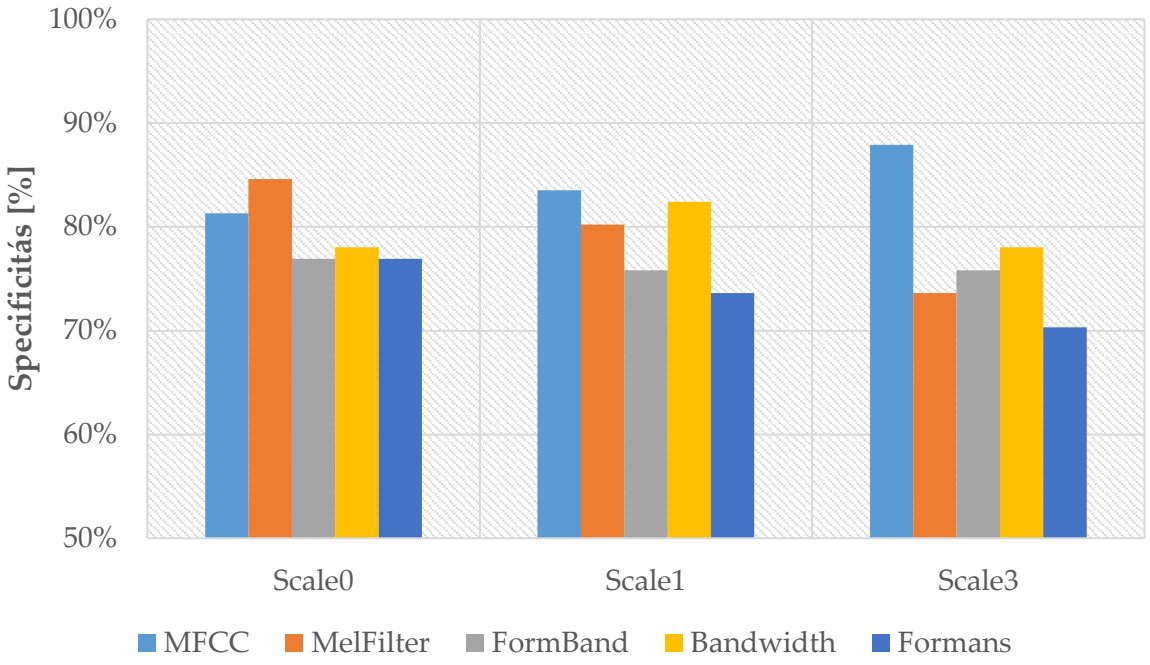


8. ábra - Az osztályozás pontossága.

