TANSZÉKVEZETŐ

DIPLOMATERVEZÉSI FELADAT

Jenei Attila
szigorló egészségügyi mérnök hallgató részére

Parkinson, depresszió és gégészeti elváltozások automatikus beszéd alapú elkülönítési lehetőségei 2D konvolúciós neurális hálókkal

Számos betegség hatással bír az emberi beszédproduktumra. A megváltozott beszéd alapján lehetőség nyílik az ilyen betegségek automatikus diagnosztizálásának támogatása. Azonban fontos a diagnosztikát támogató rendszer alapos megtervezése, úgy mint a megfelelő akusztikai-fonetikai jellemzők kiválasztása, azok feldolgozása és a megfelelő gépi tanuló eljárás alkalmazása. A dolgozat célja, hogy a hallgató megvizsgálja, hogy az egyes betegségek mekkora pontossággal ismerhetők fel és különíthetők el egymástól 2D konvolúciós neurális hálókkal a beszédjelből kinyert akusztikai-fonetikai jellemzők speciális auto és kereszt korrlációs sturktúrája alapján.

A hallgató a vizsgálatot magyar nyelven végzi, olyan hangmintákat felhasználva amik, egészséges, Parkinson kóros, depressziós és egyéb gégészeti elváltozásoktól szenvedeő személyektől származnak.

A hallgató feladatának a következőkre kell kiterjednie:

- Készítsen szakirodalom kutatást betegségek beszédjel alapú detektálása témakőrben, elsősorban ahol kereszt vagy auto korrelációt használnak az adott betegség detektálásához.
- Vizsgálja meg, .hogy 2D konvulúciós neurális hálók használatával a beszéd akusztikai-fonetikai jellemzőiből előállított speciális auto és kereszt korrelációs stuktúrák alapján lehetséges-e a Parkinson kór, depresszió és egyéb gégészeti elváltozástól szenvedő személyek felismerése és elkülönítése.
- Vizsgálja meg, hogy az eljárás paramétereinek változtatásával, hogyan változik a felismerési és elkülönítési pontosság.
- Vonjon le következtetéseket a vizsgálatokból és dokumentálja a munkáját.

Tanszéki konzulens: Kiss Gábor

Budapest, 2019. október 11

/ Dr. Magyar Gábor /
tanszékvezető

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Távközlési és Médiainformatikai Tanszék

1117 Budapest Magyar tudósok krt. 2.
Telefon: +36 1 463 2448 Fax: +36 1 463 3107
E-mail: magyar@tmit.bme.hu http://www.tmit.bme.hu

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Távközlési és Médiainformatikai Tanszék

Jenei Attila Zoltán

PARKINSON, DEPRESSZIÓ ÉS GÉGÉSZETI ELVÁLTOZÁSOK AUTOMATIKUS BESZÉD ALAPÚ ELKÜLÖNÍTÉSI LEHETŐSÉGEI 2D KONVOLUCIÓS NEURÁLIS HÁLÓKKAL

KONZULENS

Kiss Gábor

BUDAPEST, 2020

Tartalomjegyzék

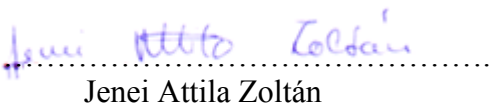
Jelölések jegyzéke **4**
Összefoglaló **5**
Summary **6**
1 Bevezetés **7**
2 Szakirodalmi áttekintés **10**
2.1 Depresszió 10
2.2 Parkinson kór 12
2.3 Diszfónia 15
2.4 Az adatbázison eddig elért eredmények 17
3 Beszédadatbázis **18**
4 Módszerek **22**
4.1 A beszéd leíró jellemzőinek meghatározása 22
4.2 A korrelációs struktúra előállítása 24
4.3 A Konvolúciós Neurális Háló (CNN) szerkezete 26
4.4 Az osztályozás leírása 30
4.5 A megvalósított folyamat paraméterei 34
5 Eredmények **41**
5.1 Beteg csoport átlagos eltérése az egészségestől a korrelációs struktúra alapján 41
5.2 Az alapháló osztályozásának leírása 42
5.3 Paraméterhangolás hatása 44
5.3.1 Korrelációs struktúra paraméterei 44
5.3.2 CNN paraméterei 45
5.4 Négy csoportos osztályozás 47
5.5 Bináris osztályozás (Egészséges vs. Beteg) 48
6 Eredmények elemzése és következtetések **51**
Köszönetnyilvánítás **54**
Táblázatjegyzék **55**
Ábrajegyzék **56**
Irodalomjegyzék **57**
Melléklet **66**

HALLGATÓI NYILATKOZAT

Alulírott **Jenei Attila Zoltán**, szigorló hallgató kijelentem, hogy ezt a diplomatervet meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző, cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy hitelesített felhasználók számára) közzé tegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Kelt: Budapest, 2020. 05. 20.



.....

Jenei Attila Zoltán

Jelölések jegyzéke

A táblázatban a többször előforduló jelölések magyar, illetve angol 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
CNN	Konvolúciós Neurális Háló (**C**onvolutional **N**eural **N**etwork)	-
DE	Depressziós csoport (**DE**pressed)	-
FD	Funkcionális diszfónia csoport (**F**unctional **D**ysphonia)	-
FN	Álnegatív (**F**alse **N**egative)	-
FP	Álpozitív (**F**alse **P**ositive)	-
FPR	Álpozitív arány (**F**alse **P**ositive **R**ate)	-
HC	Egészséges kontrollcsoport (**H**ealthy **C**ontrol)	-
OD	Organikus diszfónia csoport (**O**rganic **D**ysphonia)	-
UD	Általános hangképzőszervi elváltozásos csoport (**U**tterance **D**isease)	-
PD	Parkinson-kóros csoport (**P**arkinson **D**isease)	-
p	becslés (**p**rediction)	-
ROC	**R**eceiver **O**perating **C**haracteristic	-
SVM	Szupport Vektor Gép (**S**upport **V**ector **M**aschine)	-
TN	Valós negatív (**T**rue **N**egative)	-
TP	Valós pozitív (**T**rue **P**ositive)	-
TPR	Valós pozitív arány (**T**rue **P**ositive **R**ate)	-

Indexek, kitevők

Jelölés	Megnevezés, értelmezés
i	futó index
j	futó index
max	maximum
min	minimum

Összefoglaló

Irodalomkutatásából kiderül, hogy nincs egyezményes jellemzőhalmaz és bevett módszer a jelen betegségcsoportok felimerésére a beszédvizsgálat témakörében. Kitartott magánhangzókat, felolvasott illetve spontán beszédet összemérhető eredményességgel alkalmaznak. Azonban konszenzus fedezhető abban, hogy több információ nyerhető ki folytonos szövegek esetén. Kutatások számolnak be a jellemzőhalmazok széles halmazának alkalmazásáról, mint a formánsok, sávszélességeik, MFCC együtthatók és mel-sávos energiaértékek.

Diplomamunkámban a depresszió, Parkinson-kór, illetve az általános gépészeti elváltozás betegségének (két alkategóriájának) felismerését vizsgálom egy, a korábban még nem alkalmazott módszerrel. Az általam használt eljárásban beszédből akusztikai jellemzőket határoztam meg, amiknek képeztem részhalmazaikat.

A jellemzővektorokon eltolásokat végeztem a speciális korrelációs struktúrák létrehozásánál. Ezen mátrixok bemenetei voltak egy alaphálónak, aminek – és a korrelációs struktúráknak – finomhangolása mellett vizsgáltam az elkülönítés pontosságát.

Az eljárás előnye, hogy nem igényel bonyolultabb beszédfeldolgozási műveleteket (mint például a szegmentálás). Illetve a konvolúciós háló maga végzi a lényegi információ kinyerését a képreprezentációkból.

Alapbeállítások mellett vizsgáltam 5 jellemzőhalmaz elkülönítését 5 osztályos osztályozásban, amik közül a 14 MFCC jellemzőhalmazzal értem el a legjobb eredményeket. Javulást tapasztaltam a formánsok és sávszélességeik együttes alkalmazásával is ahhoz képest, mint külön alkalmazva őket.

Finom hangolva a korrelációs struktúra paramétereit – eltolási mérték és eltolási szám – illetve a konvolúciós háló kernel- és iterációszámát további közel 10 %-os javulást értem el a pontosságon 5 csoportnál. Két csoport összevonásával további 10 % növelést tapasztaltam. Végül a bináris kategorizálás mellett rendre a DE, PD, UD csoportokra 80 % feletti pontosságértékeket kaptam.

Az eredmények ígéretesek további kutatásokhoz, mint például jellemzőhalmazok kombinálására.

Summary

Literature research reveals that there is no conventional set of characteristics and a common method for recognizing disease groups by speech analysis. Persistent vowels, read and spontaneous speech are used with comparable effectiveness. However, consensus can be found that more information can be extracted from continuous speech. Research reports the use of a wide range of feature sets such as formants, their bandwidths, Mel-frequency cepstral coefficients (MFCC), and Mel-band energy values.

In my thesis work, I examine the recognition of depression, Parkinson's disease, and general voice disorders with a new method approach. In the procedure, I determined acoustic characteristics from speech, and then I formed their subsets. From these, I generated a correlation structure for each person, which I used for classification on convolutional neural network.

I applied shifts on the feature vectors when creating special correlation structures. These matrices were the input of convolutional neural network. I performed parameter tuning in the correlation structure and the convolutional network to examine the effect on the classification accuracy.

The advantage of this method is that it does not require more complex speech processing operations (such as segmentation). Respectively, the convolutional network itself extracts the essential information from the image representations.

Next to basic settings, I examined the separation of five feature sets in a 5-class classification, of which the 14 MFCC feature sets achieved the best results. I also found an improvement in using the formants and their bandwidths together compared to using them separately.

By tuning the parameters of the correlation structure – offset rate and offset number –, the kernel and iteration number of the convolutional network, I achieved a further nearly 10% improvement in accuracy of five groups. By merging two groups, I experienced a further 10% increase. Finally, I obtained accuracy values above 80% for the DE, PD, UD groups, respectively by binary classification.

The results are promising for further possible research such as combination of feature groups.

1 Bevezetés

Az élő szervezet (így az ember is) állandó kapcsolatban áll környezetével és önmagával. Mindennapi működés során mindkét irányból érkezhetnek az egészséget veszélyeztető hatások. Ilyen külső hatás lehet például a mérgezés, míg belső lehet az autoimmun betegség.

Betegségnek tekintünk alapvetően minden olyan állapotot, ami a szervezet normális működésétől eltér. Ez sok esetben ez a biológiai rendszer homeosztázisának megbomlását eredményezi, mint például emelkedett hőmérséklet vagy vérnyomás. Kialakulásukban, lefolyásukban, a szervezetre gyakorolt határukban számos kategóriát megkülönböztethetünk.

Több nemzetközi szintű kategorizálása létezik a felismert betegségeknek, amik közül a legtöbbet használt a World Health Organization (WHO) által létrehozott International Statistical Classification of Diseases and Related Health Problems (ICD) [1]. Ez kódok alapján tartalmazza a betegségeket, tüneteiket, társadalmi előfordulásukat és különböző előidéző körülményeiket.

A diplomamunkámban három betegségcsoporttal foglalkoztam, amik beszéd alapú felismerési lehetőségét mély tanuló eljárással – konvolúciós neurális hálóval – vizsgáltam. Ezek a betegségcsoportok a depresszió, Parkinson-kór és a gégészeti elváltozás (diszfónia). Előbbi kettő betegség az ICD szerint a „Mental and behavioural disorders” fejezetbe [2], míg utóbbi a „Symptoms, signs and abnormal clinical and laboratory findings” fejezetbe [3] tartozik.

Az említett betegségek felismerési lehetőségeinek kutatása kulcsfontosságú, ugyanis jelenleg nincs egyértelmű diagnosztikai eljárás a korai elkülönítésre. Későbbi fázisaikban – mikor már a betegségek pontosabban megállapíthatók – már az egyén életminőségének fokozott romlását idézhetik elő. Nemcsak az egyén, de a társadalom számára is hiánnyal (költséggel) jár a későn vagy nem kezelt betegség például a csökkent foglalkoztatottság vagy a teljes kiesés a munkakörből. Éppen ezért széles körben végzik a megfelelő biomarkerek kutatását a minél pontosabb és minél korábbi diagnózis felállítása érdekében.

A beszéd az emberi kommunikáció legáltalánosabb eszköze, ami nyelvi kódot közvetít. Ez a kommunikációs forma rendkívül összetett folyamatok összessége, amit

számos környezeti, illetve szervezeti tényező befolyásol. Nemcsak érzelmi hatásokra változik meg a beszéd, hanem bizonyos fiziológiai, illetve pszichiátriai tényezők, akár betegségek jelenlétében is. A beszéd mikrofonnal történő felvételével, és a beszédproduktum mélyebb vizsgálatával lehetőség nyílhat különböző betegségek felismerésére. Ezzel az egyén már betegségének akár korai stádiumában megfelelő kezelésben részesülhet, amivel életminőségének romlása visszafordítható [4], [5], [6].

A diplomamunkámban vizsgált betegségcsoportok bizonyítottan megjelennek a beszédben, beszédproduktumban. Továbbiakban a beszédből beszédjelfeldolgozás során különböző jellemzők nyerhetők ki [7]. A betegségek súlyosságától függően a jellemzőkben eltérések figyelhetők meg, amik felhasználhatók a betegségek elkülönítésére egy meghatározott (általában egészséges) populációtól. Megfelelő körülmények között akár a betegség súlyosságuk becslésére is lehetőséget adódhat [8].

A beszéd alapú vizsgálati módszerek előnye, hogy non invazív beavatkozást tesznek lehetővé. Emellett a technológia fejlődésével nagyobb tárolókapacitások, illetve információs rendszerek állnak rendelkezésre. Ezek lehetőséget biztosítanak a nagy mintaszámú beszédadatbázisok összegyűjtésére és felhasználására.

A betegségek beszéd alapú automatikus felismerésére teremtett egy egyszerűbb lehetőséget a gépi tanuló eljárások alkalmazása. Ezzel leegyszerűsödött a beszédfeldolgozás, jellemzőkinyerés és az osztályozás. Ezeket már egyre szélesebb körben alkalmazzák a beszédvizsgálati témakörben [9], [10]. A diplomamunkámban vizsgált betegségek felismerésére is számos alkalmammal használtak már korábban osztályozó algoritmusokat, amikre a ***2. fejezetben*** részletesen kitérek.

Diplomamunkámban egy, a korábbiaktól eltérő módszert mutatok be a megjelölt betegségek felismerési lehetőségeinek vizsgálatára. Az eljárás különlegessége, hogy a beszédfelvételekből kinyert jellemzőkből korrelációs mátrixokat hozok létre, amiket mint képreprezentációkat egy konvolúciós neurális háló tanítására használom.

A dolgozatban először rátérek a három betegségcsoport beszédjel alapú felismerésének – szakirodalomban eddig elért – eredményeire. A szakirodalmi kutatások irányt adnak az alkalmazandó beszédakusztikai jellemzőkre, lehetséges paraméterek alkalmazására a mély tanuló algoritmusban. Emellett bizonyos gyenge pontokra is felhívhatják a figyelmet, mint például az adatbázis sajátossága.

A szakirodalmi kutatásokon belül azokra fókuszáltam, ahol már alkalmaztak valamilyen formában autó- és keresztkorrelációs megoldásokat. Ugyanis így referenciát biztosíthatnak az eredményeim összehasonlításában.

A szakirodalmi áttekintésből kiindulva kinyertem azon akusztikai jellemzőhalmazokat, amikből előállítható korrelációs mátrix. Vizsgáltam, hogy a korrelációs struktúra alkalmazásával – paraméterállítások mellett – lehetséges-e a betegségek felismerése konvolúciós neurális hálók segítségével. A háló ezen típusának alkalmazását az indokolja, hogy a korrelációs struktúrák megfeleltethetők egy kétdimenziós képként.

Létrehoztam egy több osztályos kimenetű alapháló modellt, amivel megvizsgáltam a különböző jellemzőhalmazok elkülönítésének hatását. Ezzel már egy sorrendi eredményt kaptam arra vonatkozóan, hogy mely jellemzőhalmaz a leghatékonyabb az elkülönítésben.

Az eredmények alapján kiválasztottam egy jellemzőhalmazt, amivel az eljárás paramétereinek finom hangolását végeztem. Ezzel vizsgálható, hogy bizonyos paraméterek hogyan hatnak az elkülönítésre. Számos paraméter állítható mind a korrelációs struktúra létrehozásánál, mind a háló esetén. Az idő korlátossága miatt két-két változót választottam ki vizsgálatra. A felismerés jóságának leírására több metrikát is alkalmaztam.

Végül összefoglalom a fontosabb eredményeket és a belőlük levonható következtetéseket. A diplomamunka eredményeként egy olyan több vagy bináris betegséget felismerő osztályozási rendszer konstrukcióját várom, aminek metrikája nagyságrendileg összehasonlítható az eddigi szakirodalmi eredményekkel.

A dolgozat a következő felépítést követi: A szakirodalmi áttekintés fejezetben kitérek a betegségek ismertetésére, a diplomamunka szempontjából releváns kutatási eredményekre. Majd ismertetem az egészséges, depressziós, Parkinson kóros és hangképzőszervi elváltozásos felvételekből álló beszédadatbázist. Ezt követően a bemutatom a kinyerni kívánt jellemzőket, az autó- és keresztkorrelációs mátrixok felépítését, és az osztályozást végző konvolúciós hálót. Végül pedig rátérek az eredményekre és belőlük levonható következtetésekre. Dolgozatomat az eredmények és belőlük levonható következtetésekkel zárom.

2 Szakirodalmi áttekintés

Ebben a fejezetben bemutatom azon betegségeket, amiknek a detektálása a diplomamunkám feladata volt, továbbá ismertetem az egyes betegségcsoportoknál az eddig elérhető beszéd alapú kutatási eredményeket és következtetéseket.

Fontos megjegyezni, hogy ugyan több kutatás számol be konszenzussal bizonyos eljárásokról, mégis számszerű eredményeik nagy tartományon szórnak. A vizsgálati eredményeket nagyban befolyásolhatják az alkalmazott beszédadatbázis mennyisége és minősége, az alkalmazott módszer és a felhasznált jellemzők [11], [12], [13]. Ezért érdemes ugyanazon eljáráson további vizsgálatokat végezni annak érdekében, hogy a fennmaradó nyitott kérésekre is kielégítő válaszokat találjunk.

2.1 Depresszió

A depresszió az egyik leggyakoribb pszichiátriai betegség, amitől - a WHO felmérései szerint - több mint 300 millió ember szenved, és évente közel 800 ezren lesznek öngyilkosok [14]. Lehetséges kiváltó okai lehetnek a stresszes vagy negatív életesemények, fiziológiás betegségek, szociális problémák [15]. A betegség korai felismerése sem mindig egyértelmű, ugyanis tünetei széles körben változnak egyénenként és egy-egy egyénnél időben is [16]. Ezekből a pontos diagnózis felállításához mély szakismeret szükséges, ami az egészségügy egy kis részére hárul. A diagnosztikai folyamatot tovább nehezíti, hogy az illető teljesen elszigetelődik a társadalomtól [17].

Az egészségügyben több markert is használnak a depressziós állapot minél pontosabb meghatározásához az anamnézis felvétele mellett. Például vizsgálható lehet a szerotonin szint vagy a kortikális és limbikus rendszer kapcsolata [18].

A depresszió súlyosságának leírására leggyakrabban két skála alkalmazott. Ezek a Hamilton Depression Rating Scale (HAM-D), illetve a Beck Depression Inventory (BDI).

