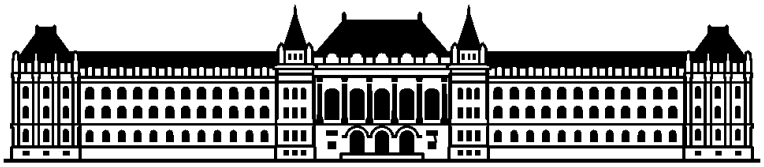


BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
TÁVKÖZLÉSI ÉS MÉDIAINFORMATIKAI TANSZÉK

# DISZFÓNIA AUTOMATIKUS FELISMERÉSE

**Ph.D. tézisfüzet**

Tulics Miklós Gábriel, MSc

Konzulens:
Vicsi Klára, DSc

BUDAPEST, MAGYARORSZÁG

# 1. Bevezetés

A diszfónia (vagy hangképzészavar) általában olyan állapotra utal, amikor egy személy rendellenesen állít elő hangot. A diszfónia a világ népességének körülbelül 30%-át érinti életének egy pontján [1, 2, 3, 4], ugyanakkor a diszfónia nem összetévesztendő a rekedtséggel. Míg rekedtségről a páciensek számolnak be, amikor hangminőségük megváltozását észlelik, addig egy diszfóniás hangot csak egy szakorvos tudja megállapítani, mivel a diszfóniás hangot rekedt, levegős, érdes vagy durva vokális tulajdonságok jellemzik, alacsonyabb fonációs funkcionalitással.

A diszfónia minden életkorú beteget érint, azonban a kutatások azt mutatják, hogy a gyermekek és az idős (65 év feletti) betegek körében magasabb a diszfónia megjelenésének a kockázata. A diszfónia gyakoribb az idősebb emberek és általában az olyan emberek között akik a foglalkozásuk során az átlagnál lényegesen többet használják a hangjukat, mint például pedagógusok, oktatók között [5, 6, 7, 8]. A fiatalok 23,4%-a szenved diszfóniától gyermekkoruk különböző időszakában [9, 10, 11, 12]. Az adatok tehát azt mutatják, hogy szinte minden negyedik gyermeknek patológiás a hangja, ugyanakkor a tanulmányok szerint a fiúk között gyakoribb a diszfónia, mint lányok körében, az arány 70-30%.

A hangképzési rendellenesség típusa szerint a diszfónia organikus vagy funkcionális jellegű. Az organikus diszfónia (OD) a beszéd egyik alrendszerének valamilyen fiziológiai változásából származik, míg a funkcionális diszfónia (FD) egy hangproblémára utal fizikai elváltozás hiányában. Az American Speech-Language-Hearing Association szerint az organikus rendellenességeket fel lehet osztani neurogén és strukturális rendellenességekre [13]. A neurogén hangképzési rendellenességek magukban foglalják azokat a hangproblémákat, amelyeket a hangképző üreg izmainak rendellenes koordinációja, szabályozása, vagy izommozgás erőssége okoz, mely hátterében neurológiai betegség állhat, például a stroke, Parkinson-kór, a szklerózis multiplex (sclerosis multiplex), a myasthenia gravis és az amyotrophicus laterális szklerózis. Strukturális organikus rendellenességek magukban foglalják a morfológiai változásokat, mint például hangszalagcsomók, polipok, gastroesophagealis reflux betegség (GERD), ciszták és hangszalag bénulás (recurrent paresis, RP) [14].

## 2. Kutatási célkitűzések

Kutatásaimmal szeretnék hozzájárulni a diszfónia beszéd alapú felismeréséhez és súlyosságának automatikus becsléséhez felnőtt és gyermekhangok esetében azáltal, hogy mélyebben megismerem a funkcionális és organikus diszfónia beszédre gyakorolt hatását. A kutatás során a következő célkitűzéseim vannak:

a) A diszfónia súlyosság automatikus becslési lehetőségeinek vizsgálata;

b) Kísérlet egy bináris osztályozásra a diszfóniás és az egészséges beszéd elválasztására, különféle gépi tanulási megközelítések segítségével;

c) A funkcionális és a organikus diszfónia automatikus elválasztásának lehetőségeinek elemzése;

d) A gyermekek egészséges és diszfóniás hangjainak automatikus elválasztásának lehetőségeinek elemzése.

Téziseim a diszfónia kutatásában az alábbiakban járul hozzá. Az összes elemzésem során magyar beszédmintákat használtam. A diszfónia elemzésének szempontjából a magyar nyelv viszonylag gyengén kutatott. A diszfóniás és egészséges hangok magyar beszédmintákkal történő automatikus osztályozásának témáját még nem vizsgálták sem felnőttek, sem gyermekek esetében.

Minden elemzésem folyamatos beszéd felhasználásával történik. A kitartott magánhangzók könnyebben használhatók, mert nem igényelnek erőforrás-igényes és nyelvfüggő szegmentálást.

Kísérletet tettem a funkcionális és organikus diszfónia automatikus osztályozására. Legjobb tudomásom szerint eddig nem volt erre irányuló kutatás. Egy diagnosztikát segítő rendszer, amely nemcsak az egészséges és a diszfóniás hangokat osztályozza, hanem a diszfónia típusát is képes meghatározni, jelentősen felgyorsíthatja azt a folyamatot, amelyben a betegeket szakemberhez irányítják. Ha a rendszer funkcionális diszfóniát észlel, a beteget foniáter szakorvoshoz vagy logopédushoz lehet irányítani. Ha azonban a rendszer organikus diszfóniát észlel, a beteget otolaringológushoz vagy onkológushoz kell irányítani. A betegek mielőbbi megfelelő szakrendelésre irányításával rengeteg értékes időt lehet megspórolni a betegek diagnózisának felállításában.

Megkíséreltem az egészséges és diszfóniás gyerekhangok elkülönítését. A végcél egy olyan szűrőrendszer létrehozása, amelyet óvodai dolgozók használhatnak. Ha egy diszfóniás hangú gyereket időben ki lehet szűrni, akkor nagyobb esélye van arra, hogy szakmai segítséget kapjon fül-orr-gégésztől vagy logopédustól.

# 3. Felhasznált adatbázisok és módszertan

## 3.1. Módszertan

Kutatásaim során statisztikai összehasonlító elemzést végeztem az egészséges és organikus, valamint funkcionális diszfóniától szenvedő emberek beszédéből származó akusztikai jellemzőkkel. A statisztikai összehasonlító elemzés segít megérteni a diszfónia beszédjellemzőkre gyakorolt hatását, és gyors jellemzőkiválasztási módszerként használható a gépi tanulási eljárásokhoz is.

Gépi tanuló eljárások használatával osztályozási és regressziós feladatokat valósítottam meg annak érdekében, hogy megvizsgáljam milyen pontossággal becsülhető meg a diszfónia, valamint a diszfónia súlyossága mekkora pontossággal állapítható meg automatikusan a beszédből kinyert jellemzők segítségével.

### 3.1.1. RBH skála

A rögzített hangfelvételeket az RBH skála alapján egy foniáter szakember sorolta be [15]. A diszfónia súlyosságát az RBH skálával lehet meghatározni, ahol az R (roughness) az érdességet, a B (breathiness) a levegősséget, míg a H (hoarseness) az általános rekedtséget hivatott mutatni. A H értéke nem lehet kisebb, mint a másik két kategória maximuma. Például ha a B=3 és az R=2, akkor a H értéke 3, nem lehet 2, vagy 1. Az egészséges hangok kódja így R0B0H0 szerint alakul. Ptok és társai igazolták, hogy az RBH skála használata megfelelő a klinikai alkalmazásokra [16]. Ezt a skálát használtam a diszfónia mértékének megkülönböztetésére az adatbázisban. A tanulmányom során az általános rekedtség (H) vizsgálata és diagnosztizálása történt.

### 3.1.2. Statisztikai módszerek

Annak érdekében, hogy megvizsgáljam az akusztikai jellemzők kapcsolatát a diszfónia súlyosságával, kiszámoltam a Pearson korrelációs együtthatót. A korreláció erősségének értelmezéséhez az (r abszolút értékére) Evans által javasolt útmutatót használtam [17]:

- 0,00-0,19 „nagyon gyenge”;
- 0,20-0,39 „gyenge”;
- 0,40-0,59 „mérsékelt”;
- 0,60-0,79 „erős”;
- 0,80-1,0 „nagyon erős”;

Annak meghatározásához, hogy a változók közötti korreláció szignifikáns-e, össze kell hasonlítani egy kiszámított p-értéket egy szignifikanciaszinttel. A korrelációs együttható szignifikanciájának vizsgálatához az $\alpha = 0{,}01$ szignifikanciaszintet használtam.

Khí-négyzet próbával, döntési táblázatok alapján hasonlítottam össze az osztályozók teljesítményét, azt feltételezve, hogy a fals negatívok és a fals pozitívok hasonló költségekkel járnak. A khí-négyzet próba nullhipotézise (H0), hogy a megfigyelt értékek és a várt értékek függetlenek, az alternatív hipotézis (H1) szerint ezek függőek. A próba során $\alpha = 0{,}05$ szignifikanciaszintet használtam.

Statisztikai analízissel ellenőriztem, hogy van-e különbség az OD és az FD csoportok rekedtségének súlyossága megoszlásában. Mivel az RBH súlyossági pontszáma ordinális, a Mann-Whitney U próbát használtam a különféle adatbázisokban szereplő súlyossági pontszámok összehasonlítására. A Mann-Whitney U próba a kétmintás t-próba nem parametrikus megfelelője. Az összes Mann-Whitney U próba során 95%-os szignifikanciaszintet ($\alpha = 0{,}05$) használtam. A próba nullhipotézise, hogy a két minta azonos eloszlásból származik.

A kutatásom során felkért négy szakember RBH-értékelésének konzisztenciáját szintén megvizsgáltam a Cronbach alfa és az osztályon belüli korrelációs együttható (Intra Class Correlation Coefficient – ICC) segítségével. Mindkét mérőszám széles körben használt a belső konzisztencia mérésére, kifejezésére.

### 3.1.3. Jellemzőkiválasztás

A bemeneti vektor dimenzióinak csökkentése érdekében a *Forward Feature Selection (FFS)* algoritmust használtam. Az FFS egy iteratív algoritmus, amely kiválasztja a legjobb jellemzőt egy előre meghatározott költségfüggvény kielégítése alapján, minden lépésben egy újabb jellemzőt hozzáadva a paraméterek halmazába. Jelen esetben a maximális pontosság szerint lettek kiválasztva a jellemzők.

### 3.1.4. Alkalmazott gépi tanuló eljárások

#### Bináris osztályozás

A bináris osztályozáshoz *SVM (Support Vector Machine)* osztályozót használtam lineáris és radiális bázisfüggvény (rbf) kernellel. Az SVM egy felügyelt gépi tanulási eljárás, amelyet főként bináris osztályozási feladatokhoz használnak [18]. Egy úgynevezett „kernel trükköt” használ az adatok átalakításához, és ezen átalakítások alapján megtalálja az optimális határt a lehetséges kimenetek között.

A második osztályozó egy *Fully-Connected mélyneurális hálózat (Fully-Connected Deep Neural Network)*, 4 rejtett réteggel, mindegyik rétegben 25 neuronnal [19]. ReLU (Rectified Linear

Unit) aktivációs függvényt használtam a rejtett rétegeken, a Softmax aktivációs függvényt a kimeneti rétegben. Optimalizációra Adam sztochasztikus optimalizációs algoritmust használtam. Költségfüggvénynek binary crossentropy-t, ami általános a bináris (két osztályú) osztályozási problémáknál. A túltanulás elkerülése végett kiejtéses regularizációt (dropout) használtam, 0,25-ös értékkel.

**Nem felügyelt klaszteranalízis**

A *k-means* egyik legegyszerűbb algoritmus, amely felügyelet nélküli tanulási módszert használ az ismert klaszterezési kérdések megoldására [20]. Ez a módszer megfelel a klaszterezési probléma gyors és egyszerű tanulmányozására, hiszen könnyen implementálható, továbbá a klaszterezési eredmények könnyen értelmezhetők. A klaszteranalízis segítségével a mintákat klasztereknek nevezett relatív csoportokba sorolhatjuk. A kutatásom során a csoportok megfelelnek a diszfónia súlyosságának hozzárendelt értékeknek. A klaszteranalízis elvégzése közben egyetlen mintáról sem rendelkezünk előzetes tudással a klasztertagságokat illetően. Ha az akusztikai jellemzők halmaza, valamint a felügyelet nélküli tanulási módszer rögzítve van, akkor összehasonlíthatunk négy klasztermodellt, amiket aszerint címkézünk, hogy a négy szakember a hangmintákat meghallgatva milyen diszfónia súlyossági kategóriába sorolta őket. Az RBH szubjektív természetének vizsgálatához k-means klaszteranalízist végeztem.

**Regresszió analízis**

A diszfónia súlyosságának automatikus meghatározásához a lineáris és az rbf kernelfüggvényű *Support vector regression (SVR)* eljárást használtam [21]. Általában a lineáris kernelű SVR kevésbé időigényes. Az rbf-kernellel rendelkező SVR jó általánosító képessége van, valamint jól tolerálja a bemeneti zajt.

### 3.1.5. Kiértékelési módszerek

A gépi tanulási algoritmusok teljesítményének becsléséhez és összehasonlításához a *Leave-one-out cross validation (LOOCV)* elnevezésű keresztvalidációs technikát használtam. Ez a k-fold keresztvalidáció egy elterjedt formája, ahol a k egyenlő a minták számával. A k-fold keresztvalidáció eljárás során egy részt használunk validálásra és a maradék k-1 részt tanításra. K darab iteráció során minden részhalmazt felhasználunk egyszer validáló részként, majd a kapott k eredmény átlagát vesszük. Ilyen módon egy sokkal pontosabb képet kapunk a modellünk teljesítményéről, mint amikor egyszerű felbontást alkalmazunk, hiszen minden adatot felhasználtunk tanításra és validálásra is. Ennek a megközelítésnek az a hátránya, hogy számítási szempontból időigényesebb, mint a k-fold keresztvalidáció (ahol k kisebb, mint az minták száma).