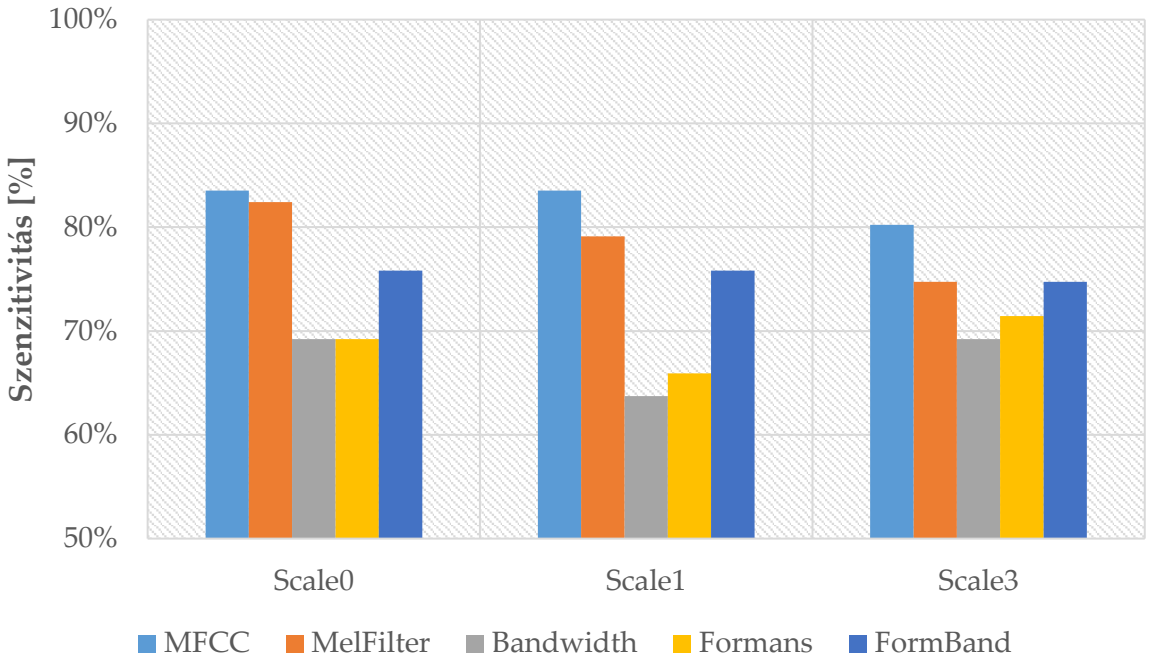
6.3.2. Specificitás és szenzitivitás

A pontossággal megegyező felépítésben ábrázoltam a szenzitivitást és a specificitást az alábbi két diagramon (***9. ábra-10. ábra***). A specificitás alapján láthatjuk százalékos értékben, hogy a valóban egészséges felvételek közül mennyit döntött az osztályozó egészségesnek. A szenzitivitásnál a depressziósnak döntött és a valóban depressiós felvételek aránya jelenik meg.

E két metrikát érdemes lehet együtt vizsgálni egymással, ugyanis, míg az egyik az egészségesekről, addig a másik a depressziós alanyokról ad információt az osztályozó tekintetében.



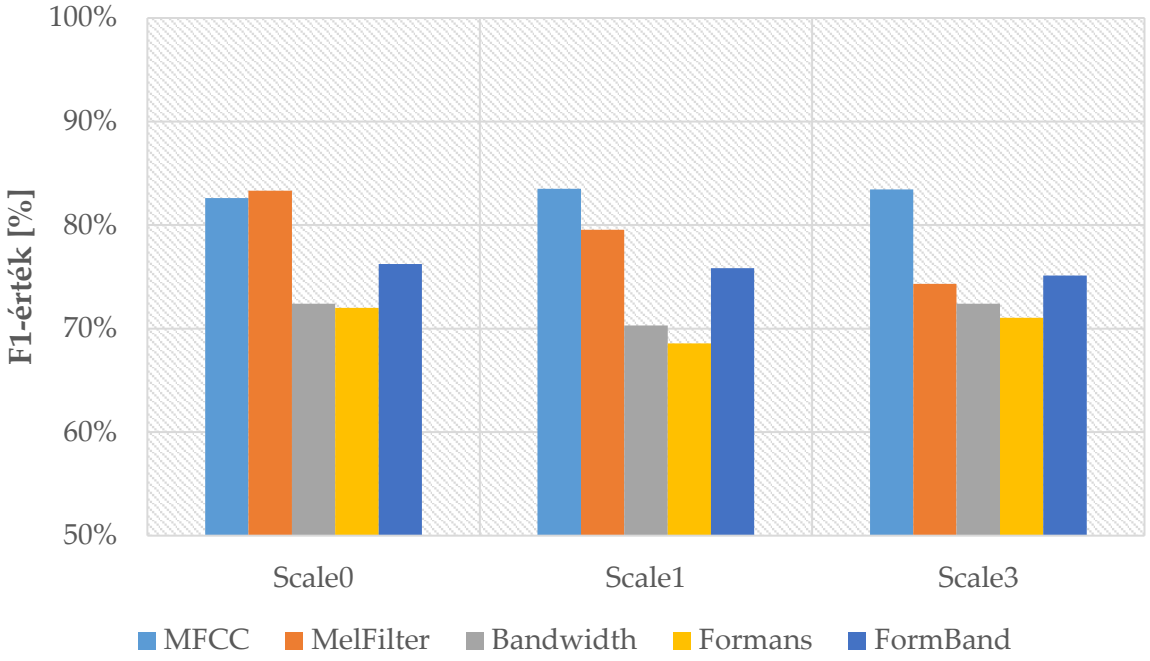
9. ábra - Specificitás alakulása.