Az BDI kérdőívet és annak skálázását 1961-ben hozta létre Aaron T. Beck [19]. Ennek több verziója született, aminek legújabb változatát 1996-ban publikálták BDI-II néven. Ez a változat 21 kérdésből tevődik össze. Ezt a későbbiekben több nyelvre is lefordították és alkalmazzák a klinikai gyakorlatban. A kérdőívet a páciens saját belátása szerint tölti ki, majd a kérdésekre kapott pontok (0 és 3 között kapható pont egy kérdésre)

összegéből besorolható egy súlyossági osztályba. A kérdőív alapján 13 pont felett depressziósnak számít az illető, azon belül is 19 pontig enyhe, 28 pontig mérsékelt és 63 pontig súlyos depressziót különböztetünk meg (a skála 0-tól 63 pontig tart). Látható, hogy egyik nagy hátránya ennek a módszernek, hogy a páciens saját magáról alkot „diagnózist”.

A HAM-D skála eredeti verziójának neve HDRS-17 volt és 1960-ban hozta létre Max Hamilton [20]. Ezt a kérdőívet már szakértő tölti ki a páciens válaszai alapján. 20 kérdésből az első 17 ér pontot és számít bele a súlyosság megállapításába, míg az utolsó három további információt nyújthat a depresszió leírására.

A depresszió beszédre gyakorolt hatását már 1921-ben megállapították [21]. Ezt követően mélyebb kutatások kezdődtek a betegség és a beszéd közötti kapcsolat vizsgálatára. A tanuló algoritmusok megjelenésével és egyre bővülő beszédadatbázisok használatával a beszédjelalapú depresszió felismerés újszerű kutatási területként bontakozott ki [18].

Korábbi kutatások alapján a depresszió felismerése 50-86 % közötti pontossággal volt lehetséges beszédjel feldolgozás alapján [22], [21], [23], [24]. A pontossági értékek természetesen függnek az alkalmazott beszédadatbázistól, az alkalmazott folyamattól és a beszédet leíró jellemzőktől. Korábbi műhelymunkában felolvasott és spontán beszéden keresztül rendre 86 és 83%-os pontosságot értek el a kutatók lineáris kernelű SVM-et alkalmazva [25]. 37 alacsony szintű jellemzőt vizsgáltak ’e’ hangokon két időablakkal. A jellemzők között szerepelt az alapfrekvencia, intenzitás, formáns frekvenciák és sávszélességeik, jitter, shimmer, mel-sávos energia értékek.

Egy 2018-as kínai kutatásban 170 mintát tartalmazó beszédadatbázisból regressziós vizsgálatot végeztek a depresszió felismerésére. Felhasználtak spektrális, prozódiai jellemzőket. Férfi mintákon 82%-os, női mintákon 75%-os pontosságot értek el [26].

Mindamellet, hogy több kutatás is bizonyította, hogy a depresszió felismerése lehetséges beszédjel feldolgozásból, mégis maradtak nyitott kérdések. Mint mely beszédet leíró jellemzők a legalkalmasabbak a felismerésre, illetve milyen szintű feldolgozás szükséges a minél pontosabb detektáláshoz.

Jelen kutatásban egy speciális autó- és keresztkorrelációs struktúrát alkalmazok, aminek megvalósítása Williamson és munkatársai [27] publikációja alapján készült.

Munkájukban alacsony szintű beszédakusztikai jellemzőket – MFCC értékeket és formáns frekvenciákat – használtak fel kombinálva. Ezekre alkalmazták az autó- és keresztkorrelációs struktúrát, aminek sajátértékeit felhasználva regressziót végeztek a depresszió súlyosságának becslésére. Eredményként nagy pontossággal voltak képesek beszülni a depresszió súlyosságát. Vizsgálatukat a német depressziós beszédadatbázison végezték [28].

Magyar vonatkozásban Lukács Roland [29] ismételte meg a vizsgálatot német [28] és magyar beszédadatbázison. Munkájában külön vizsgálta a jellemzőhalmazokat (formáns frekvenciák, MFCC értékek, mel-sávos energiaszintek), amelyekkel hasonló, esetenként jobb eredményeket ért el. Ennek valószínű okai lehetnek a beszédadatbázisban rendelkezésre álló felvételek minősége és mennyisége.

Továbbá Roland eredményeiben alig tapasztalt eltérést a magyar és német felvételek depresszió súlyosságának becslésében, amivel rávilágított egy nyelv független eljárás kialakítására.

Viszont az eljárás egyik hátrányaként említhető, hogy a sajátértékes reprezentáció nagyméretű. illetve kevésbé teszi lehetővé több jellemző együttes alkalmazását. Ebből kiindulva diplomamunkában a sajátértékek helyett a korrelációs struktúrát, mint képreprezentációt alkalmazok gépi tanuló eljárás felhasználásával.

2.2 Parkinson kór

A főleg idősebb személyeknél előforduló Parkinson-kór egy olyan neurológiai degeneratív betegség, amely következtében az agyi dopamintermelő sejtek elpusztulnak. Az ingerületközvetítő anyag hiányában a motoros pálya finom szabályozása megbomlik. Jellemző tünetei a nyugalomban jelentkező remegés (tremor), izomtónus merevség, illetve a belassult, akadozó mozgás. A betegség a hangszalagok, illetve az arc izmaira is hatással van, ezáltal megjelenik a beszédképzés során [30]. Előrehaladott állapotban a séta, beszéd és az egyszerűbb feladatok ellátása is nehézzé válhat. A diagnózis felállítására jelenleg a páciens kortörténete, tünetei illetve neurológiai vizsgálatok alapján történik. Vér- és laborvizsgálat más betegségek kizárását segítik [31]. Korai diagnosztikájának fontoságát adja, hogy jelenleg gyógyíthatatlan betegség, aminek előrehaladását és tüneteit pusztán enyhíteni lehet [32]. Legeredményesebb gyógyszeres készítményként a dopamin intermedierjét a levodopa-t alkalmazzák.

A betegség súlyosságára leggyakrabban az UPDRS (Unified Parkinson's Disease Rating Scale, Egységesített Parkinson Pontozóskála), illetve a H&Y (Hoehn és Yahr) skálát alkalmazzák. Az UPDRS egy 0-tól 176-ig terjedő pontozási skála, amit a tünetek jelentkezése és súlyossága határoz meg. A pontszám összeáll a mentális állapot/viselkedés, mindennapi élethez való aktivitással kapcsolatos tünetek és a motoros tünetek felméréséből [33]. A H&Y skála ezzel szemben pusztán 1-től 5-ig terjed, ahol az 5-ös jelöli a legsúlyosabb állapotot. Továbbá a skála nem lineáris, amiből következik, hogy 2-es H&Y nem jelent kétszer akkor súlyos tüneteket, mint az 1-es H&Y [34]. Részleteiben az 1-es H&Y egyoldali tüneteket jelent, mint például a tremor. 2-es a mindkét oldalt megjelenő tüneteket sétálási nehézségek nélkül. A 3-as kategóriában már enyhe sétálási zavar jelenik meg a kétoldali tünetek mellett. A 4-es a közepes, míg az 5-ös a súlyos sétálási zavarok melletti kétoldali tünetek kategóriája.

A Parkinson kór kialakulásánál a betegek mintegy 90%-nál elváltozások jelentkeznek a beszédproduktumban is, amivel lehetőség nyílik egy non-invazív beszéd alapú diagnosztikai eszköz kutatására [35], [36].

A korai hang alapú kutatásokban a beteg kitartott zönge hangját vizsgálták [37]. E módszer fontos információtartalmát az adta, hogy egy zönge képzéséhez a hangszalagok izmainak aktív munkája szükséges. Parkinson kór esetén bizonytalanság/akadozás figyelhető meg az izmok mozgatásában. Ezt felhasználva 90% körüli felismerési pontossággal találkozhatunk az irodalomban kitartott „a” hang esetén SVM-et és Véletlen Erdő (Random Forest) algoritmust felhasználva [38]. Adatbázisukban mindössze 43 alany felvétele szerepelt: 10 egészséges és 33 Parkinson kóros.

A kitartott „e” hangnál 92, míg a kitartott „i” hangnál 72%-os pontosságot értek el rendre 20 egészséges - 20 Parkinson-kóros, illetve 50 egészséges és 50 Parkinson-kóros felvételt felhasználva. A felsorolt eredmények bizonytalanságát adhatja a kis mintaszámú adatbázis [39].

2015-ös kutatásban lineáris kernelű SVM modellel értek el 92%-os pontosságot a Parkinson kór észlelésében. Ők az első 12 MFCC értéket használták fel kitartott „a” hang vizsgálatával. Adatbázisuk 34 személyt tartalmazott, akik közül 17-en voltak Parkinson kórosok [40].

Egy 2018-as kutatásban 75-86%-os pontosságot érték el különböző algoritmusok használatával: Döntési Fa (Decision Tree), Mesterséges Neurális Háló (Artificial Neural Network), Véletlen Erdő, SVM. Adatbázisukban 5826 résztvevő szerepelt az mPower kutatás jóvoltából [41]. Az adatbázisban okostelefonnal 10 másodpercig kitartott „a" hangot rögzítettek.

A felsorolt kutatási eredmények alapján leírható, hogy a kitartott zönge hang eredményes jelzője a Parkinson kór betegségnek. Hátránya viszont, hogy azok csak a hangszalagok működéséről és az állandósult állapotról adnak információt, míg a beszédnél kialakuló folyamatos mozgásról és működésről nem. Így megjelentek hasonló ígéretes eredményekkel kutatások, ahol a páciens egy előre megadott szövegrészt olvasott fel vagy spontán beszélt.

Amerikai és német felvételeket felhasználva vizsgálták a Parkinson-kór felismerési pontosságát a Haifai Egyetemen (Israel). 16 Parkinson kóros és 14 egészséges alany szerepelt az amerikai felvételben. A német adatbázisban 98 alany hangfelvétele szerepelt. Az alanyok fonetikailag kiegyensúlyozott szöveget olvastak fel.

Három vizsgálatot végeztek: tisztán amerikai, tisztán német felvételekkel végzett osztályozás, illetve cross-country megközelítés. Ez utóbbi esetben először a német adatbázissal tanították az osztályozót és amerikai tesztelték, illetve elvégezték a fordítottját is (amerikai adatbázissal tanították). 94% pontosságot értek el az amerikai felvételeknél formánsok, kor és nem felhasználásával. A német adatbázison 85% tudták elkülöníteni a betegséget a formánsok felhasználásával [42].

Egy újabb cikkükben konvolúciós neurális hálót alkalmaztak nyers beszéden. Adatbázisuk 43 páciens és 9 egészséges hangfelvételét tartalmazta. A résztvevők a „Rainbow passage" szöveget olvasták fel. Osztályozásnál az UPDRS skála értékeit használták (1, 1.5, 2, 2.5, 3, 4). Bináris osztályozást alkalmaztak, amikor a felsorolt 7 csoport közül mindig csak kettőt használtak fel. Ezalapján az osztályozási pontosság értéke 60 és 85% között változott [43].

Összességében a mély tanuló technikák megjelenésével leegyszerűsödhet a beszéd alapú diagnosztika kutatása. Egy-egy kitartott zönge hang helyett vizsgálat alá vehető a minnapokhoz közelebb álló spontán vagy felolvasott beszéd. Az előfeldolgozás is egyszerűsödhet, illetve fonéma szintű szegmentálás akár teljesen ki is hagyhatóvá válhat.

A kutatások alapján a zönge magánhangzók formáns frekvenciáinak bizonyos kombinációja a leggyakrabban használt jellemző a felismerésben. Általában az „a”, „u”, „i” hang első és második formáns frekvenciája az alkalmazott.

Jelen kutatások között nem találkoztam autó-és keresztkorrelációs megoldásokkal a Parkinson-kór felismerésében. Azonban tudományos cikket találtam a korreláció alapú jellemzőkiválasztással megvalósított osztályozásra [44]. A kutatásban több beszédfeladatot hajtottak végre az alanyok, mint kitartott zönge hang, szó kimondása, felolvasás és monológ. Három akusztikai jellemző modellen Szupport Vektor Gép segítségével, leave-one-out kereszt-validációval végezték az osztályozást. A jellemzőkiválasztás lényege, hogy kiválasztjuk azon jellemzőket a jellemzőhalmazból, ami a legjobban korrelálnak a felvétel osztályával (egészséges/Parkinson kóros).

2.3 Diszfónia

A diszfónia egy kortól és nemtől függetlenül jelentkező betegség, ami a beszéd minőségének változását okozza. Gyakran a rekedtség szinonimájaként használják, ami félrevezető. Ugyanis a diszfónia a hanghasználat túlzott teljesítményigényéből származó komplex funkciózavar. Ezzel szemben a rekedtség az erőltetett hangképzés indukálta átmeneti fáradás, ami legkésőbb 24 órán belül megszűnik. Mathienson etiológiai felosztása alapján a diszfónia viselkedési és organikus kategóriába sorolhatók. Történelmi mélységben viták folynak e két osztály elkülöníthetőségéről. Szakemberek szerint a viselkedési (korábban funkcionális) diszfónia átmeneti jelző, amikor a hangképzés zavara fennáll, viszont organikus elváltozás nem figyelhető meg [45]. Megnövekedett gyakorisággal figyelhető meg a hangképzésüket erősen igénybe vevő személyek, mint például operaénekesek, tanárok esetén. Közvetlen befolyásolja a beteg életminőségét, ami a társadalomtól való elszigetelődést, depressziót, szorongást is kiválthat. Daganatok kísérő tüneteként is megjelenhet, ami nem megfelelő diagnosztikája és kezelése végzetes is lehet [46], [47].

A diszfónia diagnózisának felállítását általában egy, a hangra specializálódott orvos végzi, aki szubjektíven értékeli a betegség súlyosságát [48]. Belátható, hogy ha egy specialistákból álló csoport tagjai függetlenül pontoznának egy diszfóniás hangot, úgy ezen értékek szórnának. Így szükségessé válik a konzisztens pontozás.

Magyarországon széles körben használt objektív, német eredetű skála a RBH (Roughness, Breathiness, Hoarseness) a gégészeti elváltozások leírására. A skála a hang

érdességét, levegőségét és rekedtségét pontozza 0 és 3 közötti egész számokkal, ahol 3-as jelenti a legsúlyosabb kategóriát [49].

Thomas Law és munkatársai 2012-ben kutatásukkal megállapították, hogy folyamatos beszédből megbízhatóbb értékelések születnek, mint kitartott zönge hang esetén. Továbbá a folyamatos beszéd akusztikai vizsgálatával látható az alapfrekvencia, szünetek változása, illetve több hang is elemezhetővé válik.

A leggyakrabban használt beszédakusztikai jellemzők a diszfóniával kapcsolatban a jitter, shimmer és a harmonics-to-noise (HNR) [50], [51], [52]. Egy 2008-ban megjelent tanulmány alapján megállapították, hogy statisztikailag szignifikáns különbség van a jitter és shimmer paraméterekben kitartott zönge hang esetén az egészséges és patológiás felvételek között. Viszont a folytonos beszédnél már nem sikerült szignifikáns eltérést kimutatni (5% szignifikancia szint mellett) [53]. Emellett viszont a jel-zaj viszony (Signal-to-Noise Ratio - SNR) mind kitartott hang, mind folyamatos beszéd esetén alkalmas jellemzőnek bizonyult az egészséges és patológiás elkülönítésben. Pár évvel korábban MFCC és alapfrekvencia jellemzőket is eredményesen alkalmaztak Rejtett Markov Modellen (HMM) a betegség felismerésében [54].

Thomas Dubuisson és munkatársainak munkájában megjelent a jellemzők közötti korrelációjának használata. A beszédből 87 jellemzőt (köztük alapfrekvencia, MFCC értékek, formáns frekvenciák, jitter, shimmer, HNR…) származtattak, majd közöttük létrehozták az auto- és keresztkorrelációs struktúrát. Első megközelítésként képezték a mátrixok felső háromszögének (triangulárisának) összegét és vizsgálták az összegek eloszlását, viszont ez nem volt eredményes az egészséges – diszfóniás elkülönítésben. Megoldásként jellemzőkiválasztást alkalmaztak a korrelációs struktúrán, aminek együtthatóit, mint jellemzők használták fel. Ezzel 94%-os pontosságot értek el az elkülönítésben [55].

A mély neurális hálók használata is hamar elterjedt a diszfónia osztályozásában 40 - 96%-os elkülönítési pontossággal [56], [57].

Maria E. Powell és munkatársai által publikált kutatásban konvolúciós neurális hálót alkalmaztak bináris osztályozásra akusztikai jelből létrehozott spektogrammon. 10 halmazos kereszt validációs technikát alkalmaztak. Eredményeik 58 és 90% pontosság között változtak a diszfónia alcsoportjától függően (7 alcsoportot vizsgáltak) [58].

Összefoglalva elmondható, hogy jelen betegségnél is széles tartományban változik a diszfónia felismerési pontossága (58-94%), ami ugyancsak mintaszámtól, tanuló algoritmustól és az alkalmazott eljárástól függően alakultak.

A jitter és shimmer jellemzőket használták kitartott hangok esetén, míg folytonos beszédnél energia alapú jellemzőket, mint például az MFCC értékek.

Thomas Dubuisson munkájában már alkalmazott korrelációs struktúrát, amin ő jellemzőkinyerést végzett. Továbbá Maria E. Powell kutatásában már alkalmazta a konvolúciós neurális hálót. Viszont az irodalomban e kettő kombinációjára a diszfónia esetén még nem találkoztam.

2.4 Az adatbázison eddig elért eredmények

A Távközlési és Médiainformatikai Tanszék Beszédakusztikai Laboratóriumában már születtek a három betegségcsoport felismerésére eredmények.

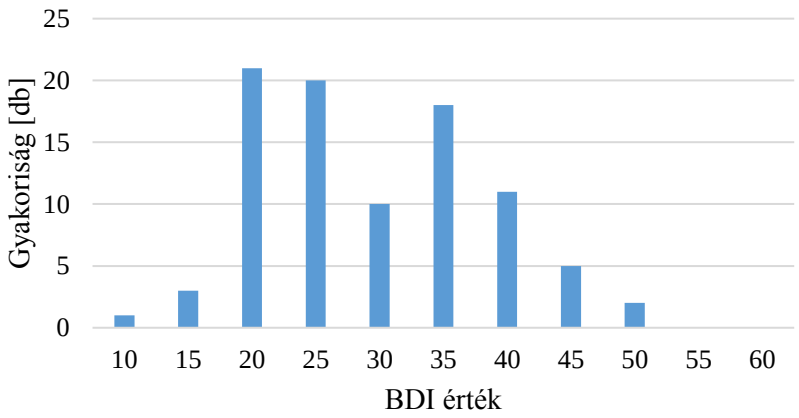
2018-ban publikált cikkükben 263 hangképzési rendellenességgel, 55 depresszióval és 76 Parkinson-kórral rendelkező páciens felismerését vizsgálták 190 egészséges kontroll felvétele mellett [59]. Minden személy az „Északi szél és a Nap” című fonetikailag kiegyensúlyozott mesét olvasta fel. Vizsgálatukban létrehoztak Williamson munkája alapján korrelációs struktúrákat, amikben négy eltolást alkalmaztak a vektorok között: 1, 2, 4 ,8. Alapfrekvencia, mel-sávos energiaértékek és az MFCC jellemzőkre 10-szer alkalmazták az eltolást a korrelációs struktúrában, míg a formáns frekvenciák esetén 30-szor. A legjobb pontosságot a formáns frekvenciákkal 65,9%, a mel-sávos energia értékekkel 74,1%, az MFCC értékekkel 69,4 % érték el a 4 csoport esetén radiál bázisfüggvényű SVM-el. Az összes jellemzőt együtt alkalmazva 77,5%-os pontosságot kaptak.

2019-ben a fent bemutatott adatbázison, más vizsgálódási eljárást használva publikáltak eredményeket 4 illetve 6 osztályos elkülönítésről [60]. Fonéma szintű szegmentálást követően mintegy 270 jellemzőt felhasználva végeztek osztályozást k legközelebbi szomszéd (k-NN), neurális háló (NN) és SVM algoritmusokkal, 10 halmazos kereszt-validációt alkalmazva. 4 osztály esetén az előző cikkben leírt 78%-os pontosságról 88%-ra javult az elkülönítés.