Annak érdekében, hogy az osztályozási, valamint klaszterezési eljárások teljesítményét leírjam, megadtam a tévesztési mátrixaikat. Munkám során *a pontosság (accuracy)*, *fedés (recall)*, valamint a *precizitás (precision)* metrikákat használom.

Regressziós feladatok pontosságának jellemzésére két leíró jellemzőt használtam. A regressziós módszerek teljesítményét *az átlagos négyzetes hibaértékének a gyökével (RMSE – root mean square error)* határoztam meg, a becsült H diszfónia súlyossági értékek és célértékként felhasznált H súlyossági értékek közötti lineáris kapcsolatot a *Pearson-korreláció* jellemezte.

## 3.2. Felhasznált Beszédadatbázisok

### 3.2.1. Magyar Diszfóniás és Egészséges Felnőtt Beszédadatbázis

A felvételek közeltéri mikrofonnal (Monacor ECM-100), alacsony zajszinttű külső hangkártyával (Creative Soundblaster Audigy 2 NX), jó minőségű A/D konverterrel (kódolás: PCM, mintavételezési frekvencia: 16 kHz, kvantálás: 16-bit) kerültek rögzítésre, csendes irodai környezetben (orvosi szobában). Minden páciens Aiszóposz meséjét, „Az északi szél és a nap”-ot olvasta fel. Ezen népmese gyakran használt a foniátriai kutatásokban, mivel a mese fonetikailag kiegyensúlyozott. Vagyis a szöveganyagát úgy szerkesztették meg, hogy az adott nyelvben előforduló minden beszédhang, valamint a leggyakoribb hangkapcsolatok szerepelnek benne. Számos nyelvre elkészült ez a szöveg, köztük a jelen esetben is használt magyarra. Az adatbázis fonéma szintű szegmentálása automatikus fonémaszegmentáló programmal történt, amit a Beszédakusztikai Laboratórium munkatársai fejlesztettek ki [22], majd szükség esetén kézileg történt ezek javítása. A szegmentálást a SAMPA fonetikus ábécéjével végeztem [23]. A tézisfüzet többi részében a fonémákat zárójelben jelölöm a SAMPA karaktereik szerint.

A hangadatbázis felvételei között számos kórképű beteg fordult elő: funkcionális diszfónia, hangszalagbénulás (recurrens paresis), a hangképző szervrendszer különböző pontjain előforduló tumorok, gasztroesophageal reflux (GERD), krónikus gégegyulladás, agyideggyulladás (bulbar paresis), amiotrófás laterálszklerózis (ALS), leukoplakia, spazmodikus diszfónia, stb. A leggyakoribb betegségek a funkcionális diszfónia (functional dysphonia - FD) és a hangszalagbénulás (recurrens paresis - RP). A diszfóniában szenvedő páciensek hangfelvételeire „Dys”-ként utalok. Összehasonlítás végett egészséges páciensekről is készültek felvételek. Az egészséges emberektől származó felvételek rögzítése a korábban említett esetekhez nem kapcsolódó kivizsgálások során kerültek felvételre. Ezeket a felvételeket „HC”-nak neveztem el (Healthy Control).

A hangfelvételek eloszlását az adatbázisban a 1. táblázat mutatja. Az adatbázis összesen 450 felvételt tartalmaz, 257 diszfóniában szenvedő páciens hangját (156 nő és 101 férfi) és 193 egészséges hangú ember felvételét (108 nő és 85 férfi).

Kutatásom során az adatbázis folyamatosan bővült új felvételekkel, ezért a dolgozat néhány állításában egy kisebb adatbázisból vontam le következtetéseimet. Minden tézispontban bemutatom az általam használt adatbázist.

**1. táblázat.** Magyar Diszfóniás és Egészséges Felnőtt Beszédadatbázis.

| Diagnózis | | | |
|---|---|---|---|
| *Biológiai nem* | Diszfónia | Egészséges | Összesen |
| *Nő* | 156 | 108 | 264 |
| *Férfi* | 101 | 85 | 186 |
| *Összesen* | 257 | 193 | 450 |

### 3.2.2. Diszfóniás és Egészséges Gyermek Beszédadatbázis

Több óvodában gyűjtöttem gyermekektől származó hangmintákat. Az összes felvételt szülői hozzájárulással készítettem, legtöbb esetben a gyermekek szüleinek jelenlétében. A felvételek során Bartos Erika „Mókus" című versét mondta el az összes gyerek. A vers választását a sok szó eleji „m" nazális hangzó indokolta, ami terápiában a hang előrehozásánál jelentős szerepet játszik. Bartos Erika verseit az 5-10 éves korosztályú gyermekek nagyon szeretik, mivel könnyen tanulhatóak. Egy felvétel átlagosan 20 másodperc hosszú. A vers leggyakoribb magánhangzója az [o] fonéma 16 előfordulással, amelyet 14 előfordulással az [O] fonéma és 9 darab előfordulással az [E] követ.

A felvételek közeltéri mikrofonnal készültek (Monacor ECM-100), alacsony zajszinttű külső hangkártyával (Creative Soundblaster Audigy 2 NX), jó minőségű A/D konverterrel (kódolás: PCM, mintavételezési frekvencia: 16 kHz, kvantálás: 16-bit) kerültek rögzítésre.

A szegmentálás a 3.2.1 automata szegmentáló segítségével lett végezve, majd kézzel lett korigálva. Az adatbázis összesen 59 felvételt tartalmaz: 25 felvétel diszfóniában szenvedő gyerektől (átlag életkor: $6{,}52(\pm 1{,}94)$ (a gyerekek többségének diagnózisa funkcionális diszfónia, míg háromnak hangszalagcsomói voltak), valamint 34 felvétel egészséges gyermekektől származik (átlag életkor: $5{,}35(\pm 0{,}54)$. A 2. táblázat összefoglalja az adatbázisban szereplő hangfelvételeket.

**2. táblázat.** Diszfóniás és Egészséges Gyermek Beszédadatbázis.

| Diagnózis | | | |
|---|---|---|---|
| *Biológiai nem* | Diszfónia | Egészséges | Összesen |
| *Lány* | 5 | 15 | 20 |
| *Fiú* | 20 | 19 | 39 |
| *Összesen* | 25 | 34 | 59 |

## 3.3. Bemeneti vektor

### 3.3.1. Bemeneti vektor akusztikai jellemzők alapján

Abban az esetben, amikor felnőttek hangját vizsgáltam, a bemeneti vektort az [E] fonémán (az olvasott szöveg leggyakoribb magánhangzója), különböző fonetikai osztályokon, valamint az egész hangfelvételen mért akusztikai jellemzőkből hoztam létre.

**Az [E] magánhangzón** a következő akusztikai jellemzőket mértem: jitter(ddp), shimmer(ddp), HNR (Harmonics-to-Noise Ratio) és a mel-frekvenciás kepsztrális együtthatók (MFCC) első komponense ($c_1$) (a továbbiakban „mfcc01"). A „ddp" rövidítés a periódusok különbségeinek különbségére utal (Difference of Differences of Periods). **Különböző fonetikai osztályokon** a Soft Phonation Index (SPI) és a Empirical mode decomposition (EMD) módszeren alapuló Intrinsic mód függvények (Intrinsic Mode Functions - IMF) entrópia energia hányadosok a következő fonetikai osztályon számoltam ki: az [E] magánhangzón, [m], [n] és [J]-vel jelölt nazális hangokon (Nasal), [E], [e:], [i], [2] és [y]-vel jelölt magas magánhangzókon (HighVowels), [O], [A:], [o] és [u] jelölt mély magánhangzókon (LowVowels), [v], [z] és [Z]-vel jelölt zöngés spiránsokon (VoicedSpirants), valamint [b], [d], [g], [dz], [dZ] és [d'] jelölt zárhangokon és afrikátákon (VoicedPlosives). Az SPI paramétert az **egész hangfájlra** is kinyertem.

**Származtatott akusztikai jellemzőkként** kiszámoltam a jellemzők átlagát, szórását és tartományukat. Így összesen 49 jellemző került vizsgálatra felvételenként. Az akusztikai jellemzők részletes leírásáról [J4] hivatkozásban lehet olvasni. Továbbiakban erre a bemeneti vektorra a „49 jellemző halmaz"-ként fogok utalni.

A gyermekek beszédének vizsgálatakor az akusztikai jellemzőket az [o] magánhangzókon (amely a vers leggyakoribb magánhangzója), különböző fonetikai osztályokon és az egész hangfelvételen mértem. **Az [o] magánhangzón** a következő akusztikai jellemzőket mértem: jitter(ddp), shimmer(ddp), HNR, 12 MFCC, alaphang (F0), első három formánsfrekvencia (F1, F2, F3), első három formánsfrekvencia tartománya (F1BW, F2BW, F3BW). Az alaphang számítását az [24] hivatkozásban ismertetett autokorrelációs módszerrel végeztem. A formáns frekvenciákat Gaussian ablak alkalmazásával mértem meg egy 150 ms hosszú jelre, 10 ms időléptékkel. Mindegyik keretnél LPC együtthatók kerültek meghatározásra. **Különböző fonetikai osztályokon** az SPI és az IMF entrópia energia hányadosok lettek kiszámolva: az [o] magánhangzón, nazális hangokon, magas magánhangzókon, mély magánhangzókon , zöngés spiránsokon, zárhangokon és afrikátákon. Az SPI paramétert az **egész hangfájlra** is kinyertem. **Származtatott akusztikai jellemzőkként** kiszámoltam a jellemzők átlagát, szórását és tartományukat. Így összesen 103 jellemző került vizsgálatra gyermek hangfelvételként.

### 3.3.2. Bemeneti vektor az ASR fonémaszintű valószínűségi értékei alapján

Az automatikus beszédfelismerők (Automatic Speech Recognizer - ASR) akusztikus modelljei felhasználhatók a diszfónia kimutatására és osztályozására szolgáló jellemzők kivonására is. A mai legmodernebb hibrid ASR akusztikus modellek egy átmeneti modellből (transition model, például ami egy Rejtett Markov modell) és egy fonémaosztályozóból (például ami egy Mély Neurális Hálózat, Deep neural network - DNN) állnak. A fonémaosztályozó felhasználható a keretek önálló módban történő osztályozására is (a felismerési hálózat és az ASR dekóder hozzáadása nélkül), a beszédkeretek egyenkénti továbbításával. Ily módon a fonémaosztályozójának DNN softmax rétegéből 10 ms-os időkereten belül megkapjuk a fonémák posterior valószínűségét. Ennélfogva csak az akusztikus modell fonémaosztályozó komponensét használjuk a becsléshez.

A kísérleteimhez használt akusztikus modellt a BABEL [25], a Magyar Referencia Beszéd Adatbázis (MRBA) [26] és a Magyar Hírközlési Beszédadatbázis [27] kevert magyar adatain tanítottam a Kaldival [28], az „nnet2" WSJ „recept" alapján. Fonémaosztályozója összekapcsolt és az LDA + MLLT formájában átalakított MFCC jellemzőkön és egy előrecsatolt mély neurális hálózaton alapul, négy, 1024 dimenziós rejtett réteggel, p-normál nemlinearitással ($p = 2$) és softmax kimenettel 2500 szenonra. A szenonokat majd fonémákká alakítja a magyar nyelvre definiált 39 elemű SAMPA fonémakészlet szerint [25].

A fonéma-valószínűségek elemzése közben az [E] magánhangzók, nazálisok, magas magánhangzók, mély magánhangzók, zöngés spiránsok, valamint zárhangok és afrikáták valószínűségének átlagait, szórásait és tartományát határoztam meg. Ez 21-dimenziós vektort eredményezett. Erre a bemeneti vektorra „ASR valószínűségi jellemzőhalmaz"-ként hivatkozom, hogy fenntartsam a kifejezések koherenciáját a nemzetközi irodalommal, bár ezek a jellemzők szűk értelemben nem ASR-jellemzők, mivel ezeket a jellemzőket egy kis akusztikus modell fonémaosztályozója állítja elő.

# 4. Eredmények

## 4.1. A diszfónia súlyosságának automatikus felmérése

### 4.1.1. Fonetikai osztály alapú korrelációs-analízis diszfónia súlyosságára

A diszfóniás beszéd diagnosztizálásakor és kezelésekor a szakorvos általában személyesen értékeli a beteg hangjának minőségét. Az értékelés szubjektív jellegű. A diszfóniás hang súlyosságát egy szakorvos értékeléseként vagy egy szakorvosok csoport értékelésének medián vagy átlagos súlyossági fokaként határozhatjuk meg [29, 30]. Ha több értékelő áll rendelkezésre, a szakemberek a diszfónia súlyosságának értékelését korábban rögzített hangminták meghallgatásával végzik. Az értékelések változhatnak az értékelő szakemberek között, ezért tanácsos konzisztenciaelemzést végezni. Law és munkatársai munkájukban [31] arra a következtetésre jutottak, hogy folyamatos beszéd esetén a szakemberek értékelésének konzisztenciája nagyobb, mint ha kitartott hangokat használnának. A legtöbb klinikai környezetben az akusztikai mérések kitartott hangok alapján történnek, azonban a folyamatos beszéd vizsgálatának számos előnye van a kitartott magánhangzók elemzésével szemben, mivel a folyamatos beszéd tartalmaz számos alapfrekvencia variációt, szüneteket, így lehetőség van a beszédhangok különböző változatainak elemzésére.