10. ábra - Szenzitivitás alakulása.

6.3.3. F1-érték

Az F1-értéknél a Melfilter az egyetlen, aminél a csökkenő jelleg látszik, míg a többi értékre stagnálás látható az eltolás növelése mellett.



11. ábra - F1-érték alakulása a depressziós csoportnál.

7. AZ EREDMÉNYEK ELEMZÉSE ÉS KÖVETKEZTETÉSEK

A struktúra előállításánál alkalmazott eltolásokra látható, hogy minél nagyobb az eltolás mértéke, annál „élesebb" az intenzitáskép. Ez azt jelenti, hogy nagy eltolásnál időben távol kerül egymástól két érték, amik valószínűleg kevésbé fognak korrelálni egymással. Ez alapján nagyobb eltolás alkalmazása az osztályozás javításához nem járul hozzá.

Az osztályozási pontosságot bemutató ábráról látható, hogy a MelFilter jellemző látványosan romlott az egyre nagyobb eltolások mellett, míg a többi jellemző stabilnak mutatkozott. Az alkalmazott jellemzők közül az MFCC bizonyult a legnagyobb pontosságot eredményező jellemzőnek. Továbbá megfigyelhető, hogy a formáns, sávszélesség kombinálása javított az osztályozási pontosságon ahhoz képest, mintha csak önmagukban szerepeltek volna. Így a továbbiakban érdemes lehet több jellemző bevonása a korrelációs mátrixba.

A specificitás tekintetében az MFCC jellemző növekvő jelleget mutat, ami alapján leírható, hogy az eltolás növelésével képes volt pontosabban felismerni az egészséges mintákat, míg ezzel párhuzamosan a depressziós felismerése kis mértékben csökkent.

A szenzitivitás a formánsok és a sávszélességek önmagukban való osztályozásánál a legalacsonyabb. Azaz több egyént döntött egészségesnek, annak ellenére, hogy depressziós volt, mint fordítva.

A formáns és a sávszélesség kombinációja látszólag kiegyensúlyozta az álpozitív és az álnegatív értékeket.

Az F1- érték jó mérőszám a specificitás és a precizitás kiegyensúlyozottságának mérésére. Alapvetően nem veszi figyelembe a TN cellát. Az ábra alapján az MFCC jellemző stabil és 80 % fölötti értéket képvisel mindhárom eltolás esetén. A többi jellemző csökken vagy stagnál az eltolás növelésével.

Jövőbeli célom az osztályozó algoritmus olyan irányú fejlesztése, hogy az a diagnosztikát támogassa. Erre lehetséges irány a depresszió súlyosságának megbecslése, illetve további betegségcsoportok bevonása az osztályozási folyamatba.

8. FELHASZNÁLT FORRÁSOK

[1] G. Kiss és K. Vicsi, „Comparison of read and spontaneous speech in case of automatic detection of depression”, *8th IEEE Int. Conf. Cogn. Infocommunications, CogInfoCom 2017 - Proc.*, köt. 2018-Janua, 2018.

[2] World Health Organization, „Depression”, 2018. [Online]. Elérhető: https://www.who.int/news-room/fact-sheets/detail/depression. [Elérés: 09-okt-2019].

[3] Jordan Bates, „5 Unexpected Reasons Why Modern Life Depresses Many People”, *Highexistance*. [Online]. Elérhető: https://highexistence.com/5-reasons-modern-life-depression/. [Elérés: 09-okt-2019].

[4] G. S. Malhi és J. J. Mann, „Depression”, *Lancet*, köt. 392, sz. 10161, o. 2299–2312, 2018.

[5] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, és D. D. Mehta, „Vocal and facial biomarkers of depression based on motor incoordination and timing”, *AVEC 2014 - Proc. 4th Int. Work. Audio/Visual Emot. Challenge, Work. MM 2014*, o. 65–72, 2014.

[6] Németh Géza, „A beszéd fizikai jellemzése”, in *A magyar beszéd*, V. Klára, Szerk. Budapest: Akadémia Kiadó, 2010, o. 38–50.

[7] J. P. Teixeira, C. Oliveira, és C. Lopes, „Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters”, *Procedia Technol.*, köt. 9, o. 1112–1122, 2013.