3 Beszédadatbázis

Az egészséges és beteg személyektől származó hangfelvételek gyűjtését és a beszédadatbázis felállítását a Távközlési és Médiainformatikai Tanszék Beszédakusztikai Laboratóriumának munkatársai végezték. Minden felvétel elkészítése előtt a páciens (és a kontroll alany is) beleegyező nyilatkozatot írtak alá, amiben hozzájárultak a hangjuk kutatási célokra való felhasználásához. Az adatbázisokat és felvételeiket a diplomamunka elvégzésére használtam fel.

A kutatáshoz egyrészt a folyamatosan bővülő Magyar Depressziós Beszédadatbázist használtam. A depressziós személyektől származó beszédminták gyűjtésében a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikájának munkatársai segítettek. Olyan felvételekkel dolgoztam, ahol a vizsgált személyek orvosi igazolás alapján nem szenvedtek a depresszión kívül más olyan betegségben, ami befolyásolná a beszédüket. A felvételek gyűjtésénél a labor munkatársai törekedtek arra, hogy a beszélők lefedjék a depresszió súlyosságának különböző fokozatait. Depressziós beszédfelvételek körülbelül egyenletes eloszlással szerepeltek a BDI-II (Beck Depression Inventory-II) által definiált depressziós súlyosság szerinti kategóriák között, úgy, mint az enyhe depresszió, közepes depresszió és súlyos depresszió (***3.1. ábra***).



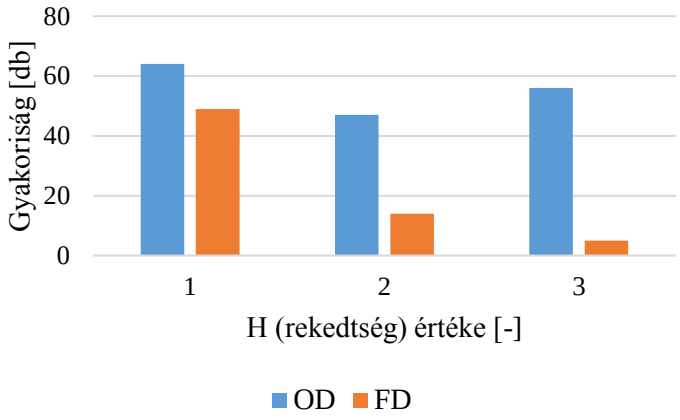
3.1. ábra: Depresszió súlyosságának eloszlása a felhasznált felvételek között.

A munkám során 91 felvételt használtam fel, ezeket a továbbiakban **DE**-vel (**DE**pressed) jelölöm. Ez 58 női és 33 férfi felvételt tartalmazott, átlagos életkoruk 40,9 év, szórása 13,9 év. 4 nő és 1 férfi esetében nem álltak rendelkezésre kor adatok.

A patológiás hangminták felvételeit az Országos Onkológiai Intézet Fül-orr-gégészeti osztályának járóbeteg ellátói részlegéről gyűjtötték. A felvételek között előforduló betegségek a funkcionális diszfónia, recurrens paresis, hangképző szervrendszeri tumorok, gasztroesophageal reflux, krónikus gégegyulladás, bulbar paresis, amiotrófiás laterálszklerózis, leukoplakia, spazmodikus diszfónia. Ezeket összességében két nagyobb csoportba válogattuk: organikus diszfónia (**OD**) és funkcionális diszfónia (**FD**).

Az OD-ből 167 felvételt (74 férfi és 93 nő) használtam fel, míg az FD-ből 68-at (20 férfi, 48 nő). Átlagos életkoruk rendre 51,6 és 55,8 év, szórásuk 14,4 és 16,1 év. E két csoportnál kevés koradat, mindössze 26 (FD) és 19 (OD) személynél állt rendelkezésemre.

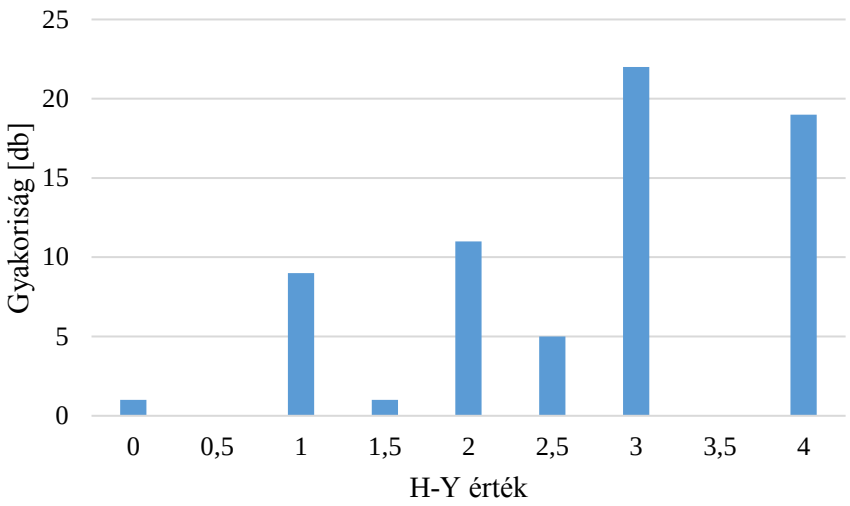
A páciensek rekedtségi értéke alapján a felhasznált felvételek eloszlása a ***3.2. ábra***-n látható.



3.2. ábra: A dizfóiás felvételek eloszlása rekedtségi (H) érték alapján.

A Parkinson-kórral (**PD**) diagnosztizált betegek hangfelvételei két helyről lettek gyűjtve: a Semmelweis Egyetemről (25 felvétel) és a Virányos Klinikáról (55 felvétel).

A Parkinson-kórral diagnosztizált páciensek felvételei közül összesen 80-at (43 férfi, 37 nő) használtam fel, amik H&Y érték szerinti eloszlása a ***3.3. ábra***-n látható. Átlagos életkoruk 64,6 év, aminek szórása 9,3 év.



3.3. ábra: Felhasznált Parkinson-kóros felvételek H-Y érték szerinti megoszlása.

A beteg hangfelvételek mellett egészséges felvételekből álló adatbázist is létrehoztak a laborban, mint kontroll csoport (**HC**). A felvételek készítésénél az egészséges személyek saját nyilatkozatuk szerint semmilyen olyan betegségben nem szenvedtek, amik befolyást gyakorolnának a beszédükre. Egészséges beszédmintákból 140 felvételt (85 nő, 55 férfi) használtam a diplomamunkám során. Átlagos életkoruk 50,3 év, aminek szórása 17,9 év.

Az adatbázisok felvételeiben az „Északi szél és a Nap” című közel 1 perces mesét olvasták fel a személyek. A hanganyagok 44,1 és 16 kHz mintavételezési frekvenciával kerültek felvételre csíptethető mikrofonnal csendes helységben. A felvételek 16 biten lettek eltárolva.

Összefoglalva a felhasznált beszédminták mennyisége a ***3.1 táblázatban*** látható. A negyedik oszlopban feltűntettem a százalékos előfordulásukat a felhasznált teljes halmazhoz képest.

Az arányszámokból látható, hogy az adatbázisban a csoportok mintaszáma nem kiegyenlítettek. A legnagyobb részarányt az általános hangképzőszervi elváltozásban szenvedő alanyok száma képviseli, azon belül is az organikus diszfónia aránya magasabb a funkcionális diszfóniánál.

A nemek számát tekintve szinte minden csoport esetében a női alanyok voltak többen, a Parkinson-kór kivételével.

3.1. táblázat: A dimplomamunkában alkalmazott beszédminták mennyisége.

Megnevezés		**Jelölése**		**Mennyisége [db]**		**Arány [%]**		**Életkor [év] (szórás)**	**Nemek [nő/férfi]**
Egészséges		HC		140		25,6		50,3 (17,9)	85/55
Depressziós		DE		91		16,7		40,9 (13,9)	58/33
Parkinson-kór		PD		80		14,7		64,6 (9,3)	37/43
Általános hangképző-szeri elváltozás	**Funkcionális diszfónia**	UD	FD	235	68	44,9	12,4	55,8 (14,4)	48/20
	Organikus diszfónia		OD		167		30,6	51,6 (16,1)	93/74
		Összesen		546		100			321/225

Az életkor szempontjából az látható, hogy a depressziós adatbázis átlag életkora alacsonyabb, a Parkinson-kóros adatbázisé pedig nagyobb, mint az egészséges kontroll átlag életkoránál. A két diszfónia csoport átlag életkora hozzávetőlegesen megegyezik az egészségesével, viszont a kevés adat miatt ezt nem lehet biztosra állítani.

4 Módszerek

Ebben a fejezetben először ismertetem az általam fejlesztett módszer kidolgozásához szükséges irodalmi hátteret, majd leírom a folyamat tényleges megvalósítását, végül a vizsgálati feladatokat A hangfelvételtől az osztályozásig tartó folyamat három nagyobb blokkra választható szét: beszédakusztikai jellemzők kinyerése, korrelációs struktúra létrehozása és az osztályozás (CNN).

4.1 A beszéd leíró jellemzőinek meghatározása

Az emberi kommunikáció egyik jelentős részét képezi a beszéd, ami hanghullámok formájában – általában levegő közegben – terjed. A beszédet felépítő hangokat a tüdőből kiáramló levegő hozza létre az által, ahogy keresztüláramlik az elernyedt vagy megfeszített hangszalagokon, illetve arc/orr üregein. A hang további formálását valósítja meg a nyelv és száj aktív munkája [61].

Magánhangzók és zöngés mássalhangzók képzésénél a hangszalagok rezgésbe jönnek, létrehozva egy közel fűrészfog jellegű hangnyomás-idő függvényt (suttogás esetében nem alakul ki zönge). Ez tartalmaz alap és felhangokat, mely felhangok az alaphang egész számú többszörösei. Az így kialakult színképet tovább befolyásolja a változó méretű hangképző csatorna. Ez az üregrendszer több rezonanciahellyel rendelkezik (rezonanciafrekvencia képző helyek), amik környezetében bizonyos részhangok intenzitása megnő, másoké lecsökken [62], [63].

Az ilyformán létrehozott akusztikai beszédjel adott egyénre jellemző, mivel hordozza annak sajátosságait. A beszéd kialakítása bonyolult, irányított fiziológiai folyamatok összességét kívánja. Ez alapján kis változás a beszédképzési folyamat bármely pontjában jelentős változást képes okozni az egyén beszédproduktumában.

A beszéd időben folyamatosan változó jel, aminek feldolgozása bonyolult. Ennek egyik oka, hogy biológiai produktumként függ annak pillanatnyi állapotától. Például az egyén kitartott zöngehangja is különböző periódusokat tartalmaz időről időre. Továbbá a beszéd több olyan elemet tartalmaz, amik miatt nem nevezhető stacionárius jelnek.

A beszéd vizsgálata történhet egy meghatározott időablak alkalmazásával, mely kicsíny időtartományon belül közel stacionáriusnak tekinthető a beszédjel. Rövid

ablakszélességnél az időbeni változásokat lehet figyelemmel követni, míg hosszabb időablak mellett a frekvenciafelbontás láthatóbb.

Ebben az időablakban elvégezhető az úgynevezett spektrumelemzés, amivel egy beszédszakaszra meghatározott időablakon belül teljesítmény/energiaspektrum határozható meg. Ezzel tulajdonképpen egy teljesítmény/energia – frekvencia diagram állítható elő meghatározott időablakra Fourier transzformáció segítségével. A beszéd adott szakaszára nézve a módszer megadja a beszéd frekvenciakomponensek időbeli változását [62], [64].

Az időablakos megoldással meghatároztam a következő beszédakusztikai jellemzőket [62], [63], [64]:

- ***Mel sávos energia értékek (Mel-Band Energy Values):*** Az emberi hallásmechanizmust pontosabban leíró mel frekvencia értékekké számolható át a hagyományosabb értelemben vett frekvencia az ***1. egyenlet*** alapján. A mel skálán sávok határozhatók meg (mel-sáv), amiken, mint szűrőkön engedhető át a beszéd teljesítmény/energia színképe.

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

- ***Mel Frequency Cepstral Coefficient (MFCC):*** meghatározása a teljesítményspektrumból úgy történik, hogy meghatározott mel szűrőn belül összegezhetők az energiaértékek. Ezt mel spektrumnak is szokás nevezni. Majd logaritmusának diszkrét koszinusz transzformációját kell számolni.
- ***Formáns frekvencia:*** A hangképző csatornák által létrehozott kiemelés, ahol a képzett hangok egy részének megnő, másoknak lecsökken az intenzitásértéke. A beszédjelből adott ablakszélesség mellett elvégezhető a Fourier transzformáció, amivel teljesítmény - frekvencia diagram állítható elő. Ezen frekvenciaösszetevők burkológörbéjének maximumhelyei nevezhetők formáns frekvenciáknak.
- ***Formáns frekvencia sávszélessége:*** A teljesítmény - frekvencia diagramban a rezonancia frekvenciacsúcstól (formáns frekvencia teljesítményértéke) -3 dB csökkenésnél a burkoló görbe szélessége frekvencia mértékegységben a formáns frekvenciák sávszélességei.

4.2 A korrelációs struktúra előállítása

Az ***4.1 fejezet*** szerint előállított idővektorok ábrázolhatók egymás függvényeként. A két beszédakusztikai jellemző ponthalmazának kapcsolatát az alább bemutatott korrelációs együttható értékkel jellemeztem.

A korreláció egy olyan statisztikai módszer, ami két változó közötti lineáris összefüggést ír le. A korrelációs számszerű leírására korrelációs együtthatót alkalmaznak, ami egy dimenzió nélküli mérőszám. A változók között erős lineáris kapcsolatot feltételezünk, ha a korrelációs együttható értéke közel van a -1 vagy +1 értékhez. 0 korrelációs érték esetén a két változót lineárisan függetlennek tekintjük egymástól A szakirodalomban leggyakrabban a Pearson-féle korrelációt alkalmaznak, amit a ***2. egyenlet*** ír le [68].

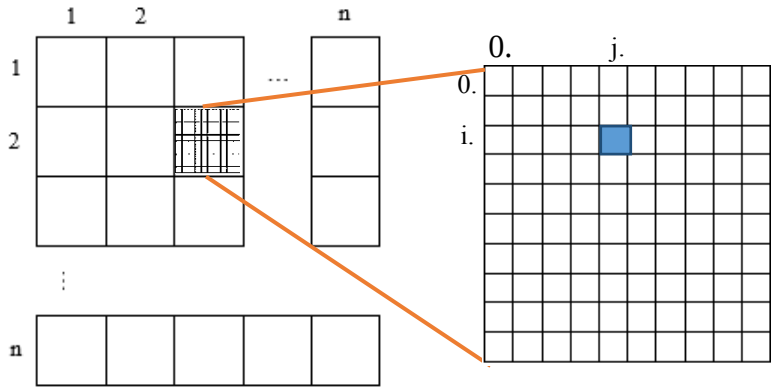
$$r = \frac{\sum_{i=1}^{n}(x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x - \bar{x})^2 \cdot (y - \bar{y})^2}} = \frac{\sum_{i=1}^{n}(x - \bar{x}) \cdot (y - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \tag{2}$$

Az x és y változók, $\bar{x}$ és $\bar{y}$ átlagos értékeik, σ_x és σ_y a változók tapasztalati szórásai, *n* a változók száma.

A korreláció mértéke lehet

- nagyon erős pozitív (negatív): 1,00 - 0,90 ([-1,00] - [-0,90]),
- erős pozitív (negatív): 0,90 - 0,70 ([-0,90] - [-0,70]),
- mérsékelt pozitív (negatív): 0,70 - 0,50 ([-0,70] - [-0,50]),
- alacsony pozitív (negatív): 0,50 - 0,30 ([-0,50] - [-0,30]),
- elhanyagolható: 0,30 - 0,00.

A beszédakusztikai jellemzők által meghatározott korrelációs együttható értékek mátrixos formába rendezhetők. Kis módosítással a korrelációs együttható értékekből létrehozott – autó és keresztkorrelációs – mátrix felépítését az ***4.1. ábra*** mutatja be. A rajta lévő számok 1-től n-ig jelölik a korrelációs mátrix bemeneti jellemzővektorok számát (például 1: az első MFCC, 2: a második MFCC … stb.). Az így felépülő struktúra szimmetrikus, főátlóiban autó-, mellékátlóiban keresztkorrelációs együttható értékek szerepelnek.



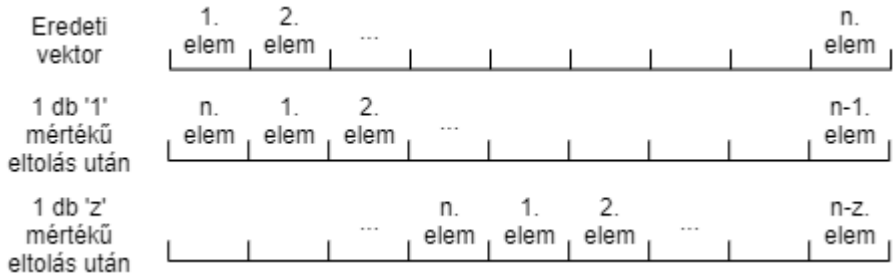
4.1. ábra: A korrelációs struktúra szerkezete (ábra bal oldala). A struktúra celláiban almátrixok találhatók a vektoreltolásoknak megfelelően (ábra jobb oldala).

A korrelációs struktúra alapvetően $n \cdot n$ darab cellából épülne fel (***4.1. ábra*** bal oldala). Viszont a jelen vizsgálatban elemeltolásokat végeztem a beszédakusztikai jellemzők vektoraiban. Így a cella helyét olyan négyzetes altmátrix veszi át, aminek nagysága az eltolások számával megegyező (*dt*).

Így minden almátrix nagysága $dt \times dt$, a teljes struktúráé pedig $(n \cdot dt) \times (n \cdot dt)$. Egy almátrixot kiválasztva a struktúrából egyértelműen megállapítható, hogy mely két jellemzővektor alkotja a felhasznált jellemzővektor halmazból.

Az almátrixok felépítése (***4.1. ábra*** jobb oldala) úgy történik, hogy az almátrix első cellája (0. sor, 0. oszlop) a két eredeti (eltolás nélküli) jellemzővektor korrelációs együttható értékét tartalmazza. Az almátrix első sorának második cellájánál már az első jellemzővektor elemeit meghatározott mértékkel eltoltam. Az eltolás műveletét az ***4.2. ábra*** szemlélteti. Minden következő cellában az elemeltolást megismételtem. Így egy 10 × 10-es almátrixban az első sor utolsó cellájához az első jellemzővektor elemeit 9-szer toltam el meghatározott mértékben.

Hasonlóan elvégezve a második jellemzővektorral is, az almátrix egy tetszőleges i. sorában és j. oszlopában található korrelációs együttható értéke megállapítható. Mégpedig úgy, hogy az első jellemző vektor *i*-szer vett eltolása és a második jellemzővektor *j*-szer vett eltolása melletti korrelációs együttható értékét tartalmazza (***4.1. ábra*** jobb oldala).



4.2. ábra: Vektorok elemeltolási módszerének szemléltetése 1-es mértékű eltolás esetén.

Az így kialakult teljes korrelációs struktúra is szimmetrikus. Főátlójában 1-es értékek szerepelnek, mivel azokban az esetekben mindkét jellemzővektor, ugyanakkora mértékben, ugyanannyiszor lettek eltolva és korreláltatva egymással.