Fontos feladat a releváns akusztikus jellemzők azonosítása, annak érdekében, hogy a diszfóniás hangok súlyosságát automatikusan megbecsülhessük. A következő tézisek ezzel a kérdéssel foglalkoznak.

Kisméretű beszédadatbázis felhasználása esetén nagyon fontos a beszédjellemzők lehető legnagyobb mértékű optimalizálása, ahelyett, hogy sok akusztikai jellemzőt használnánk, azzal a kockázattal, hogy nemkívánatos zaj kerüljön a rendszerbe. A hipotézisem a következő: a szakember által megítélt kórosan elváltozott hang súlyossága (RBH) korrelál az akusztikai jellemzők megváltozásának fokával.

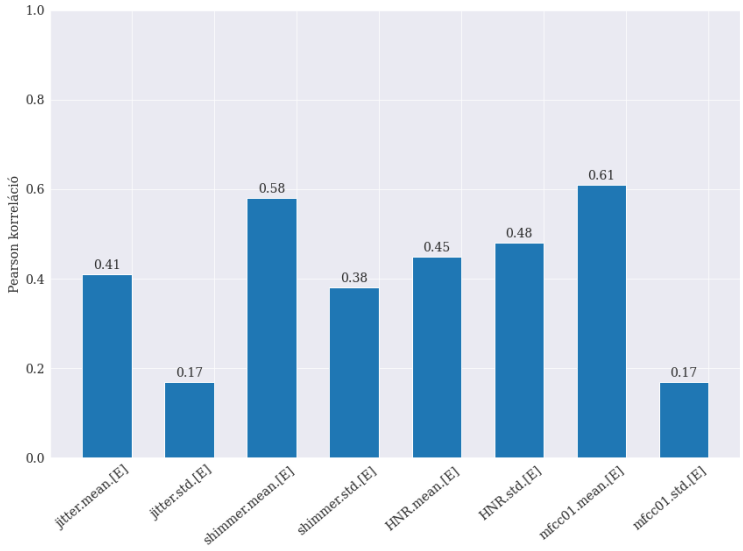
**3. táblázat.** Egészséges és diszfóniás emberektől származó hangfelvételek eloszlása az adatbázisban, H értéktől függően.

| | **H értéke** | | | | **Összesen** |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | |
| **Férfi** | 67 | 34 | 20 | 32 | 153 |
| **Nő** | 69 | 72 | 27 | 21 | 189 |
| **Összesen** | 136 | 106 | 47 | 53 | 342 |

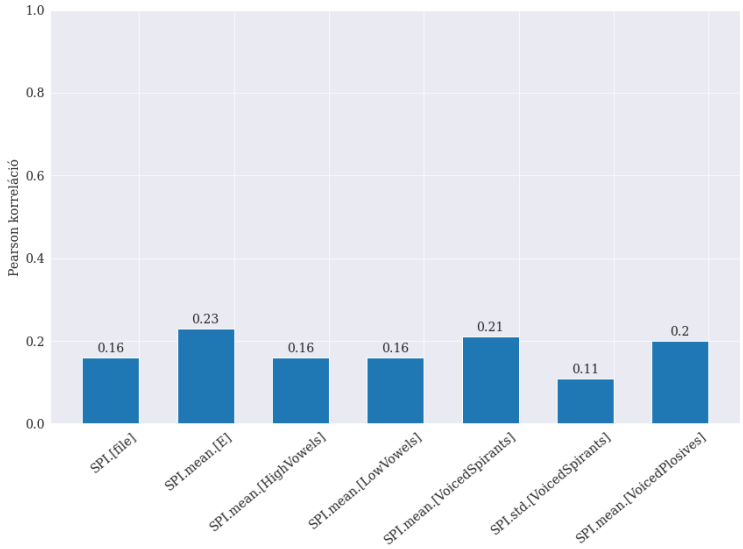
Első tézisem során korrelációs elemzést végeztem (a 3.3.1. fejezetben bemutatott) akusztikai jellemzők és egy szakember által meghatározott rekedtségi súlyosságok között. Az említett szakember kezelte a hangmintához tartozó pácienseket és határozta meg a diagnózisaikat. A szakember

pácienskonzultációk során közvetlenül hallgatta és értékelte a páciensek hangminőségét. A Pearson-féle korrelációs együtthatót minden esetben kiszámítottam, ahol a korreláció szignifikáns volt az akusztikai jellemző és a szubjektív értékelés közötti, 0,01 (kétoldalú) szignifikanciaszinten.

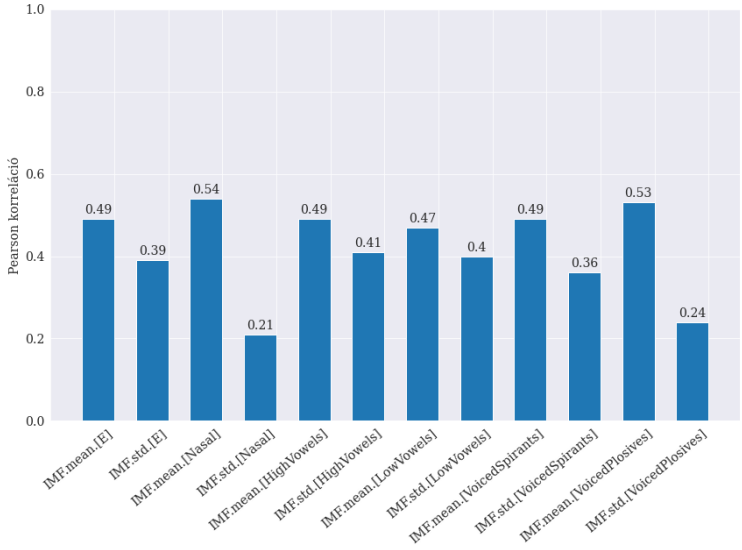
Az elemzést a 3.2.1. alfejezetben bemutatott beszédadatbázis részhalmazán végeztem. A kísérletben használt hangfelvételek (RBH skála) H szerinti eloszlását a 3. táblázat mutatja. A 0 H értékkel rendelkező felvételek egészséges személyektől származnak. Ebben a kísérletben összesen 136 egészséges embertől származó, valamint 206 diszfóniás hangot produkáló embertől származó hangfelvétellel dolgoztam.




**1. ábra.** Pearson korreláció általánosan használt akusztikai jellemzőkkel.




**2. ábra.** Pearson korreláció SPI-vel különböző fonetikai osztályokon mérve.




**3. ábra.** Pearson korreláció IMF entrópia alapú energia hányadosokkal, különböző fonetikai osztályokon mérve.

Az 1. ábrán bemutatott eredmények alapján kijelenthető, hogy a jitter (ddp), shimmer (dda), HNR és az mfcc01 korrelálnak a diszfónia súlyosságával. Az ábrákon a Pearson-korrelációk abszolút értékei vannak feltüntetve, ahol a korreláció szignifikáns a 0,01 szintnél. Minél nagyobb a korrelációs együttható abszolút értéke, annál erősebb a kapcsolat az akusztikai jellemzők és a rekedtség súlyossága között. Evans javaslata szerint a jitter.std.[E] és mfcc01.std.[E] korrelációja a rekedtség súlyosságával „nagyon gyengének”, a shimmer.std.[E] korrelációja „gyengének”, a jitter.mean.[E], shimmer.mean.[E], HNR.mean.[E] és HNR.std.[E] korrelációja „mérsékeltnek”, míg az mfcc01.mean.[E] korrelációja a rekedtség súlyossága „erősnek” tekinthető.

Amikor az SPI-t különböző fonetikai osztályokon mértem, a Pearson-korrelációs együtthatók 0,11 és 0,23 közötti értéket vettek fel, ezek „nagyon gyenge” és „gyenge”, viszont szignifikáns korrelációk az akusztikai jellemzők és a rekedtség súlyossága között. Az eredményeket a 2. ábra mutatja.

Amint a 3. ábra mutatja, a különböző fonetikai osztályokon mért IMF entrópia alapú energia hányadosok is szignifikánsan korrelálnak a diszfónia súlyosságával. Az IMF.std.[E], IMF.std.[Nasal], IMF.std.[VoicedSpirants], IMF.std.[VoicedPlosives] „gyenge” korrelációt mutatnak, míg az IMF.mean.[E], IMF.mean.[Nasal], IMF.mean.[HighVowels], IMF.std.[HighVowels], IMF.mean.[LowVowels], IMF.std.[LowVowels], IMF.mean.[VoicedSpirants] és IMF.mean.[VoicedPlosives] jellemzők „mérsékelt” korrelációt mutatnak a diszfónia súlyosságával.

**I. A. Tézis** [C4] *Megmutattam, hogy a jitter(ddp), shimmer(dda), Harmonics-to-Noise Ratio (HNR), mfcc01, Soft Phonation Index (SPI) és IMF entrópia alapú energia hányadosok különböző fonetikai osztályokon mérve szignifikáns korrelációt mutatnak 0,01 szinten a diszfónia súlyosságával a Magyar Diszfóniás és Egészséges Felnőtt Beszédadatbázist használva.*

### 4.1.2. Nem felügyelt és felügyelt tanulási módszerek a szakemberek négy fokozatú értékelésének modellezésére

Fontos kérdés az is, hogy a korrelációs elemzéssel kiválasztott akusztikai jellemzők alkalmasak-e a szakemberek négy fokozatú értékelésének (RBH szubjektív skála szerinti) modellezésére. Ebben a vizsgálatban két beszédadatbázist használtam: a Kezdeti Diszfóniás és Egészséges Beszédadatbázist, valamint a Kiválasztott Diszfóniás és Egészséges Beszédadatbázist. Egy nem felügyelt tanulási módszert, a k-means algoritmust használtam a Kiválasztott Diszfóniás és Egészséges Beszédadatbázison. Olyan adathalmazokon szokás ezt az algoritmust használni, ahol az adataink nem rendelkeznek címkékkel. Az algoritmus célja, hogy csoportokat, úgynevezett klasztereket találjon az adathalmazban, ahol csoportok számát a k jelöli. Mielőtt elvégeztem a nem felügyelt tanulási módszert, kétosztályos osztályozást hajtottam végre annak megállapítására, hogy a választott akusztikai jellemzők tartalmaznak-e elég gazdag információt ahhoz, hogy az egészséges és diszfóniás hangokat meg lehessen különböztetni egymástól, miközben csökken a bemeneti vektor dimenziója.

A Kezdeti Diszfóniás és Egészséges Beszédadatbázis, összesen 263 beszédfelvételt tartalmaz, 127 felvételt egészséges alanyoktól (62 férfitól és 65 nőtől származó hangfelvétel) és 136 felvételt tartalmaz funkcionális vagy organikus diszfóniában szenvedő betegektől (66 férfitól és 70 nőtől származó hangfelvétel). Mindegyik felvétel különböző ember hangját tartalmazza. A beteget kezelő szakember határozta meg a diagnózist. A szakember közvetlenül betegkonzultáció során hallgatta meg és értékelte a beteg beszédének minőségét. A kétosztályos osztályozási kísérlet ezen az adatbázison lett végrehajtva.

A Kiválasztott Diszfóniás és Egészséges Beszédadatbázisban szereplő hangok kórosságának súlyosságát négy szakember határozta meg. Ez a beszédadatbázis összesen 148 felvételt tartalmaz, és a felügyelet nélküli klaszter- és regressziós elemzéshez használtam fel. A négy szakember egyike állította fel a diagnózist és értékelte a beteg beszédének minőségét a konzultációk során, míg a másik három szakember nem került személyes kapcsolatba a páciensekkel, csupán a hangfelvételek visszahallgatása alapján határozták meg a diszfónia súlyosságát. Mind a négy szakembernek nagy tapasztalata van a hangképzési rendellenességekkel rendelkező páciensek kezelésében.

Kétosztályos osztályozást végeztem a Kezdeti Diszfóniás és Egészséges Beszédadatbázison LOOCV technikával és SVM osztályozóval. A kétosztályú osztályozás célja, hogy megtudjam,

hogy a kiválasztott akusztikai jellemzők tartalmaznak-e elegendő információt az egészséges és a diszfóniás hangok megkülönböztetésére, miközben jellemzőkiválasztással csökken a bemeneti vektor dimenziója. Az osztályozási kísérleteket során több kombinációt is kipróbáltam, lineáris és rbf kernelt is használtam. Az SVM osztályozó C hiperparaméterének alapértelmezett értéke 1, míg rbf kernel esetén a $\gamma$ alapértelmezett értéke 1/az akusztikai jellemzők száma. Az optimális hiperparaméterek kiválasztásához rácsos keresést (grid search) használtam.

A legmagasabb pontosságot, 89%-ot, rbf kernel használatával értük el. A jellemzőkiválasztó algoritmus csökkentette a bemeneti dimenziót 18 akusztikai jellemzőre („18 jellemző szett"), miközben magasabb pontosságot ért el, mint az alapértelmezett beállítás. Az FFS algoritmus által kiválasztott akusztikai jellemzők a következők: jitter.mean.[E], shimmer.mean.[E], HNR.mean.[E], mfcc01.mean.[E], jitter.std.[E], shimmer.std.[E], HNR.std.[E], mfcc01.std.[E],SPI.std.[E], SPI.mean.[Nasal], SPI.std.[Nasal], SPI.std.[LowVowels], SPI.mean.[VoicedSpirants], SPI.std.[VoicedSpirants], IMF.std.[E], IMF.mean.[Nasal], IMF.mean.[VoicedPlosives], IMF.std.[VoicedPlosives].