[8] M. Farrús, J. Hernando, és P. Ejarque, „Jitter and shimmer measurements for speaker recognition”, in *Proceedings of the Interspeech 2007*, 2007, o. 778–781.

[9] Abdel-rahman Mohamed, „Deep Neural Network acoustic models for ASR”, University of Toronto, 2014.

[10] A. G. Asuero, A. Sayago, és A. G. González, „The correlation coefficient: An overview”, *Crit. Rev. Anal. Chem.*, köt. 36, sz. 1, o. 41–59, 2006.

[11] L. Roland, „Keresztkorreláció elemzés depressziós beszédanyagon”, 2016.

[12] R. Y. Yang és R. Rai, „Machine auscultation: enabling machine diagnostics using convolutional neural networks and large-scale machine audio data”, *Adv. Manuf.*, köt. 7, sz. 2, o. 174–187, 2019.

[13] R. Yamashita, M. Nishio, R. K. G. Do, és K. Togashi, „Convolutional neural networks: an overview and application in radiology”, *Insights Imaging*, köt. 9, sz. 4, o. 611–629, 2018.

[14] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, és D. P. Subha, „Automated EEG-based screening of depression using deep convolutional neural network”, *Comput. Methods Programs Biomed.*, köt. 161, o. 103–113, 2018.

[15] D. Huynh, „Applying Dropout to Prevent Shallow Neural Networks from

Overtraining", Lund University, 2017.

[16] Wikipedia, „Softmax function", 2019. [Online]. Elérhető: https://en.wikipedia.org/wiki/Softmax_function. [Elérés: 24-aug-2019].

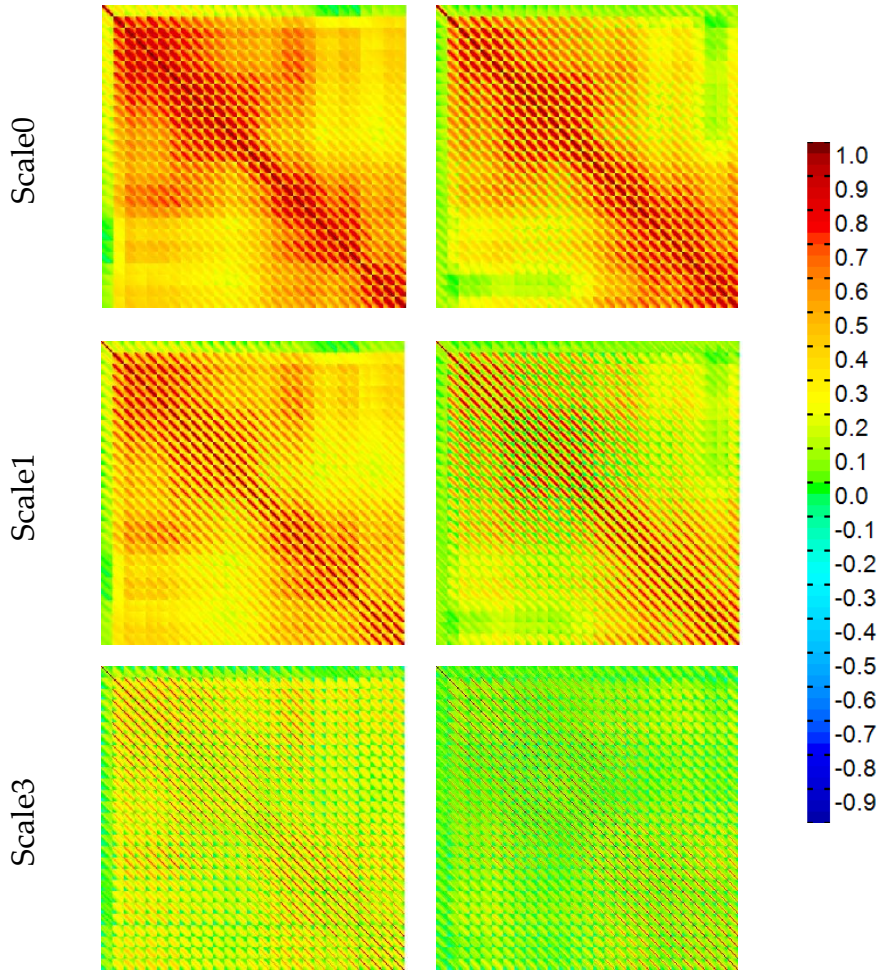
[17] G. E. Nasr, E. A. Badr, és C. Joun, „Cross Entropy Error Function in Neural Networks: Forecasting Gasoline Demand.", *FLAIRS Conf.*, sz. January, o. 381–384, 2002.