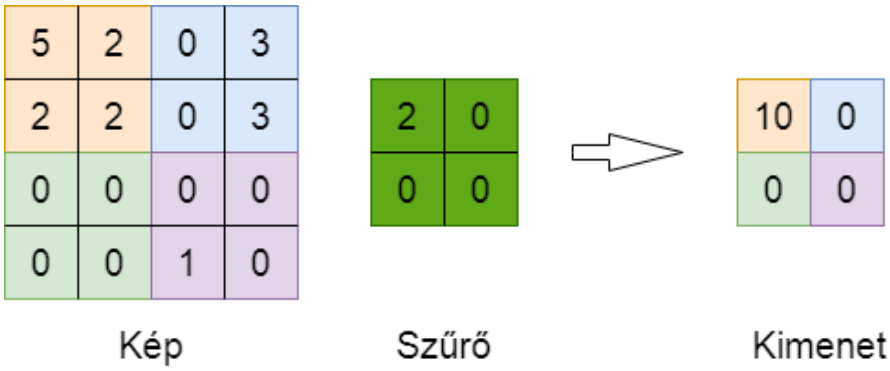
4.3 A Konvolúciós Neurális Háló (CNN) szerkezete

Az előző fejezetben létrehozott korrelációs struktúra reprezentálható egy $(n \cdot dt) \times (n \cdot dt)$ nagyságú képként. Az így kialakított „képek” bemenetként használhatók egy konvolúciós háló esetén, ami automatikus osztályozást valósíthat meg. A konvolúciós neurális háló általános alkalmazásának célja, hogy a bemenetére adott képen találjon információval bíró részleteket – mintázatokat –, ami alapján az adott probléma (osztályozás, regresszió, … stb.) maximálisan megoldható. Az eljárás előnye jelen esetben sok más algoritmussal szemben, hogy a háló feladata a fontos információ kiszűrése a korrelációs mátrixból, amivel majd az elkülönítést végzi.

A konvolúciós hálókból már léteznek előre megalkotott és tanított struktúrák, mint például az AlexNet vagy a VGG-16 [69]. Viszont ezek nem specifikusak a jelen diplomamunkában bemutatott problémára. Így a munkámhoz rétegenként létrehoztam a saját alapmodellem a következő elemek felhasználásával [27], [28], [29], [30]:

Konvolúciós réteg: különböző kernelek (szűrők) segítségével információkinyerést végez a bemenetére kapcsolt képeken. A kimeneti kép egy új pixelértéke előáll a szűrő elemeinek és a letapogatott képrészlet elemeinek lineáris kombinációjaként. Ezen módon az új kép összes pixele előállítható, mialatt a szűrő átlapolva vagy át nem lapolva végiggördül a bemeneti képen. Az ***4.3. ábra*** az át nem lapoló szűrő példáját hivatott bemutatni, ahol a szűrő minden képszegmens első pixelértékének kétszeresét nyeri ki.

A tanítás során a súlyok úgy állítódnak be, hogy a legeredményesebb felismerést szolgáltató filter(ek) és hozzá tartozó neuron(ok) jut(nak) érvényre. A modellben a kernel mérete és léptetése beállítható.



4.3. ábra: Szűrők alkalmazása a konvolúciós neurális hálóban. Az át nem lapoló szűrő minden 2 × 2-es képszegmensnek az első cellájának kétszeresét viszi tovább.

Aktivációs függvény: neuronok (vagy rétegek) közötti információáramlást valósít meg. Ezzel meghatározott függvénykapcsolatot teremthetünk a neuronok/rétegek kimenetei és a következő neuronok/rétegek között. Az egyik ilyen legegyszerűbb aktivációs függvény például a lineáris függvény, ami lineáris megfeleltetést biztosít a bemenet é a kimenet között. Munkám során kétféle aktivációs függvényt használtam fel. ***ReLU:*** egység meredekségű aktivációs függvény, amely 0 alatti bemenetei értékeket 0-val feleltet meg. 0 feletti értékeket változtatás nélkül átenged. ***SoftMax függvény (3. egyenlet):*** más néven normalizált exponenciális függvény, ami a bemeneti vektorának elemeit valószínűségi értékekké alakítja (0 és 1 között korlátos). Általában a neurális háló végén található, és az osztályozó kimeneti értékeit alakítja át.

$$f_{SoftMax} = \frac{e^{x_i}}{\sum_{i=1}^{n} e^{x_i}} \tag{3}$$

DropOut: a tanítás során meghatározott mennyiségű neuront véletlenszerűen kiválasztva átmenetileg figyelmen kívül hagy a hálóból. E módszerrel a rendszer összetettsége csökkenthető. Továbbá a háló túltanulásának elkerülésére is alkalmazzák a gyakorlatban. A modellben 0 és 1 közötti értékkel lehet megadni.

MaxPooling réteg: egy olyan diszkretizációs folyamat, ami a bemenetére kapott kép dimenzióját csökkenti. Ezzel hasonlóan csökkenthető a rendszer számításigénye, illetve szintén egy lehetőség a túltanulás lehetőségének csökkentésére. A MaxPooling egy olyan (át nem lapoló) szűrőt használ, ami a bemeneti kép körülhatárolt szegmenséből a maximum értéket kiválasztja. Ehhez a szűrő mérete és padding („kipárnázás”) típusa állítható be. Jelentése ez esetben az, hogyha a bemeneti kép mérete nem pontosan fér bele a szűrőbe, akkor a különbözet kipótolható például a szomszédos pixelértékekkel. Ez a ’same’ típusú kipárnázás, amit a modellemben is alkalmazok.

Flatten réteg: az előtte álló réteg több dimenziós kimenetét egy egydimenziós vektorba rendezi.

Dense réteg: Ez egy úgynevezett „Fully Connected Neural Network” réteg, amin belül egy réteg összes neuronja összeköttetésben áll a következő réteg összes neuronjával. Paraméterként a kimeneti „tér” adható meg, ami a mi esetünkben a csoportok száma lesz.

A háló elkészülését követően azt tanítani szükséges. Ez alatt az algoritmus inicializálja a szabad paramétereit, majd meghatározott számú iterációig azokat úgy állítja, hogy a lehető legközelebb kerüljön egy optimumhoz. Ez az optimum lehet az osztályozási pontosság maximalizálása, de lehet akár az eredeti és becsült érték közötti hiba minimalizálása is regressziós probléma esetén. A diplomamunkában felügyelt módszerű tanítást alkalmazok az osztályozás megoldására [74]. Ennek értelmében minden korrelációs struktúrához egy címke tartozik, amit az algoritmus láthat tanulás alatt.

A tanított hálózat eredményességét a definiált problémára egy független halmazzal mérhetjük (ami nem vett részt előzetesen a tanításban). Ezzel az eljárással vélekedni tudunk arról, ha az algoritmus alul- vagy túltanult, esetleg éppen megfelelően működik.

Mivel a hálózat a tanításból „ismeri” meg a kategorizálási problémát, ezért szükségszerű, hogy a tanító halmazt a lehető legnagyobb elemszámúra válasszuk meg. Jelen alkalmazásban ezért teljes kereszt validációt használtam, ami egy speciális esete a K-halmaz validációnak. Ez utóbbi alapján a bementet képező teljes mintahalmazból (egészséges és beteg alanyok korrelációs struktúrái) egy részhalmazt különítek el. Ez a teljes halmaz $\frac{1}{K}$-ad mennyisége, ami a háló tesztelését végzi. A visszamaradt $\frac{K-1}{K}$-ad mennyiségű mintát pedig tanításra használódik fel. K alkalommal megismételve az eljárást, K független eredmény nyerhető ki. Minden ciklusban a teszt halmaz és tanító halmaz egymáshoz képest diszjunktak. Teljes kereszt validációnál a K érték az adathalmaz elemszámával egyenlő. Így a teszthalmaz minden ciklusban egy darab minta.

A tanítás során a tanulási arány (learning rate) állításával van lehetőségünk szabályozni, hogy mekkora mértékben változzanak a hálók súlyai. Általánosan a tanulás aránya a tanítás elején a legnagyobb, majd csökken az iterációk számával, ahogy a hálózat egyre többet „látja” a tanító mintát. Viszont a nem megfelelő beállítása a modell helytelen tanulását eredményezheti. Például, ha a tanulás aránya már a tanítás elején alacsony,

akkor a modell „megakadhat” egy lokális szélsőértékre, aminél egy optimálisabb pont is elérhető lett volna. Ennek másik oldala, ha a paraméter túl nagy, ugyanis ekkor a modell nem fog az optimumba konvergálni. Ezért állítását érdemes lehet egy olyan függvénnyel összekapcsolni, ami a tényleges eltérést (hibát) méri.

Ennek a hibának a leírására költségfüggvényeket alkalmazhatunk, amiket optimalizációs függvénnyel kombinálva a tanulás aránya automatikusan szabályozható. A neurális hálóval többosztályos osztályozást végeztem, amihez az alábbi formulával a kereszt entrópia hibafüggvény meghatározható (***4. egyenlet***) [75].

$$Keresztentrópia = -\sum_{i=1}^{N} cl_i \cdot log(p_i) \tag{4}$$

Az egyenletben szereplő *cl* jelöli az osztályt vektor formában (adott osztály értékénél 1-es, többinél 0), míg *p* egy valószínűségi értéket.

4.4 Az osztályozás leírása

A konvolúciós háló osztályozásának leírására a kimenetekből először tévesztési mátrixot/táblázatot hoztam létre. A táblázat számszerűen tartalmazza, hogy az osztályozó melyik csoportra döntötte a tesztmintákat. A ***4.1. táblázat*** jelölései az alábbiak bináris osztályozást feltételezve [76]:

- ***TP (valós pozitív):*** Eredetileg pozitív mintát pozitívnak határozott.
- ***TN (valós negatív):*** Eredetileg negatív mintát negatívnak határozott.
- ***FP (álpozitív):*** Eredetileg negatív mintát pozitívnak határozott.
- ***FN (álnegatív):*** Eredetileg pozitív mintát negatívnak határozott.

4.1. táblázat: Tévesztési mátrix felépítése bináris osztályozásra. A + jelöli a pozitív, - a negatív mintát.

		Eredeti	
		+	-
Becsült	+	TP	FP
	-	FN	TN

A táblázat értékei alapján a klinikai gyakorlatra is levonhatók következtetések és megfontolások. Az álnegatív értékek jelölik azt, hogy a pozitívnak diagnosztizált páciens nem kap kezelést. Ez a legrosszabb esetben a páciens életébe kerülhet. Ezzel szemben az álpozitív esetben az negatív alany átesik további vizsgálatokon, amire külön nincs szükség. Ez utóbbinak költségvonzata van az egészségügyben.

Jelen esetben a bináris mellett többosztályos osztályozást is megvalósítok, amiből közvetlen többosztályos téveszti mátrixokat származtatok. Ennek reprezentációja látható a ***4.2. táblázat***-ban**.** A táblázatban *p* jelöli a becsült mennyiséget, aminek első indexe az eredeti osztály, második indexe a becsült osztály. Mátrixként szemléltetve a főátló tartalmazza az osztályozó helyes döntését. Ez esetben már kevésbé egyértelmű, hogy mi is nevezhető valós pozitív vagy éppen álpozitív esetnek.

4.2. táblázat: Tévesztési mátrix többosztályos osztályozás leírására Az osztályok a *Beszédadatbázis* fejezet szerintiek.

		Eredeti				
		DE	FD	HC	OD	PD
Becsült	DE	$p_{DE\|DE}$	$p_{FD\|DE}$	$p_{HC\|DE}$	$p_{OD\|DE}$	$p_{PD\|DE}$
	FD	$p_{DE\|FD}$	$p_{FD\|FD}$	$p_{HC\|FD}$	$p_{OD\|FD}$	$p_{PD\|FD}$
	HC	$p_{DE\|HC}$	$p_{FD\|HC}$	$p_{HC\|HC}$	$p_{OD\|HC}$	$p_{PD\|HC}$
	OD	$p_{DE\|OD}$	$p_{FD\|OD}$	$p_{HC\|OD}$	$p_{OD\|OD}$	$p_{PD\|OD}$
	PD	$p_{DE\|PD}$	$p_{FD\|PD}$	$p_{HC\|PD}$	$p_{OD\|PD}$	$p_{PD\|PD}$

A táblázat értékeiből – a binárishoz hasonlóan – levezethetők jellemzők, mint a pontosság, felidézés, precizitás és az F1 érték. Ezek előnye, hogy tömörebb módon (egy értékkel) írják le az osztályozást. Számolási képleteiket alább ismertetem a többosztályos osztályozás esetében (***5 - 9. egyenletek***).

Az osztályozási pontosság a tévesztési mátrix főátló elemeinek összege és az összes elem mennyiségének hányadosa [77]. Irányszám arra vonatkozóan, hogy az osztályozó mekkora arányban dönt helyesen. Hátránya, hogy nem veszi figyelembe az adathalmazcsoportok mintaszám eloszlását.

$$Pontosság\ [\%] = \frac{p_{DE|DE} + p_{FD|FD} + p_{HC|HC} + p_{OD|OD} + p_{PD|PD}}{összes\ elem\ mennyisége} \cdot 100 \qquad (5)$$

A felidézés (X csoportra vonatkoztatva) a helyesen döntött elemek mennyisége és az eredeti csoport elemszámának hányadosa. Megmutatja, hogy adott csoport eredeti mennyiségéből mennyit döntött helyesen az algoritmus [78]. Kiemelten alkalmazzák azon területeken, ahol az álnegatív eseteknek nagy költségük van.

$$Felidézés_X\ [\%] = \frac{p_{X|X}}{p_{X|DE} + p_{X|FD} + p_{X|HC} + p_{X|OD} + p_{X|PD}} \cdot 100 \qquad (6)$$

A precizitás (X csoportra vonatkoztatva) a helyesen döntött elemek száma és az adott csoportra becsült minták mennyiségének hányadosa. Objektív mérőszáma annak, hogy az összes X-nek becsült mintából mennyi volt ténylegesen X. Fontos mérőszám, ha nagy a költségvonzata az álpozitív eseteknek [79].

$$Precizitás_X\ [\%] = \frac{p_{X|X}}{p_{DE|X} + p_{FD|X} + p_{HC|X} + p_{OD|X} + p_{PD|X}} \cdot 100 \quad (7)$$

Az F1 érték egy összetettebb mérőszám, ami felhasználja a precizitás és a felidézés értékét. A pontosság melletti használatát indokolja, ha egyenetlen a csoportok mintaszáma. A ***8. egyenlet*** az adott csoportra értelmezett F1 érték számolási menetét mutatja [80].

$$F1\ érték_X\ [\%] = \frac{2 \cdot Precizitás_X \cdot Felidézés_X}{Precizitás_X + Felidézés_X} \cdot 100 \quad (8)$$

A teljes adatbázisra vetített F1 érték (későbbiekben makro F1 érték) számolható a csoportokhoz tartozó F1 értékek átlagával. A ***9. egyenlenél*** látható ez, ahol *K* az osztályok mennyisége.

$$F1\ érték_{teljes}\ [\%] = \frac{1}{K}\sum_{i=1}^{K} F1\ érték_i \quad (9)$$

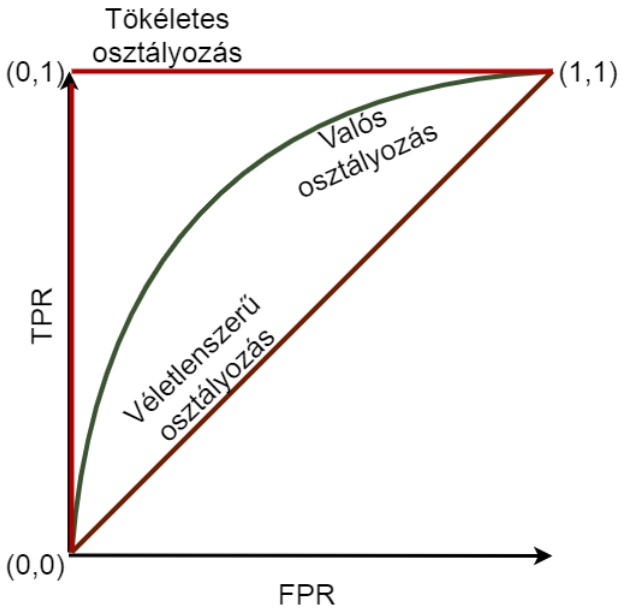
Bináris osztályozás esetén különböző komparátorértékek (döntési határok) megválasztásával szabályozható az osztályozó affinitása a pozitív és a negatív csoportokra. Minden döntési határ mellett meghatározható a valós pozitív arány (TPR - ***10. egyenlet***) és az ál pozitív arány (FPR - ***11. egyenlet***), amiket ábrázolva előállítható a ROC diagram (***4.4. ábra***) [81].

A valós pozitív arány meghatározható a helyesen pozitívnak döntött minták és a pozitív csoport mintaszámának arányával.

$$TPR = \frac{TP}{pozitív\ csoport\ mintaszáma} \quad (10)$$

Az álpozitív arány meghatározható az álpozitívnak döntött minták és a negatív csoport mintaszámának arányával.

$$FPR = \frac{FP}{negatív\ csoport\ mintaszáma} \quad (11)$$



4.4. ábra: Általános ROC diagram: véletlenszerű (barna szín), valós (zöld), tökéletes (piros) osztályozás.

Nagy komparátorérték mellett mind a FPR, mind a TPR alacsony, ugyanis alig történik döntés pozitív osztályba (diagram bal alsó része). A döntési korlát csökkentésével a görbéken felfelé haladunk és egyre több valós pozitív mintát könyvelhetünk el. Viszont egyre nagyobb mértékben a FPR is megjelenik. Bizonyos határérték alatt elérhetjük ugyan a 100%-os valós pozitív minta felismerését, viszont a helytelenül pozitívnak döntött minták száma is jelentősen megnő.

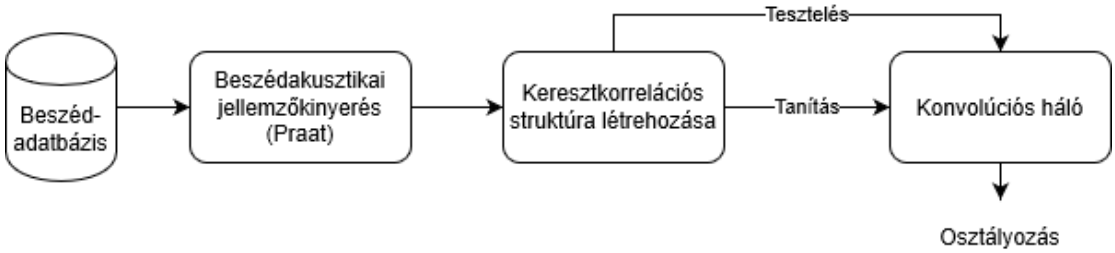
A diagramban a (0,0) pontból az (1,1) pontba húzott 45 fokos meredekségű egyenes jelképezi a véletlenszerű osztályozást. A valós osztályozó göbéje minél nagyobb területet foglal el a diagramból és minél közelebb helyezkedik el a (0,1) ponthoz, annál jobban megközelíti a tökéletes osztályozót.

A pontosság relatív javulását a triviális osztályozáshoz (baseline) képest határoztam meg. A triviális osztályozásnak az adott csoport eredeti mintaszámának a teljes adathalmazhoz képesti arányát definiáltam (lásd. ***3.1. táblázat*** arány oszlop). Ezt felhasználva a pontosság relatív javulását X csoportra nézve a ***12. egyenlettel*** határoztam meg.

$$Pontosság_{rel.|X}\,[\%] = \frac{Pontosság - X\ mintaszáma}{100 - X\ mintaszáma} \cdot 100 \qquad (12)$$

4.5 A megvalósított folyamat paraméterei

A betegségek felismerési lehetőségeinek vizsgálatához az ***4.5. ábra*** szerinti fő lépéseket valósítottam meg. Ez alapján, a felvételeken jellemzőkinyerést végeztem, a jellemzővektorok adott részhalmazának felhasználásával korrelációs struktúrát alakítottam ki, végül pedig az osztályozó algoritmust segítségével többosztályos, illetve bináris osztályozást végeztem. A lépéseket és a megadott paramétereket ebben az alfejezetben részletesen ismertetem. A folyamat mögötti infrastruktúrát a tanszék biztosította számomra.