Érdekes kérdés, hogy ezen kiválasztott jellemzők modellezni tudják-e a szakemberek egyéni értékeléseit. A klaszterelemzés célja az adatokban rejlő rejtett struktúrák feltárása, hasonló elemek csoportosításával. Esetemben a rejtett struktúra a felvételek 'igaz rekedtségi címkéje', amit egy ideális szakember adna. A klaszterelemzés ezt az ideális szakembert kívánja utánozni, hogy a felvételekhez ideális rekedtségi címkézést párosítson. Természetesen jobban bízunk a szakember minősítésében, mint a klaszterező algoritmusban, de ha a megtalált klaszterek valóban közel állnak egy szakember minősítéséhez, akkor ezt az értékelést igaznak nevezhetjük. Ezért érdemes négy klaszter modellt összehasonlítanom, amelyekben egy-egy szakember ítélete jelöli az igaz címkéket. Ha az akusztikai jellemző halmaza és a felügyelet nélküli tanuló algoritmus rögzített, akkor összehasonlítható a négy szakember értékeléseinek klaszter modellje külön-külön. Az RBH skála szubjektív voltának vizsgálata céljából klaszteranalízist végeztem.

A k-means klaszterezés célja, hogy a megfigyeléseket $k$ klaszterekbe tegye, ahol $k$ a felhasználó által inputként meghatározott klaszterek száma. A klaszterek számát négyre állítottam. A klaszteranalízis a megfigyeléseket A, B, C és D klaszterekbe rendezte. A klasztereket a súlyossági értékekhez az abszolút hibák minimális átlaga alapján rendeltem: A klasztert H = 0-hoz, B klasztert H = 1-hez, C klasztert H = 2-höz, D klasztert H = 3-hoz.

Az egyes szakemberek tévesztési mátrixait a 4., 5., 6. és 7. táblázat reprezentálja. A H döntések átlagos pontossága az egyes szakemberek esetén sorrendben: 49%, 44%, 45%, 47%, az átlagos pontosság 46.25%, 2.22% szórással. Ha a négy osztály kiegyensúlyozott eloszlású lenne, akkor egy véletlen osztályozás pontossága 25%. Ebből a vizsgálatból azt a következtetést vonhatom le, hogy a bemenő jellemzővektor alkalmas a diszfónia súlyosságának egyedi értékelésére.

**4. táblázat.** Tévesztési mátrix Szakember 1 értékelése alapján.

| | | **Szakember 1 (szerint a H érték)** | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | Osztályprecizitás |
| **Prediktált címke** | 0 | **12** | 1 | 2 | 1 | 75% |
| | 1 | 13 | **33** | 5 | 3 | 61% |
| | 2 | 9 | 25 | **10** | 5 | 20% |
| | 3 | 2 | 4 | 6 | **17** | 59% |
| Osztályfedés | | 33% | 52% | 43% | 65% | |

**5. táblázat.** Tévesztési mátrix Szakember 2 értékelése alapján.

| | | **Szakember 2 (szerint a H érték)** | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | Osztályprecizitás |
| **Prediktált címke** | 0 | **11** | 3 | 2 | 0 | 69% |
| | 1 | 5 | **26** | 23 | 0 | 48% |
| | 2 | 6 | 24 | **16** | 3 | 33% |
| | 3 | 0 | 2 | 15 | **12** | 41% |
| Osztályfedés | | 50% | 47% | 29% | 80% | |

**6. táblázat.** Tévesztési mátrix Szakember 3 értékelése alapján.

| | | **Szakember 3 (szerint a H érték)** | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | Osztályprecizitás |
| **Prediktált címke** | 0 | **11** | 2 | 2 | 1 | 69% |
| | 1 | 2 | **20** | 25 | 7 | 37% |
| | 2 | 3 | 15 | **16** | 15 | 33% |
| | 3 | 0 | 1 | 9 | **19** | 66% |
| Osztályfedés | | 69% | 53% | 31% | 45% | |

**7. táblázat.** Tévesztési mátrix Szakember 4 értékelése alapján.

| | | **Szakember 4 (szerint a H érték)** | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | Osztályprecizitás |
| **Prediktált címke** | 0 | **12** | 2 | 2 | 0 | 75% |
| | 1 | 7 | **24** | 18 | 5 | 44% |
| | 2 | 6 | 18 | **17** | 8 | 35% |
| | 3 | 0 | 6 | 6 | **17** | 59% |
| Osztályfedés | | 48% | 48% | 40% | 57% | |

Meghatároztam a klaszterek által meghatározott súlyossági értékek és az egyes szakemberek értékelése között a Pearson-korrelációt, valamint a klaszterek által meghatározott súlyossági értékek és a szakemberek átlagos RBH-percepciós értékelésének Pearson-korrelációját. Az eredményeket a 8. táblázat foglalja össze. Minden korreláció „mérsékelt" viszonyt mutat. Az átlagos korreláció 0,52 és 0,01 szórás. A legmagasabb korrelációs érték a klaszterek által definiált súlyossági értékek és a szakemberek értékelésének átlaga között lett: 0,59. Mivel a talált klaszterek a legjobban a négy szakember átlagával korrelálnak, ez az igazi címke, amelyet a regressziós elemzésekhez használok célcímkeként.

**8. táblázat.** Klaszterek által definiált súlyossági értékek és a szakemberek értékelése közötti Pearson korrelációk.

| | Szakember 1 | Szakember 2 | Szakember 3 | Szakember 4 | Értékelések átlaga |
|---|---|---|---|---|---|
| **Pearson korreláció** | 0,51 | 0,54 | 0,53 | 0,51 | 0,59 |

**I. B. Tézis** [C4, J4] *Megmutattam, hogy a kiválasztott akusztikai jellemzőkkel végzett k-means klaszterezési eljárással kapott klaszterek jól korrelálnak a diszfónia súlyosságával. A klaszterek által definiált súlyossági értékek és a szakemberek értékelésének átlaga között 0,59-os Pearson korreláció volt mérhető.*

### 4.1.3. A diszfónia súlyosságának automatikus becslése regressziós elemzéssel

Ezt az elemzést a Kiválasztott Diszfóniás és Egészséges Beszédadatbázison végeztem. A szakemberek átlagos H perceptuális értékelését használtam célváltozónak a regressziós modellhez, ezért érdemes megvizsgálni a négy szakember értékeléseinek konzisztenciáját.

Fontos elemezni, hogy szakemberek értékelései konzisztensek-e. A szakemberek belső konzisztenciájának értéke képet ad arról, hogy mekkora a legnagyobb korrelációs érték, amelyet regressziós

modellünktől elvárhatunk. Nem várjuk el, hogy a regressziós modell jobb eredményt érjen el, mint egy jól képzett szakember. A szakemberek belső konzisztenciájának („megbízhatóságának") mérésére Cronbach alfa és Intra Class Correlation Coefficient (ICC) metrikát használtam. A szakemberek döntései közötti érdekes különbségek ellenére a belső konzisztencia mérésekor jó belső konzisztenciát mutattak a metrikák (Cronbach Alpha = 0,89, ICC = 0,89).

A regresszió jelentős előnyt jelent a klaszterelemzéshez képest, mivel a becslés szinte folyamatosan követi a függvényt. Ez a tulajdonság jelentősen javíthatja a modell minőségét. Az adatmennyiség kicsiny volta miatt leave-one-out cross validation technikát alkalmaztam. A regressziós módszerek jóságát a négyzetes középérték hibával (RMSE - Root Mean Square Error) és a Pearson korreláció (a cél és a prediktált H értékek közötti) értékével vizsgáltam. Az optimális hiperparaméterek megtalálásához grid search-öt használtam.

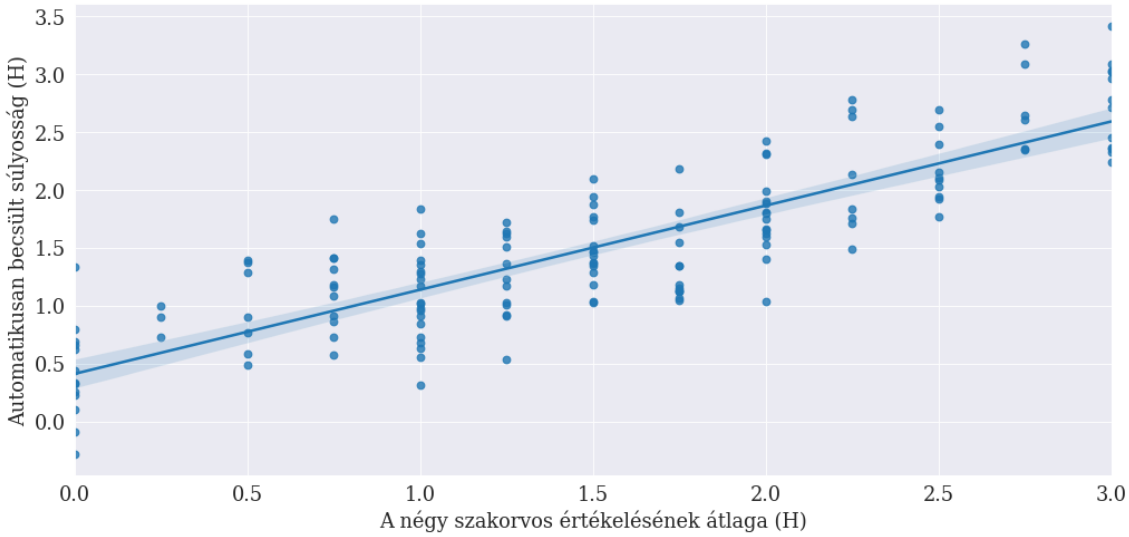
Ebben az elemzésben lineáris és rbf kernelű SVR regressziós eljárást használtam. A legjobb modell elérése érdekében több paraméter szett is kipróbálásra került bemenő jellemzővektorként: a klaszteranalízisben használt 18 jellemző szett, valamint az FFS algoritmus kimenete lineáris és rbf kernelfüggvényű regresszió esetén külön. Ahogy korábban említésre került, a négy szakember értékelésének átlaga volt a célváltozó. A 9. táblázat foglalja össze az eredményeket.

**9. táblázat.** Regresszióelemzés eredményei – a négy szakember értékelésének átlaga célváltozóként.

| **Akusztikai jellemző halmaz** | **Regresszió típusa** | **Korreláció** | **H RMSE értéke** | **Hiperparaméterek** |
|---|---|---|---|---|
| **18 jellemző szett** | lineáris kernel | 0,84 | 0,50 | C = 1 |
| **FFS eredménye, 8 jellemző szett** | lineáris kernel | **0,85** | **0,46** | C = 1 |
| **18 jellemző szett** | rbf kernel | 0,81 | 0,50 | C = 2, $\gamma$ = 0,125 |
| **FFS eredménye, 14 jellemző szett** | rbf kernel | **0,85** | **0,45** | C = 4, $\gamma$ = 0,25 |

Az FFS algoritmus az eredeti 33-dimenziós bemenetet mindössze nyolc jellemzőre csökkentette lineáris kernel használatával. A következő akusztikai jellemzők kerültek kiválasztásra: mfcc01.mean.[E], shimmer.mean.[E], SPI.std.[LowVowels], HNR.std.[E], SPI.mean.[HighVowels], IMF.mean.[Nasal], SPI.std.[VoicedPlosives], IMF.std.[LowVowels]. Ez a kombináció adta a legmagasabb, 0,85-os korrelációs értéket a négy szakember értékelésének átlagával.

Az rbf kernel használatakor az FSS algoritmus 14 jellemzőt választott ki, ezek a következők voltak: shimmer.mean.[E], HNR.mean.[E], mfcc01.mean.[E], HNR.std.[E], SPI.mean.[E], SPI.std.[E], SPI.std.[Nasal], SPI.mean.[HighVowels], SPI.mean.[LowVowels], SPI.std.[LowVowels], SPI.mean.[VoicedPlosives], IMF.mean.[Nasal], IMF.mean.[VoicedPlosives], IMF.std.[VoicedPlosives]. Ebben az esetben értem el a legalacsonyabb RMSE-értéket (0,454). Érdemes megmelíteni, hogy az FFS-el meghatározott modellek csak kevéssel jobb az eredmény, mint a 18 jellemző szett esetében.




**4. ábra.** Automatikusan becsült diszfóniasúlyosság a perceptuálisan megítélt H érték alapján, lineáris kernelű SVR-rel és 8 jellemzővel.

Ez szemlélteti a javasolt megközelítés becslési hatékonyságát, függetlenül a páciens patológiás hátterétől és a diszfóniájának súlyosságától. Mivel az ICC érték a 4 szakember között 0,89-t eredményezett, egy elméleti célnak is tekinthető, amelyet elérni akarunk. Ennek fényében a regressziós modellel kapott 0,85 korrelációs értéket szinte tökéletesnek tekinthetjük.

A 4. ábra mutatja a diszfónia automatikusan prediktált súlyosságát az eredeti perceptuális referencia értékekhez képest. Az ábra a lineáris kernelű SVR regressziós modelljét mutatja be, melynek jellemzőhalmazát az FFS algoritmus eredményezte. Az ábra ismételten bemutatja a javasolt megközelítés becslési hatékonyságát, függetlenül a páciens patológiás hátterétől és a diszfóniájának súlyosságától. Látható, hogy a modell jó predikciót ad a H1 súlyossági osztályra, súlyosabb esetekben pedig alulról becsli azokat.

**I. C. Tézis** [C4, J4] *Megmutattam, hogy a diszfónia súlyosságának automatikus becslése lehetséges csak nyolc akusztikai jellemző felhasználásával, lineáris kernelű SVR-rel. Ezzel a módszerrel 0,85 Pearson-korreláció és 0,46 RMSE érhető el a Kiválasztott Diszfóniáa és Egészséges Beszédadatbázison.*

## 4.2. Diszfóniás és egészséges beszéd automatikus osztályozása

### 4.2.1. Az SVM és a DNN osztályozók összehasonlítása akusztikai jellemzők esetén bemeneti vektorként

A diszfóniás és egészséges beszéd bináris osztályozása céljából a kutatók általában sokféle beszédből származó akusztikai jellemzőket használnak gépi tanulási algoritmusok bemeneti vektoraként [32, 33].