[18] G. Kovács, „Statisztikai modellek értékelő eljárásai", Eötvös Lóránd Tudományegyetem, 2015.

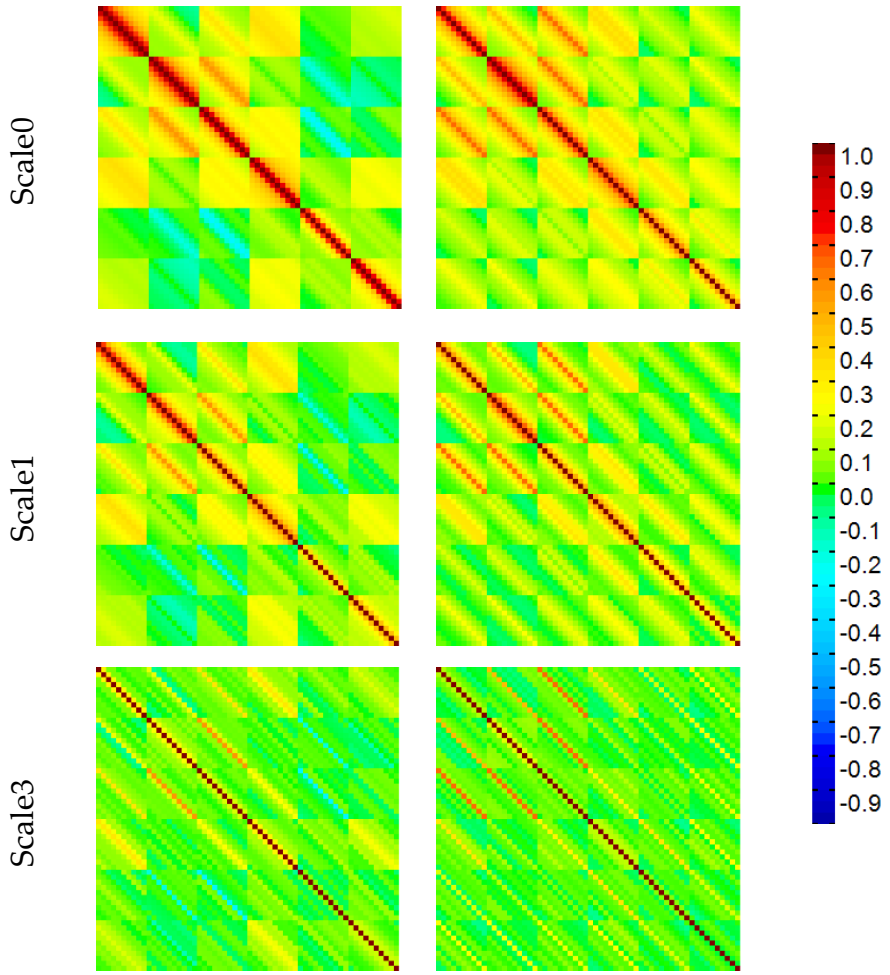
[19] J. Buhagiar, N. Strisciuglio, N. Petkov, és G. Azzopardi, „Automatic segmentation of indoor and outdoor scenes from visual lifelogging", *Front. Artif. Intell. Appl.*, köt. 310, sz. May, o. 194–202, 2018.

[20] Python Software Foundation, „Python", 2019. [Online]. Elérhető: https://www.python.org/. [Elérés: 22-okt-2019].

9. MELLÉKLET



12. ábra – Korrelációs struktúrák Melfilter jellemzőnél depressziós (bal) és egészséges (jobb) alanyra.



13. ábra – Korrelációs struktúrák Formáns - Sávszélesség jellemzőnél depressziós (bal) és egészséges (jobb) alanyra.