4.5. ábra: A folyamat blokkvázlata. Elemei: jellemzőkinyerés, korrelációs struktúra létrehozása, osztályozás megvalósítása (tanítás, tesztelés).

A beszédadatbázis felvételeiből jellemzőket nyertem ki egy tanszéki program segítségével. Ez a program a Praat szabad felhasználású beszédelemző software segítségével jellemzőket határoz meg. Ezt megelőzően a felvételeket amplitúdó szerinti csúcsértékre normalizáltam. Az akusztikai jellemzők kiszámolásánál egységesnek, 50 milliszekundumos időablakot állítottam be, amivel a frekvenciafelbontás jobban vizsgálható. A programban az ***4.3. táblázat*** szerinti beállításokat használtam.

Az ilyen módon kinyert jellemzővektorokat egy általam írt C# program segítségével összerendeztem, így személyenként egy fájlban tároltam egy-egy jellemzővektor halmazt (mint MFCC, mel-sávos energia értékek, formáns … stb.). Ebben a fájlban eltároltam az alany/páciens azonosítóját és csoportját is. Továbbá, ahol a jellemzőkinyerő program nem tudott értéket meghatározni (oda "—undefined—„ jelölést rakott), azt a cellát eltávolítottam a vektorból, oly módon, hogy ugyanazon indexen (időablaknál) a többi jellemzővektorból is törtöltem az érték akkor is, ha az számérték volt. Így az egyszerre vizsgált jellemzővektorok időben nem csúsztak el egymáshoz képest.

4.3. táblázat: Jellemzőkinyerés során alkalmazott paraméterek.

Jellemző	Paraméterek	Praat parancs
Formánsok	Maximális frekvencia: 5500 [Hz]. Ablak típusa: Gauss Formánsok száma: 3	*Formant: Get value at time*
Formánsok sávszélességei	Sávszélességek száma: 3	*Formant: Get bandwidth at time*
Mel-sávos energiaértékek	Filterek száma: 27 Minimum frekvencia: 100 [Hz]. Lépésköz: 100 mel.	*Sound: To MelFilter*
MFCC	Értékek száma: 14 Minimum frekvencia: 100 [Hz] Lépésköz: 100 mel.	*Sound: To MFCC*

Következő lépésben előállítottam a korrelációs mátrixokat a következő paraméterek megadásával.

- Eltolások száma (dt): 5, 10, 15.
- Eltolások mértéke: 1, 2, 8.
- Jellemzővektorok száma (lásd lejjebb részletesen): *n*

Az *n* érték a jellemzővektorok száma a felhasznált halmazban. A diplomamunka során összesen 5 különböző jellemzővektor halmazt vizsgáltam. Ezek az alábbiak:

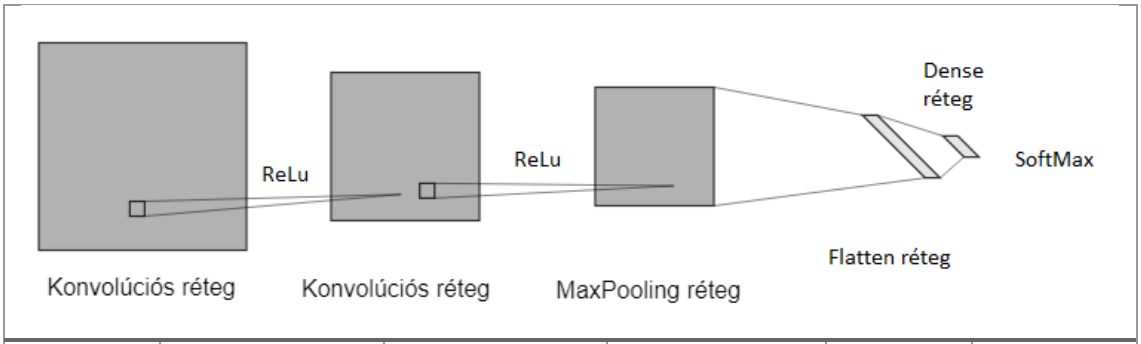
- Mel-sávos energia értékeket tartalmazó jellemzővektorok (*n* = 27). Továbbiakban **MelFilter**.
- MFCC együtthatókat tartalmazó jellemzővektorok (*n* = 14). Továbbiakban **MFCC**.
- Formáns frekvenciákat tartalmazó jellemzővektorok (*n* = 3). Továbbiakban **Formáns**.
- Formáns frekvenciák sávszélességét tartalmazó jellemzővektorok (*n* = 3). Továbbiakban **Sávszélesség**.
- Formáns frekvenciák és azok sávszélességeit tartalmazó jellemzővektorok (n = 6). Továbbiakban **Form-Sáv**.

A program kimenete így $(dt \cdot n) \times (dt \cdot n)$ méretű korrelációs mátrixok. Minden személyre egy korrelációs mátrix jött létre egy beállítás (*dt*, eltolási mérték, *n*) mellett.

Az osztályozó algoritmusnak egy egyszerű hálószerkezetet használtam, amit Python kódban írtam meg. Tensorflow környezetet használtam, amit kiegészítettem több programkönyvtárral, mint numpy, keras, pandas [82].

A program első része a korrelációs mátrixok előkészítését valósította meg. A fájlok tartalmát beolvastam és adatállományt hoztam létre belőlük. Ennek minden sora egy-egy személyre vonatkozott. Ezeket a sorokat véletlenszerűen megkevertem. Az adatállományból a korrelációs értékeket kinyertem és normalizáltam 0-1 értékek közé, majd újra mátrixformába rendeztem. A csoportokat kategorikus címkékké alakítottam. A CNN-t egy szekvenciális modellel hoztam létre, amihez az ***4.3 fejezetben*** ismertetett elemeket adtam hozzá. Az alapháló modell paramétereit és szerkezetét az ***4.4. táblázat*** tartalmazza.

4.4. táblázat: CNN rétegei és beállított paraméter értékei az alapháló modellen.



	Konvolúciós (2DConv)	Konvolúciós (2DConv)	MaxPooling	Flatten	Dense
Kernel száma	32	32	-	-	-
Kernel méret	10 · 10	bementi méret fele	2 · 2	-	-
Kernel lépésköze	10	bementi méret fele	2	-	-
Aktivációs függvény	ReLU	ReLU	ReLU	-	SoftMax
DropOut	25 %	25 %	25 %	-	-

Az első konvolúciós réteg esetén a bemeneti mátrix méretét mindig az éppen vizsgált jellemzőhalmaz mérete ($(n \cdot dt) \times (n \cdot dt)$) határozta meg. A kernel méretét az almátrixok méretének választottam meg. Ehhez állítottam a kernel lépésközt is, így át

nem lapolva hajtja végre a háló a konvolúciót. A második konvolúciós réteg kernel méretét és lépésközét úgy választottam meg, hogy a MaxPooling réteg bemenetére minden esetben 2×2 –es méretű mátrixok kerüljenek. A réteg kimenete így kernelenként egy skalár érték. A kernelek számának mindkét konvolúciós réteg esetében 32-t vettem fel. ReLU aktivációs függvényeket alkalmaztam az első három réteg kimenetére.

A DropOut elemmel véletlenszerűen a tanítás alatt 25%-át a neuronoknak figyelmen kívül hagytam az első három rétegnél, így csökkenthető a túltanulás veszélye a neurális hálónál. Az értéket szakirodalom alapján egységesnek választottam.

A Dense réteg bemenetére a kernelek értékei kerültek. Ez az alapháló modell esetén 32 db érték. A háló kimeneti értékeit SoftMax függvénnyel alakítottam valószínűségi értékekké. Minden tesztelem esetén így egy $1 \cdot x$ dimenziós vektort kaptam eredményül, ahol az x a csoportok számát jelöli. Ezek közül a legnagyobb értékhez tartozó címkére döntött az algoritmus.

A létrehozott modellben a tanulási arányt automatikus állításához ADAM féle optimalizációt alkalmaztam, a keras [83] szerinti általános paraméterekkel.

A tanítás során az algoritmus egyszerre kapta meg az 545 mintát minden iterációban, míg 1 elem végezte a tesztelést. A teljes kereszt validáció értelmében ezt 546-szor ismételtem. A futtatás végén az 546 tesztelem eredményéből építettem fel a tévesztési mátrixot. Bináris osztályozás esetén a tanító halmaz nagyságát és a teljes kereszt validáció szerinti ismétlések számát az éppen aktuális csoportok elemszáma határozta meg. A tanítás során nem különítettem el harmadik – validációs – halmazt az alacsony mintaszám miatt. Az alapháló modellben 100 iterációt állítottam be.

A háló létrehozását és a felparaméterezését követően az alábbi vizsgálatokat végeztem el:

1. Korrelációs struktúrák vizsgálata: A beteg és az egészséges csoportok korrelációs struktúráját átlagoltam. Majd minden beteg csoport átlagos korrelációs struktúrájából kivontam az egészséges csoport átlagolt struktúráját. Ezt minden jellemzőhalmazra elvégeztem $dt = 10$ és 1-es eltolási mérték beállítása mellett. Ezzel vizuálisan megjeleníthető az adott betegség átlagos eltérése az egészséges populációtól megfelelő jellemzőhalmaz esetén.

2. **Az 5 jellemzőhalmaz vizsgálata:** Az alapháló modellen alkalmaztam külön-külön az 5 jellemzőhalmazzal meghatározott korrelációs struktúrákat. Ezek dimenziójának alakulását a teljes alapháló modellen az ***4.5. táblázat*** tartalmazza. Ezzel sorrendi eredményeket kaptam arra vonatkozóan, hogy az alapbeállítások mellett melyik jellemző a legmegfelelőbb az elkülönítésben.

4.5. táblázat: Jellemzőhalmazok dimenziójának alakulása az alapháló modellen alap struktúrabeállításokkal (*dt* = 10, 1-es eltolási mérték).

		Kimenetek				
Jellemző halmazok	A háló bemenete	Konvolúciós (2DConv)	Konvolúciós (2DConv)	MaxPooling	Flatten	Dense
Formáns	32 · 30 × 30	32 · 3 × 3	32 · 2 × 2	32 · 1 × 1	32	5
Sávszélesség	32 · 30 × 30	32 · 3 × 3	32 · 2 × 2	32 · 1 × 1	32	5
Formáns + sávsz.	32 · 60 × 60	32 · 6 × 6	32 · 2 × 2	32 · 1 × 1	32	5
MelFilter	32 · 270 × 270	32 · 27 × 27	32 · 2 × 2	32 · 1 × 1	32	5
MFCC	32 · 140 × 140	32 · 14 × 14	32 · 2 × 2	32 · 1 × 1	32	5

3. A struktúra, illetve háló paramétereinek hatása az elkülönítésre: A korrelációs struktúrán az eltolások számának és az eltolások mértékének állítását végeztem. Az eltolás mértékének változtatásával a neurális háló modell paramétereit nem kell változtatni. Viszont az eltolások számának változtatásával az első konvolúciós réteg kernel méretét és lépésközét minden esetben ehhez igazítottam (***4.6. táblázat***). Az eltolás számának a 10-es alapbeállításon kívül az 5-ös és 15-ös méretet vizsgáltam. Az eltolás mértékének a 4-et és 8-at vettem fel az 1-es alapbeállítás mellé.

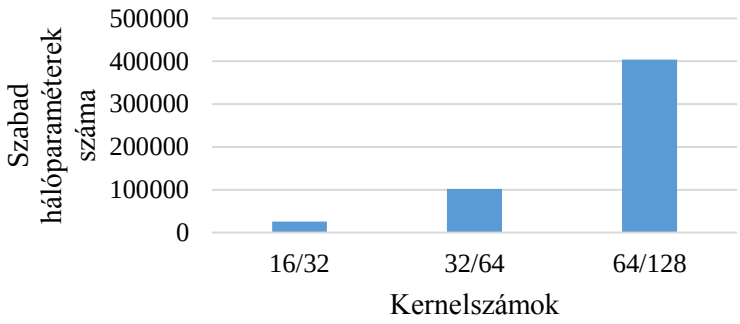
4.6. táblázat: Az ELSŐ konvolúciós réteg paramétereinek alakulása 5, 10, 15 *dt* érték szerint az alapháló modellben (*csak amik változnak*).

dt	**5**	**10**	**15**
Kernel méret	5 × 5	10 × 10	15 × 15
Kernel lépésköze	5	10	15

A háló esetében a tanulás alatti iteráció számát és a kernelek mennyiségét változtattam. Iterációszámnak 25, 50, 75, 100, 125, 150 értéket, illetve a kernelek számának 16, 32, 64, 128 értéket állítottam be.

A kernelszámokat úgy választottam meg, hogy az első konvolúciós rétegnek minden esetben feleannyi legyen, mint a második konvolúciós rétegnek a kernelszáma [84]. Ennek oka, hogy a második réteg végzi a lényegi mintakinyerést, illetve így a háló paramétereinek száma csökkenthető (gyorsabb futást eredményez), mintha ugyanakkora mennyiségű kernelt használnék. Így rendre 16/32, 32/64 és 64/128 kernelszámot alkalmaztam, ahol az első szám az első konvolúciós réteg, a második a második konvolúciós réteg kernelszáma.

Az ***4.6. ábra*** szemlélteti a háló szabad paramétereinek alakulását a kernelszám megválasztásával. Ezen szabad paraméterek beállítása történik meg a tanulás alatt. A kernelszámok duplázásával hatványos változást tapasztaltam a paraméterekben, ami hatványosan jelenik meg a tanítás idejében is.



4.6. ábra: Hálómodell szabad paramétereinek alakulása a kernelszám megválasztásával.

4. Az OD és az FD összevont vizsgálata: Az organikus és a funkcionális diszfónia összevonásával általános gégészeti megbetegedés csoportot hoztam létre, amivel a 4 osztályos osztályozást vizsgáltam a finomhangolt hálómodellen.

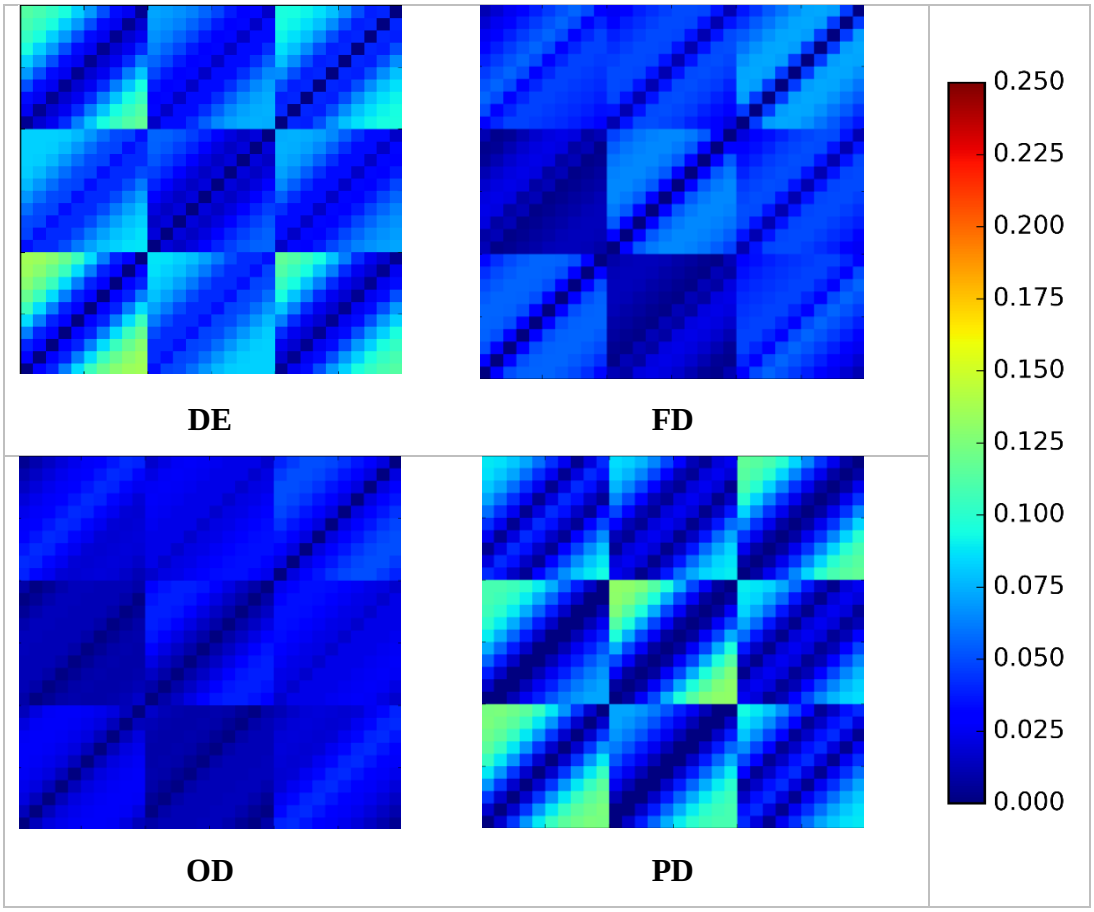
5. Bináris osztályozás: A finomhangolt hálómodellem változtatás nélkül egyszerre két osztályt vizsgálok, amiből az egyik mindig a HC csoport. Ezzel tesztelhető, hogy hogyan teljesít az algoritmus 1-1 betegség külön-külön való felismerésében. DE, PD és UD osztályozáshoz a komparátorérték állításával létrehoztam a ROC diagramot. Erre a pozitív mintára az algoritmus által adott valószínűségi értéket (0 és 1 közötti érték) használtam fel. Fontos megjegyezni, hogy a bináris osztályozást az 5 csoportra

finomhangolt osztályozóval végeztem el. Így ezen eredmények iránymutatók, viszont nem optimalizáltak specifikusan az adott beteg-egészséges csoportok elkülönítésére.

5 Eredmények

5.1 Beteg csoport átlagos eltérése az egészségestől a korrelációs struktúra alapján

A formáns jellemzőhalmazzal létrehozott struktúrát választottam – méretéből adódóan – bemutatásra. Az átlagos egészséges csoporthoz viszonyított átlagos betegségcsoport mátrixok az ***5.1. ábra***-n láthatók. (A többi jellemzőhalmaz átlagos korrelációs struktúrája az egészségeshez képest a ***Melléklet 1. ábra***-án található.) A struktúrák 30 × 30-as méretűek, amiken a kék szín jelzi az azonosságot a HC csoporttal, míg a piros szín felé az egyre nagyobb eltérést.



5.1. ábra: Az átlagolt beteg csoportok az egészségeshez viszonyítva.

Minden mátrix esetén megfigyelhető, hogy a teljes struktúra főátló elemei 0 értékűek. Ez abból következik, hogy minden csoport minden mintájában a főátló csupa egyeseket tartalmaz. Ez az átlagolás és a kivonás után nulla lesz. Ez az almátrixok

főátlóira is közelítőleg érvényes a formánsok esetén. A legnagyobb mértékben (max. 0,150) a mellékátlóban figyelhető meg az eltérések a HC csoporttól.

A DE és a PD csoportnál az eltérés maximális értéke 0,150-nél van (zöldes-sárga jelölés). A diszfónia csoportoknál viszont kisebb eltérés (maximum 0.100) tapasztalható a formáns frekvenciák halmazának használatával.

Más jellemzőhalmaz átlagos korrelációs struktúráját vizsgálva az egészséges kontrollhoz képest megállapítható, hogy változatos színmintát mutatnak. Például a formánsok sávszélességeivel – a DE kivételével – az almátrixok főátlóban jelennek meg eltérések (maximum 0.100). Illetve Melfilter vagy MFCC jellemzőhalmazzal a diszfónia csoportokra kiemelkedő eltérést (0.250 - piros színt) is kaptam. Az egészséges csoporthoz viszonyított átlagos korrelációs struktúrák ezekre a jellemzőhalmazokra a ***Melléklet 1. ábra***-n láthatók.