Osztályozási feladatokra elterjedt az SVM használata [34, 35], mivel képesek kis adatminták kezelésére, de Deep Learning technikákat is egyre gyakrabban használnak [36, 37, 38, 39, 40, 41]. Mélyneurális hálózatokat (Deep neural networks - DNNs) különféle feladatokra használnak, általában nagy adathalmazokon.

Ebben a kísérletben két osztályozási megközelítést alkalmaztam. Az első osztályozó SVM volt, a második osztályozó pedig a 3.1.4. fejezetben bemutatott mélyneurális hálózat. FFS algoritmust használtam a bemeneti vektor dimenzióinak csökkentése érdekében, abban az esetben, amikor SVM-et alkalmaztam osztályozóként. Bővebben az FFS algoritmusról a 3.1.3. fejezetben írtam. Az SVM optimális hiperparamétereit grid search-el állapítottam meg.

A kísérletben használt beszédadatbázisról a 3.2.1. fejezetben és 1. táblázatban mutattam be. Az adatbázis összesen 450 hangfelvételt tartalmaz, 257 diszfóniában szenvedő betegekből (156 nőtől és 101 férfitól), valamint 193 egészséges hangú embertől (108 nőtől és 85 férfitól).

A bemeneti vektort a 3.3.1. fejezetben leírt akusztikai jellemzőkből hoztam létre, így 49 akusztikai jellemzőt használtam bemeneti vektorként.

LOOVC validációs technikával végzett kétosztályos osztályozás eredményét a 10. táblázat foglalja össze.

**10. táblázat.** Bináris osztályozás eredményei HC és a Dys osztályok között, akusztikai jellemzőkkel bemeneti vektorként.

| **Bemeneti vektor** | **FFS** | **Jellemzők száma** | **Osztályozó és beállításai** | | **Hiper-paraméterek** | **LOOCV pontosság** |
|---|---|---|---|---|---|---|
| Akusztikai jellemzők | Igen | 11 | SVM | lineáris kernel | C = 1 | 85% |
| Akusztikai jellemzők | Igen | 9 | SVM | rbf kernel | C = 256; $\gamma$ = 0,0625 | 85% |
| Akusztikai jellemzők | Nem | 49 | SVM | lineáris kernel | C = 4 | 83% |
| Akusztikai jellemzők | Nem | 49 | SVM | rbf kernel | C = 1024; $\gamma$ = 0,00098 | 84% |
| Akusztikai jellemzők | Nem | 49 | DNN | dropout | 0,25 | **88%** |

A táblázat első oszlopa a bemeneti vektor típusát írja le. A következő oszlop, hogy volt-e FFS jellemzőkiválasztó alkalmazva vagy sem, továbbá, hogy végül hány jellemző került az osztályozó bemenetére. A következő oszlopok az osztályozót és beállításait írja le, majd az utolsó oszlop az osztályozás pontosságát. SVM esetében minden esetben grid search-el lettek meghatározva a hiperparaméterek. DNN esetében a dropout értéke 0,25-re lett állítva.

A DNN osztályozó használatával magasabb osztályozási pontosságot értem mint SVM osztályozóval. A pontosság relative 3,53%-kal növekszik, így a legmagasabb pontosság 88%-ra adódik. SVM esetében nincs számottevő különbség lineáris és az rbf kernel között (85% és 85% pontosság).

Az FFS és lineáris kernelű SVM osztályozás tévesztési mátrixát mutatja be a 11. táblázat, míg a DNN-es osztályozását a 12. táblázat. SVM esetében a HC osztály precizitása 84.83%, míg a Dys osztály precizitása 84.56%. A HC osztály fedése 78.24%, míg a Dys osztály fedése 89.49%.

**11. táblázat.** Tévesztési mátrix FFS és lineáris kernelű SVM használata esetén.

| | **Tényleges HC** | **Tényleges Dys** | Osztályprecizitás |
|---|---|---|---|
| **Pred. HC** | 151 | 27 | 84,83% |
| **Pred. Dys** | 42 | 230 | 84,56% |
| Osztályfedés | 78,24% | 89,49% | |

**12. táblázat.** Tévesztési mátrix DNN használata esetén.

| | **Tényleges HC** | **Tényleges Dys** | Osztályprecizitás |
|---|---|---|---|
| **Pred. HC** | 189 | 48 | 79,75% |
| **Pred. Dys** | 4 | 209 | 98,12% |
| Osztályfedés | 97,93% | 81,32% | |

A 88%-os osztályozási pontosságot elért modell tévesztési mátrixa a 12. táblázatban látható. Ahogy a táblázat mutatja, a HC osztály precizitása 79.75%, míg a Dys osztály precizitása 98.12%. A HC osztály fedése 78.24%, míg a Dys osztály fedése 89.49%. Ez azt jelenti, hogy azok az esetek, amikor a két osztályt helyesen prediktálta a modell nem kiegyensúlyozott. A HC osztály fedése 97.93%, míg a Dys osztály fedése 81.32%.

A pontosság nem az egyetlen abszolút mérték, amellyel jellemezzük osztályozónkat. A pontosság metrika sok fontos információt fed el, ezért óvatosan kell kezelni. Szimmetrikus tévesztési mátrixokat preferáljuk abban az esetben, ha az osztálytévesztési súlyok egyenlőek. Az orvosi alkalmazásokban általában a tévesztési mátrix nem egy szimmetrikus mátrix. Egy beteg ember egészségesnek minősítése súlyosabb hiba, mint ha egy egészséges embert betegnek minősítünk. Ez azt jelenti, hogy a Dys osztály fedése szintén nagyon fontos szempont egy osztályozás értékelésében. Minél alacsonyabb annak a kockázata, hogy egy diszfóniában szenvedő személyt egészségesnek prediktálunk, annál jobb.

Kedvezőbb eset ha a modellünk néhány egészséges embert diszfóniásra ítél, mint ha egészségesnek ítélné a diszfóniában szenvedőket. Ebben az értelemben jobb módszernek tűnik az FFS és lineáris kernelű SVM használata a DNN-el szemben, hiszen első esetben a Dys osztály fedése 89.49%, míg az utóbbi esetben 81.32%.

Ha a fals negatívoknak és fals pozitívoknak hasonló a költségük, két döntési táblázatot szerkeszthetünk a helyes ítéletek száma és téves ítéletek száma függvényében, valamint khí-négyzet próbát végezhetünk. A próba során a $p$ érték 0,09 értéket vett fel, tehát nem volt statisztikai különbség a 85%-os pontosságot elérő lineáris kernelű SVM és 88%-os pontosságot elérő DNN között.

**II. A. Tézis** [C2, C9] *Megmutattam, hogy a diszfóniás és egészséges hangok bináris osztályozása a magyar nyelvre lehetséges. 88%-os osztályozási pontosság érhető el akusztikai jellemzőkből álló bemeneti vektorral, LOOCV validációs technikával és fully-connected rétegű mélyneurális hálózat alkalmazásával a Magyar Diszfóniás és Egészséges Felnőtt Beszédadatbázist használva.*

### 4.2.2. ASR valószínűségi jellemzőhalmaz használata bemeneti vektorokként a DNN osztályozóhoz

Az automatikus beszédfelismerést (Automatic speech recognition - ASR) hagyományosan egy akusztikai és egy nyelvi modell létrehozására bontják [42]. Az akusztikai modell leggyakrabban egy rejtett Markov-modell (Hidden Markov Model - HMM) hibridje, amely feladata, hogy megkönnyítse a lehető legjobb illeszkedést. Fonémamodelleket hoz létre, ami felelős az osztályozandó tényleges keret(ek) és fonémák közötti hasonlósági mértékek biztosításáért. Noha ez a modellkészlet ritkán felel meg a tiszta fonémamodelleknek, az egyszerűség kedvéért ezt „fonémamodellnek" nevezem, főleg mivel a fonémamodell kimenetéből kinyerhetőek a fonéma posterior valószínűségek. Nyilvánvaló, hogy ezeket a fonéma posterior valószínűségeket fel lehet használni a keretek osztályozására, vagy a kiejtés jóságának (Goodness of Pronunciation (GOP) score) meghatározására [43], amelyet felhasználhatunk a beszédértékelésben a kiejtés automatikus értékeléséhez.

Egyértelműen felhasználhatók a GOP értékek vagy a fonéma posterior valószínűségek diszfónia osztályozására. A kutatók jelentős csoportja osztja azt az álláspontot, hogy az ASR akusztikai modell fonéma posterior értékei alkalmasak a diszfónia felismerésére. Például a [44] és [45] cikkek szerzői ASR fonéma posterior értékeket használnak beszédzavarok súlyosságának becslésére. Azonban az ASR fonémamodelljének tanítása más célt szolgál, mint amit a diszfónia felismerése megkövetel. Az ASR-nek tolerálnia kell a nagy beszélők közötti és beszélőn belüli variabilitást és azt, hogy a tanítóminták között diszfóniás hangok is szerepelnek-e, nem ellenőrzött. Röviden: egy fonémamodell nem arra van tanítva, hogy megkülönböztesse a diszfóniás és egészséges beszédet. Az ötlet, ami miatt továbbra is használják a kutatók azon a feltételezésen nyugszik, hogy a diszfóniás beszéd

nem szabványos, ezért a fonéma posterior eloszlása nem lesz csúcsos, hanem laposabb lesz (azaz a fonémaosztályozó bizonytalan lesz a döntésében).

Azt állítom, hogy ezt a módszert óvatosan kell kezelni diszfónia esetében. Kérdéses, hogy a általános beszédfelismerésre tanított akusztikai modell használható-e egyáltalán diszfóniás beszéd felismerésére. Az automatikus beszédfelismerő rendszerek tanításához nagy mennyiségű tanítóadatra van szükség különböző nemű, korú, akcentusú, iskolai végzettségű stb. személyektől. A tanító beszédadatbázisok tartalmazhatnak különböző nyelvjárású beszélők hanját, dohányosok hangját, fiatalokét és idősekét egyaránt. Diszfóniás hangok akár véletlenül is nagyobb számban szerepelhet az adatbázisban. A beszédfelismerés általános célja a rekedt, náthás, szomorú, vidám, öreg és fiatal beszéd egyenértékű felismerése. Le akartam ellenőrizni, hogy a fonéma-valószínűségek külön használata, vagy együttes használata akusztikai jellemzőkkel nagyobb osztályozási pontossághoz vezet-e, mint ha csak akusztikai jellemzőket használnánk bemeneti vektorként. Legjobb tudomásom szerint ezt a kérdést sosem vizsgálták a diszfóniás beszéd szempontjából.

A posterior valószínűségekből alkotott bemeneti vektort a 3.3.2. fejezetben leírtak alapján készítettem el, így 21-dimenziós vektort használtam a DNN bemeneteként. A kapott eredményt összehasonlítottam a 4.2.1. fejezetben kapottakkal. Az eredmények azt mutatják, hogy abban az esetben magasabb osztályozási pontosságot érek el, ha akusztikai jellemzőket használok az osztályozó bemeneteként, mint amikor az ASR valószínűségi jellemzőhalmazt. 88%-os pontosságot értem el első esetben és 60%-os pontosságot, ha az ASR valószínűségi jellemzőhalmazt használtam a DNN bemeneti vektoraként.

**13. táblázat.** Bináris osztályozás eredményei HC és a Dys osztályok között, DNN osztályozóval és a bemeneti vektorok összehasonlításával.

| **Bemeneti vektor** | **FFS** | **Jellemzők száma** | **Osztályozó és beállításai** | | **Hiper-paraméterek** | **LOOCV pontosság** |
|---|---|---|---|---|---|---|
| Akusztikai jellemzők | Nem | 49 | DNN | dropout | 0,25 | 88% |
| ASR valószínűségi jellemzőhalmaz | Nem | 21 | DNN | dropout | 0,25 | 60% |
| Közös jellemzőhalmaz | Nem | 70 | DNN | dropout | 0,25 | **89%** |

Mivel az ASR valószínűségi jellemzőhalmazzal való osztályozás eredménye elmaradt az akusztikai jellemzőkkel kapott eredményétől, megvizsgáltam, hogy a két bemeneti vektor kombinációja (továbbá „közös jellemzőhalmaz") növeli-e az osztályozás pontosságát. Amikor a közös jellemzőhalmaz került a neurális hálózat bemenetére, az osztályozás pontossága 89%-ra nőtt. Az osztályozás eredményeit a 13. táblázat tartalmazza. Noha ez jobb, mint pusztán az akusztikai jellemzők használata, az ASR valószínűségi jellemzőhalmaz használata nem befolyásolja jelentősen az osztályozást.