5.2 Az alapháló osztályozásának leírása

Az alaphálóval, a struktúra alapbeállításai mellett az 5 jellemzőhalmazzal elért eredményeket tévesztési mátrixba rendeztem (***5.2. ábra***).

	DE	FD	HC	OD	PD
DE	32	0	13	2	17
FD	0	0	0	0	0
HC	19	21	54	25	8
OD	21	45	64	131	23
PD	19	2	9	9	32

Sávszélesség

	DE	FD	HC	OD	PD
DE	29	3	12	3	10
FD	0	0	0	0	0
HC	6	9	33	19	1
OD	31	50	82	133	29
PD	25	6	13	12	40

Formáns

	DE	FD	HC	OD	PD
DE	51	3	16	5	13
FD	0	0	0	0	0
HC	15	7	64	26	11
OD	12	53	51	127	14
PD	13	5	9	9	42

Form – Sáv.

	DE	FD	HC	OD	PD
DE	32	1	19	4	6
FD	0	1	0	1	0
HC	29	25	89	30	11
OD	24	40	28	125	31
PD	6	1	4	7	32

Melfilter

	DE	FD	HC	OD	PD
DE	35	1	8	3	7
FD	0	7	2	5	0
HC	26	26	103	30	9
OD	18	30	18	118	22
PD	12	4	9	11	42

MFCC

5.2. ábra: Az 5 jellemzőhalmazzal az alapháló modellen, a korrelációs struktúra alapbeállításai mellett (*dt* = 10, 1-es eltolási mérték) elért téveszési mátrixok. Az oszlopok az eredeti csoportokat, a sorok a háló döntésit mutatják.

Az oszlopok az eredeti csoportot jelölik, a sorok az algoritmus döntéseit. A főátlóban a helyesen döntött minták száma szerepel a megfelelő csoportra nézve.

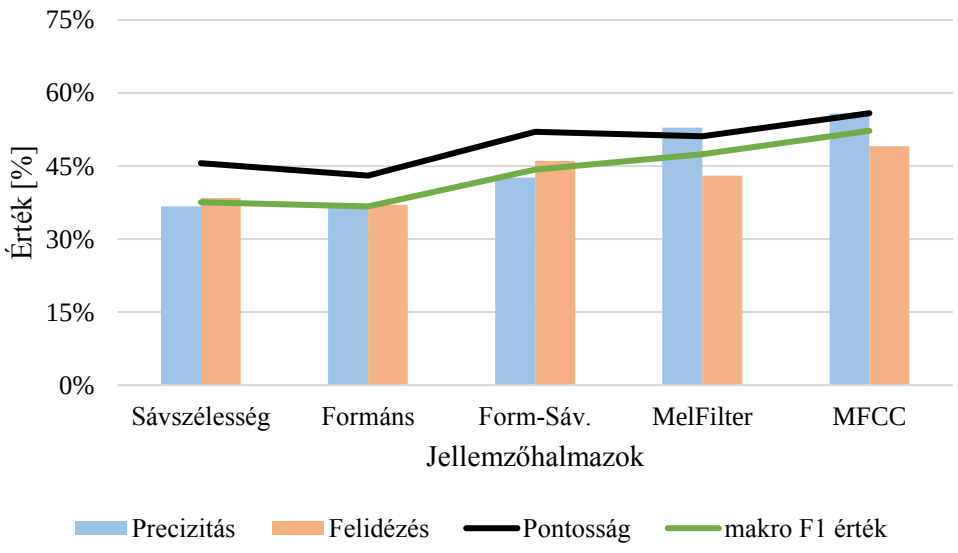
A formáns jellemzőhalmazzal több PD mintát ismert fel az algoritmus, mint a sávszélességek alkalmazásával. A sávszélesség esetén viszont több helyesen döntött HC minta szerepel. Se a sávszélességekkel, se a formánsokkal nem történt FD-re való döntés. Legnagyobb mintaszámban OD-re és HC-re döntötte inkább az FD-et.

Kombináltan alkalmazva a formánsok és a sávszélességek jellemzőhalmazait, javult a DE, HC és a PD felismerése. Viszont FD-re továbbra sem történt döntés.

A MelFilter jellemzőhalmazzal a sávszélességekéhez hasonló eredményt kaptam. Az OD-re döntött helyes minták száma csökkent, a HC csoportot viszont az előzőkénél nagyobb számban ismert fel.

Az MFCC jellemzőhalmaz alkalmazásánál megfigyelhető, hogy a legnagyobb mértékben sikerült helyesen egészséges mintát elkülöníteni az összes jellemzőhalmaz közül. A legtöbb helyesen döntött PD mintát az MFCC-vel (illetve a kombinált jellemzőegyüttessel) ismerte fel a háló.

A tévesztési mátrixokból az osztályozást leíró metrikákat számoltam az ***5.3.* ábra**-n.



5.3. ábra: Alapháló eredményéből számolt metrikák az 5 jellemzőhalmaz esetén (*dt*=10, eltolási mérték 1).

Oszlopdiagram formában a felidézés és a precizitás, vonaldiagramként a pontosság és a makro F1 érték látható. A függőleges tengelyen százalékos értékek vannak feltűntetve.

A precizitás és a felidézést minden csoportra meghatároztam, majd átlagukat ábrázoltam. Fontos megjegyezni, hogy a formáns, sávszélesség és kombinációjuk esetén – ahogy az a tévesztési mátrixokból is látható – nem számolható precizitás érték. Így ezen esetekben az 5 helyett 4 csoport átlagával dolgoztam az ***5.3. ábra***-n. A precizitás és felidézés értékek a sávszélesség, formánsok és kombinációjuk eredményeiben pár százalékon (± 2%) belül együtt mozognak. A Melfilter és az MFCC halmazokkal már rendre 9,5 illetve 6,7% az eltérés a két metrika között.

A pontosság értéke az összes jellemzőhalmaz mentén 43 és 56% között alakult, míg az makro F1 érték 36 és 52% között. Ez a két metrika legnagyobb értékét az MFCC jellemzőhalmazzal (pontosság = 55,9%, makro F1 érték = 52,2%) érte el, amit a MelFilter jellemzőhalmaz (pontosság = 51,1%, makro F1 érték = 47,4%) követett.

Külön-külön használva a formánsokat és a sávszélességeiket rendre 43,0 és 45, 5 % pontosságot kaptam, míg együttesen alkalmazva őket 52,0% pontosság lett az eredmény. Hasonlóan a makro F1 értékük is javult a kombinálás hatására.

Az eredmények alapján az MFCC jellemzőhalmazt választottam ki a további vizsgálatokra.

5.3 Paraméterhangolás hatása

5.3.1 Korrelációs struktúra paraméterei

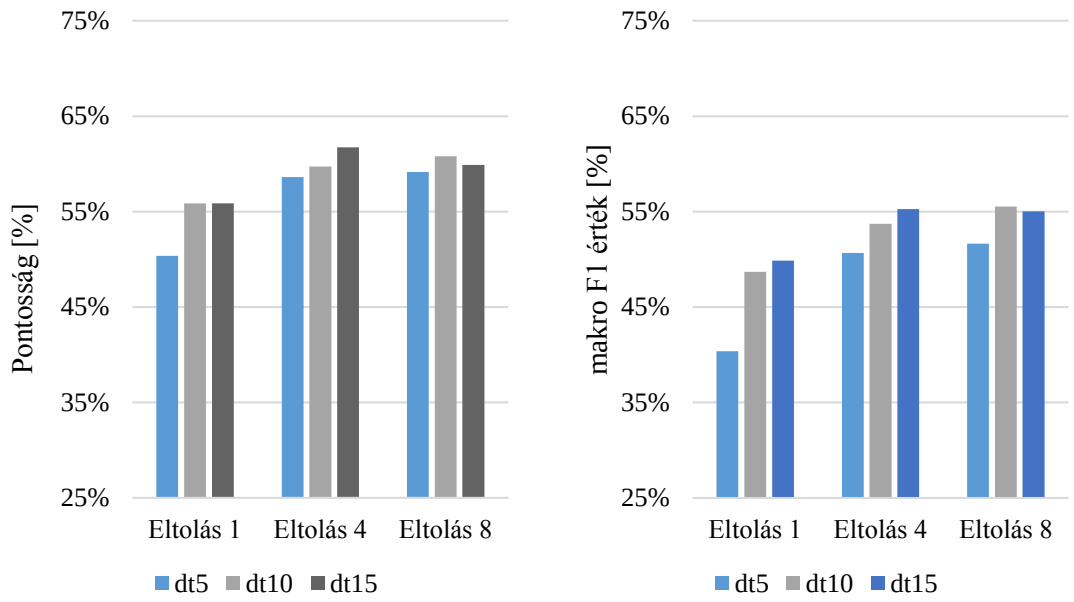
Az eltolás számának (dt = 5, 10, 15) és mértékének (1, 4, 8) változtatásával kapott értékeket mutatja be az ***5.4. ábra***. A pontosság az ***5.4. ábra*** bal oldalán, a makro F1 érték a jobb oldalán látható. A kategóriatengelyen az eltolások mértéke szerint ábrázoltak az elért értékek. A függőleges tengelyen a metrikák százalékos értéke olvasható le.

A pontosság esetén 1-es eltolást alkalmazva látható, hogy érdemes nagyobb eltolási számú struktúrát használni (50,4%-ról 55,9%), viszont a 10 és a 15 eltolási számok alkalmazása között már nem tapasztaltam változást.

Nagyobb eltolási mértéket használva már mindhárom eltolási szám esetén más pontossági értéket kaptam: 4-es eltolás mellett növekedett a pontosság az eltolás

számának növelésével (58,6%-ról 61,7%-ra), míg a 8-as eltolásnál a 10 eltolási szám rendelkezett a legnagyobb értékkel (60,8%). Viszont az is tapasztalható, hogy a 8-as eltolási mértékkel már nem sikerült elérni javulást a 4-es eltolási mértékhez képest.

A makro F1 érték metrikán keresztül is hasonlóak a tapasztalatok. 1-es és 4-es eltolási mérték mellett az eltolások számának növekedésével javulást értem el. Viszont a 8-as eltolási mérték mellett már jelentősebb változás nem volt megfigyelhető.



5.4. ábra: Az eltolás számának és mértékének változtatásával elért eredmények (pontosság a bal oldali diagramon, a makro F1 érték a jobb oldali diagramon) az alapháló modellen az 14 MFCC értékkel

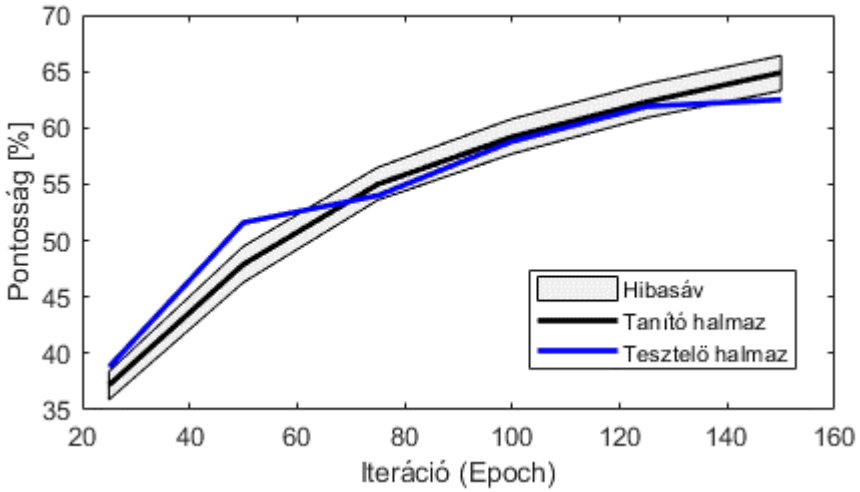
Az eredmények alapján az 5-ös eltolási számot és a 8-as eltolási mértéket választottam ki a további vizsgálatok elvégzésére. Nagyobb eltolási mértéken ugyanis mindhárom eltolási szám alkalmazása hasonló eredményeket adott, viszont a tanítás ideje nagyobb mértékben, mint arányosan nőtt a *dt* érték növelésével.

5.3.2 CNN paraméterei

A konvolúciós háló esetén számos paraméter állítása lehetséges, amiből a tanulás alatti iterációszámot és a két konvolúciós réteg kernelszámát választottam ki vizsgálatra. Iterációszámnak rendre 25, 50, 75, 100, 125 és 150 értéket vettem fel (***5.5. ábra***).

A meghatározott iterációban az 545 tanítómintából pontosságértéket számoltam, amit 1 minta tesztelt a tanítás végén. A teljes kereszt-validáció értelmében 546 db pontosságértéket kaptam a tanításból (minden iterációban) és 1 db pontosságértéket a

tesztelésből (546 tesztelemből). A tanítás alatti pontosságértékből átlagot (fekete görbe) és szórást (szürke sáv) számoltam az adott iterációkban. Erre ábrázoltam a teszthalmaz pontosság eredményét (kék görbe).



5.5. ábra: MFCC jellemzőhalmazzal *dt*=5 és 8-as eltolási mérték mellett elért eredméynt az alaphálómodellen.

Az ***5.5. ábra*** alapján mind a teszt halmaz, mind a tanítóhalmaz pontossága növekszik, majd 125 iteráció felett a tanítóhalmaz növekedése mellett a teszt halmaz kevésbé növekszik. Ebből következik, hogy az iterációszám növelésével (125 iteráció fölé) már nem javítható az osztályozó pontossága. A következőkben így 100-ról 125-re állítottam az iterációszámot a hálómodellben.

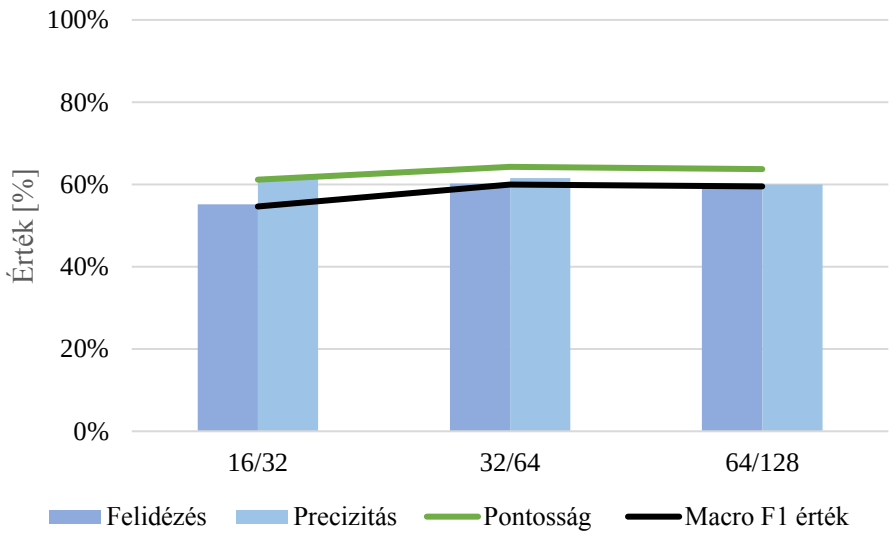
Második paraméter, ami állításra került a konvolúciós hálóban, a két konvolúciós réteg kernelszáma volt. Ezeket kettő hatványaiként vettem fel rendre úgy, hogy a második konvolúciós réteg kernelszáma kétszerese legyen az elsőnek. Ez alapján 16/32, 32/64 és 64/128 beállítást alkalmaztam 125 iterációm mellett (*dt* = 5, 8-as eltolási mérték a MFCC jellemzőre).

Az eredmények az ***5.6. ábra***-n láthatók, ahol a kategóriatengelyen a kernelszámok vannak feltűntetve. A függőleges tengelyen a különböző metrikák százalékos értékei láthatók. Oszlopdiagramként a felidézés és a precizitás, vonaldiagramon a pontosság és a makro F1 érték feltűntetett.

A pontosság értéke a vizsgált tartományon 61,2%-ról 64,3%-ra változott. Ez a maximális érték a 32/64 kernelszám állításával jött létre. A makro F1 értéknek is hasonlóan a 32/64 kernelszámnál volt a maximuma (60,0%). Legkisebb értéke a makro F1 értéknek 16/32 kernelszámnál 54,6%. A csoportok mentén átlagolt precizitás szűk

tartományon, 60,0% és 61,8% között változott a kategóriatengely mentén. A felidézés alacsonyabb értékről indulva (55,2%) növekedett a kernelszámok növelésével. A maximális értékét 32/64 kernelszámnál érte le (60,0%).

Összefoglalva a háló két paraméterének állításával kapott eredményeket elmondható, hogy legfeljebb 125 iteráció használata érdemes jelen konstrukcióban. A kernelszámok jelen módszer szerinti megválasztásával a 32/64 kernelszám beállítása hozta a legnagyobb értékeket a pontosságban a makro F1 értékben is.



5.6. ábra: A metrikák alakulás a kernelszámok változtatásával az alapháló modellen, 125 iterációszám (*dt* = 5, 8-as eltolási mérték) mellett. x/y esetén az x az első konvolúciós rétegben, y a másodikban alkalmazott kernelszám.

5.4 Négy csoportos osztályozás

Kiindulva az alapháló és az alapbeállítású korrelációs struktúrájú osztályozás eredményéből, az organikus és a funkcionális diszfónia nehezen különíthetők el egymástól. Ezért e két csoportot összevonva, mint általános hangképzőszervi elváltozás (UD) is vizsgáltam. Ez már a pontosság szerinti legjobb beállítások mellett lett elvégezve (125-ös iteráció, 32/64 kernelszám, *dt*=5, 8-as eltolási mérték, MFCC).

Az ***5.7. ábra***-n összehasonlítottam az 5 (bal oldali ábra) és a 4 (jobb oldali ábra) csoportos osztályozás tévesztési mátrixait. Az oszlopokban az eredeti minták, a sorokban az algoritmus becslései szerepelnek.

	DE	FD	HC	OD	PD
DE	54	2	12	5	3
FD	2	13	5	14	0
HC	18	12	94	19	9
OD	10	40	15	126	4
PD	7	1	14	3	64

	DE	HC	PD	UD
DE	47	11	4	4
HC	19	90	8	19
PD	5	11	63	3
UD	20	28	5	209

5.7. ábra: Az 5 és 4 csoportos osztályozás eredménye. Az ábra bal oldalán az 5, a jobb oldalán a 4 osztályos tévesztési mátrix. Az oszlopok az eredeti mintákat, a sorok a döntést jelentik.

A depresszió felismerése 7 mintával, az egészségesé 4 mintával, míg a Parkinson kórnak 1 mintával csökkent a 4 csoport osztályozása esetén az 5 csoportoshoz képest. Az általános hangképzőszervi elváltozás felismerése viszont mintaszámban javult. Az 5 csoport eseten 64,3%, a 4 csoport esetén 74,9% osztályozási pontosságot sikerült elérni.

Az 5 csoportnál a pontosság relatív javulása rendre 47,6 (DE); 51,8 (FD); 38,6 (HC); 33,7 (OD) és 49,6 (PD) % volt a módszereknél definiált triviális osztályozáshoz képest. Ez az osztályok mentén átlagosan 44,3%-os relatív javulást jelent. A 4 csoportra nézve 58,2 (DE); 49,3 (HC); 60,3 (PD); 31,9 (UD) % javulás írható le. Ez áltagosan 49,9 % javulást jelent. Látható, hogy a 4 csoportos osztályozásnál minden csoportra nagyobb relatív javulást értem el az 5 csoportos osztályozáshoz képest.

A makro F1 értéket meghatározva megállapítható, hogy 4 csoport használatával növekedett, az 5 csoportos 60,0%-ról 71,7%-ra változott.

Az eredményeket az ***5.1. táblázat***-ban foglalom össze, ahol feltűntettem az osztályozás pontosságát, makro F1 értékét és a pontosság átlagos százalékos javulását. A 4 csoportos osztályozás pontossága az adatbázison eddig elért eredményekkel (lásd ***2.4 fejezet***) összehasonlítható, ahol MFCC jellemzőhalmazt alkalmaztak. Leírható, hogy az én eredményeim a 2018-as eredményeket (69,4%) meghaladták.

5.1. táblázat: A két diszfóniás csoport összevonásának eredményei.

	5 csoportra	4 csoportra
Pontosság [%]	64,3	74,9
makro F1 érték [%]	60,0	71,7
Pontosság átlagos rel. javulása [%]	44,3	49,9

5.5 Bináris osztályozás (Egészséges vs. Beteg)

Bináris osztályozásra az MFCC jellemzőhalmazt használtam fel 5-ös eltolási számmal és 8-as eltolási mértékkel. Megtartottam a legjobb beállítási paramétereket a

hálónál: 32/64-es kernelszám, 125 iteráció a tanítás során. Tévesztési mátrixba foglalt eredmények az ***5.8. ábra***-n láthatók rendre a depressziós, Parkinson-kóros és az általános gégészeti elváltozásos csoportra. Mellettük a másik csoport mindig az egészséges kontroll volt.

	DE	HC
DE	69	18
HC	22	122

	HC	PD
HC	132	12
PD	8	68

	HC	UD
HC	107	25
UD	33	210

5.8. ábra: Rendre a DE, PD és UD bináris osztályozása HC-hez képest.

A DE-HC osztályozásra a tévesztési táblázat alapján a pontossági érték 82,7%, az F1 érték 81,7%. Közel ugyanannyit döntött tévesen egészségesnek (22 minta), mint tévesen depressziósnak (18 minta). A HC-PD osztályozásnál a két érték rendre 90,9 és 90,1%. Ez esetben 8 egészségest döntött Parkinson-kóros betegnek, míg 12 Parkinson-kórost viszont egészségesnek. Végül a HC-UD bináris elkülönítésre rendre 84,5 és 83,3 % az eredmény.

A bináris osztályozás összefoglalását az alábbi (***5.2. táblázat***) tartalmazza. A DE és PD esetben mind a precizitás, mind a felidézés százalékos értékben kisebb, mint az egészségeshez tartozó párjaik. Az UD csoportnál viszont az egészséges csoportra szerepelnek a kisebb értékek.

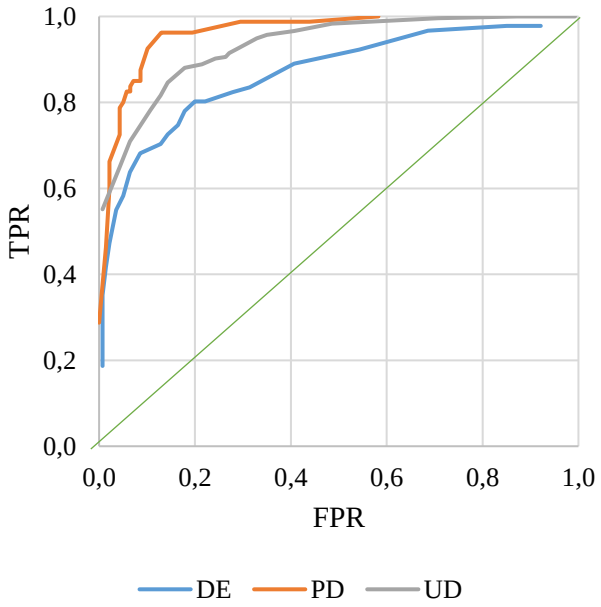
5.2. táblázat: Bináris osztályozás eredménye a DE, PD és UD betegségekre.

	Pontosság [%]	Precizitás [%]		Felidézés [%]		Precizitás [%]	Felidézés [%]	F1 érték [%]
		Beteg	HC	Beteg	HC	(átlag)	(átlag)	
DE	82,7	79,3	84,7	75,8%	87,1	82,0	81,5	81,7
PD	90,9	89,5	91,7	85,0%	94,3	90,6	89,6	90,1
UD	84,5	86,4	81,1	89,4%	76,4	83,7	82,9	83,3

Különböző komparátorértékek (döntési határok) beállításával szabályozható az osztályozó beteg, illetve egészséges csoportra döntésének affinitása. Ennek eredményét ábrázolja az ***5.9. ábra*** mindhárom betegségcsoport bináris osztályozására.

Az ábra alapján a PD csoport görbéje közelíti meg legjobban az (0,1) pontot (tökéletes osztályozás). Ezt az UD, majd a DE követi. A 45 fokos egyenes a véletlenszerű osztályozást ábrázolja.

Az ***5.3. táblázat***-ban 5 komparátorérték mellett feltűntettem az algoritmus pozitív döntéseit a három betegség bináris osztályozásából. Az első a valós pozitív, a második az álpozitív minták száma. A sorok az adott betegségcsoport bináris osztályázását szemléltetik. Az első sorban a komparátorértékek vannak.



5.9. ábra: ROC görbék a DR, PD és UD bináris osztályozásból. Az egység meredekségű egyenes a véletlen döntés görbéje.

0,95 határ mellett pusztán 1 helytelenül felismert minta látható mindhárom betegségcsoport esetén, viszont alacsony a helyesen felismert pozitív minták száma is. Lecsökkentve 0,0125-re a komparátorértéket elérhető a PD-vel és az UD-vel az összes pozitív minta felismerése (PD: 80, UD:234), viszont jelentős az álpozitív minták száma (PD: 81, UD:139).

5.3. táblázat: 5 komparátorérték mellett pozitívnak döntött elemek száma. x/y jelöléssel az x valós pozitív, y az álpozitív mintaszám.

Komp.	***0,95***	***0,75***	***0,50***	***0,25***	***0,0125***
DE	17/1	50/5	66/20	75/39	89/129
PD	23/0	58/6	66/9	74/14	80/81
UD	129/1	191/18	212/37	226/57	234/139

6 Eredmények elemzése és következtetések

A formánsok és sávszélességeik átlagos korrelációs struktúrája a DE és PD csoportnál mutatott kiugróbb elváltozást. A DE esetén a közérzet, érzelem megváltozása okozhatja e jellemzők jelentősebb megváltozását. A PD csoport átlagos életkora pedig befolyásolhatja az egészséges átlagtól való eltérést a korrelációs struktúrákban. A két jellemző együttes alkalmazása viszont javított a DE és a HC felismerésén. Ennek lehetséges magyarázata, hogy a DE csoportnál mind a sávszélességnek, mind a formánsoknak megvan a maga mintázata, amik kombinálása még egyedibb mintázatot nyújt a CNN-nek.

FD minták becslése nem történt a formánsok és sávszélességeik alkalmazásával, inkább az OD-re és a HC-re választott az algoritmus. Ezt befolyásolhatta, hogy a két csoport (OD és FD) súlyossága a rekedtség alapján átfedtek egymással. Illetve a minták többsége az 1-es (enyhe) rekedtségi kategóriába tartoznak, így kevésbé különülhetnek el az egészséges mintáktól.

A MelFilter jellemzőhalmazzal már minden csoportnál tapasztaltam eltérő mintákat az egészségeshez képest. Az OD és FD esetén viszont ugyanazon területek tértek el az egészségestől azzal a különbséggel, hogy az OD-nál nagyobb volt az eltérés.

Az MFCC jellemzőhalmazzal az eltérések kis részben fedtek csak át a betegségcsoportok között. A tévesztési mátrix alapján is látható, hogy ezzel a jellemzőhalmazzal ismertem fel helyesen a legtöbb mintát.

Az alapháló modell alkalmazásánál az MFCC jellemzőhalmaz teljesített a legjobban (55,9% pontosság, 52,2% makro F1 érték). Ez azért lehetséges, mert az MFCC egy lényegkiemeléses eljárás a MelFilter felhasználásával. A MelFilter jellemzővel a második legjobb eredményt kaptam (47,4% pontosság 51,1% makro F1 érték). Elmaradását az MFCC-hez képest valószínűleg az okozza, hogy a 27 mel-sáv 8 kHz-ig mindent tartalmaz, beleértve az elkülönítéshez hozzá nem járuló, de zavaró jeleket is.

Továbbá az eredmények rámutattak arra, hogy bizonyos jellemzők kombinálása – ez esetben formáns és sávszélességeik – javíthat az osztályozó algoritmus elkülönítésén ahhoz képest, mintha külön alkalmaznánk őket.

Az eltolási szám növelése növelte a pontosságot és a makro F1 értéket is. Ennek lehetséges oka lehet, hogy az első konvolúciós réteg több mintából végezhette el a konvolúciót, ami az egyénre specifikusabb értéket nyújthatott.

Az eltolások mértékének növelésével is nőtt a metrikák értéke, viszont 8-as eltolásnál már kismértékű csökkenést tapasztaltam. A csökkenés adódhat abból, hogy jelentősebb eltolási mérték alkalmazásánál eltűnik a két jellemzővektor közötti korrelációs kapcsolat. Így a struktúrában az erős korrelációk a főátlóra korlátozódnak, ami hátrányosan befolyásolja az elkülönítést.

125 iteráció mellett kaptam azt a legnagyobb pontosságot, ahol még a tanító és tesztelő halmaz eredménye együtt, hibasávon belül haladtak. Nagyobb iterációnál a tanító halmazokra való túltanulás veszélye nő, a teszthalmaz eredménye már kisebb mértékben változik. Kisebb mintaszámnál a teszt halmaz pontossága dinamikusabban változik a tanítóhalmazhoz képest. Ezt feltételezhetően az alul tanulás okozhatja.

A kernelszámok változtatása a makro F1 értékben hozott jelentősebb változást a pontosság értékéhez képest. Legnagyobb értékük (pontosság: 64,3%, makro F1 érték: 60,0%) 32/64 kernelszámnál volt. Ennél nagyobb kernelszám használata nem javított az elkülönítésben, viszont az algoritmus futását jelentősen megnövelte. Ennek lehetséges oka, hogy adott kernelszámon felül már nem hozható létre olyan mintázatú kernel, ami hozzájárulhatna az elkülönítéshez.

Az OD és FD összevonásával összességében javult a helyes felismerésük, mint általános gégészeti elváltozás csoport a legjobb beállítások megtartása mellett. A triviális osztályozáshoz képest az minden az UD melletti csoportnak javult a felismerése. Az összevonás indoka, hogy a két diszfóniás csoport nehezen elkülöníthető egymástól a beszéd alapján.

A három betegségcsoport közül a PD-t sikerült a legnagyobb pontossággal felismerni (90,9%). Ez az érték magasnak tekinthető abból a szempontból, hogy az osztályozó nem bináris osztályozásra lett paraméter hangolva. A DE és UD csoportokra rendre 82,7 és 84,5% pontosságok értem el.

Létrehozva a ROC diagramot látható, hogy a PD görbéje közelíti meg leginkább a tökéletes osztályozó viselkedését. Ezt az UD, majd a DE követi.

Mindhárom betegségcsoport bináris elkülönítésére megállapítható, hogy a szakirodalomban megtalálható pontossági eredményekkel összevethetők, tartományuk

felső részébe tehetők (80% feletti eredmény). A 2018-as labormunka 4 csoportos eredményével hasonlóan összevethetők az eredmények, MFCC-vel elért eredményüknél (69,4%) jobb pontosságot sikerült elérnem (74,9%).

Köszönetnyilvánítás

Mindenekelőtt köszönöm témavezetőmnek, Kiss Gábornak, hogy lehetővé tette számomra a betegségek beszédjel alapú felismerésének vizsgálatával foglalkozó tanulmány elkészítését, észrevételeivel és ötleteivel segítette munkámat.

Külön köszönetet szeretnék mondania Dr. Sztahó Dávidnak és Tulics Miklós Gábrielnek, akikhez fordulhattam szakterületüket érintő kérdéseimmel (Parkinson-kór, általános gégészeti elváltozás).

Továbbá köszönettel tartozom az Egészségügyi mérnöki mesterképzés valamennyi oktatójának, akik hozzájárultak ismereteim bővítéséhez mind a mérnöki, mind az egészségügyi területen.

Végül pedig köszönöm barátnőmnek és családomnak, hogy szeretetükkel és türelmükkel támogattak egyetemi tanulmányaim alatt.

Táblázatjegyzék

3.1. táblázat: A dimplomamunkában alkalmazott beszédminták mennyisége. 21
4.1. táblázat: Tévesztési mátrix felépítése bináris osztályozásra. A + jelöli a pozitív, - a negatív mintát. 30
4.2. táblázat: Tévesztési mátrix többosztályos osztályozás leírására Az osztályok a *Beszédadatbázis* fejezet szerintiek. 31
4.3. táblázat: Jellemzőkinyerés során alkalmazott paraméterek. 35
4.4. táblázat: CNN rétegei és beállított paraméter értékei az alapháló modellen. 36
4.5. táblázat: Jellemzőhalmazok dimenziójának alakulása az alapháló modellen alap struktúrabeállításokkal (*dt* = 10, 1-es eltolási mérték). 38
4.6. táblázat: Az ELSŐ konvolúciós réteg paramétereinek alakulása 5, 10, 15 *dt* érték szerint az alapháló modellben (*csak amik változnak*). 38
5.1. táblázat: A két diszfóniás csoport összevonásának eredményei. 48
5.2. táblázat: Bináris osztályozás eredménye a DE, PD és UD betegségekre. 49
5.3. táblázat: 5 komparátorérték mellett pozitívnak döntött elemek száma. x/y jelöléssel az x valós pozitív, y az álpozitív mintaszám. 50

Ábrajegyzék

3.1. ábra: Depresszió súlyosságának eloszlása a felhasznált felvételek között. 18
3.2. ábra: A dizfóiás felvételek eloszlása rekedtségi (H) érték alapján. 19
3.3. ábra: Felhasznált Parkinson-kóros felvételek H-Y érték szerinti megoszlása. 20
4.1. ábra: A korrelációs struktúra szerkezete (ábra bal oldala). A struktúra celláiban almátrixok találhatók a vektoreltolásoknak megfelelően (ábra jobb oldala). 25
4.2. ábra: Vektorok elemeltolási módszerének szemléltetése 1-es mértékű eltolás esetén. 26
4.3. ábra: Szűrők alkalmazása a konvolúciós neurális hálóban. Az át nem lapoló szűrő minden 2×2-es képszegmensnek az első cellájának kétszeresét viszi tovább. 27
4.4. ábra: Általános ROC diagram: véletlenszerű (barna szín), valós (zöld), tökéletes (piros) osztályozás. 33
4.5. ábra: A folyamat blokkvázlata. Elemei: jellemzőkinyerés, korrelációs struktúra létrehozása, osztályozás megvalósítása (tanítás, tesztelés). 34
4.6. ábra: Hálómodell szabad paramétereinek alakulása a kernelszám megválasztásával. 39
5.1. ábra: Az átlagolt beteg csoportok az egészségeshez viszonyítva. 41
5.2. ábra: Az 5 jellemzőhalmazzal az alapháló modellen, a korrelációs struktúra alapbeállításai mellett (*dt* = 10, 1-es eltolási mérték) elért téveszési mátrixok. Az oszlopok az eredeti csoportokat, a sorok a háló döntésit mutatják. 42
5.3. ábra: Alapháló eredményéből számolt metrikák az 5 jellemzőhalmaz esetén (*dt*=10, eltolási mérték 1). 43
5.4. ábra: Az eltolás számának és mértékének változtatásával elért eredmények (pontosság a bal oldali diagramon, a makro F1 érték a jobb oldali diagramon) az alapháló modellen az 14 MFCC értékkel 45
5.5. ábra: MFCC jellemzőhalmazzal *dt*=5 és 8-as eltolási mérték mellett elért eredményt az alaphálómodellen. 46
5.6. ábra: A metrikák alakulás a kernelszámok változtatásával az alapháló modellen, 125 iterációszám (*dt* = 5, 8-as eltolási mérték) mellett. x/y esetén az x az első konvolúciós rétegben, y a másodikban alkalmazott kernelszám. 47
5.7. ábra: Az 5 és 4 csoportos osztályozás eredménye. Az ábra bal oldalán az 5, a jobb oldalán a 4 osztályos tévesztési mátrix. Az oszlopok az eredeti mintákat, a sorok a döntést jelentik. 48
5.8. ábra: Rendre a DE, PD és UD bináris osztályozása HC-hez képest. 49
5.9. ábra: ROC görbék a DR, PD és UD bináris osztályozásból. Az egység meredekségű egyenes a véletlen döntés görbéje. 50
1. ábra (Melléklet): Beteg csoportok átlagos struktúrái az egészségeshez viszonyítva. 66

Irodalomjegyzék

[1] World Health Organization, „ICD-10 Version:2019”, *World Health Organization*, 2019. [Online]. Elérhető: https://icd.who.int/browse10/2019/en. [Elérés: 20-febr-2020].

[2] World Health Organization, „Mental and behavioural disorders”, in *International Statistical Classification of Diseases and Related Health Problems 10th Revision*, 2019.

[3] World Health Organization, „Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)”, in *International Statistical Classification of Diseases and Related Health Problems 10th Revision*, 2019.

[4] D. Dixit, V. Mittal, és Y. Sharma, „Voice Parameter Analysis for the disease detection”, *IOSR J. Electron. Commun. Eng.*, köt. 9, sz. 3, o. 48–55, 2014, doi: 10.9790/2834-09314855.

[5] K. C. Fraser, F. Rudzicz, és E. Rochon, „Using text and acoustic features to diagnose progressive aphasia and its subtypes”, *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, sz. August, o. 2177–2181, 2013.

[6] J. Deng, N. Cummins, M. Schmitt, K. Qian, F. Ringeval, és B. Schuller, „Speech-based diagnosis of autism spectrum condition by generative adversarial network representations”, *ACM Int. Conf. Proceeding Ser.*, köt. Part F1286, o. 53–57, 2017, doi: 10.1145/3079452.3079492.

[7] K.-C. Wang, „Time-frequency feature representation using multi-resolution texture analysis and acoustic activity detector for real-life speech emotion recognition.”, *Sensors (Basel).*, köt. 15, sz. 1, o. 1458–1478, jan. 2015, doi: 10.3390/s150101458.

[8] S. Dávid, K. Gábor, T. M. Gábriel, és V. Klára, „Betegségek automatikus szétválasztása időben eltolt akusztikai jellemzők korrelációs struktúrája alapján”, in *XV. Magyar Számítógépes Nyelvészeti Konferencia*, 2019, o. 203–212.

[9] V. Berisha, R. Utianski, és J. Liss, „Towards A Clinical Tool For Automatic Intelligibility Assessment.”, *Proc. ... IEEE Int. Conf. Acoust. Speech, Signal*

Process. ICASSP, o. 2825–2828, 2013, doi: 10.1109/ICASSP.2013.6638172.

[10] L. Verde, G. De Pietro, és G. Sannino, „Voice Disorder Identification by Using Machine Learning Techniques”, *IEEE Access*, köt. 6, o. 16246–16255, 2018, doi: 10.1109/ACCESS.2018.2816338.

[11] S. Dham, A. Sharma, és A. Dhall, „Depression Scale Recognition from Audio, Visual and Text Analysis”, 2017.

[12] L. He és C. Cao, „Automated depression analysis using convolutional neural networks from speech”, *J. Biomed. Inform.*, köt. 83, sz. May, o. 103–111, 2018, doi: 10.1016/j.jbi.2018.05.007.

[13] G. Schlotthauer, M. E. Torres, és M. C. Jackson-Menaldi, „A Pattern Recognition Approach to Spasmodic Dysphonia and Muscle Tension Dysphonia Automatic Classification”, *J. Voice*, köt. 24, sz. 3, o. 346–353, 2010, doi: 10.1016/j.jvoice.2008.10.007.