**14. táblázat.** Tévesztési mátrix DNN és a közös jellemzőhalmaz használata esetén.

| | **Tényleges HC** | **Tényleges Dys** | Osztályprecizitás |
|---|---|---|---|
| **Pred. HC** | 167 | 25 | 86,98% |
| **Pred. Dys** | 26 | 232 | 89,92% |
| Osztályfedés | 86,53% | 90,27% | |

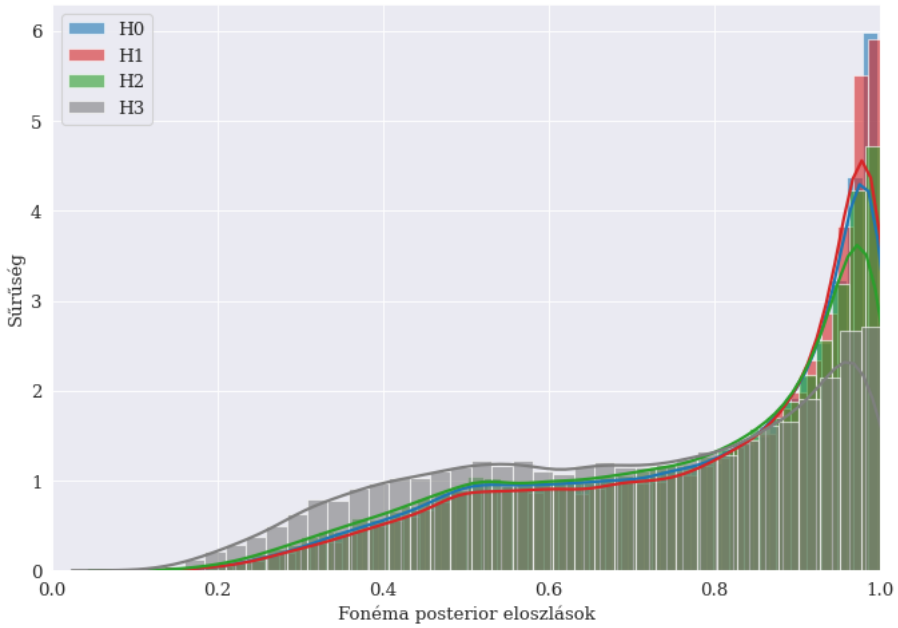
A bemenetén közös jellemzőhalmazt használó osztályozó tévesztési mátrixa a 14. táblázatban látható. A HC osztály fedése 86,98%, míg a Dys osztály fedése 89,92%. A HC osztály fedése 86,53%, míg a Dys osztály fedése 90,27%.

Annak ellenére, hogy a Dys osztály fedése a közös jellemzőhalmaz használatakor magasabb, mint amikor csak akusztikai jellemzőket használtam (90,27% az első esetben és 81,32% a másodikban), a közös jellemzőhalmaz eredményeit összehasonlítva azzal az esettel, amikor csak akusztikai jellemzőket használtam lineáris kernelű SVM esetén (a 11. táblázatban bemutatva) a fedésben való növekedés nem számottevő.

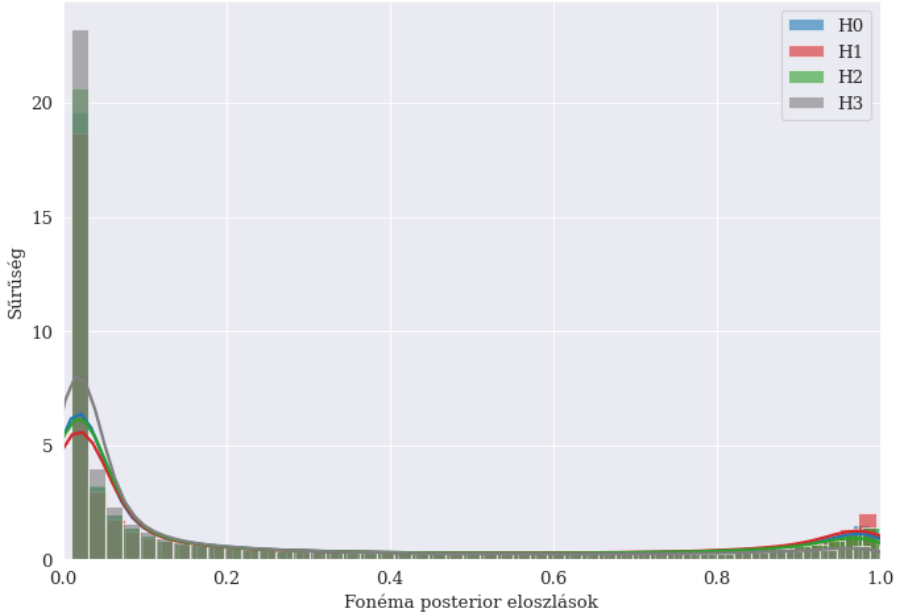
Ha a fals negatívoknak és fals pozitívoknak hasonló a költségük és khí-négyzet próbát hajtunk végre, akkor nincs szignifikáns különbség a 85%, 88% és 89%-es osztályozó között. A próba során a $p$ paraméter 0,07 értéket vette fel, a csak akusztikai jellemzőket felhasználó lineáris kernelű SVM és közös jellemzőhalmazt használó DNN között, valamint 0,91-es értéket vette fel a a csak akusztikai jellemzőket felhasználó DNN és a közös jellemzőhalmazt használó DNN között.

Annak ellenőrzésére, hogy az ASR fonéma-valószínűségek miért nem javították a osztályozás pontosságát, kiszámoltam egy adott fonéma posterior eloszlását a négy súlyossági fok (H) függvényében. Első megközelítésemben kiszámoltam a beszédfelismerő softmax keretinek legnagyobb valószínűségi értékeit (ahol a fonéma 'megnyeri' a keretét) és azokat ábrázoltam a négy súlyossági fok függvényében, második megközelítésemben kiszámoltam az összes posterior értéket, ahol a fonéma valamelyik keretben megjelent. Az eredményeket az 5. és 6. ábra mutatja az [E] fonéma esetében. Látható, hogy a különböző súlyossági kategóriák nem különülnek el jól egymástól. Más fonémák hasonló tendenciát mutatnak.

Ezek alapján arra a következtetésre lehet jutni, hogy nem érdemes kiszámítani az ASR fonéma-valószínűségeket, mivel azoknak nincs jelentős hatása az osztályozási pontosságra nézve, viszont jelentősen bonyolíthatja és lelassíthatja a jelenlegi javasolt rendszert. A kutatást elsősorban az mozdíthatja előre, ha több adathoz jutunk és további akusztikai jellemzőket keressünk. A mellett érvelek, hogy az ASR fonémamodellek tanítása során alkalmazott eltérő objektív kritériumok és ellenőrizetlen tanítóadatok a diszfónia vonatkozásában igazolják, hogy a fonéma-valószínűségi értékek miért nem tudtak segíteni az eredmények javításában.




**5. ábra.** Fonéma posterior eloszlások a legnagyobb valószínűségű [E] fonémák esetén.




**6. ábra.** Fonéma posterior eloszlások minden [E] fonéma esetén

Ezen eredmények alapján a következő tézisek fogalmazhatók meg.

**II. B. Tézis** [C2, C9] *Megmutattam, hogy az általános célra tanított ASR fonémaszintű posterior valószínűségi értékek használata kevésbé hatékony az egészséges és diszfóniás hangok automatikus osztályozásában, mint az akusztikai jellemzők közvetlen használata. Mélyebb elemzések gyenge kapcsolatot mutattak ki a fonémaszintű posterior valószínűségi értékek és a diszfónia súlyossága között.*

**II. C. Tézis** [C2, C9] *Megmutattam, hogy azASR fonémaszintű valószínűségi értékek hozzáadása az akusztikai jellemzők halmazához nem javítja jelentősen az egészséges és diszfóniás hangok automatikus osztályozási pontosságát.*

### 4.3. Funkcionális és organikus diszfónia automatikus osztályozása

Ahogyan a bevezetőben is említettem (1. fejezet) a diszfóniát szokás funkcionális és organikus diszfónia csoportokra osztani.

Barth [46] és Stern [47] szerint funkcionális fonációs rendellenességről beszélünk, ha a rendelkezésére álló diagnosztikai eszközök nem észlelnek szerves elváltozásokat. D. Weiss megfogalmazása szerint a „funkcionális" jelző csak átmeneti állapotot jelent és csak addig érvényes, amíg a tudomány eszközei nem tudják feltárni a betegség valódi szervi okait [48]. Gundermann vitatja ezt a nézetet, és kijelenti, hogy nem helyénvaló a „funkcionális" kifejezést a „organikus hiánya" helyett alkalmazni [49]. A szervi rendellenesség lehet a funkcionális rendellenesség kiindulópontja. Ez fordítva is igaz, hiszen funkcionális zavar eredményezhet szervi/organikus elváltozást. A fentebb bemutatott irodalom alapján a két kategória nem tűnik mindig egymást kizárónak.

Érdekes kutatási kérdés, hogy szét lehet-e automatikusan választani egymástól a funkcionális és organikus diszfóniát. Ha a funkcionális diszfóniát meg lehetne határozni nagy valószínűséggel egy diagnosztikát segítő rendszer segítségével, a beteget gyorsan foniáterhez vagy logopédushoz lehetne irányítani. Ha viszont a rendszer organikus diszfóniát észlel, a beteget otolaringológushoz vagy onkológushoz lehetne irányítani. Egy ilyen rendszer sok időt takaríthat meg, a pácienseket időben el lehetne irányítani a megfelelő szakrendelésre.

Az FD és az OD meghatározása és szétválasztása körül vita van, a két kategória nem mindig kölcsönösen kizáró. Természetes, hogy a két csoport jobban osztályozható egy olyan beszédadatbázison, ahol a rekedtség súlyosságának eloszlása statisztikailag eltérő a két csoport között, például ha az OD csoport statisztikailag szignifikánsan magasabb fokú rekedtségi súlyosságot mutat, mint az FD csoportban lévő hangfelvételek. Ilyen módon az osztályozó a hangfelvételeket a betegség típusa helyett a rekedtség súlyosságát osztályozhatja alacsony és magas rekedtségi osztályba. Ami

valójában a célunk, hogy a diszfóniában szenvedő emberek hangfelvételeit funkcionális és organikus diszfónia csoportokba soroljuk. Ennek a kísérletnek az elvégzéséhez létrehoztam a Kiválasztott Diszfóniás Beszédadatbázist, amiben a rekedtség súlyosságának megoszlása az OD és az FD csoportok között nem volt szignifikáns különbség.

Ebben a tézisben megkíséreltem az funkcionális és organikus diszfónia automatikus osztályozását SVM segítségével, továbbá megpróbálom azonosítani azokat az akusztikai jellemzőket, amelyek a legmegfelelőbbek az osztályozás szempontjából.

A kísérletben használt beszédadatbázis a 3.2.1. fejezetben ismertetett adatbázis egy szűrt változata. A Kiválasztott Diszfóniás Beszédadatbázist úgy hoztam létre, hogy a nemek és a rekedtség súlyosságának megoszlása az OD és az FD csoportokban egyenlő legyen. Az adatbázis összesen 164 felvételt tartalmaz, 82 felvétel organikus, 82 felvétel funkcionális diszfóniában szenvedő betegektől. Az adatbázis 122 női hangfelvételt (61 OD és 61 FD) és 42 férfi hangfelvételt (21 OD és 21 FD) tartalmaz. Az OD-csoportban az átlag rekedtségi fok (H paraméter az RBH szubjektív skálán) 1,5 volt 0,7 szórással, míg az FD csoport esetében az átlag 1,4 volt 0,7 szórással. A Kiválasztott Diszfóniás Beszédadatbázis leírását a 15. táblázat tartalmazza.

**15. táblázat.** A Kiválasztott Diszfóniás Beszédadatbázis

| | **Női hangfelvételek** | **Férfi hangfelvételek** | **H súlyosság** | **Női H súlyosság** | **Férfi H súlyosság** |
|---|---|---|---|---|---|
| **OD** | 61 | 21 | 1,5 ($\pm 0,7$) | 1,5 ($\pm 0,6$) | 1,5 ($\pm 0,8$) |
| **FD** | 61 | 21 | 1,4 ($\pm 0,7$) | 1,3 ($\pm 0,7$) | 1,5 ($\pm 0,8$) |

A kísérlet során ugyanazt a 49 akusztikai jellemzőt használtam bemeneti vektorként, mint amit a3.3.1. fejezetben írtam le.

A Kiválasztott Diszfóniás Beszédadatbázist úgy állítottam elő, hogy ne legyen jelentős különbség a rekedtség súlyosság megoszlásában az OD és az FD csoportokban. Mivel a súlyossági pontszámok ordinálisak, a Mann-Whitney U próbát használtam, hogy eldöntsem, a két minta azonos eloszlásból származik-e. A 3.1.2. fejezet részletesebben leírja a Mann-Whitney U próbát. 95%-os ($\alpha = 0,05$) szignifikanciaszintet használtam a próba elvégzése során.

A Mann-Whitney U próba elvégzésekor, a kiszámított p-érték a teljes adathalmazon a két minta között 0,65 lett. Nők esetében a p érték 0,34, férfiak esetében 0,65 (p-érték $> \alpha$). Az OD-csoport rekedtségének eloszlása azonosnak tekinthető az FD-csoport rekedtségének megoszlásával. Más szavakkal, az OD és az FD populációk rekedtségének eloszlása nem különbözik eléggé, ahhoz, hogy statisztikailag szignifikáns különbséget mutassanak. Így a Kiválasztott Diszfóniás Beszédadatbázis alkalmas további osztályozási vizsgálatokra, mivel kizárja annak lehetőségét, hogy az osztályozó a rekedtség súlyosságát osztályozza két csoportba.

A kísérletben FFS jellemzőkiválasztó algoritmust használtam SVM osztályozóval és LOOCV keresztvalidációs technikával. Az SVM-ről és az FFS algoritmusról a 3.1.4. és 3.1.3. fejezetekben írtam bővebben. Bemeneti vektorként azt a 49 akusztikai jellemzőt tartalmazó halmazt használtam, amit a 3.3.1. fejezetben mutattam be.

A Kiválasztott Diszfóniás Beszédadatbázison végzett osztályozási eredményeket a 16. táblázat mutatja. A legmagasabb osztályozási pontosság 71%-ra adódott lineáris kernelű SVM használatakor. Az FFS algoritmus 5 akusztikai jellemzőt választott ki: a názálisokon mért SPI szórását, a magas magánhangzókon mért IMF entrópia energiahányados szórását és tartományát, valamint a spiránsokon mért IMF entrópia energiahányados átlagát és szórását. Abban az esetben, amikor FFS jellemzőkiválasztó algoritmust használtam magasabb pontosságot értem el, mint amikor mind a 49 akusztikai jellemző került a bemenetre. Ez igaz lineáris és rbf kernelű SVM esetén is.

**16. táblázat.** Bináris osztályozás eredményei OD és FD osztályok között a Kiválasztott Diszfóniás Beszédadatbázison.

| **FFS** | **Jellemzők száma** | **Osztályozó és beállításai** | | **Hiper-paraméterek** | **LOOCV pontosság** |
|---|---|---|---|---|---|
| Nem | 49 | SVM | lineáris kernel | C = 0,125 | 66% |
| Igen | 5 | SVM | lineáris kernel | C = 1 | **71%** |
| Nem | 49 | SVM | rbf kernel | C = 128; $\gamma$ = 0,0005 | 67% |
| Igen | 10 | SVM | rbf kernel | C = 2; $\gamma$ = 0,00781 | 66% |

**17. táblázat.** Tévesztési mátrix FFS és lineáris kernelű SVM használata esetén.

| | **Tényleges FD** | **Tényleges OD** | Osztályprecizitás |
|---|---|---|---|
| **Pred. FD** | 61 | 27 | 69% |
| **Pred. OD** | 21 | 55 | 72% |
| Osztályfedés | 74% | 67% | |

A 71%-os osztályozási pontosságot elérő beállítás tévesztési mátrixa a 17. táblázatban látható. Az OD osztály precizitása 72%, míg az FD osztály precizitása 69%. Az FD osztály fedése 74%, míg az OD osztály fedése 67%. Ez az osztályozási pontossági érték megbízhatóbb, mint ha a 3.2.1. fejezetben bemutatott beszédadatbázis diszfóniás felvételein végeztem volna el az osztályozást, mivel ezt az eredményt nem befolyásolja a két csoport közötti rekedtség súlyosságának különbsége.

Az eredmények egyértelműen mutatják, hogy a két diszfónia típus automatikusan szétválasztható egymástól.

**III. A. Tézis** [C3] *Megmutattam, hogy az organikus és a funkcionális diszfónia automatikus elválasztása akusztikus jellemzők felhasználásával lehetséges 71% -os osztályozási pontossággal, lineáris kernelű SVM használatával, a Kiválasztott Diszfóniás Beszédadatbázison.*

## 4.4. Diszfóniás gyerekhangok automatikus osztályozása

Ennek a fejezetnek az a célja, hogy az egészséges hangú gyerekhangokat megkíséreljе automatikusan megkülönböztetni a diszfóniában szenvedő gyerekek hangjától. A kísérletben használt akusztikai jellemzőket a 3.3.1. fejezetben mutattam be.

A bináris osztályozás elvégzéséhez a lineáris és az rbf kernelfüggvényű SVM osztályozót használtam. Először mind a 103 akusztikai jellemzőt felhasználtam az osztályozó bemenetén, majd FFS algoritmussal csökkentettem a bemeneti vektor dimenzióját. Rbf kernelű SVM esetén C hiperparaméterének alapértelmezett értéke a jellemzők számával egyenlő, míg a $\gamma$ alapértelmezett értéke 1/az akusztikai jellemzők száma. Jellemzőkiválasztás során nem tudhatjuk előre, hogy hány jellemzőt választ majd ki az algoritmus, így a hiperparamétereket intuíció alapján lehet beállítani. Intuíció alapján C értéket 10-nek, míg $\gamma$ értékét 0,1-nek választottam. LOOCV validációs technikát használtam minden esetben. Az osztályozási eredményeket a 18. táblázat foglalja össze.

**18. táblázat.** Bináris osztályozás eredményei egészséges és diszfóniás gyerekhangok között a Diszfóniás és Egészséges Gyermek Beszédadatbázison.

| **FFS** | **Jellemzők száma** | **Osztályozó és beállításai** | | **Hiper-paraméterek** | **LOOCV pontosság** |
|---|---|---|---|---|---|
| Nem | 103 | SVM | lineáris kernel | C = 1 | 88% |
| Nem | 103 | SVM | rbf kernel | C = 124; $\gamma$ = 0,008 | 86% |
| Igen | 8 | SVM | lineáris kernel | C= 1 | **93%** |
| Igen | 8 | SVM | rbf kernel | C = 10; $\gamma$ = 0,1 | **93%** |

Mint ahogy a táblázatból kiderül, a legmagasabb osztályozási pontosság 93% lett lineáris és rbf kernel használata esetén. A jellemzőkiválasztó algoritmus 8 jellemzőre csökkentette a bemeneti vektor dimenzióját, miközben nagyobb pontosságot ért el, mint amikor az összes jellemzőt használtam.

Az FFS-t és lineáris kernelt használó osztályozás SVM eredmény tévesztési mátrixát a 19. táblázat mutatja be. A HC osztály precizitása 94%, míg a Dys osztály precizitása 92%. A HC osztály fedése 94%, míg a Dys osztály fedése 92%.

Egy sikeres osztályozás tévesztési mátrixa szimmetrikus abban az esetben, ha az osztálytévesztési súlyok egyenlőek. Ellenkező esetben, egy aszimmetrikus tévesztési mátrixnál azt sejthetjük, hogy az osztályozó elfogult egyik osztállyal szemben. A 93%-os osztályozó tévesztési mátrixa annyira

szimmetrikus, amennyire csak lehet. Viszont, ahogy korábban említettem, orvosi rendszerekben a tévesztési mátrix általában nem szimmetrikus, hiszen az osztálytévesztések nem egyenlőek. Ha egy egészséges gyermeket diszfóniásnak prediktálunk az kevésbé rossz, mint ha egy diszfóniás gyereket egészségesnek ítélnénk.

A kutatásban fontos, hogy a ténylegesen egészségesnek (HC-nak) ítélt, illetve ténylegesen diszfóniás (Dys) felvételek száma minimális legyen. Ez csak kétszer fordul elő, 92%-os Dys osztályfedést eredményezve.

**19. táblázat.** Tévesztési mátrix FFS és lineáris kernelű SVM használata esetén.

| | **Tényleges HC** | **Tényleges Dys** | Osztályprecizitás |
|---|---|---|---|
| **Pred. HC** | 32 | 2 | 94% |
| **Pred. Dys** | 2 | 23 | 92% |
| Osztályfedés | 94% | 92% | |

Megállapíthatjuk, hogy az alkalmazott bemeneti vektorok nagy biztonsággal különböztetik meg az egészséges és diszfóniás gyerekek hangjait. Ebből az eredményből úgy tűnik, hogy a diszfónia korai szakaszban hatékonyabban szűrhető, de ilyen állítások megfogalmazásához sokkal több adatot kell gyűjteni.

**IV. A. Tézis** [J2] *Megmutattam, hogy az egészséges és a diszfóniás gyermekek hangjainak automatikus osztályozása lehetséges 93%-os osztályozási pontossággal lineáris és az rbf kernelfüggvényű SVM-el a Diszfóniás és Egészséges Gyermek Beszédadatbázison.*

# 5. Eredményeim alkalmazhatósága

Az eredmények azt mutatják, hogy gyakorlatilag már megvalósítható egy olyan diagnosztikát segítő rendszer kifejlesztése, amely képes a diszfóniás és az egészséges beszéd megkülönböztetésére. Fontos azonban megjegyezni, hogy míg a rendszer jól használható előszűrésre, a pontos diagnózis felállítása továbbra is az orvos felelőssége.

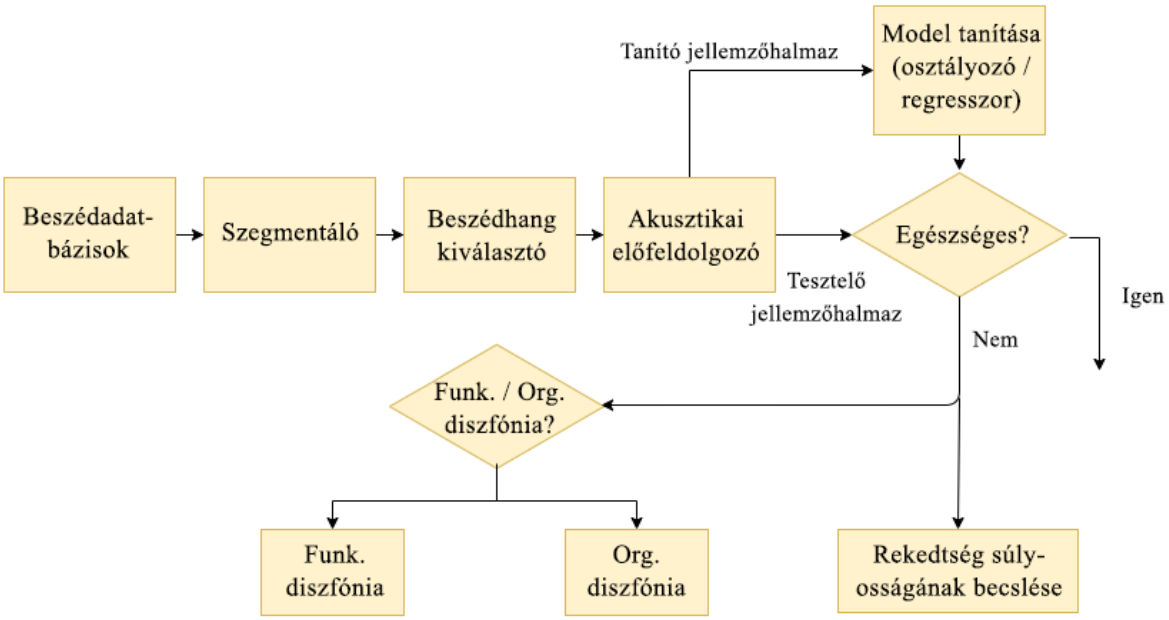
A felnőttek számára javasolt rendszer több lépésből áll: a betegek beszédfelvételeit beszédadatbázisokba rendezzük (Magyar Diszfóniás és Egészséges Felnőtt Beszédadatbázis). A felvételeket normalizáljuk és szegmentáljuk fonéma szinten. Az analizálandó fonémák kiválasztása után akusztikai jellemzőket nyerünk ki a hangfelvételekből és vektorba rendezzük őket. A jellemzővektort a rendszer az előzetes ismeretek alapján binárisan osztályokba sorolja (egészséges vagy diszfóniás). Ha a hangfelvétel az analízis során egészségesnek minősül, a folyamatnak itt vége. Ha diszfóniásnak minősül, a diagnosztikát segítő rendszer felismerné a diszfónia típusát (funkcionális vagy organikus diszfónia), és a betegség súlyosságát is megbecsülné regressziós modell alapján.

Az előzetes ismereteket egy gondosan felépített beszédadatbázissal, és optimális osztályozási-, és regressziós modellel szerezzük, amik a 3.2.1., 4.1., 4.2., 4.3. és 4.4. fejezetekben lettek részletezve.

Az új beszédminták esetében az osztály (egészséges vagy diszfóniás) és a diszfónia súlyossága ismeretlen. A hangfelvétel előfeldolgozási módszere ugyanaz mint korábban: akusztikai jellemzőket mérünk a fonémák szintjén, majd egy tesztelő jellemzőhalmazt alakítunk ki. Ez a tesztelő jellemzőhalmaz egy összehasonlító egységbe kerül és elvégezzük rajta az osztályozási vagy regressziós műveleteket. Ez a folyamat az 7. ábrán látható.

Ha a funkcionális diszfóniát meg lehetne határozni nagy valószínűséggel a diagnosztikát segítő rendszer segítségével, a beteget gyorsan foniáterhez vagy logopédushoz lehetne irányítani. Ha viszont a rendszer organikus diszfóniát észlel, a beteget otolaringológushoz vagy onkológushoz lehetne irányítani. Egy ilyen rendszer sok időt takaríthat meg, a pácienseket időben el lehetne irányítani a megfelelő szakrendelésre. A vizsgálatban javasolt végrendszer fiatal orvosoknak vagy háziorvosoknak is segítene hatékonyan kiszűrni a diszfóniás betegeket és automatikusan megállapítani a betegségük súlyosságát.

A diszfónia korai felismerését szolgáló diagnosztikát segítő rendszer a gyermekek esetében is a fent leírt logikát követné, azzal a különbséggel, hogy az organikus és funkcionális okok még nem megkülönböztethetőek. Mivel az osztályozási eredmények a gyermekek hangjai esetében ígéretesek, további beszédminták gyűjtése javasolt, hogy általánosítani lehessen az osztályozó modellt nagyobb adatkészleten. Hosszú távon érdemes kifejleszteni egy eszközt, ami automatikusan észleli a diszfóniás hangokat gyerekek körében. A mobiltelefonok alkalmasak lennének ennek a módszernek a bevezetésére és a gyakorlati alkalmazására. Az egészségügyi alkalmazásokat általában




**7. ábra.** A felnőttek számára javasolt diagnosztikát segítő rendszer kerete.

okostelefonokhoz vagy táblagépekhez, esetenként okosórákhoz tervezik. Ezek lehetővé teszik a felhasználók számára, hogy akkor és ott férjenek hozzá információhoz, amikor szükségük van rá, csökkentve az információ keresésére pazarolt időt. Ezek az eszközök olcsók, könnyen használhatóak és hordozhatóak.

A hangmintákat, a metaadatokat, a mért akusztikai jellemzők értékeit és az osztályozó kimenetét össze lehet gyűjteni és fel lehet tölteni egy felhőszerverre. Ily módon hosszú távon monitorozhatjuk a gyermekek hangminőségét. A cél egy olyan szűrőrendszer felépítése, amelyet óvodai dolgozók használhatnak. Ha egy diszfóniás hangú gyermeket időben kiszűrünk, akkor nagyobb esélyünk van arra, hogy a gyermek szakszerű segítséget kapjon egy fül-orr-gégésztől vagy egy logopédustól.