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

[15] J. W. Kanter, A. M. Busch, C. E. Weeks, és S. J. Landes, „The nature of clinical depression: symptoms, syndromes, and behavior analysis.”, *Behav. Anal.*, köt. 31, sz. 1, o. 1–21, 2008, doi: 10.1007/bf03392158.

[16] G. S. Malhi és J. J. Mann, „Depression”, *Lancet*, köt. 392, sz. 10161, o. 1–13, 2018, doi: 10.1016/S0140-6736(18)31948-2.

[17] L. Ge, C. W. Yap, R. Ong, és B. H. Heng, „Social isolation, loneliness and their relationships with depressive symptoms: A population-based study.”, *PLoS One*, köt. 12, sz. 8, o. e0182145, 2017, doi: 10.1371/journal.pone.0182145.

[18] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, és T. F. Quatieri, „A review of depression and suicide risk assessment using speech analysis”, *Speech Commun.*, köt. 71, o. 10–49, 2015, doi: 10.1016/j.specom.2015.03.004.

[19] A. T. BECK, C. H. WARD, M. MENDELSON, J. MOCK, és J. ERBAUGH, „An Inventory for Measuring Depression”, *Arch. Gen. Psychiatry*, köt. 4, sz. 6, o. 561–571, 1961.

[20] M. HAMILTON, „A rating scale for depression.”, *J. Neurol. Neurosurg. Psychiatry*, köt. 23, sz. 1, o. 56–62, febr. 1960, doi: 10.1136/jnnp.23.1.56.

[21] D. N. Klein, R. Kotov, és S. J. Bufferd, „Personality and Depression: Explanatory Models and Review of the Evidence”, *Annu. Rev. Clin. Psychol.*, köt. 7, sz. 1, o. 269–295, ápr. 2011, doi: 10.1146/annurev-clinpsy-032210-104540.

[22] K. Gábor, S. Dávid, M. Gábriel, és E. Anna, „Language independent detection possibilities of depression by speech”, in *Recent Advances in Nonlinear Speech Processing*, köt. 48, sz. 977126, A. Esposito, M. Faundez-Zanuy, A. M. Esposito, G. Cordasco, T. Drugman, J. Solé-Casals, és F. C. Morabito, Szerk. Springer International Publishing, 2016, o. 103–114.

[23] L.-S. A. Low, M. C. Maddage, M. Lech, L. B. Sheeber, és N. B. Allen, „Detection of Clinical Depression in Adolescents’ Speech During Family Interactions”, *IEEE Trans. Biomed. Eng.*, köt. 58, sz. 3, o. 574–586, márc. 2011, doi: 10.1109/TBME.2010.2091640.

[24] C. Solomon, M. F. Valstar, R. K. Morriss, és J. Crowe, „Objective Methods for Reliable Detection of Concealed Depression”, *Front. ICT*, köt. 2, sz. APR, o. 1–16, ápr. 2015, doi: 10.3389/fict.2015.00005.

[25] 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, doi: 10.1109/CogInfoCom.2017.8268245.

[26] H. Jiang *és mtsai.*, „Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features”, *Comput. Math. Methods Med.*, köt. 2018, o. 1–9, szept. 2018, doi: 10.1155/2018/6508319.

[27] 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, doi: 10.1145/2661806.2661809.

[28] M. Valstar *és mtsai.*, „AVEC 2013”, in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge - AVEC ’13*, 2013, köt. 550, o. 3–10, doi: 10.1145/2512530.2512533.

[29] L. Roland, „Keresztkorreláció elemzés depressziós beszédanyagon”, Budapest Műszaki és Gazdaságtudományi Egyetem, 2016.

[30] E. A. C. Pereira és T. Z. Aziz, „Parkinson’s disease and primate research: Past, present, and future”, *Postgrad. Med. J.*, köt. 82, sz. 967, o. 293–299, 2006, doi: 10.1136/pgmj.2005.041194.

[31] National Institute of Neurological Disorders and Stroke (NIH), „Parkinson’s Disease Information Page”, 2019. [Online]. Elérhető: https://www.ninds.nih.gov/Disorders/All-Disorders/Parkinsons-Disease-Information-Page. [Elérés: 07-márc-2020].

[32] Parkinson’s Foundation, „Treatment”, *Parkinson’s Foundation*. [Online]. Elérhető: https://www.parkinson.org/Understanding-Parkinsons/Treatment. [Elérés: 07-márc-2020].

[33] F. M. Ivey, L. I. Katzel, J. D. Sorkin, R. F. Macko, és L. M. Shulman, „The Unified Parkinson’s Disease Rating Scale as a predictor of peak aerobic capacity and ambulatory function.”, *J. Rehabil. Res. Dev.*, köt. 49, sz. 8, o. 1269–1276, 2012, doi: 10.1682/jrrd.2011.06.0103.

[34] J. M. Rabey és A. D. Korczyn, „The Hoehn and Yahr Rating Scale for Parkinson’s Disease”, in *Instrumental Methods and Scoring in Extrapyramidal Disorders*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1995, o. 7–17.

[35] J. Jankovic, „Parkinson’s disease: Clinical features and diagnosis”, *J. Neurol. Neurosurg. Psychiatry*, köt. 79, sz. 4, o. 368–376, 2008, doi: 10.1136/jnnp.2007.131045.

[36] G. M. Earhart, T. Ellis, A. Nieuwboer, és L. E. Dibble, „Rehabilitation and Parkinson’s Disease”, *Parkinsons. Dis.*, köt. 2012, o. 1–3, 2012, doi: 10.1155/2012/371406.

[37] C. Dromey, L. O. Ramig, és A. B. Johnson, „Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study”, *J. Speech Hear. Res.*, köt. 38, sz. 4, o. 751–764, 1995, doi: 10.1044/jshr.3804.751.

[38] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, és L. O. Ramig, „Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson’s Disease”, *IEEE Trans. Biomed. Eng.*, köt. 59, sz. 5, o. 1264–1271,

2012, doi: 10.1109/TBME.2012.2183367.

[39] E. Vaiciukynas, A. Verikas, A. Gelzinis, és M. Bacauskiene, „Detecting Parkinson's disease from sustained phonation and speech signals", *PLoS One*, köt. 12, sz. 10, o. 1–16, 2017, doi: 10.1371/journal.pone.0185613.

[40] A. Benba, A. Jilbab, A. Hammouch, és S. Sandabad, „Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease", in *2015 International Conference on Electrical and Information Technologies (ICEIT)*, 2015, o. 300–304, doi: 10.1109/EITech.2015.7163000.

[41] B. M. Bot *és mtsai.*, „The mPower study, Parkinson disease mobile data collected using ResearchKit", *Sci. Data*, köt. 3, sz. 1, o. 160011, dec. 2016, doi: 10.1038/sdata.2016.11.

[42] H. Hazan, D. Hilu, L. Manevitz, L. O. Ramig, és S. Sapir, „Early diagnosis of Parkinson's disease via machine learning on speech data", *2012 IEEE 27th Conv. Electr. Electron. Eng. Isr. IEEEI 2012*, sz. July 2015, 2012, doi: 10.1109/EEEI.2012.6377065.

[43] A. Frid, H. Hazan, D. Hilu, L. Manevitz, L. O. Ramig, és S. Sapir, „Computational diagnosis of Parkinson's disease directly from natural speech using machine learning techniques", *Proc. - 2014 IEEE Int. Conf. Softw. Sci. Technol. Eng. SWSTE 2014*, sz. June, o. 50–53, 2014, doi: 10.1109/SWSTE.2014.17.

[44] T. Bocklet, E. Noth, G. Stemmer, H. Ruzickova, és J. Rusz, „Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis", in *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, 2011, o. 478–483, doi: 10.1109/ASRU.2011.6163978.

[45] P. Carding és I. Horsley, „An evaluation of voice therapy in non-organic dysphonia", *Eur. J. Disord. Commun.*, köt. 27, o. 137–158, 1992, doi: 10.3109/13682829209012036.

[46] R. J. Stachler *és mtsai.*, „Clinical Practice Guideline: Hoarseness (Dysphonia) (Update)", *Otolaryngol. - Head Neck Surg. (United States)*, köt. 158, o. 1–42, 2018, doi: 10.1177/0194599817751030.

[47] Y. Maryn, D. Morsomme, és M. De Bodt, „Measuring the Dysphonia Severity Index (DSI) in the Program Praat", *J. Voice*, köt. 31, sz. 5, o. 644.e29-644.e40,

2017, doi: 10.1016/j.jvoice.2017.01.002.

[48] Y.-R. Chien, M. Borský, és J. Guðnason, „Objective Severity Assessment from Disordered Voice Using Estimated Glottal Airflow”, in *Interspeech 2017*, 2017, köt. 2017-Augus, sz. August 2017, o. 304–308, doi: 10.21437/Interspeech.2017-138.

[49] T. Haderlein, C. Schwemmle, M. Döllinger, V. Matoušek, M. Ptok, és E. Nöth, „Automatic Evaluation of Voice Quality Using Text-Based Laryngograph Measurements and Prosodic Analysis”, *Comput. Math. Methods Med.*, köt. 2015, o. 1–11, 2015, doi: 10.1155/2015/316325.

[50] J. P. Teixeira és P. O. Fernandes, „Acoustic Analysis of Vocal Dysphonia”, *Procedia Comput. Sci.*, köt. 64, o. 466–473, 2015, doi: 10.1016/j.procs.2015.08.544.

[51] H. T. Lathadevi és S. P. Guggarigoudar, „Objective acoustic analysis and comparison of normal and abnormal voices”, *J. Clin. Diagnostic Res.*, köt. 12, sz. 12, o. MC01–MC04, 2018, doi: 10.7860/JCDR/2018/36782.12310.

[52] J. Teixeira, C. Oliveira, és C. Lopes, „Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters”, *Procedia Technol.*, köt. 9, o. 1112–1122, dec. 2013, doi: 10.1016/j.protcy.2013.12.124.

[53] Y. Zhang és J. J. Jiang, „Acoustic Analyses of Sustained and Running Voices From Patients With Laryngeal Pathologies”, *J. Voice*, köt. 22, sz. 1, o. 1–9, 2008, doi: 10.1016/j.jvoice.2006.08.003.

[54] A. A. Dibazar, T. W. Berger, és S. S. Narayanan, „Pathological Voice Assessment”, in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, o. 1669–1673, doi: 10.1109/IEMBS.2006.259835.

[55] T. Dubuisson, T. Dutoit, B. Gosselin, és M. Remacle, „On the use of the correlation between acoustic descriptors for the normal/Pathological voices discrimination”, *EURASIP J. Adv. Signal Process.*, köt. 2009, 2009, doi: 10.1155/2009/173967.

[56] J. I. Godino-Llorente és P. Gomez-Vilda, „Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors”, *IEEE Trans. Biomed. Eng.*, köt. 51, sz. 2, o. 380–384, 2004, doi:

10.1109/TBME.2003.820386.

[57] R. Linder, A. E. Albers, M. Hess, S. J. Pöppl, és R. Schönweiler, „Artificial Neural Network-based Classification to Screen for Dysphonia Using Psychoacoustic Scaling of Acoustic Voice Features”, *J. Voice*, köt. 22, sz. 2, o. 155–163, 2008, doi: https://doi.org/10.1016/j.jvoice.2006.09.003.

[58] M. E. Powell *és mtsai.*, „Decoding phonation with artificial intelligence (DeP AI): Proof of concept”, *Laryngoscope Investig. Otolaryngol.*, köt. 4, sz. 3, o. 328–334, 2019, doi: 10.1002/lio2.259.

[59] D. Sztahó, G. Kiss, M. G. Tulics, és K. Vicsi, „Automatic Separation of Various Disease Types by Correlation Structure of Time Shifted Speech Features”, in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 2018, o. 96–99.

[60] D. Sztaho, G. Kiss, M. G. Tulics, B. Hajduska-Der, és K. Vicsi, „Automatic discrimination of several types of speech pathologies”, in *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 2019, o. 1–6, doi: 10.1109/SPED.2019.8906556.

[61] F. Susanne és P. Pascal, „Understanding speech pruduction: the pilos approach”, *Rev. française Linguist. appliquée*, köt. XIII, sz. 2, o. 35–44, 2008.

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

[63] B. H. Story, „Vowel and consonant contributions to vocal tract shape”, *J. Acoust. Soc. Am.*, köt. 126, sz. 2, o. 825–836, 2009, doi: 10.1121/1.3158816.

[64] J. Saini és R. Mehra, „Power Spectral Density Analysis of Speech Signal using Window Techniques”, *Int. J. Comput. Appl.*, köt. 131, sz. 14, o. 33–36, 2015, doi: 10.5120/ijca2015907549.

[65] K. K. Bhuvanagiri és S. K. Kopparapu, „Modified Mel Filter Bank to Compute MFCC of Subsampled Speech”, sz. October 2014, okt. 2014.

[66] S. K. Kopparapu és M. Laxminarayana, „Choice of Mel filter bank in computing MFCC of a resampled speech”, in *10th International Conference on Information Sciences, Signal Processing and their Applications, ISSPA 2010*, 2010, sz. May 2010, o. 121–124, doi: 10.1109/ISSPA.2010.5605491.

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

[68] M. M. Mukaka, „Statistics corner: A guide to appropriate use of correlation coefficient in medical research”, *Malawi Med. J.*, köt. 24, sz. 3, o. 69–71, szept. 2012.

[69] R. Karim, „Illustrated: 10 CNN Architectures”, *Towars Data Science*, 2019. [Online]. Elérhető: https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d. [Elérés: 07-máj-2020].

[70] K. O’Shea és R. Nash, „An Introduction to Convolutional Neural Networks”, sz. November, 2015.

[71] 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, doi: 10.1007/s40436-019-00254-5.

[72] 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, doi: 10.1007/s13244-018-0639-9.

[73] 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, doi: 10.1016/j.cmpb.2018.04.012.

[74] V. Nasteski, „An overview of the supervised machine learning methods”, *HORIZONS.B*, köt. 4, sz. December, o. 51–62, dec. 2017, doi: 10.20544/HORIZONS.B.04.1.17.P05.

[75] 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.

[76] P. Sharma, „Decoding the Confusion Matrix”, *Towars Data Science*, 2019. [Online]. Elérhető: https://towardsdatascience.com/decoding-the-confusion-matrix-bb4801decbb. [Elérés: 01-ápr-2020].

[77] Banso D. Wisdom, „Understanding the Confusion Matrix (II)”, *dev.to*, 2019. [Online]. Elérhető: https://dev.to/overrideveloper/understanding-the-confusion-

matrix-264i. [Elérés: 01-ápr-2020].

[78] G. Kavita, „How to compute precision and recall for a multi-class classification problem”, *Kavita-Ganesan*. [Online]. Elérhető: https://kavita-ganesan.com/how-to-compute-precision-and-recall-for-a-multi-class-classification-problem/#.XoRMeUAzbIV. [Elérés: 31-márc-2020].

[79] W. Koehrsen, „Beyond Accuracy: Precision and Recall”, *Towars Data Science*, 2018. [Online]. Elérhető: https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c. [Elérés: 30-márc-2020].

[80] Service Amazon Web, „Multiclass Model Insights”, *docs.aws.amazon.com*, 2020. [Online]. Elérhető: https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html. [Elérés: 31-márc-2020].

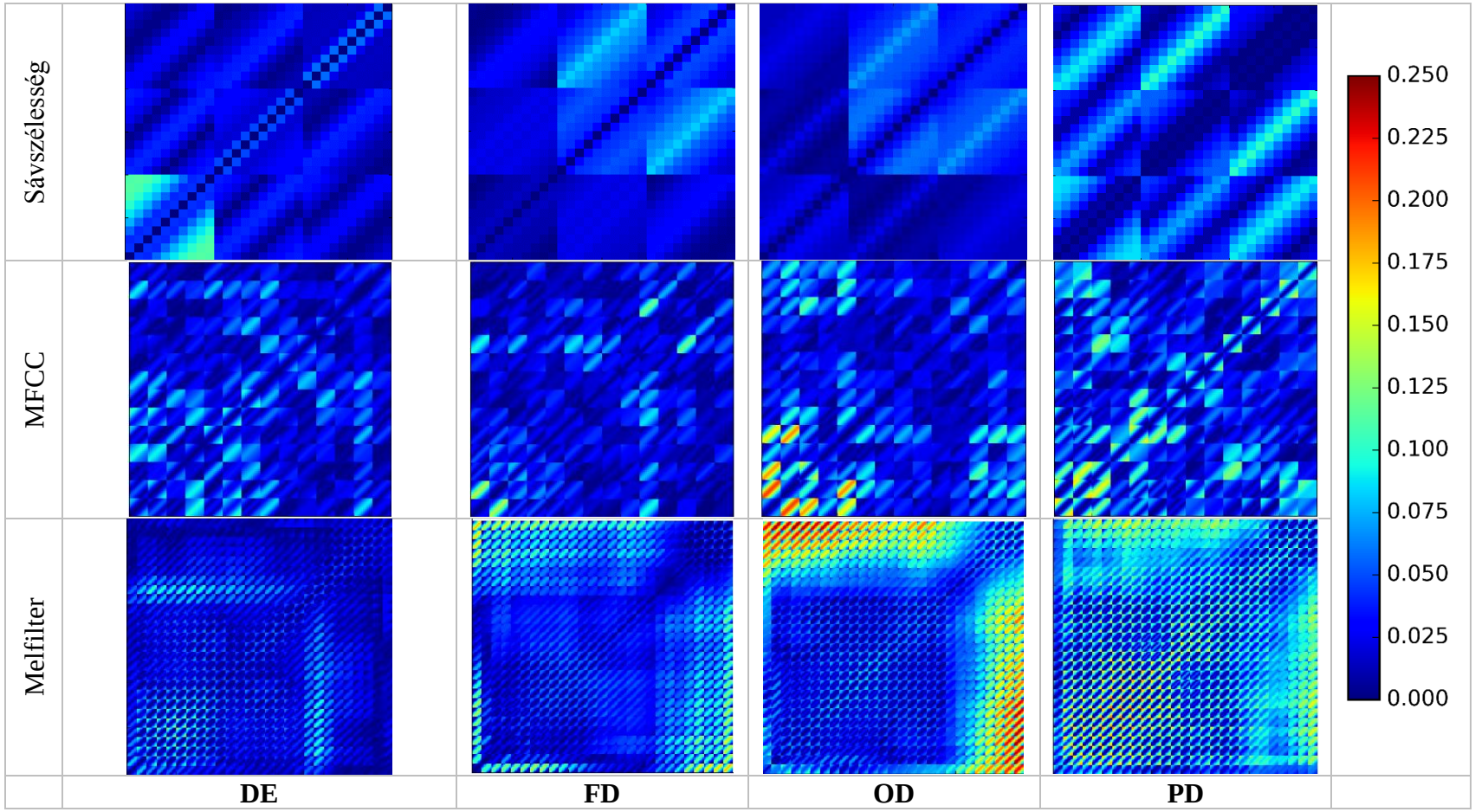
[81] T. Fawcett, „An introduction to ROC analysis”, *Pattern Recognit. Lett.*, köt. 27, sz. 8, o. 861–874, jún. 2006, doi: 10.1016/j.patrec.2005.10.010.

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

[83] Keras, „Usage of optimizers”, *keras.io*. [Online]. Elérhető: https://keras.io/optimizers/. [Elérés: 02-ápr-2020].

[84] R. Adrian, „Keras Conv2D and Convolutional Layers”, *https://www.pyimagesearch.com/*, 2018. [Online]. Elérhető: https://www.pyimagesearch.com/2018/12/31/keras-conv2d-and-convolutional-layers/. [Elérés: 01-máj-2020].

Melléklet



1. ábra (Melléklet): Beteg csoportok átlagos struktúrái az egészségeshez viszonyítva.