## Hivatkozások


[1] R. J. Stachler, D. O. Francis, S. R. Schwartz, C. C. Damask, G. P. Digoy, H. J. Krouse, S. J. McCoy, D. R. Ouellette, R. R. Patel, C. C. W. Reavis *et al.*, „Clinical practice guideline: hoarseness (dysphonia)(update)," *Otolaryngology–Head and Neck Surgery*, vol. 158, no. 1_suppl, pp. S1–S42, 2018.

[2] S. M. Cohen, J. Kim, N. Roy, C. Asche, and M. Courey, „Prevalence and causes of dysphonia in a large treatment-seeking population," *The Laryngoscope*, vol. 122, no. 2, pp. 343–348, 2012.

[3] S. M. Cohen, „Self-reported impact of dysphonia in a primary care population: An epidemiological study," *The Laryngoscope*, vol. 120, no. 10, pp. 2022–2032, 2010.

[4] R. Reiter, T. K. Hoffmann, A. Pickhard, and S. Brosch, „Hoarseness—causes and treatments," *Deutsches Ärzteblatt International*, vol. 112, no. 19, p. 329, 2015.

[5] K. Jones, J. Sigmon, L. Hock, E. Nelson, M. Sullivan, and F. Ogren, „Prevalence and risk factors for voice problems among telemarketers," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 128, no. 5, pp. 571–577, 2002.

[6] J. Long, H. N. Williford, M. S. Olson, and V. Wolfe, „Voice problems and risk factors among aerobics instructors," *Journal of Voice*, vol. 12, no. 2, pp. 197–207, 1998.

[7] E. Smith, H. L. Kirchner, M. Taylor, H. Hoffman, and J. H. Lemke, „Voice problems among teachers: differences by gender and teaching characteristics," *Journal of Voice*, vol. 12, no. 3, pp. 328–334, 1998.

[8] T. Davids, A. M. Klein, and M. M. Johns III, „Current dysphonia trends in patients over the age of 65: is vocal atrophy becoming more prevalent?" *The Laryngoscope*, vol. 122, no. 2, pp. 332–335, 2012.

[9] N. Bhattacharyya, „The prevalence of pediatric voice and swallowing problems in the united states," *The Laryngoscope*, vol. 125, no. 3, pp. 746–750, 2015.

[10] M. C. Duff, A. Proctor, and E. Yairi, „Prevalence of voice disorders in african american and european american preschoolers," *Journal of Voice*, vol. 18, no. 3, pp. 348–353, 2004.

[11] P. N. Carding, S. Roulstone, K. Northstone, A. S. Team *et al.*, „The prevalence of childhood dysphonia: a cross-sectional study," *Journal of Voice*, vol. 20, no. 4, pp. 623–630, 2006.

[12] E.-M. Silverman and C. H. Zimmer, „Incidence of chronic hoarseness among school-age children," *Journal of Speech and Hearing Disorders*, vol. 40, no. 2, pp. 211–215, 1975.

[13] „American speech-language-hearing association - voice disorders," https://www.asha.org/practice-portal/clinical-topics/voice-disorders/, accessed: 2020-04-03.

[14] R. Jani, S. Jaana, L. Laura, and V. Jos, „Systematic review of the treatment of functional dysphonia and prevention of voice disorders," *Otolaryngology—Head and Neck Surgery*, vol. 138, no. 5, pp. 557–565, 2008.

[15] J. Wendler, A. Rauhut, and H. Kruger, „Classification of voice qualities," *Journal of Phonetics*, vol. 14, no. 3-4, pp. 483–488, 1986.

[16] M. Ptok, C. Schwemmle, C. Iven, M. Jessen, and T. Nawka, „On the auditory evaluation of voice quality," *HNO*, vol. 54, no. 10, pp. 793–802, 2006.

[17] J. D. Evans, *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co, 1996.

[18] C. Cortes and V. Vapnik, „Support vector machine," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] G. Horváth, „Neurális hálózatok és műszaki alkalmazásaik," *Műszaki Kiadó*, 1998.

[20] J. MacQueen *et al.*, „Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[21] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, „Support vector regression machines," in *Advances in neural information processing systems*, 1997, pp. 155–161.

[22] G. Kiss and K. Vicsi, „Mono-and multi-lingual depression prediction based on speech processing," *International Journal of Speech Technology*, vol. 20, no. 4, pp. 919–935, 2017.

[23] V. Klára, „Sampa computer readable phonetic alphabet," 2008.

[24] P. Boersma, „Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the institute of phonetic sciences*, vol. 17, no. 1193. Amsterdam, 1993, pp. 97–110.

[25] P. Roach, S. Arnfield, W. Barry, J. Baltova, M. Boldea, A. Fourcin, W. Gonet, R. Gubrynowicz, E. Hallum, L. Lamel *et al.*, „Babel: An eastern european multi-language database," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, vol. 3. IEEE, 1996, pp. 1892–1893.

[26] K. Vicsi, A. Kocsor, C. Teleki, and L. Tóth, „Hungarian speech database for computer-using environments in offices," in *Proc. 2nd Hungarian Conf. on Computational Linguistics*, 2004, pp. 315–318.

[27] C. Teleki, V. Szabolcs, T. S. Levente, and V. Klára, „Development and evaluation of a hungarian broadcast news database," in *Forum Acusticum*, 2005.

[28] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, „The kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and understanding*, no. CONF. IEEE Signal Processing Society, 2011.

[29] Y.-R. Chien, M. Borskỳ, and J. Guðnason, „Objective severity assessment from disordered voice using estimated glottal airflow." in *Interspeech*, 2017, pp. 304–308.

[30] I. Laaridh, W. Kheder, C. Fredouille, and C. Meunier, „Automatic prediction of speech evaluation metrics for dysarthric speech," in *Interspeech*, 2017.

[31] T. Law, J. H. Kim, K. Y. Lee, E. C. Tang, J. H. Lam, A. C. van Hasselt, and M. C. Tong, „Comparison of rater's reliability on perceptual evaluation of different types of voice sample," *Journal of Voice*, vol. 26, no. 5, pp. 666–e13, 2012.

[32] N. Adiga, C. Vikram, K. Pullela, and S. M. Prasanna, „Zero frequency filter based analysis of voice disorders." in *Interspeech*, 2017, pp. 1824–1828.

[33] M. Markaki, Y. Stylianou, J. D. Arias-Londoño, and J. I. Godino-Llorente, „Dysphonia detection based on modulation spectral features and cepstral coefficients," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 5162–5165.

[34] F. Kazinczi, K. Mészáros, and K. Vicsi, „Automatic detection of voice disorders," in *International Conference on Statistical Language and Speech Processing*. Springer, 2015, pp. 143–152.

[35] V. Klára, I. Viktor, and M. Krisztina, „Voice disorder detection on the basis of continuous speech," in *5th European Conference of the International Federation for Medical and Biological Engineering*. Springer, 2011, pp. 86–89.

[36] P. Harar, J. B. Alonso-Hernandezy, J. Mekyska, Z. Galaz, R. Burget, and Z. Smekal, „Voice pathology detection using deep learning: a preliminary study," in *2017 international conference and workshop on bioinspired intelligence (IWOBI)*. IEEE, 2017, pp. 1–4.

[37] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, „Detection of pathological voice using cepstrum vectors: A deep learning approach," *Journal of Voice*, 2018.

[38] H. Wu, J. J. Soraghan, A. Lowit, and G. Di Caterina, „A deep learning method for pathological voice detection using convolutional deep belief networks." in *Interspeech*, vol. 2018, 2018.

[39] J. I. Godino-Llorente and P. Gomez-Vilda, „Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 380–384, 2004.

[40] L. Salhi, M. Talbi, and A. Cherif, „Voice disorders identification using hybrid approach: Wavelet analysis and multilayer neural networks," *World Academy of Science, Engineering and Technology*, vol. 45, no. 21, pp. 330–339, 2008.

[41] V. Srinivasan, V. Ramalingam, and P. Arulmozhi, „Artificial neural network based pathological voice classification using mfcc features," *International Journal of Science, Environment and Technology*, vol. 3, no. 1, pp. 291–302, 2014.

[42] C.-H. Lee, L. R. Rabiner, R. Pieraccini, and J. G. Wilpon, „Acoustic modeling for large vocabulary speech recognition," *Computer Speech & Language*, vol. 4, no. 2, pp. 127–165, 1990.

[43] S. M. Witt and S. J. Young, „Phone-level pronunciation scoring and assessment for interactive language learning," *Speech communication*, vol. 30, no. 2-3, pp. 95–108, 2000.

[44] T. Lee, Y. Liu, Y. T. Yeung, T. K. Law, and K. Y. Lee, „Predicting severity of voice disorder from *DNN-HMM* acoustic posteriors." in *Interspeech*, 2016, pp. 97–101.

[45] Y. Liu, T. Lee, P. Ching, T. K. Law, and K. Y. Lee, „Acoustic assessment of disordered voice with continuous speech based on utterance-level *ASR* posterior features." in *Interspeech*, 2017, pp. 2680–2684.

[46] E. Barth, *Einführung in die Physiologie, Pathologie und Hygiene der menschlichen Stimme*. G. Thieme, 1911.

[47] H. Stern, „Klinik und therapie der krankheiten der stimme," *Mschr Ohrenheilk*, vol. 58, pp. 1–53, 1924.

[48] D. Weiss, „Der begriff des funktionellen mit besonderer berücksichtigung der sprach-und stimmheilkunde," *Mschr Ohrenheilk*, vol. 68, pp. 830–832, 1934.

[49] H. Gundermann, *Die Berufsdysphonie: Nosologie der Stimmstörungen in Sprechberufen unter besonderer Berücksichtigung der sogenannten Lehrerkrankheit.* Thieme, 1970.

# Publikációk

### Nemzetközi folyóiratok

[J1] Szaszák, Gy., **Tulics, M. G.**, & Tündik, M. Á., „Analyzing FO discontinuity for speech prosody enhancement," *Acta Univ. Sapientiae Elect. Mech. Eng*, vol. 6, no. 1, pp. 59–67, 2014. (6/3=2 pont)

[J2] **Tulics, M. G.**, & Vicsi, K., „Automatic classification possibilities of the voices of children with dysphonia," *Infocommunications Journal* Vol. X. No.3. pp. 30-36., 7 p. 2018. (4 pont)

[J3] Kovács, A., **Tulics, M. G.**, Tündik, M. Á., Moró, A., Gróf, A., „Magmanet: Ensemble of 1d convolutional deep neural networks for speaker recognition in hungarian," *Phonetician*, vol. 115, pp. 72–86, 2018. (6/5=1.2 pont)

[J4] **Tulics, M. G.**, & Vicsi, K. (2019). „The automatic assessment of the severity of dysphonia," International Journal of Speech Technology, 1-10. (6 pont)

[J5] Szántó, D., Jenei, A. Z., **Tulics, M. G.**, & Vicsi, K., „Developing a Noise Awareness Rising Web Application within the „Protect your Ears" project," *Infocommunications Journal* 2020. Elfogadott

### Magyar folyóiratok

[J6] Sztahó, D, Kiss, G, **Tulics, M G**, Czap, L, Vicsi, K, „Számítógéppel támogatott prozódiaoktató program," *Alkalmazott Nyelvészeti Közlemények* 9 : 1 pp. 144-153. , 10 p. (2014) (2/4=0.5 pont)

### Nemzetközi konferenciák

[C1] **Tulics, M. G.**, Kazinczi, F., & Vicsi, K., „Statistical analysis of acoustical parameters in the voice of children with juvenile dysphonia," In *International Conference on Speech and Computer* (pp. 667-674). Springer, Cham. 2016. (3/2=1.5 pont)

[C2] **Tulics, M. G.**, Szaszák, Gy., Mészáros, K. & Vicsi, K., „Artificial Neural Network and SVM based Voice Disorder Classification," In *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE. 2019. (3/3=1 pont)

[C3] **Tulics, M. G.**, Lavati, L. J., Mészáros, K. & Vicsi, K., „Possibilities for the automatic classification of functional and organic dysphonia," In *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*. 2019. (3/3=1 pont)

[C4] **Tulics, M. G.**, & Vicsi, K., „Phonetic-class based correlation analysis for severity of dysphonia," In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 000021-000026). IEEE. 2017. (3 pont)

[C5] Kiss, G., **Tulics, M. G.**, Sztahó, D., Esposito, A., & Vicsi, K., „Language independent detection possibilities of depression by speech," In *Recent advances in nonlinear speech processing* (pp. 103-114). Springer, Cham. 2016. (3/4=0.75 pont)

[C6] Sztahó, D., **Tulics, M. G.**, Vicsi, K., & Valálik, I., „Automatic estimation of severity of Parkinson's disease based on speech rhythm related features," In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 000011-000016). IEEE. 2017. (3/3=1 pont)

[C7] Sztahó, D., Kiss, G., **Tulics, M. G.**, & Vicsi, K., „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)* (pp. 1-4). IEEE. 2018. (3/3=1 pont)

[C8] Sztahó, D., Kiss, G., **Tulics, M. G.**, Dér-Hajduska,B. & Vicsi, K., „Automatic discrimination of several types of speech pathologies," In *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*. 2019. (3/4=0.75 pont)

[C9] **Tulics, M. G.**, Szaszák, Gy., Mészáros, K. & Vicsi, K., „Using ASR Posterior Probability and Acoustic Features for Voice Disorder Classification," In *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE. 2020. Elfogadott

**Magyarországi konferenciák**

[C10] **Tulics, M. G.**, Jászai, H., & Vicsi, K., „A diszfónia súlyosságának automatikus becslése, a szakértői értékelések szubjektív jellegének figyelembevételével," In: *Vincze, Veronika (szerk.) XIV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2018)* Szeged, Magyarország : Szegedi Tudományegyetem Informatikai Tanszékcsoport, pp. 206-218. , 13 p. 2018. (1/2=0.5 pont)

**Összes publikációs pontszám: 24.2 pont**

**Független idézetek: 13**